Article:

Fuller, G.W. orcid.org/0000-0001-8532-3500, Cross, M.J., Stokes, K.A. et al. (1 more author) (2019) King-Devick concussion test performs poorly as a screening tool in elite rugby union players: a prospective cohort study of two screening tests versus a clinical reference standard. British Journal of Sports Medicine, 53 (24). pp. 1526-1532. ISSN 0306-3674

https://doi.org/10.1136/bjsports-2017-098560

VALIDATION OF THE KING DEVICK CONCUSSION SCREENING TOOL IN ELITE RUGBY PLAYERS: A PROSPECTIVE COHORT STUDY

Fuller GW[1]* Cross M* Stokes K, Kemp SPT

* Contributed equally as first author

Corresponding author[1]: Dr Gordon Ward Fuller

Work address: Emergency Medicine Research in Sheffield Group, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

Telephone: +44 (0) 7968 280641 / +44 (0) 1484 684640

E-mail: g.fuller@sheffield.ac.uk

Dr Matthew Cross, Rugby Football Union, Rugby House, Twickenham Stadium, 200 Whitton Road, Twickenham, Middlesex, TW2 7BA, UK; Department for Health, University of Bath, Claverton Down, Bath, BA2 7AY, UK

Email: MatthewCross@rfu.com

Professor Keith Stokes, Department for Health, University of Bath, Claverton Down, Bath, BA2 7AY, UK

Email: k.stokes@bath.ac.uk

Dr Simon P T Kemp, Rugby Football Union, Rugby House, Twickenham Stadium, 200 Whitton Road, Twickenham, Middlesex, TW2 7BA, UK

Email: SimonKemp@RFU.com

Key words:

Brain Concussion, Sensitivity and Specificity, Rugby, King-Devick Test

Word count: 3,531

ABSTRACT

Background

The King-Devick (KD) Test is an objective clinical test of eye movements that has been used to screen for concussion. This study aimed to characterise the accuracy of the KD test as an off-field evaluation for concussion after a suspicious head impact event.

Methods

A prospective cohort study was performed in elite English Rugby Union competitions between September 2016 and May 2017. The study population comprised consecutive players identified with a head impact event with the potential to result in concussion. The KD test was administered off field, alongside the World Rugby Head Injury Assessment (HIA-1) screening tool and results were compared to a pre-season baseline. Accuracy was measured against a reference standard of confirmed concussion, based on the clinical judgement of the team doctor after serial assessments.

Results

A total of 145 head injury events requiring off-field medical room screening assessments were included in the primary analysis. The KD test demonstrated a sensitivity of 59.6% (95%CI 49.0-69.6), and a specificity of 39.2% (95%CI 25.8-53.9), to identify players subsequently diagnosed with concussion. Area under the receiver operating characteristic curve for prolonged KD test times was 0.51 (95%CI 0.41-0.61). The World Rugby HIA-1 off field screening tool sensitivity did not differ significantly from the KD test (sensitivity 74.8%, 95%CI 65.6-82.5, p=0.08); but specificity was significantly higher (91.3%, 95%CI 82.0-96.7, p<0.001). Combining the KD test and the World Rugby HIA-1 multi-modal screening assessment achieved a significantly higher sensitivity of 92.6% (95% CI 85.9-96.7%), with significantly lower specificity of 33.3% (95%CI 20.8-47.9%), compared to the HIA-1 test alone.

Conclusions

The KD test had limited accuracy as a remove-from-play sideline screening test for concussion. Combining the KD test with the HIA-1 multi-modality screening instrument provided improved sensitivity for identifying concussion, but at the expense of lower specificity.

INTRODUCTION

Concussion is a common and high profile injury in collision sports. Short term sequlae include somatic, cognitive and neurological symptoms or signs, increased risk of injury, and reduced athletic performance.[1] Possible long term consequences may include an increased risk of neurodegenerative disorders or depression.[1] Early detection of suspected concussion and removal of the affected player will likely mitigate these potential adverse effects and facilitate further evaluation, management and safe return-to-play.

Due to the variability and subtlety of symptoms and signs, and pressure on athletes to continue playing, identification of sports-related concussion is challenging and injuries may go unrecognized or be ignored.[2] Elite sports, including Rugby Union, have consequently introduced management systems to identify and manage head impact events with the potential for concussion during matches.[3] These typically involve brief, off-field, multi-modality initial screening for a possible concussion, rather than definitive diagnosis of a head injury. However, a recent systematic review supporting the 5th Consensus statement on Concussion in Sport was unable to make an evidence-based recommendation for any single screening test.[4]

Visual and eye movement neuronal pathways are widely distributed throughout the brain and may become impaired following brain trauma.[5] The King-Devick (KD) test, an oculomotor test originally designed for reading evaluation, has been promoted as a concussion screening tool.[6] Preliminary studies have demonstrated a worsening of performance from baseline in concussed patients.[6] However, a recent systematic review concluded that *'The quality of evidence is not yet sufficient to warrant clinical recommendations for the use of oculomotor based vision measurement either as an indicator of mild traumatic brain injury or as a measure of recovery following mTBI'*.[7]

The aim of this study was therefore to validate the KD test for identifying players with concussion in elite adult male Rugby Union. The primary objective was to characterise the accuracy of the KD test for identifying concussion. Secondary objectives included comparing the sensitivity and specificity of the KD and World Rugby Head Injury Assessment HIA-1 off-field screening test and evaluating the joint performance of the KD and HIA-1 screening tests.

