
Article:
Fuller, G.W., Kemp, S.P.T. and Raftery, M. (2017) The accuracy and reproducibility of video assessment in the pitch-side management of concussion in elite rugby. Journal of Science and Medicine in Sport, 20 (3). pp. 246-249. ISSN 1440-2440

https://doi.org/10.1016/j.jsams.2016.07.008

The accuracy and reproducibility of video assessment in the pitch-side management of concussion in elite rugby

Abstract

Objectives: To investigate the accuracy and reliability of side-line video review of head impact events to aid identification of concussion in elite sport

Design: Diagnostic accuracy and inter-rater agreement study

Methods: Immediate care, match day and team doctors involved in the 2015 Rugby Union World Cup viewed 20 video clips showing broadcaster's footage of head impact events occurring during elite Rugby matches. Subjects subsequently recorded whether any criteria warranting permanent removal from play or medical room head injury assessment were present. The accuracy of these ratings were compared to consensus expert opinion by calculating mean sensitivity and specificity across raters. The reproducibility of doctor's decisions was additionally assessed using raw agreement and Gwets AC1 chance corrected agreement coefficient.

Results: Forty rugby medicine doctors were included in the study. Compared to the expert reference standard overall sensitivity and specificity of doctors decisions were 77.5% (95%CI 73.1-81.5%) and 53.3% (95%CI 48.2-58.2%) respectively. Overall there was raw agreement of 67.8% (95% CI 57.9-77.7%) between doctors across all video clips. Chance corrected Gwets AC1 agreement coefficient was 0.39 (95% CI 0.17-0.62), indicating fair agreement.

Conclusions: Rugby World Cup doctors' demonstrated moderate accuracy and fair reproducibility in head injury event decision making when assessing video clips of head impact events. The use of real-

time video may improve the identification, decision making and management of concussion in elite sports.

Key Words: Brain concussion, Rugby, Video analysis, Reproducibility of results, Sensitivity and specificity

Introduction

Concussion is a common injury in elite sport , with concussion rates of 10.5 concussions per 1000 player-match-hours reported in professional Rugby Union.[1] Continued participation following concussion may risk further injuries, exacerbate concussive symptoms, or predispose to second impact syndrome.[2]

The in-game assessment and management of head impact events has been a contentious issue.[3] Media attention has focused on high profile incidents where players clearly demonstrating symptoms or signs of concussion have controversially been allowed to continue playing by the player's medical team.[4] Conversely, inaccurate evaluation and incorrect withdrawal of non-injured players could have detrimental effects on team performance.

Elite adult Rugby Union, in common with other collision sports, has introduced systems to improve the in-game management of head impact events with the potential to cause concussion. The World Rugby Head Injury Assessment (HIA) process (formally the pitch-side concussion assessment, or PSCA, process) was developed by a working group of international concussion specialists and was informed by expert opinion, recent consensus statements, and review of scientific literature. The HIA process, recently updated for the 2015 Rugby World Cup, has been described previously and is summarised in the web appendix.[5]

Side-line video review of head impact events has been introduced in several elite sports, including American Football, Rugby Union and Ice Hockey, to identify possible concussions and inform removal from play decisions.[6] Video has three main roles in the World Rugby HIA process – identification of suspicious head impact events that warrant off-field assessment, identification of signs that confirm a concussion leading to a permanent removal, and a final video review prior to a decision that clears a player to return to play.

Despite the introduction of video review into side-line concussion assessment by many sports little is known about the accuracy or reliability of these assessments. This study aimed to characterise the performance of rugby medicine doctors when evaluating video clips of head impact events. Specific objectives were to determine the accuracy of match day and team doctors opinions against an expert reference standard, and to describe the reproducibility of video clip assessments in the World Rugby HIA process.

