DR. DEXTER SEOW (Orcid ID : 0000-0003-1769-9244)

Article type : Review

Prediction models for musculoskeletal injuries in professional sporting activities: a systematic review

Running head: Review of injury prediction models

Dexter Seow[1], MB BCh
Ian Graham[1], PhD, MPhys
Andrew Massey[1], MB BCh BAO, FFSEM, MRCGP, MSc, BSc (Hons)
[1]Liverpool Football Club, Liverpool, United Kingdom

Address Correspondence to:
Dexter Seow
Liverpool Football Club
Deysbrook Lane, L12 8SY,
Liverpool, United Kingdom
Email: dexterseow@rcsi.ie
Phone: +44 7712 325376

ACKNOWLEDGEMENTS

Nil.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/TSM2.181

ABSTRACT

The purpose of this systematic review was two-fold: (1) identify prediction models for musculoskeletal injuries during the participation of professional sporting activities and (2) evaluate these models by their predictive performances. A systematic review of the PubMed and Embase databases was performed using specific search terms selected according to the PRISMA guidelines. Ten studies met the eligibility criteria and were included. The most commonly employed data component for data preprocessing was body composition data (most commonly, body mass and height), followed by player profile (most commonly, age followed by position). The most common machine learning technique for data processing was the decision tree, followed by logistic regression. The median AUC of the best performing models indicated per study was 0.75 (0.16), median sensitivity/recall was 0.78 (0.15), median specificity was 0.81 (0.27) and median precision was 0.53 (0.13). The performance of prediction models in the literature has been poor, caused by a fundamental difficulty in discovering real effects in small sample sizes with low injury rates. A better understanding of how training and game exposure is associated with the data components for data preprocessing and ultimately associated with injury is vital for the future development of robust injury prediction models.

Keywords: *injury forecast, injury prediction, injury projection*

INTRODUCTION

Musculoskeletal injuries are common occurrences in sports.[1,2,3] Injuries can be debilitating,[4,5] alter team structure leading to reduced cohesion between players[6] and add costly expenses.[7,8] Prediction models can, in principle, permit the ability to ameliorate these complications, providing an advantage for players (health/mental state), coaches (team selection) and club (minimization of costs needed for treatment, rehabilitation and time loss attributed to player's injury). Therefore, the development of prediction models is attractive to the clinician to detect high-risk individuals.

The development of prediction models in sport has emerged in recent years, with this concept being relatively new. Prediction models apply data mining and analysis techniques (data preprocessing, processing, sampling techniques) to machine learning to estimate injury probability.[9,10] These statistical approaches have otherwise been applied to previous medical diagnosis studies with excellent results indicated.[11,12]

The construction of a high-performance prediction model in sport has yet to be established, predominantly due to the presence of uncertainty in desired prediction model components and construct. A systematic review of the literature can provide valuable learning points for future prediction modelling. Therefore, the purpose of the study was two-fold: (1) identify prediction models for musculoskeletal injuries during the participation of professional sporting activities and (2) evaluate these models by their predictive performances.

METHODS

Study design, search strategy and study identification

A systematic review of the PubMed and Embase databases was performed by two authors (D.S. and A.M.) using specific search terms selected according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[13] from inception to December 4, 2019. Titles, abstracts and full-texts were screened using specific eligibility criteria and citations of full-text studies for review were further screened for eligible studies. Studies were included by the consensus of the two authors with any disagreement resolved mutually.

Search terms

The search terms were: (injury OR injuries) AND (prediction OR predictor OR forecast OR projection) AND sport. No field restrictions were applied.

Eligibility criteria

The inclusion criteria were: 1) studies trialled prediction model(s) for musculoskeletal injuries resulting from professional sporting activities, 2) full-text studies and 3) written in English. The exclusion criteria were: 1) animal studies, 2) case reports, 3) study populations with chronic and/or systemic disease, 4) diagnostic studies, 5) in vitro studies, 6) reviews and 7) therapeutic studies.

Definition of a musculoskeletal injury

A musculoskeletal injury was defined as: "any physical complaint sustained by a player that results from a sports match or training, irrespective of the need for medical attention or time loss from sporting activities." This definition is a modified version of the original by Fuller et al.[14] The modification replaced "football" with "sports/sporting" in the definition to account for all sporting activities. The definition accounts for both contact and non-contact injuries.