METHODS

Study design, setting and study population

A prospective cohort study was performed in the top two English elite domestic league competitions (Premiership and Championship, 24 teams) in a single season between September 2016 and May 2017 to determine the accuracy of the KD screening test for concussion. To maximise internal validity the study followed expert recommendations on the conduct and reporting of diagnostic accuracy and reliability studies.[8-10]

The source population comprised consecutive male adult players entering the World Rugby Head injury Assessment (HIA) process after identification of meaningful head impact events with the potential to cause concussion. The HIA process has been described previously.[11] Briefly, players overtly demonstrating signs of concussion (e.g. loss of consciousness, tonic posturing, ataxia) are immediately and permanently removed from the remainder of the match, without undergoing further off-field concussion screening. Where the consequences of a head impact event are not clear, players undergo an off-field screening assessment for possible concussion with the multi-modality HIA-1 screening instrument, comprising Maddock's questions, Tandem gait test, immediate and delayed recall, a symptom checklist, and brief evaluation of clinical signs. Any abnormality in the HIA-1 screening test mandates removal from play. The main study population included players undergoing off-field HIA-1 screening, as these are the players which could potentially benefit from KD testing within the HIA process. However, in other elite sports all players undergo off-field screening following head impact events regardless of presenting signs. Players immediately and permanently removed from play were therefore also included in a subsequent combined analysis to increase the potential generalisability of the findings.

Index test

The KD Test is an objective clinical test of rapid eye movements, primarily evaluating brain pathways involved in saccadic eye movements, attention and language.[6 12] The test involves reading aloud a series of random single-digit numbers displayed in rows on three successive screens in a tablet application following familiarisation based on a practice screen. Athletes begin at the top left of each screen and read as quickly as possible from left to right across each row. The spacing between the rows of numbers becomes narrower on each successive screen requiring increased concentration and more accurate eye movements to avoid errors. The time taken is automatically kept for each test and the KD summary score for the entire test is based on the cumulative time taken to read all 3 test screens. The number of uncorrected errors, defined as any addition, omission or reversal of the number pattern, is also recorded. A pre-season baseline KD performance is established by the better of two consecutive trials. Post head impact event results are then compared to the subject's baseline. Any worsening of time and/or errors committed indicates an abnormal result.

Reference standard

All players entering the World Rugby HIA process undergo detailed medical assessments post-match (HIA-2 assessment) and after 2 nights rest (HIA-3 assessment) to monitor clinical progress and confirm a diagnosis of concussion by the team doctor. The HIA-2 assessment consists of a clinical evaluation including the SCAT-3 instrument. The HIA-3 assessment comprises a clinical evaluation, supported by an expanded SCAT-3 symptom checklist, a cognitive assessment (typically a computerised neuro-cognitive tool such as CogSport) and a balance assessment using the balance error scoring system and tandem gait balance tests. The reference standard, against which the accuracy of the KD test was tested, was a clinical diagnosis of concussion during the 48 hours post-injury, based on abnormal HIA-2 and/or HIA-3 assessments, determined by the team doctor.

Data collection and procedures

Medical staff completed a web-based training session led by King-Devick Technologies prior to participating in the study. Following this training, ongoing technical support was provided by King-Devick Technologies, with study-specific support given by the research team. Players from included teams received baseline KD testing pre-season by recording the best time (fastest) of two consecutive trials in a representative off-field setting during a training session. Following a meaningful head impact event the KD test was repeated. The KD test was performed by the team doctor in a dedicated medical room, after completion of the usual World Rugby HIA-1 screening test or following immediate and permanent removal with clear signs of concussion.[11] KD test time and errors were recorded using a proprietary tablet application. The KD test was used non-operationally and results were not displayed immediately, but due to the KD application design were accessible to clinicians. Team doctors were instructed not to look at results, or allow findings to influence return to play decisions. KD data were recorded contemporaneously using tablets and the web-based proprietary KD software platform. HIA process data are routinely collected at the point of assessment using the tablet based, web-hosted, CSx data platform;[13] with data subsequently linked to the World Rugby and RFU HIA databases. KD and HIA data were linked deterministically using unique player identifiers.

Analyses

Sample characteristics, and the distribution of baseline KD scores, were initially examined using descriptive statistics. Repeatability of baseline KD testing was evaluated using repeatability coefficients.[14] The accuracy of an abnormal KD test result (prolonged time from baseline and/or errors) for detecting concussion was then assessed in the primary analyses. Each case was coded according to the index test and reference standard result, with a 2x2 contingency table constructed to determine true positives, false positives, true negatives and false negatives. Prevalence of concussion, sensitivity and specificity, positive and negative predictive values, positive and negative likelihood ratios and diagnostic odds ratios with their 95% confidence intervals (CI), were subsequently calculated. The accuracy of prolonged KD time in isolation was examined by calculating the area under the receiver operating curve. Youden's J statistic was used to attempt to identify an optimal threshold cut-point for prolonged KD times.[15 16] Accuracy of errors alone was also examined independently through calculation of sensitivity and specificity. This primary analysis was initially performed for players requiring off-field screening following head impact events where the consequences were not clear; but was also repeated in a combined sample also including players immediately and permanently removed from play after demonstrating clear signs of concussion.

