Article category:
Original research article

Sub-discipline:
Sports Medicine (Injury screening)

Sources of outside support:

There was no specific funding for the research study, though the research was supported by World Rugby, which employs some of the author group, as described in Competing Interests

Competing interests:

Three of the authors (RT, EF, MH) are full-time and part-time employed by World Rugby in roles of research and medicine. JP and GF have served as independent advisors on a working group on concussion administered by World Rugby, for which expenses are covered.

Ethics approval and consent to participate

The research plan for this study was approved by the World Rugby Institutional Ethics committee (REF 19007). Players had provided written informed consent for all data gathered as part of the World Rugby Concussion management programme to be used for research in a de-identified manner

Availability of data and material

Original participant data belongs to the players and the clubs/unions that generate such data. This may be provided upon request to third parties. World Rugby (the corresponding author) may facilitate the provision of that data, in terms of permissions and contacts, though there is not a single point of contact, since the data are generated globally from multiple teams and Unions.

Author' contributions and declaration

MR conceived the study. MR, RT and EF designed the study. RT and MH performed the analyses. All authors made substantial contributions to the study design, data processing and interpretation. RT drafted the article and all other authors revised it critically for important intellectual content. RT is the guarantor. All authors had full access to all of the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. The manuscript has not been published elsewhere and is not being considered for publication elsewhere.

Sport Concussion Assessment Tool: baseline and clinical reference limits for concussion diagnosis and management in elite rugby union

Word count: 2985

Abstract word count: 247

Number of Figures: 2

Number of Tables: 1

Supplementary tables and figures: 3

Abstract

Objectives

Rugby Union has adapted the Sports Concussion Assessment Tool (SCAT) into an abridged off-field concussion screen and the complete SCAT is used during diagnostic screens performed after head impact events. No firm guidelines exist as to what should be considered "abnormal" and warrant further evaluation. This study evaluates SCAT performances in 13479 baseline SCAT assessments, and proposes clear reference limits for each sub-component of the SCAT5. Baseline reference limits are proposed to guide management of baseline testing by identifying abnormal sub-modes, enhancing the clinical validity of baseline screens, while clinical reference limits are identified to support concussion diagnosis when no baseline is available.

Design

Cross sectional census sample

Methods

13 479 baseline SCATs from 7 565 elite adult rugby players were evaluated. Baseline reference limits were identified for each sub-mode as the sub-mode result achieved by approximately 5% of the population, while clinical references limits corresponded to the sub-mode score achieved by as close as possible to 50% of the cohort.

Results

Players reported symptoms 35% (95% CI 1.29 – 1.42) more frequently during SCAT5 than SCAT3 baseline assessments (mean 1.4 ± 2.7 vs 1.0 ± 2.4). Ceiling effects were identified for many cognitive sub-tests within the SCAT. Baseline and Clinical reference limits corresponding to the worst performing 5th percentile and 50th percentile were described.

Conclusion

Targeted baseline re-testing should be repeated when abnormal sub-modes are identified according to proposed baseline reference limits, while a more conservative clinical reference limit supports concussion diagnosis during screens in diagnostic settings.

Keywords:

Concussion, SCAT, Rugby Union, neurological screening, concussion management, injury

Practical implications

- SCAT5 screening should remain part of the overall management of sports related concussion
- The clinical utility of baseline screening can be enhanced if clinicians view such screening as a means to identify abnormalities as part of annual medical screening
- Clinicians who undertake regular baseline screening should use pre-identified reference limits to identify abnormal tests that warrant further investigation, either repeating tests or investigating contributing factors described here
- In the clinical setting, the application of clinical reference limits that correspond to sub-test scores achieved by half the cohort provide a more conservative method of identifying abnormal tests and removing players with suspected concussions

List of abbreviations

SCAT – Sports Concussion Assessment Tool

HIA - Head Injury Assessment

Introduction

The Sports Concussion Assessment Tool (SCAT) was first developed in 2004 using tests from eight existing tools by the Concussion in Sport Group, as a standardised assessment tool for acute concussion [1].

Rugby Union has adapted and implemented the SCAT into an abridged off-field concussion screening tool for the professional game (Head Injury Assessment). The complete SCAT5 is used during diagnostic screens performed within three hours of the head impact event (HIA2 screen) and after two nights' rest (HIA3 screen) [2].

World Rugby requires mandatory completion of a baseline SCAT in professional players, usually performed in the pre-season, with subsequent diagnostic results evaluated relative to these uninjured baseline results.