Methods

This was a cross-sectional diagnostic accuracy and inter-rater agreement study investigating the assessment of head impact event video clips in elite Rugby. The study population consisted of all immediate care, match day and team doctors involved in game day concussion management during the 2015 Rugby World Cup. All participants were rugby medicine doctors, experienced in providing pitch-side medical care in elite Rugby, who had completed World Rugby's on-line course in concussion management, and accreditation in immediate care in Rugby.[7-9]

The study was performed immediately prior to the October 2015 Rugby World Cup. In preparation for this tournament a face-to-face training session was conducted to re-inforce HIA protocols, including the identification of suspicious head impact events, criteria for immediate and permanent removal, and indications for a medical room head injury assessment. This study was performed immediately after the training session.

Video recordings of all Elite Rugby matches are reviewed for head impact events by World Rugby as part of the HIA process. A selection of 20 video clips were identified *a priori* by World Rugby's Chief Medical Officer. The clips included a range of head injury severities, in addition to trivial head impact events with no evidence or suspicion of concussion. Twenty clips were chosen to provide a sufficient range of clinical presentations within realistic time constraints.

Videos showed the incident in real time with the relevant player identified by jersey colour and number. Slow motion replays were then shown from various angles according to the availability of broadcaster's footage. Sound and commentary was removed. Clips lasted a mean of 61.5 seconds (range 21-129). The resulting clips therefore mimicked the availability of video within the HIA process as much as possible.

The clips were then evaluated by a team of 7 international concussion experts in a web-based consensus process. All experts were established academics in the field of sports concussion and held national-level

positions in Sports Medicine. Expert opinion on each video clip was determined by consensus. In cases where a majority consensus of greater than 70% was not reached the World Rugby Chief Medical Officer held a casting vote.

Following the training session, clips were then shown to Rugby World Cup doctors with each participant rating the clips individually without conferring. A single replay was available on request. The first 10 clips simulated the side-line identification of suspicious head injury events. Participants were asked whether the clip showed any HIA process criteria for removal from play (either permanent removal or medical room head injury assessment), or if the head impact event was not significant. The remaining 10 videos simulated post-removal medical room video review and asked participants to indicate whether permanent removal or further head injury assessment was indicated. For each clip participants were also asked to identify which HIA process criterion for permanent removal or head injury assessment was the most important in influencing their decision. HIA process criteria for permanent removal and head injury assessment are presented in the web appendix.

The statistical analysis proceeded in 2 stages. Firstly, the accuracy of match and team doctors decisions were compared to the consensus expert opinion. Participant responses were pooled across video clips and classified according to the reference standard. Mean sensitivity and specificity was then calculated across raters. Secondly, the reproducibility of doctor's decisions was assessed. Raw agreement was initially determined, but to account for the fact that some agreement would be expected even if the participants were guessing, a chance corrected agreement coefficient was also calculated. Gwets AC1 coefficient was calculated as the preferred measure of agreement due to theoretical considerations of increased stability, robustness to marginal probabilities, and lack of dependence on rating prevalence.[10-12] Landis and Koch's benchmark values were used to interpret the magnitude of agreement coefficients with: 0–0.20 indicating slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1 almost perfect agreement.[13] Results were calculated with their 95% confidence intervals (CI), both

overall, and separately for each stage of the HIA video process. Statistical analyses were carried out in Stata version 13 (StataCorp, College Station, USA) and AgreeStat 2011.3 (advanced Analytics, Gaithersburg, MD, USA).

A study protocol defining an *a priori* analysis plan received ethical approval from the University of Sheffield. Prior consent was available for the use of each video clip and individual consent was obtained from all participants prior to the study. Research funding was provided by the World Rugby. Reporting is in accordance with EQUATOR guidelines for reliability studies.[14]

Results

All forty immediate care, match day and team doctors involved in the 2015 Rugby World Cup participated in the study, rating all videos with no missing data.

Expert consensus was reached on 16 of the 20 videos. A casting vote from the World Rugby Chief Medical Officer was therefore necessary for 4 videos where agreement amongst experts ranged from 43-57%.