A number of secondary analyses were performed to: compare agreement and accuracy between the KD test and current World Rugby HIA-1 screening tests (Raw agreement / Fleiss's kappa and McNemar's test respectively); demonstrate the combined performance of the KD and HIA-1 screening tests when performed in parallel (sensitivity and specificity); and evaluate the reproducibility of pre- and post-season KD testing (Bland-Altman limits of agreement analysis).[17] Additional sensitivity analyses investigated the potential influence of clustered (clustered sandwich estimator for standard errors) and missing data (scenarios with different assumptions for cases with missing data).[18]

Sample size, statistics, ethics, and funding

A sample size calculation of 207 players undergoing off-field concussion screening assessments was calculated for the primary analysis, using Bruderer's method based on a conventional α of 0.05 and the following assumptions from previous HIA data: a prevalence of concussion of 30% in players with meaningful head impact events requiring HIA-1 concussion screening assessment;[3 19] a sensitivity of

90%; a specificity of 75% for prolonged KD test times to identify concussion; and a desired precision of ±7.5% for the 95% CI of the sensitivity estimate. This sample size would provide a 95% CI precision of ±7.0% for specificity and ±0.05 for an AUROC of 0.83.

Available case analyses were performed with sample size determined by the number of players with complete data for each analysis. Statistical analyses were carried out in Stata version 13.1 (StataCorp, College Station, USA) with a conventional significance level (α) of 0.05 used. The study protocol received ethical approval from the University of Bath. All players provided informed consent for participation prior to the start of the season. All data were anonymised. The KD application, technical support, and KD test data-management were freely provided by King-Devick Technologies. Statistical analyses were performed independently of the RFU and KD at the University of Sheffield according to a pre-specified protocol.

RESULTS

Derivation and characteristics of study participants

A total of 274 consecutive head impact events with the potential to cause concussion were detected in 261 players (13 players with 2 head impact events) during 264 matches in the 2016/2017 season. Of these 73 incidents (occurring in 67 players) were associated with overt signs or symptoms of concussion requiring immediate and permanent removal from play. The remaining 201 incidents (occurring in 196 players), where it was unclear if a meaningful head impact event had occurred, underwent off-field medical room screening assessments. Figure 1 presents a flow chart describing the derivation of study participants.

The mean age of the complete sample was 27.6 years (SD 2.6), with a mean height of 187 cm (SD 6.9) and mean weight 105 kg (SD 11.8). 61.1% were forwards with 38.9% backs. A wide range of mechanisms of injury were observed with head contact during tackles predominating (n=225, 60.7%, either being tackled or tackling). The distribution of baseline KD results was slightly positively skewed, with a median time of 44.3 seconds (IQR 38.5-50.9, range 28.3-73.9 seconds, n=207). The KD test demonstrated a small improvement on average in pre-season testing, with a mean improvement of 1.75 seconds across the two baseline trials (paired t test, $p<0.001$). The repeatability coefficient was 13.9 seconds, indicating that the absolute difference between the two baselines differs up to this value on 95% of occasions. The second baseline trial was slower in 24.1% of players. Player characteristics are shown in Table 1.

Baseline KD test data was missing in 20.1% of the 261 included players. Across the overall sample of 274 head impact events, variable-wise missing data rates were: HIA-1 test: 10.5%, KD index test: 27.4%; reference standard 7.3%. Case-wise missing data rate for each analysis are shown in Table 2.