A number of sporting organisations have abandoned compulsory baseline testing for use in concussion diagnosis, instead using only normative data [3]. The time required to complete baseline assessments, the possibility that their use may not improve the sensitivity or specificity of concussion diagnosis, and the difficulty of confirming player effort during baseline testing [1,4,5], are practical and theoretical considerations for those involved in sports concussion diagnosis and management.

In the absence of a baseline performance, screen results may be compared to normative data derived from a sport-, sex- and age-matched population [6]. In research published to date, including in rugby players [6,7], SCAT performances have typically been categorised into ranges as per the Wechsler classification [8], but without commitment to clinically relevant cut-offs that indicate when a concussion diagnosis should be made.

While diverging views on the merits of baseline testing for SRC exist, baseline testing is a clinically useful annual interaction between players and team doctors, offering ancillary benefits. One must consider, therefore, whether baseline utility can be improved. This might be achieved by enhancing the validity of baseline SCAT tests through content modification, or by changing the baseline SCAT process to improve clinical utility. Unusually poor sub-mode performances may indicate poor effort or an underlying issue at baseline, perhaps triggering repeat testing. This approach is not unique, with computerised cognitive tests also using normative data to trigger repeat testing and ensure engagement [4].

The primary aim of this study was to analyse SCAT baseline performance in a large (n = 13479, from 7565 players) cohort of professional rugby players to identify clear baseline reference limits that indicate abnormal sub-mode performance, and thus require re-testing at baseline. A secondary aim was to apply the baseline cohort data to identify a distinct clinical reference limit to support concussion diagnosis during the HIA1, HIA2 and HIA3 phases in the event that baseline data are absent for a player after a head impact during play.

We propose a baseline reference limit that corresponds to the sub-test score or performance achieved by the worst-performing 5% of the cohort, while the clinical reference limit is proposed to be the sub-mode score attained by as close as possible to 50% of the cohort.

Finally, we apply the baseline limits to propose an approach to abnormal sub-component results that will optimize the baseline SCAT collection process (Appendix A). This is intended to improve player effort and baseline validity by ensuring that results falling outside of expected ranges are subject to scrutiny at baseline, rather than later.

Methods

A cross sectional study was performed using data from the World Rugby Head Injury Assessment (HIA) database, which contains baseline and diagnostic concussion screen results from the professional game. In order to use the HIA process, a competition must adhere to mandatory competition player welfare standards [World Rugby Player Welfare Site] that ensures a standardised approach to concussion detection and management as well as data collection. The source population thus comprises the majority of eligible professional male players in domestic and international competitions, as well as International Women's squads that underwent mandatory baseline SCAT assessment between 2015 and 2019.

The SCAT assessments were administered prior to commencement of the relevant competition season or tournament, according to methods described previously [6]. A total of 14803 baseline screens from 7630 players were present in the database.
For the present analysis, we excluded baseline SCATs performed post-exercise, and thus analysed 13479 resting SCAT assessments (5757 SCAT3 and 7722 SCAT5) from 7565 players. We recognise that there may be learning effects in players with multiple tests. These potential effects will be evaluated in subsequent research studies. We chose to include all resting baseline tests to maximize the external validity of the study, since the annual requirement to perform these SCATs means that most players will perform multiple SCATs in their careers.

Descriptive data for each sub-component are presented as means, standard deviations, medians and ranges. Distributions of continuous variables were visualised using density histograms and summarised using mean (M), median (Md), Standard deviation (SD), interquartile range (IQR) and range.

The research plan for this study was approved by the World Rugby Institutional Ethics committee (REF 19007). Players had provided written informed consent for all data gathered as part of the World Rugby Concussion management programme to be used for research in a de-identified manner

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of the research.

Descriptive statistics for each sub-test were presented as Means, Standard Deviation, Medians and Interquartile Ranges (IQR). A baseline reference limit was identified as that score that was achieved by approximately the worst-performing 5% of players in the cohort. That is, the 5th/95th percentile guided the identification of a sub-test result that would achieve as close to 5% abnormal results as possible.

A clinical reference limit was identified using a similar method, but at the 50th percentile, rather than the 5th/95th percentile. Classifications were defined based on direction of scoring for abnormality in each sub-test, with higher symptom scores and modified Balance Error Scoring System (mBESS) errors referred to as abnormally high, and lower cognitive test performances referred to as abnormally low.