Compared to the expert reference standard overall sensitivity and specificity of doctors decisions were 77.5% (95%CI 73.1-81.5%) and 53.3% (95%CI 48.2-58.2%) respectively. For the side-line identification of suspicious head impact events sensitivity was 87.0% (95%CI 81.5-91.3%) and specificity: 39.0% (95%CI 32-46%). Sensitivity was relatively lower for medical room video review, determining whether permanent removal or head injury assessment was appropriate, at 68.0% (95%CI 61.1-74.4%), but with higher specificity of 67.5% (95%CI 60.5-73.9%). 2x2 contingency tables, pooling participants responses across relevant video clips and classifying them according to the reference standard, are presented separately for the accuracy of side-line and medical room video review in Tables 1 and 2.

Overall there was raw agreement of 67.8% (95% CI 57.9-77.7) between doctors across all video clips. Chance corrected Gwets AC1 agreement coefficient was 0.39 (95% CI 0.17-0.62), indicating fair agreement. Figure 1 presents the level of agreement across each video clip.

Agreement was relatively higher for the side-line identification of suspicious head impact events with raw agreement of 71.2% (95%CI 54.5-88.0%), and a Gwets AC1 coefficient of 0.53 (95%CI 0.14-0.93), denoting moderate agreement. Less agreement was observed for medical room video review determining whether permanent removal or head injury assessment was indicated. Raw agreement was 64.3% (95%CI 50.1-78.0), with a Gwets AC1 coefficient of 0.29 (95%CI 0.015-0.56) demonstrating fair agreement.

There was similarly lower agreement on which of the HIA process criterion was most important in each scenario, with a mean raw agreement between participants of 30.6% (95%CI 25.9-35.3%) across the clips, and a Gwets AC1 coefficient of 0.26 (95%CI 0.2-0.31). No discernible association was evident between the specific criterion and the degree of agreement. Notably, only one of the five videos with the lowest inter-rater agreement corresponded to a clip where a firm expert consensus was not reached.

Discussion

Rugby World Cup doctors' demonstrated moderate accuracy and fair reproducibility when assessing video clips of head impact events. There was high sensitivity (87%) and moderate agreement (Gwets AC1 coefficient 0.53) when detecting suspicious head injury events. Sensitivity for identifying permanent removal criteria in the medical room was lower (68%) with fair reproducibility evident (Gwets AC1 coefficient 0.29).

There are a number of limitations that could adversely affect the validity of the results. Firstly, it is questionable whether a true reference standard can be defined for each video clip, challenging calculation of diagnostic accuracy metrics. Video clips of head impact events are inherently subjective as illustrated by the minority of video clips where considerable disagreement was present amongst the expert panel. However, derivation of a reference standard through a formal consensus process is an established method to obtain a credible gold standard in the absence of an objective reality.[15] Secondly, although comparing favourably to other published reproducibility studies, the sample size is relatively low, resulting in imprecise results consistent with either poor or moderate agreement. These findings should therefore be considered as preliminary, requiring confirmation in larger studies. Thirdly, although every effort was taken to imitate the HIA video review process, controlled study procedures cannot fully simulate the real-life application of video technology. On-field observations, competitive pressures or crowd distractions could all conceivably influence actual performance of video review. Finally, we pre-specified the relatively liberal Landis and Koch scale for benchmarking agreement coefficients. Although well established and widely used, this may overstate agreement compared to other benchmarks e.g. Fleiss' or McHugh's proposed scales.[16, 17]

The accuracy and reproducibility of video review in sports concussion assessment has not been extensively studied. Makdissi and Davis (2016) examined video clips of 82 concussion cases sustained

during professional Australian Football matches, demonstrating a similar level of fair to poor inter-rater agreement for a range of clinical signs indicating head injury (kappa statistics 0.13 to 0.37).[18] Gardner and colleagues (2016) evaluated 162 video clips where players were temporarily substituted for assessment of suspected concussion during the 2014 Australian National Rugby League season.[19] Reported agreement between two raters was relatively higher, both overall (kappa 0.6, 95% CI 0.56-0.64) and for specific observable signs (kappa statistics 0.36 – 0.7). However, the comparability of these results, with videos scrutinised by two researchers and multiple replays allowed, is uncertain.