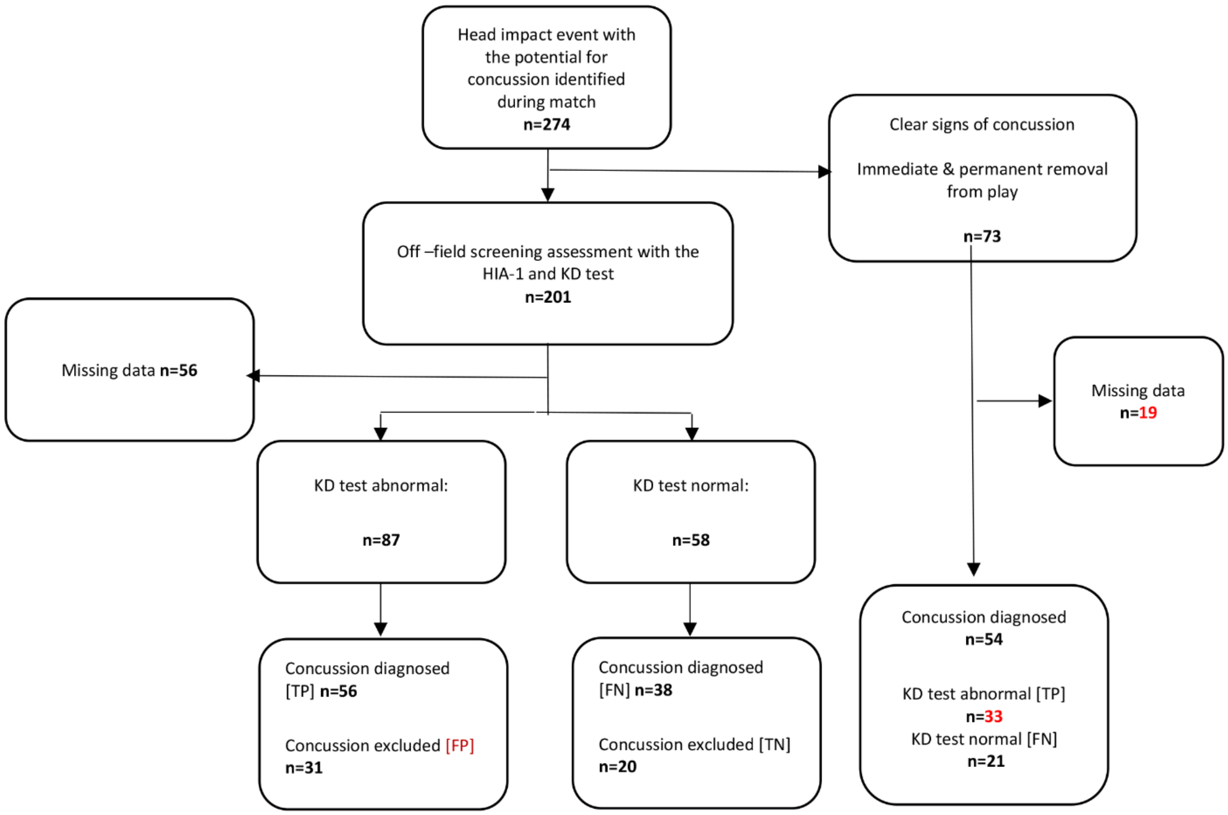

Figure 1. Derivation of study participants in primary analysis

Table 1. Characteristics of participants in primary analysis

	Frequency n=	Summary statistics
Total number of Players	261	
Player demographics:		
Age		27.6 (SD 2.6)
Weight		105 kg (SD 11.8)
Height		187 cm (SD 6.9)
Position: Forward	131	50.2%
Back	84	32.1%
Unknown	36	13.7%
Mechanism of head impact event*:		
Tackling	120	43.8%
Being Tackled	46	16.8%
Ruck/maul	28	10.2%
Scrum	0	0.0%
Accidental collision	20	7.3%
Unknown	60	21.9%
Baseline KD test		
Time	207	44.3s (IQR 38.5-50.9)

* A total of 274 consecutive head impact events with the potential to cause concussion were detected in 261 players (13 players with 2 head impact events). SD=Standard deviation, KD=King-Devick test.

Primary analysis

Of the 201 incidents requiring off-field medical room screening assessments, there was missing data on index test or reference standard results, in 56 (28%), leaving 145 head impact events for inclusion in an available case analysis. Ninety Four of the included events had a confirmed final clinical diagnosis of concussion, giving a target disorder prevalence of 65.0% (95% CI 56.0-72.6 %). The distribution of post head-impact event KD test times did not differ significantly between concussed and non-concussed players (median increase in KD test time from baseline +1.15 seconds, IQR -3.9 – +5.0 versus +0.7 seconds, IQR -2.8 – +6.4 respectively, p=0.62). The proportion of KD test errors was also not significantly different between concussed and non-concussed players 11.7% v 13.7%, p=0.72.

Of concussed players, 56 had an abnormal KD tests (true positives) resulting in a sensitivity of 59.6% (95% CI 49.0-72.6%). Fifty one players were reference standard negative with no confirmed concussion, of which 20 cases were classified as true negatives with normal KD results. The specificity to correctly identify players without concussion in this study group was therefore 39.2% (95%CI 25.8-53.9%). The positive and negative predictive values of the KD test were 64.4% (95% CI 53.4-74.4) and 34.5% (22.5-48.1%) respectively. Figure 2 and Table 2 summarises the performance of the KD test and presents point estimates of metrics of test accuracy with their precision. There were no obvious distinguishing features of false negative cases.

The ability of prolonged KD test times to discriminate between concussion and no concussion in players undergoing off-field screening following suspicious head impact events was not significantly different than chance; with a receiver operating characteristic curve close to the identity line and an area under the curve of 0.51 (95%CI 0.41-0.61). No optimal cut-point for prolonged KD test time was evident, with a Youden's Index of 0.11 (95%CI 0.0- 0.28) at the best empirical cut-point of 2.15 seconds. The KD test conventionally measures both time and number of errors. However, ignoring errors, sensitivity and specificity of prolonged KD time alone for concussion were 54.3% (95%CI 43.7-64.6%) and 45.1% (95% CI 31.1-59.7%) respectively. Sensitivity of errors in isolation was low (11.7%, 95%CI 6.0-20.0%), but specificity was higher at 86.3% (95%CI 73.7-94.3%).

Of the 73 incidents where players were immediately and permanently removed from play with clear signs of concussion, 54 (including 19 incidents with loss of consciousness, 4 with tonic posturing, 4 with ataxia, and 17 with confusion) underwent immediate off-field KD testing. From these 21 players (38.9% 95%CI 26.6-52.8%) passed the KD test with a quicker than baseline time and no errors. Across the combined available case sample of consecutive meaningful head impact events with the potential to cause concussion (including both incidents with clear signs of concussion and those where the consequences of the head impact event were unclear, n=199) the sensitivity and specificity of KD test for diagnosing concussion was 60.1 (95%CI 51.8-68.1) and 39.2 (95% CI 25.8-53.9) respectively.