A modification in SCAT5 compared with SCAT3 involved the method of assessing symptoms. In SCAT5, a player is handed the symptom sheet to read aloud, and instructed to 'rate his/her symptoms based on **how he/she TYPICALLY feels'**. These have been termed 'trait' symptoms [9]. During SCAT3, and when the SCAT is applied post-injury, the instruction to players is to identify '**how they feel now'**, so-called 'state' symptoms [10]. We explored whether this change affected symptom results by calculating proportion ratios, with 95% confidence limits calculated according to the delta method. The proportion ratio was calculated as the proportion of players reporting a symptom during SCAT5, divided by the proportion of players reporting that symptom during SCAT3 assessments. Effect sizes for proportion ratios were judged on the following threshold values: trivial – PR <1.11; small ≥1.11 PR <1.43; moderate ≥1.43 PR <2.00; large ≥2.00 PR <3.30; very large ≥3.30 PR <10.00; and extremely large PR ≥10.00 [11]. Z-scores were produced for each comparison to test against the null hypothesis that no difference would exist in symptom reporting frequency between the two SCAT modalities. Statistical significance was accepted at $P<0.05$.

All Statistical analyses were conducted using SPSS (V.23 for Windows, IBM Corp, Armonk, NY, USA). Statistical significance was accepted at $\alpha<0.05$.

Results

The sub-component scales, cases, means, standard deviation, medians, interquartile ranges and 5th/95th percentiles for each SCAT components are shown in the supplementary materials (Table 1 supplementary material).

65.2% of players were asymptomatic during baseline testing. Five or more symptoms were reported by 9.1% of players, indicative of an “unusually high” number, while the 95th percentile corresponded to seven symptoms.

The percentage of baseline assessments in which each symptom was reported is shown in Table 1. To support clinical management and insight, symptoms are grouped into categories of Physical, Cognitive, Vestibulo-ocular and Psychological [3,12].

Table 1 here *

Fatigue, neck pain, trouble sleeping and nervous/anxious were the most commonly reported symptoms, accounting for 52.0% of all symptoms reported.

Symptom endorsement was higher during SCAT5 (1.4 ± 2.7) than SCAT3 (1.0 ± 2.4). Players report at least one symptom 35% more frequently during the SCAT5 assessment than the SCAT3 assessment (proportion ratio 1.35, 95% CI 1.29 – 1.42, $P<0.001$, Figure 1). All individual symptoms were reported more frequently in SCAT5 than SCAT3 (proportion ratios = 1.12 – 1.55), although effects were small to moderate in size.

Figure 1 here *

85.2% of players scored perfectly (five out of five) for Orientation questions, while 99.9% of players answered at least three questions correctly (equating to 9,992 per 10,000 assessments).

Date was most frequently answered incorrectly (12.7%) followed by Month, Time and Day (all 0.8%), with Year least frequently incorrect (0.1%).

Given the change in Immediate Memory assessment from the SCAT3 (a five-word list) to the SCAT5 (a ten-word list was added as an option), Immediate Memory performance was evaluated separately for SCAT3 and SCAT5 assessments (Figure 1, Supplementary material).

When using the five-word list (n = 8437), 65.9% of players scored 15 out of 15 (Median 15, IQR 14 – 15), with only 2.9% of players scoring fewer than twelve (Figure 1, Supplementary material). The ten-word list (n = 5042) results were more normally distributed (Mean 21.5, Median 21, IQR 19 to 24, Table 1 supplementary material). The 5th percentile for the 10-word list corresponded to a score of 15 out of 30.

The mean concentration score was 4.1 ± 1.0 (out of a maximum of five), with 44.2% of players scoring perfectly. The 5th percentile corresponded to a score of two out of 5, with 91.7% of players achieving a score of three out of five. Months in reverse was correctly answered in 91.7% of baseline assessments.

Delayed Recall was assessed using either the five- or ten-word list. A similar ceiling effect was observed using the five-word list, with 97.5% of players scoring two or more out of five. Using the ten-word list, the 5th percentile corresponded to a score of four out of ten, with 96.6% of players recording at least four correct answers.

99.9% of players completed the tandem gait test in under 17 seconds. The 95th percentile corresponded to a time of 13.3 seconds.

Errors during double leg balance were rare, with 97.4% of players performing the assessment without any errors. An average of 1.9 errors were made during the single leg balance assessment, with 29.9% of players performing without error. The 95th percentile corresponded to six errors.