Data extraction and categorisation

Data were extracted onto Microsoft® Excel datasheet version 16.16.1 (Microsoft® Excel for Mac, Redmond, WA). Data extracted specific to study and player characteristics were: study design, prediction model focus, sport, participation level, population (n), number of players injured, number of injuries, mean age, mean height, mean body mass and follow-up. Data extracted specific to prediction models were three-fold:

- Data preprocessing: data component collection timing and data components.
- Data processing: machine learning technique, sampling technique and prediction window.
- Performance of best performing model indicated per study: true positive (n, correctly identified injuries, TP), true negative (n, correctly identified non-injuries, TN), false positive (n, incorrectly identified injuries, FP), false negative (n, incorrectly identified non-injuries, FN), specificity (TN/(FP+TN)), sensitivity/recall (TP/(TP+FN)), precision (TP/(TP+FP)) and area under the curve (AUC).

Assessment of evidence

Assessment of evidence was two-fold: 1) level of evidence and 2) risk of bias. The level of evidence was assessed based on the criteria from The Oxford Centre for Evidence-based Medicine.[15] The risk of bias was assessed using the Newcastle-Ottawa Scale. The Newcastle-Ottawa Scale is a 9-point/star scale based on an 8-item checklist of three domains (selection, comparability and outcome or exposure), designed to examine the methodological risk of bias in non-randomised clinical studies.[16] For each item assessed, the best-suited responses were selected. The responses that were deemed of a high risk of bias were awarded a star. Studies with 7 to 9 stars were graded to have a risk of bias as 'very low,' 5 to 6 stars as 'low,' 4 stars as 'satisfactory' and 0 to 3 stars as 'high.'

Statistical analysis

All statistical analysis was performed using R version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria). Descriptive statistics were calculated for all continuous and categorical variables. Continuous variables were reported as mean ± standard deviation for normally distributed data and as median with interquartile range (IQR) for non-normally distributed data. Categorical variables were reported as frequencies with percentages.

RESULTS

Literature search

A literature search based on the search strategy yielded 2589 studies for review. The removal of duplicates resulted in 2037 studies screened by title and abstract, with 26 studies selected for full-text review. Two studies were identified from the citations of the full-text studies reviewed.[18,22] The literature search resulted in 10 studies that met the eligibility criteria and therefore, included in the systematic review (Figure 1).[17-26]

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses Flow Diagram
[Insert Figure 1]

Study and player characteristics

The included studies were published between 2010 to 2019, with 7 studies of the level of evidence II (70%)[17,19-22,25,26] and 3 studies of the level of evidence III (30%).[18,23,24] The risk of bias assessed by the Newcastle-Ottawa Scale indicated that all 10 studies of 'very low' risk of bias. Item-specific scores of the Newcastle-Ottawa Scale are illustrated in Appendix 1. The definition of a musculoskeletal injury was reported to be specifically a 'time-loss' injury in 7 studies.[17,19-21,24-26] The remaining 3 studies did not state if the injury definition was exclusive to time-loss injuries.[18,22,23] The total population (n) was 1009 players; Talukder et al. did not report population (n)[18] and Ruddy et al. reported player overlaps between seasons.[20] The prediction models were focused on hamstring injuries in 3 studies,[20,22,26] non-contact injuries in 2 studies,[17,21] all injuries in 1 study,[18] lower-body injuries in 1 study,[25] lower body muscular injuries in 1 study,[19] lower body non-contact injuries in 1 study[24] and long bone injuries in 1 study.[23] Sports were involved at the professional level in 9 studies[17-24,26] and in special forces of the military in 1 study.[25] Military activities were included in this systematic review as their required high physical demands were considered to be comparable to that of professional sporting activities.[25] Sports at the profession level included Australian football (3 studies),[20,22,24] soccer (3 studies),[19,21,26] basketball (1 study),[18] endurance sport (1 study),[23] handball (1 study),[19] rugby league (1 study);[17] López-Valenciano et al. involved players in both soccer (n = 98) and handball (n = 34).[19] The median mean age was 24.4 (2.4) years, median mean height was 185.4 (7.9) cm and median mean body mass was 87.3 (3.4) kg. The distribution of sex was all-males in 8 studies.[17-21,23,24,26] The remaining 2 studies did not report their distribution of sex.[22,25] The median mean follow-up was 9.0 (9.9) months.