Secondary analyses

Of the 201 incidents requiring off-field concussion screening, 21 had missing index test or reference standard data for assessment of HIA-1 screening test accuracy, giving an available case sample of 180 head impact events. Sensitivity of the HIA-1 screening test was higher than the KD test 74.8% (95% CI 65.6-82.5%), although this did not reach statistical significance (McNemar's test, p=0.08). Conversely, HIA-1 specificity was significantly better than the KD test at 91.3% (95% CI 82.0-96.7%, McNemar's test p<0.001).

The HIA-1 and KD tests, conducted in parallel, showed no agreement beyond chance (raw agreement 46.2%, Fleiss's kappa -0.08, p=0.34). Combining HIA-1 and KD test performance, into a parallel joint off-field assessment, generated a sensitivity of 92.6% (95% CI 85.9-96.7%) with a specificity of 33.3% (95%CI 20.8-47.9%, n=159). This combined sensitivity was significantly better compared to either the KD test (McNemar's test, p=<0.001) or HIA-1 test (McNemar's test, p<0.001) alone. Combined specificity was significantly lower than the HIA-1 test alone (McNemar's test, p<0.001), but did not differ significantly from the KD test used in isolation (McNemar's test, p=0.25). Separately combining KD errors or prolonged time individually with HIA-1 screening results revealed a sensitivities of 80.6% (95%CI 71.8-87.5) and 88.9% (95%CI 81.4-94.1), and specificities of 76.5% (95%CI 62.5-87.2) and 37.3% (95%CI 24.1-51.9), respectively (n=159).

Bland-Altman limits of agreement analysis revealed a mean improvement of 1.69 seconds (95%CI -3.2 to -0.1 seconds) and 95% limits of agreement of -11.4 to + 8.0 seconds between baseline and post-season tests in non-concussed players (single team, n=40). Sixty five percent (95%CI 48.4-78.6%) of these healthy players 'failed' their post season KD test with a slower time.

Scenario analyses investigating the potential influence of missing data indicated that KD performance remained lower than the HIA-1 test even when assuming a missing data pattern most favourable to KD test performance (i.e. All missing KD tests results being correct, prevalence of concussion 60%). The 'best case' sensitivity and specificity for the KD test estimates were 70.3% (95% 61.6-78.1%) and 57.5% (95% CI 45.4-69.0). Further sensitivity analyses exploring clustered data did not alter point estimates and negligibly affected 95% CI interval coverage.