Tandem stance errors averaged 0.8, with a 95th percentile at three errors. Collectively, total errors ranged between zero and 22, with a mean of 2.8 and a 95th percentile corresponding to eight errors.

A schematic summary of the identified baseline and clinical reference limits for each sub-mode in the SCAT5 is shown in Figure 2. The baseline reference limit (top panel) is derived from the 5th and 95th percentiles, and is that score at which as close as possible to 5% of players achieve an abnormally poor result. The clinical reference limit corresponds a sub-mode score as close as possible to the 50th percentile.

Figure 2 here *

The baseline and clinical reference limits are further summarized into clinical guidelines in Table 2 of the supplementary material, showing the sub-component result that would warrant further investigation (during baseline) and which would support a diagnosis of concussion (during diagnostic settings such as World Rugby's HIA1 off-field screen and the HIA2 and HIA3 assessments).

Discussion

This study used a large dataset of baseline SCAT3 and SCAT5 assessments in professional rugby players to identify reference limits for each sub-component in the SCAT assessment.

We propose that the baseline reference limits identified here will enhance the clinical utility of the SCAT baseline testing. Any scores outside of the baseline reference limits indicate that re-testing be undertaken and, if abnormalities persist, further clinical evaluation. In addition, during rugby matches when no player baseline is available, the identified clinical reference limits may be used to support concussion diagnosis and to guide return-to-play decisions.

Given that the baseline reference limit is identified at the sub-mode score as close as possible to the $5^{th}/95^{th}$ percentile, while the clinical reference limits correspond to scores near the 50^{th}percentile, every sub-mode score requirement during clinical settings is more challenging than during baseline (Figure 2). For example, six or more single leg errors constitutes an abnormal baseline test, whereas an abnormal clinical screen at HIA1, HIA2 or HIA3 occurs at two or more errors (Figure 2 and Table 2 Supplementary material). We recognize that this will produce more abnormal clinical tests than previously, since the thresholds have been reduced compared to historical thresholds. However, since baseline testing is now mandatory, the real impact of this change will be small because normative data should rarely be applied. We also deem it to be more conservative, and thus preferable, because fewer false negatives will occur. Thus, despite the risk of increased cases of false positives, we deem the proposed clinical reference limits to be preferred in cases where no baseline screen is present.

The baseline reference limits that guide re-testing of abnormal baseline screens are based on the premise that these scores are achieved by the worst-performing 5% of players (Figure 2). We then propose specific guidance to evaluate these results (Appendix A), with advice on repeating any abnormal baseline sub-components, followed by clinical steps that may identify contributing factors and possible confounders for persistent abnormal results.

The process we outline here will also address concerns such as player effort, effective implementation and data reliability, since repeating tests that are identified using the baseline reference limits will ensure greater concentration and performance. This approach also supports concussion education, allows the physician to obtain a better understanding of individual player's medical profiles and ensures more accurate post-injury diagnosis. Player welfare will also be improved with the recommendation to investigate reported baseline 'trait' symptoms. Each sub-component outcome is described briefly.

We found that symptom endorsement is greater using SCAT5 than SCAT3 (Figure 1), possibly as a result of different instructions for how symptoms should be collected.

The distinction between a trait and a state symptom is key to the collection and diagnostic utility of valid and reliable baseline SCAT5 symptoms, which should be recorded only if typically present. After a head impact, only symptoms that are new or altered should indicate a concussive event.

For this reason, symptom reference limits have not been proposed (Figure 2). In return-to-play and diagnostic settings, clinicians should interpret the presence of new symptoms as indicative of a concussion, while symptoms claimed by the player to be typically present (trait symptoms) should be questioned to identify if these symptoms have changed. A 'trait' symptom that has worsened should be interpreted as indicative of a concussion.

In the general population, a variety of medical conditions may cause concussion-like symptoms. For example, headaches may be cervicogenic in origin, dizziness may be related to viral infection or cardiac disease, and sleep disorders may relate to underlying depression or anxiety [12]. The most commonly endorsed symptom, fatigue (19.6% of SCAT5s, Table 1), is often load-related, but the clinician should also consider illness (e.g. anaemia), and psychologically-related fatigue. These may require investigation using tests such as Profile of Mood States (POMS) [13], medical work-up and endocrinological review considered if mood is unaffected.