In the World Rugby HIA process, independent match-day doctors can supplement side-line observation with video reviews of incidents to identify any suspicious head impact events. The finding of favourable sensitivity and moderate agreement for this stage of side-line recognition is reassuring, suggesting that a real-time video review process integrated into the HIA process has the potential to identify head injuries or incidents requiring removal from play or further assessment. The high media profile of the Rugby World Cup and the potential consequences to practitioners for not removing players who are subsequently deemed to be concussed may be associated with a reduced threshold for further off-pitch evaluation. This could explain the lower specificity for identifying non-significant head impact events.

Once a suspicious head impact event has been identified, the HIA process includes a second video review undertaken off-field in the medical room with the match-day and team doctors. This decides whether the player should be immediately and permanently removed (without a further head injury assessment) or the player should have a HIA1 screening head injury assessment. The less favourable accuracy and reproducibility apparent for medical room video review may be of lesser concern, as the player has already been removed from the field of play. The minority of players who were not identified for permanent removal (video false negative) will all undergo a head injury assessment and a final further review of the video as additional safeguard prior to any return to play. Previous research has demonstrated the high sensitivity of the head injury assessment as a triage process suggesting it is

unlikely that such cases would return to play.[5] Furthermore, permanent removal of play for a minority of incidents where a HIA screening assessment was indicated (video false positive) may be acceptable in the context of a 'safety first' approach.

The lack of reproducibility on the single most prominent HIA process criterion present in each scenario is unsurprising as agreement tends to decrease as the number of rating categories increases. Furthermore, more than one HIA process criteria was evident in several clips. Ultimately, the final decision may matter more than the exact underlying reason. Conversely, the lack of conformity in decision making highlights the importance of robust training of doctor in the video review process, specifically focusing on definitions of each assessment criterion.

The sample of World Cup match-day and team doctors should ensure good external validity. However, it is possible that accuracy and agreement could be lower in other tournaments utilising less experienced medical staff, if specific concussion education is not provided to video-reviewing medical personnel, or if the effects of concussion management training wane. Furthermore, an advanced 'Hawk-Eye/Smart Replay' video system was recently integrated into the HIA process during certain competitions (including the RWC).[20] This technology allows the user to fully control replays from 20-30 camera feeds, which might be expected to increase the sensitivity of head impact evaluation compared to the standard video review with limited broadcaster feeds investigated here. Generalizability to other sports implementing differing systems for reviewing video, using non-medical practitioners and exhibiting other patterns of play is less certain.

Conclusion

Elite contact and collision sports are played at a fast pace in a disorganised environment, where the view of medical staff is often obscured, challenging the evaluation of head impact events. This study suggests that Rugby medicine doctors who have received specific training show moderate accuracy and fair

reproducibility in head injury event decision making when reviewing video footage. Despite these encouraging results it is likely that further improvement in identification and decision making could be achieved with development of education resources and greater operational clarification of permanent removal from play criteria.

Practical Implications

- Side-line video review is an additional modality to help in the identification and assessment of potential concussive episodes
- Sports physicians show moderate accuracy and fair reliability when reviewing video clips of head impact events
- Introduction of side-line video review could improve the management of concussion in elite sport

Acknowledgments

The authors would like to match day doctors, team doctors and intermediate care doctors participating in the 2015 Rugby World Cup for their assistance with the study. Funding was received by World Rugby.