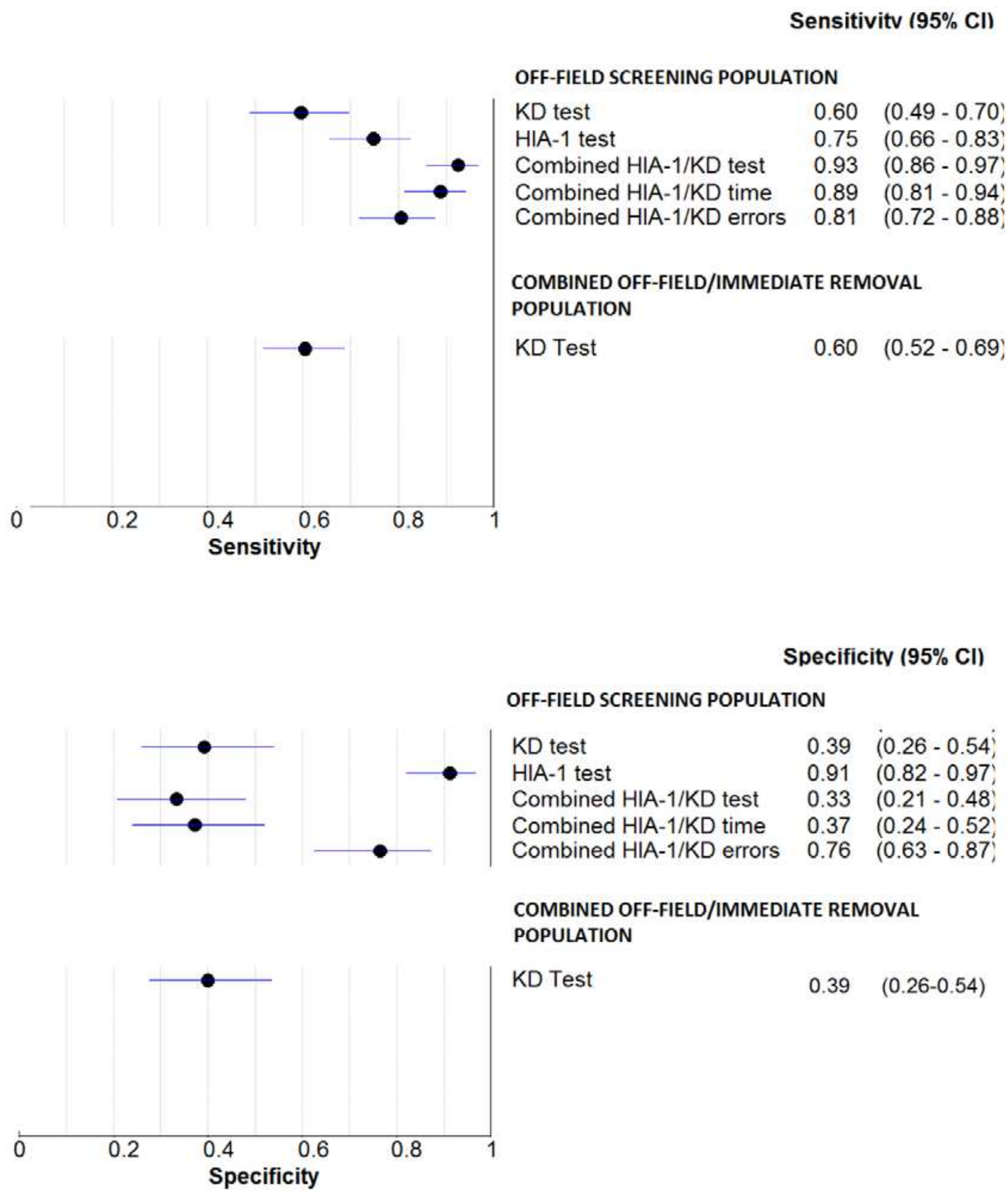

Figure 2. Forrest plots summarising sensitivity and specificity results for different screening tests and study populations

Table 2. Statistical metrics describing the accuracy of KD and HIA-1 tests

Analysis	Sample n=	Missing data	TP	FN	FP	TN	Sensitivity (%, 95% CI)	Specificity (%, 95% CI)	LR (+) (95% CI)	LR (-) (95% CI)	PPV (%, 95% CI)	NPV (%, 95% CI)
Players undergoing off-field screening (total n=201)												
KD test	145	56	56	38	31	20	59.6 (49.0-69.6)	39.2 (25.8-53.9)	0.98 (0.7-1.3)	1.03 (0.7-1.6)	64.4 (53.4-74.4)	34.5 (22.5-48.1)
HIA-1 test	180	21	83	28	6	63	74.8 (65.6-82.5)	91.3 (82.0-96.7)	8.6 (4.0-18.6)	0.3 (0.2-0.4)	93.3 (95.9-97.5)	69.2 (58.7-78.5)
Combined HIA-1/KD*	159	42	100	8	34	17	92.6 (85.9-96.7)	33.3 (20.8-47.9)	1.4 (1.1-1.7)	0.2 (0.1-0.5)	74.6 (66.4-81.7)	68.0 (46.5-85.1)
Combined HIA-1/KD time only*	159	42	96	12	32	19	88.9 (81.4-94.1)	37.3 (24.1-51.9)	1.42 (1.1-1.8)	0.3 (1.2-0.6)	75.0 (66.6-82.2)	61.3 (42.2-78.2)
Combined HIA-1/KD errors only*	159	42	87	21	12	39	80.6 (71.8-87.5)	76.5 (62.5-87.2)	3.4 (2.0-5.7)	0.3 (0.2-0.4)	87.9 (79.8-93.6)	65 (51.6-76.9)
Players immediately removed from play and those undergoing off-field screening (total n=274)												
KD test	199	75	89	59	31	20	60.1 (51.8-68.1)	39.2 (25.8-53.9)	0.99 (0.8-1.3)	1.02 (0.7-1.5)	74.2 (65.4-81.7)	25.3 (16.2-36.4)

TP=True positive, FN=False negative, FP=True positive, TN=True negative; KD=King-Devick test; HIA-1=Head Injury Assessment-1 off-field screening test; LR=Likelihood ratio; PPV=Positive predictive value; NPV= Negative predictive value; CI= Confidence interval.

*Tests applied in parallel. Abnormal result on either test denotes a positive result. Players included in the available case analysis if one test result abnormal and other result missing.

DISCUSSION

Summary of results

The KD test demonstrated a sensitivity of 59.6% and specificity of 39.2% for the presence of clinically diagnosed concussion in elite Rugby players. Given the reported prevalence, team doctors would be between 35% and 48% sure that a player did not have concussion following a negative KD test at the 95% confidence level. This performance compared less favourably with the World Rugby HIA-1 off field screening tool (sensitivity 74.8%, $p=0.08$; specificity 91.3%, $p<0.001$). Combining the KD test and the HIA-1 tool in parallel provided a multi-modal assessment with a higher sensitivity of 92.6%, but lower specificity of 33.3% than the HIA-1 test alone ($p<0.001$).

Strengths and limitations

This study is the largest prospective investigation of the KD screening test for sports related concussion published to date, and has a number of strengths. Consecutive players were recruited following suspicious head impact events avoiding the bias inherent in a diagnostic case-control study designs commonly used in previous KD studies. The index tests and reference standard were independently applied with no potential for incorporation, partial or differential verification biases. Furthermore, the reference standard was determined after serial examinations by experienced sport's medicine physicians minimising the risk of reference standard misclassification.