Neck pain, reported in 16.3% of baseline SCAT5 tests, warrants further investigation due to a possible role in prolonged concussion recovery and persistent post-concussion symptoms [14,15]. Cervical muscles are thought to play a significant role in chronic headaches [16,17]. Mechanical neck pain is common but other causes such as cervical disc pathology, shoulder pathology and medical conditions require exclusion.

Trouble falling asleep may be considered as a sign of heavy training load and functional over-reaching [18], possible increased use of ergogenic agents such as caffeine and taurine [19] or a potential indicator for undiagnosed mental health issues. Sleep hygiene assessment should be considered because quality and quantity of sleep are recognised components of an athlete's recovery and preparation [20].

Finally, anxiety was reported in 9.7% of SCAT5 tests at rest and may be a potential indicator of a mental health condition. This symptom requires further investigation which may include specific neuro-psychological screen or immediate referral for psychological evaluation.

We found that the use of the ten-word lists at least partly overcomes ceiling effects [1,21] during Immediate Memory and Delayed Recall sub-tests (Figure 2, supplementary material). This may improve the clinical performance of these sub-components for concussion diagnosis.

Clinically abnormal cognitive tests that persist at re-test baseline SCAT indicate further assessments, either via computerised psychometric assessment or formal neuropsychological assessment. In players with a history of previous concussions, post-concussion syndrome should be considered. In all instances medical illnesses need to be considered.

Numbers of balance errors are higher than previously identified, with a 95th percentile corresponding to six errors for single leg stance, and 4.6% of players making four errors during tandem stance.

Abnormal balance results that persist on retesting may indicate chronic ankle ligamentous instability, a common complaint in field sport populations [22]. Many such athletes regularly strap their ankles, in this case, baseline testing should be repeated under similar conditions. Further lower limb orthopaedic causes should be investigated and if indicated, vestibular-ocular assessment, and a thorough neurological examination is also recommended.

Among the strengths of this study are its size, among the largest documented number of baseline assessments in athletes, allowing for robust conclusions and normative ranges and reference limits to be created. The method of collection, using the CSx platform, allows immediate data collection with minimal missing data. Study conduct and reporting is consistent with STROBE guidelines for observational studies [23]

There are some limitations to the present study. The inclusion of multiple tests per player may introduce learning effects, which we acknowledge. However, we chose to include these tests because during the diagnostic screens after head impacts, those same learning effects are present, and thus any normative limits derived from baseline testing should be generated using all tests for external validity. The potential for learning effects will be explored in future research.

We also cannot account for individual player circumstances and characteristics, including previous concussions and other injuries, and acknowledge that these may affect baseline performances and thus normative ranges. Future research will also explore how head impact events and diagnosed concussions affect subsequent baseline performance. Finally, intra- and inter-observer reliability was not assessed.

Our recommendation is that individual baseline SCAT be retained as part of the overall management of sports related concussion. We have identified what we propose as reference limits for abnormal sub-test results during baseline and during clinical settings when baseline data are absent. We recommend that the baseline reference limits guide the re-testing of abnormal sub-modes, and possible investigation of persistent abnormal performances. This approach should ensure collection of a more reliable and valid individual baseline SCAT test and allows the clinician to use baseline testing as a screening tool for both concussive and non-concussive related injury.

Recognising that this new approach will add to the workload of the team medical staff our next analysis will investigate the necessity for annual baseline SCAT by reviewing cases where multiple baseline SCATs are available over time. This analysis will also review the impact of previous concussive events on baseline SCAT modes and identify if exercise and rest influence each SCAT mode. This subsequent analysis will support recommendations regarding the necessity for annual part or full baseline SCAT collection.

Acknowledgements

We would like to acknowledge and sincerely thank the team doctors and medical practitioners for their help in facilitating collection of Sports Concussion Assessment Tool (SCAT) data.

References

1. Echemendia RJ, Meeuwisse W, McCrory P, et al. The Sport Concussion Assessment Tool 5th Edition (SCAT5): Background and rationale. *Br J Sports Med* 2017;51:848–850.

2. Raftery M, Kemp S, Patricios J, et al. It is time to give concussion an operational definition: a 3-step process to diagnose (or rule out) concussion within 48 h of injury: World Rugby guideline. *Br J Sports Med* 2016;50:642–643.

3. Harmon KG, Drezner JA, Gammons M, et al. American Medical Society for Sports Medicine position statement: concussion in sport. *Br J Sports Med* 2013;47:15–26.