References

1. Moore IS, Ranson C, Mathema P. Injury Risk in International Rugby Union: Three-Year Injury Surveillance of the Welsh National Team. *Orthopaedic journal of sports medicine.* 2015; 3(7):2325967115596194.
2. Brolinson PG. Management of sport-related concussion: a review. *Clinical journal of sport medicine : official journal of the Canadian Academy of Sport Medicine.* 2014; 24(1):89-90.
3. Helmy A, Agarwal M, Hutchinson PJ. Concussion and sport. *BMJ (Clinical research ed.).* 2013; 347:f5748.
4. Cusimano MD, Sharma B, Lawrence DW, Ilie G, Silverberg S, Jones R. Trends in North American newspaper reporting of brain injury in ice hockey. *PLoS One.* 2013; 8(4):e61865.
5. 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):529-535.
6. NFL Head NaSC. NFL Head, Neck and Spine Committee's Protocols Regarding Diagnosis and Management of Concussion. USA: National Football League; 2015.
7. Rugby W. Concussion Management for Doctors and Healthcare Professionals. http://playerwelfare.worldrugby.org/?documentid=module&module=22. Accessed 2016 8th February 2016.
8. Rugby W. Medical Protocols for Match Day Medical Staff. http://playerwelfare.worldrugby.org/?documentid=module&module=23. Accessed 2016 8th February 2016.
9. Rugby W. Concussion Management for Elite Level Match Day Medical Staff. http://playerwelfare.worldrugby.org/?documentid=module&module=24. Accessed 2016 8th February 2016.

10. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC medical research methodology.* 2013; 13:61.
11. Gwet K. *Handbook of Inter-Rater Reliability: How to Estimate the Level of Agreement Between Two or Multiple Raters.*, Gaithersburg, MD, STATAXIS Publishing Company; 2001.
12. Gwet K. Inter-Rater Reliability: Dependency on Trait Prevalence and Marginal Homogeneity. . *Statistical Methods for Inter-Rater Reliability Assessment.* 2002:1-9.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33(1):159-174.
14. 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.
15. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health technology assessment (Winchester, England).* 2007; 11(50):iii, ix-51.
16. McHugh ML. Interrater reliability: the kappa statistic. *Biochemia medica.* 2012; 22(3):276-282.
17. Fleiss JL, Levin BA, Paik MC. *Statistical methods for rates and proportions,* 3rd ed. / Joseph L. Fleiss, Bruce Levin, Myunghee Cho Paik. ed, Hoboken, N.J. ; [Chichester], Wiley-Interscience; 2003.
18. Makdissi M, Davis G. The reliability and validity of video analysis for the assessment of the clinical signs of concussion in Australian football. *Journal of science and medicine in sport / Sports Medicine Australia.* 2016.
19. Gardner AJ, Iverson GL, Stanwell P, Moore T, Ellis J, Levi CR. A Video Analysis of Use of the New 'Concussion Interchange Rule' in the National Rugby League. *Int J Sports Med.* 2016; 37(04):267-273.

20. Hawk-Eye. Hawk-Eye in Rugby Union. http://www.hawkeyeinnovations.co.uk/sports/rugby-union.

Table 1. 2x2 contingency table summarising the accuracy of Rugby World Cup doctors' for suspicious head impact events

	Reference standard		Total
Doctors' assessment	Suspicious head impact event*	Not suspicious	
Suspicious head impact event	174 [TP]	122 [FP]	296
Not suspicious	26 [FN]	78 [TN]	104
Total	200	200	**400**

*A suspicious head impact event exhibited a HIA process criterion for permanent removal or off-pitch head injury assessment. †Non-suspicious head impact events exhibited no HIA process criteria. TP: True positive, FP: False positive, FN: False negative, TN: True negative

Table 2. 2x2 contingency table summarising the accuracy of Rugby World Cup doctors' for permanent removal criteria

	Reference standard		**Total**
Doctors' assessment	Permanent removal*	Head injury assessment†	
Permanent removal	136 [TP]	65 [FP]	201
Head injury assessment	64 [TN]	135 [TN]	199
Total	200	200	**400**

*****One or more HIA permanent removal criteria present (see web appendix) †No HIA permanent removal criteria present and one or more medical room Head Injury Assessment criteria present.

Figure 1. Overall agreement across each of the 20 video clips.

Light grey shading represent the percentage majority opinion. Dark grey shading represents the percentage minority opinion. The dashed black line indicates the 50% agreement that would be expected if participants were rating video clips randomly.