Conversely, there are a number of limitations which could challenge internal validity. Firstly, there were missing data on baseline, off-field tests, and reference standard results. These were predominantly secondary to non-systematic reasons such as missing baseline KD times in injured, absent, or transferred players. Furthermore, there were no distinguishing characteristics of excluded head impact events and diagnostic accuracy metrics for the HIA-1 off-field screen are consistent with previous studies. Sensitivity analyses indicated that the KD test may have improved diagnostic accuracy metrics if there were systematic reasons for missing data. However, the findings of the comparison between the KD test and HIA-1 screening tool would not be materially altered, even in a best case scenario assuming a missing data pattern most favourable to KD test performance. Taken together this suggests that the findings are robust to selection bias. Secondly, there is the possibility of diagnostic review bias. Although KD test results were not initially displayed, it was possible for team doctors to access this data later, or form a subjective opinion based on qualitative KD test performance, potentially influencing their diagnostic assessment. Unfortunately due to operational and competitive imperatives completely separate index and reference standard assessment was not possible. The KD test was conducted after the HIA-1 tool, but prior to communicating return to play decisions. There was minimal agreement between HIA-1 and KD test results, indicating that it is unlikely that interpretation of the KD test was influenced by the preceding findings, however it is possible that pending return to play decisions were perceived by players, influencing their subsequent KD test performance. Thirdly, as acknowledged in the Berlin consensus document, the diagnosis of concussion may be challenging. Misclassification of the reference standard by inaccurate clinical assessment could therefore lead to errors in the reported accuracy metrics. Furthermore reference standard misclassification could have arisen from players deliberately concealing symptoms to avoid missing games through graduated return to play protocols. Finally, the study is relatively underpowered with imprecise results.

Comparison with previous studies

Three systematic reviews have previously examined the performance of the KD test in sports-related concussion, including 10 individual studies.[4 7 20] More recently Molloy and colleagues performed a diagnostic case-control study in semi-professional Rugby Union.[21] Baseline KD results from the current study (44.3s) were consistent with the 43.8 seconds (95%CI 40.1- 47.5) reported in the recent meta-analysis by Galetta 2016.[20] The observed improvement in times between baseline KD trials was also very similar to those previously reported. Published KD accuracy results were imprecise and heterogeneous, with sensitivity estimates ranging from 53% to 100%.[4] These studies were at high or unclear risk of bias secondary to case-control study designs, test review bias, inaccurate reference standards, or inappropriate interval between index test and reference standard; making comparison of results difficult. Galetta and colleagues performed an individual patient meta-analysis using original data from a sub-set of 9 diagnostic case-control studies.[6] This pooled analysis reported a value for the sensitivity of the KD time for detecting concussion on the side-lines at 86% (96/112 concussed athletes with any worsening of baseline KD time; 95% CI: 78- 92). Pooled specificity was 90% (181/202 non-concussed control athletes with no worsening of baseline KD times, 95% CI 85-93).[20] Differences in study methodology are likely to explain the discordance with the current findings, for example diagnostic case-control studies are known to exaggerate diagnostic accuracy metrics.[9]

Interpretation of results

The source population from the top tiers of professional English Rugby should ensure that these results are generalisable throughout elite Rugby Union competitions. External validity to the elite level of other sports with different frameworks for evaluating head impact events is less certain. The extent to which video review or observable signs of concussion are used to immediately diagnose and remove players with concussion, will influence the predictive values of the KD test and could introduce spectrum effects. However, given the reported performance of the KD test these factors are unlikely to substantially alter the conclusions, and KD accuracy remained low in players removed with clear signs of concussion e.g. loss of consciousness or tonic posturing. In lower levels of competition where trained medical staff are not available, off-field concussion screening tests are not recommended, and a 'recognise and remove' strategy is more appropriate, with immediate withdrawal from play when there is any degree of suspicion of concussion.[22] Previously administered as test cards, the KD test is now currently available only as a proprietary tablet application. Preceding studies have suggested differential baseline performance between these formats, and although unlikely, it is possible that diagnostic accuracy could also vary across these configurations.[23]

The KD test requires vision, eye movements (saccades, convergence and accommodation), attention and language function. Neuronal pathways for these systems are widely distributed throughout cortical and subcortical cerebral areas, cerebellum and the brainstem; vulnerability to functional or structural damage in concussion could imply content validity for the KD test.[24] However, concussion usually manifests as a diverse range of somatic, cognitive, behavioural or emotional symptoms; and/or physical signs such as loss of consciousness and ataxia. It would therefore be surprising if a single, simple, test would be able to detect such a complex pathology.

Incorporation of the KD test within a multi-modal assessment, evaluating several clinical domains, could offer a more rational approach. With simultaneous, parallel, testing a net gain in sensitivity usually occurs at the expense of a net loss in specificity.[25] Likewise, although a favourable sensitivity of 93% was achieved when combining the KD test and the HIA-1 tool, specificity was reduced to 33%. A key concept in off-field assessment is rapid screening for a suspected concussion, rather than the definitive diagnosis of a head injury, and perfect accuracy is therefore implausible.[4] Furthermore, it is unlikely that false negative and false positive cases are equally important. The competitive, clinical, legal, reputational and health implications of varying sensitivities and specificities of different combinations of tests therefore requires careful consideration.

Conclusions

This study suggests that the KD test may have limited accuracy as a remove-from-play sideline screening test for concussion. Combination of the KD test within a multi-modality screening assessment may provide satisfactory sensitivity for identifying concussion, but at the expense of lower specificity. Further research is necessary to confirm these findings and identify alternative effective concussion screening strategies.