4. Walton SR, Broshek DK, Freeman JR, et al. Valid but Invalid: Suboptimal ImPACT Baseline Performance in University Athletes. *Med Sci Sport Exerc* 2018;50:1377–1384.

5. Chin EY, Nelson LD, Barr WB, et al. Reliability and Validity of the Sport Concussion Assessment Tool-3 (SCAT3) in High School and Collegiate Athletes. *Am J Sports Med* 2016;44:2276–2285.

6. Fuller GW, Govind O, Tucker R, et al. Sport concussion assessment tool-Third edition normative reference values for professional Rugby Union players. *J Sci Med Sport* 2018;21:347–351.

7. Fuller GW, Raftery M. Sport Concussion Assessment Tool-Fifth Edition Normative Reference Values for Professional Rugby Union Players. *Clin J Sport Med* 2018;Publish Ahead of Print:1.

8. Hänninen T, Tuominen M, Parkkari J, et al. Sport concussion assessment tool – 3rd edition – normative reference values for professional ice hockey players. *J Sci Med Sport* 2016;19:636–641.

9. Asken BM, Houck ZM, Bauer RM, et al. SCAT5 vs. SCAT3 Symptom Reporting Differences and Convergent Validity in Collegiate Athletes. *Archives of Clinical Neuropsychology* 2019;32:963–11.

10. Asken BM, Houck ZM, Bauer RM, et al. SCAT5 vs. SCAT3 Symptom Reporting Differences and Convergent Validity in Collegiate Athletes. *Archives of Clinical Neuropsychology* 2019;32:963–11.

11. Hopkins W, 2010. Linear models and effect magnitudes for research, clinical and practical applications. *Sportscience* 2010;14:49–48.

12. Harmon KG, Clugston JR, Dec K, et al. American Medical Society for Sports Medicine position statement on concussion in sport. *Br J Sports Med* 2019;53:213–225.

13. McNair DM, Lorr M, Droppleman LF. *Profile of Mood States*. 1971.

14. Ellis MJ, McDonald PJ, Olson A, et al. Cervical Spine Dysfunction Following Pediatric Sports-Related Head Trauma. *J Head Trauma Rehabil* 2019;34:103–110.

15. Ellis MJ, Leddy J, Willer B. Multi-Disciplinary Management of Athletes with Post-Concussion Syndrome: An Evolving Pathophysiological Approach*. Front Neurol* 2016;7:136.

16. Lyngberg AC, Rasmussen BK, Jørgensen T, et al. Has the prevalence of migraine and tension-type headache changed over a 12-year period? A Danish population survey. *Eur J Epidemiol* 2005;20:243–249.

17. Madsen BK, Søgaard K, Andersen LL, et al. Efficacy of strength training on tension-type headache: A randomised controlled study. *Cephalalgia* 2018;38:1071–1080.

18. Leeder J, Glaister M, Pizzoferro K, et al. Sleep duration and quality in elite athletes measured using wristwatch actigraphy. *J Sports Sci* 2012;30:541–545.

19. Bonnar D, Bartel K, Kakoschke N, et al. Sleep Interventions Designed to Improve Athletic Performance and Recovery: A Systematic Review of Current Approaches. *Sports Med* 2018;48:683–703.

20. Fullagar HHK, Skorski S, Duffield R, et al. Sleep and athletic performance: the effects of sleep loss on exercise performance, and physiological and cognitive responses to exercise. *Sports Med* 2015;45:161–186.

21. Echemendia RJ, Broglio SP, Davis GA, et al. What tests and measures should be added to the SCAT3 and related tests to improve their reliability, sensitivity and/or specificity in sideline concussion diagnosis? A systematic review. *Br J Sports Med* 2017;51:895–901.

22. Attenborough AS, Hiller CE, Smith RM, et al. Chronic ankle instability in sporting populations. *Sports Med* 2014;44:1545–1556.

23. Elm von E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. In: Vol 12. 2014:1495–1499.

Figure titles and legends

Figure 1: Proportion ratios (x/÷ 95% confidence intervals) for symptoms reported in SCAT5 relative to symptoms reported in SCAT3 symptoms. Effect sizes and P-values also shown.

Figure 2: Schematic representation of the identified baseline (top panel) and clinical (bottom panel) reference limits for SCAT5 sub-modes. Baseline reference limits are to be applied at baseline testing, indicating abnormal sub-modes that require re-testing. Clinical reference limits are applied during screens when baseline data are absent in clinical settings, and correspond to a sub-mode score nearest the 50^{th} percentile