Table 1. Study and player characteristics

[Insert Table 1]

Data preprocessing

Data components employed for data preprocessing included body composition data, cardiopulmonary assessments, injury history, neuromuscular assessments, player profile, players statistics, psychological assessments and workload data. The most commonly employed data component was body composition data (7 studies),[18-21,23,25,26] followed by player profile, i.e. age/position (6 studies)[18-22,26] then injury history[19-21,25,26] and workload data (5 studies each)[17,18,21,22,24].

The most commonly employed subcomponents of body composition data were body mass and height. DEXA variables were only employed in a single study.[23] The most commonly employed subcomponent of the player profile was age, followed by position. Injury history referred to any or specific injury history. This included injury history in the previous season, anterior cruciate ligament injury history or hamstring injury in previous 12 months. The most commonly employed subcomponent of workload data was both the modified rating of perceived exertion scale and total distance followed by high speed running, minutes of play and number of games played.

Table 2. Data preprocessing
[Insert Table 2]

Data processing

Data processing involved the use of machine learning techniques, sampling technique and cross-validation to generate a model to predict musculoskeletal injuries within a specific time window. The most commonly constructed machine learning technique was a decision tree followed by logistic regression, then both random forest and supporting vector machine. Sampling techniques were reported in 6 studies with the synthetic minority oversampling technique most commonly employed.[18-22,26] Cross-validation was reported in 9 studies with 10-fold cross-validation, the most common method. The most common prediction window was the season end (6 studies).[17,19-21,24,26]

Table 3. Data processing
[Insert Table 3]

Predictive performances

The most commonly reported measure of predictive performance was AUC (9 studies)[18-26] followed by sensitivity/recall (5 studies)[17-19,21,26] then both specificity[17-19,26] and precision (4 studies each).[17,19,21,25] The median AUC of the best performing models indicated per study was 0.75 (0.16), median sensitivity/recall was 0.78 (0.15), median specificity was 0.81 (0.27) and median precision was 0.53 (0.13). Talukder et al. reported the prediction model with the most superior AUC,[18] whereas Gabbett et al. reported the prediction model with the most superior sensitivity/recall, specificity and precision.[17]

Table 4. Performance of best performing model indicated per study
[Insert Table 4]

DISCUSSION

The performance of prediction models in the literature has been poor. The included studies were of a level of evidence 2 and 3 that were graded of 'very low' risk of bias according to the Newcastle-Ottawa Scale. The prediction models contended with the variability of small number statistics. Injuries are rare and can occur multiple times within the one individual and therefore, are Poisson distributed. The performance of prediction models in the current literature has been poor, with the underlying difficulty that is to discover the real effects in small sample sizes and low injury rates. Consider Gabbett's et al.'s data (2010) of 2.3 injuries per 1,000 hours and 13,103 hours of exposure.[17] Assume that half those hours are spent in normal training and a half in "risky" training. The injury risk of "risky" training would have been theoretically needed to double for its effect to be reliably detected.

Baseline testing versus monitoring approaches

The included studies can be split into two different approaches: "baseline testing" and "monitoring." López-Valenciano et al., Ruddy et al., Carbuhn et al., Connaboy et al. and Ayala et al. adopted the baseline testing approach,[19,20,23,25,26] while Gabbett, Talukder et al., Rossi et al., Carey et al. and Colby et al. adopted the monitoring approach.[17,18,21,22,24] In the baseline testing approach, players were tested before a season starts at a singular time point and injuries were recorded throughout the season. In the monitoring approach, data was to be collected daily or weekly with a player's injury risk able to vary based on their recent data. In principle, the monitoring approach can include variables that are typically measured in the baseline testing approach, but this was not done in studies that employed the monitoring approach. Many of these measures can be considered temporal and therefore, a one-off baseline testing score may not be a true representation months later.

The five studies that employed the baseline testing approach did not have strong predictive performance. In the baseline testing approach, training and game exposure was not tracked or reported. Differences in exposure and many other possible factors (i.e. fitness level, strength, fatigue) inherently generate differences in injury occurrence that would drown out the signals that had existed in the pre-season tests performed. This can be reflected in the studies that employed the baseline testing approach, moderate accuracy for their best-performing model was reported with many models required to be fitted to achieve this moderate accuracy and that Ruddy et al. further reported the inability to predict hamstring injuries.[20]