REFERENCES

1. McCrea HJ, Perrine K, Niogi S, et al. Concussion in sports. Sports health 2013;**5**(2):160-4

2. Helmy A, Agarwal M, Hutchinson PJ. Concussion and sport. BMJ (Clinical research ed.) 2013;**347**:f5748

3. 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

4. Patricios J, Fuller GW, Ellenbogen R, et al. What are the critical elements of sideline screening that can be used to establish the diagnosis of concussion? A systematic review. British journal of sports medicine 2017

5. Ciuffreda KJ, Kapoor N, Rutner D, et al. Occurrence of oculomotor dysfunctions in acquired brain injury: a retrospective analysis. Optometry (St. Louis, Mo.) 2007;**78**(4):155-61

6. Galetta KM, Liu M, Leong DF, et al. The King-Devick test of rapid number naming for concussion detection: meta-analysis and systematic review of the literature. Concussion 2015

7. Hunt AW, Mah K, Reed N, et al. Oculomotor-Based Vision Assessment in Mild Traumatic Brain Injury: A Systematic Review. The Journal of head trauma rehabilitation 2015

8. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Annals of internal medicine 2003;**138**(1):W1-12

9. 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

10. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Journal of clinical epidemiology 2011;**64**(1):96-106

11. Fuller CW, Fuller GW, Kemp SP, et al. Evaluation of World Rugby's concussion management process: results from Rugby World Cup 2015. British journal of sports medicine 2016

12. Technologies K-D. King-Devick Test. Secondary King-Devick Test 2017. https://kingdevicktest.com/.

13. CSx. Secondary CSx 2017. http://csx.co.nz/.

14. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology 2008;**31**(4):466-75

15. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation 2007;**115**(5):654-7

16. Youden WJ. Index for rating diagnostic tests. Cancer 1950;**3**(1):32-5

17. Bland JM, Altman DG. Measuring agreement in method comparison studies. Statistical methods in medical research 1999;**8**(2):135-60

18. Genders TS, Spronk S, Stijnen T, et al. Methods for calculating sensitivity and specificity of clustered data: a tutorial. Radiology 2012;**265**(3):910-6

19. Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine 1996;**3**(9):895-900

20. Galetta KM LM, Leong DF , Ventura RE , Galetta SL, Balcer LJ. The King-Devick test of rapid number naming for concussion detection: meta-analysis and systematic review of the literature. Concussion 2016;**1**(2):null

21. Molloy JH, Murphy I, Gissane C. The King-Devick (K-D) test and concussion diagnosis in semi-professional rugby union players. Journal of science and medicine in sport / Sports Medicine Australia 2017;**20**(8):708-11

22. Echemendia RJ, Meeuwisse W, McCrory P, et al. The Concussion Recognition Tool 5th Edition (CRT5): Background and rationale. British journal of sports medicine 2017;**51**(11):870-71

23. Raynowska J, Hasanaj L, Silverio A, et al. Agreement of the Spiral-Bound and Computerized Tablet Versions of the King-Devick Test of Rapid Number Naming for Sports Related Concussion (I13.001). Neurology 2016;**86**(16 Supplement)

24. Ventura RE, Balcer LJ, Galetta SL. The neuro-ophthalmology of head trauma. The Lancet. Neurology 2014;**13**(10):1006-16

25. Pepe MS, Thompson ML. Combining diagnostic test results to increase accuracy. Biostatistics (Oxford, England) 2000;**1**(2):123-40

FIGURE 1 LEGEND

Figure 1. Derivation of study participants in primary analysis

FIGURE 2 LEGEND

Figure 2. Forrest plots summarising sensitivity and specificity results for different screening tests and study populations

ACKNOWLEDGEMENTS

The authors would like to acknowledge the invaluable support provided by team doctors and physiotherapists from the participating teams who provided KD and HIA data used within this study. The study was funded by the Rugby Football Union and supported by King-Devick Technologies.

COMPETING INTERESTS

There are no competing interest.

CONTRIBUTORSHIP

Dr Gordon Fuller conceived and designed the study in collaboration Dr Matt Cross, Dr Simon Kemp, and Professor Keith Stokes. Dr Matt Cross coordinated data collection and was responsible for data management. Dr Gordon Fuller processed, analysed and interpreted the data, and wrote and prepared the manuscript for publication. Analyses were checked for accuracy by Dr Matt Cross. All authors critically revised the manuscript for important intellectual content and gave final approval of the version to be published.

FUNDING INFORMATION

The study was funded by the Rugby Football Union and supported by King-Devick Technologies.

ETHICAL APPROVAL INFORMATION

The study protocol received ethical approval from the University of Bath.

DATA SHARING STATEMENT

No data sharing agreements in place at present.

What are the new findings?

- The King-Devick test has been promoted as a remove-from-play sideline screening test for sports related concussion
- This is the largest prospective investigation of the KD screening test for sports related concussion published to date
- The diagnostic accuracy design with novel inclusion of consecutive head impact events with the potential to cause concussion maximises internal validity
- The KD test demonstrated limited accuracy as a stand-alone off-field screening test for concussion.

How might it impact on clinical practice in the near future?

- The King-Devick test should be used with discretion as a stand-alone remove-from-play sideline screening test.
- Combination of the KD test within a multi-modality screening assessment may improve sensitivity, but at the expense of lower specificity.