The monitoring approach associated recent player data to injury risk. Gabbett's study employed a training load (defined as training time multiplied by perceived difficulty) to predict injury and found a strong relationship between training load and injury risk.[17] This is an intuitive result – the greater the training exposure, the greater the injury exposure and therefore, the greater the likelihood of injury. However, not studied was the relative impact of training/game duration and training/game difficulty that may have had a greater role in injury prediction. There may also be currently unknown factors that have a role in injury prediction (i.e. sleep, nutrition). Talukder et al. collected data from the past 14 days to predict injuries in the next 7 days.[18] They focus on players playing at least 15 minutes per game, and this filtering attempted to aid exposure standardisation. Rossi et al., Carey et al. and Colby et al. adopted a similar approach having collected players' recent training history using acute: chronic workload ratio (ACWR), mean standard deviation workload ratio (MSWR) and/or exponentially weighted moving averages (EWMA).[21,22,24] The limitations of ACWR are prominent, notably that there it suffers from sparse data bias. [27] Additional limitations are also concerning such as unmeasured confounding, time-dependent confounding and recurrent injuries.[27] Therefore, the use of ACWR may not be a valid tool for injury prediction.

This systematic review has indicated that the best performing prediction model employed the monitoring rather than the baseline testing approach. This was not surprising as changes in exposure and other important variables by session could be accounted for. All these changes were unobserved by the baseline testing approach. This suggested that the employment of a monitoring approach in future prediction models would be optimal. Otherwise, the prediction model may be at greater risk of false-negative predictions, as injury occurrence is an event dominated by exposure. The effect that in-season injury, including the period of rehabilitation, has on future injury risk within the season is also poorly understood. This can influence both the exposure and injury cases if they are only measured across the population with individual effects not accounted for. This is particularly important in the subset of the population that typically accounts for a large proportion of all injuries.

Statistical techniques

The included studies employed techniques that handled "unbalanced" data. Most of the observations used for modelling were of cases where no injury occurred. If a prediction model predicted "no injury" for most cases, the model is likely to perform well, as an injury is an uncommon event. Oversampling of the injury cases, undersampling of the non-injury cases, or assignment of different costs to false negatives and false positives can be performed to avoid this. Assigning a different cost to false positives and false negatives can be the most intuitive approach but would require thought to define the real cost of a false positive compared to a false negative. Over and undersampling generated similar results, but decisions must be made on the level of oversampling to use and the method to use. These decisions indirectly stress the different costs of false positives and false negatives.

A probabilistic model may solve the problem of unbalanced data. A probabilistic model assigns a probability to each prediction rather than an injury/no-injury category. Of the included studies, Gabbett et al., Carey et al., Carbuhn et al. and Colby et al. employed a probabilistic technique: logistic or Poisson regression.[17,22-24] The strength of assigning a probability to each observation is that different costs can be separately assigned to false positives and false negatives after fitting. They may better provide a picture of injury likelihoods.

Decision tree methods have several advantages: they can handle mixed types of data, missing values, are robust to outliers and can deal with irrelevant variables. All these features useful when dealing with real-world data. The main disadvantage of decision trees has been their predictive power, as this has been typically low.[28] Predictive performance can be improved by employing boosting techniques as in López-Valenciano et al.[19] and Ayala et al.[26] or using a random forest approach as in Talukder et al.[18] and Ruddy et al.[20] However, the increased accuracy came with lower interpretability as the model predictions were unable to be longer traced in a single model tree. A prediction model must succinctly describe the relative importance of different variables in its predictions and describe how changing combinations of variables will change injury risk. The included studies have employed prediction models that have been explainable.

Confounding

In studies that employed the monitoring approach, the results appeared capable to be predictive of injury. This is not the same as the prevention of injury and a variable that is associated with higher injury risk may not cause injury. Gabbett's (2010) demonstration that training load per week increases injury risk is a case in point: higher training loads led to higher injury risk, but that increased risk may not be preventable if the training is necessary to prepare for the game.[17] Similarly, measurements of ACWR and MSWR may be associated with an injury.[29] However, if there are consistent patterns of ACWR and MSWR in the lead up to a game, this association may not be causal. Bornn et al. discussed this confounding problem with respect to ACWR; a variable claimed to be predictive of injury risk.[30] This study has demonstrated that scheduling effects, the necessity of playing games on specific dates can associate ACWR with injury, even though no causal effect may have had existed. Therefore, care must be made to distinguish between prediction and prevention. There may be an association of ACWR and injury risk, but a change in ACWR may be inherently pre-determined by the heavily loaded work schedule. Calculating ACWR using exponentially weighted moving averages is a step forward to account for the heavily loaded work schedule as it has been shown to better reflect the decaying nature of fitness and fatigue over time.[29] Clinicians and support staff must realise that training load is only a part contributor to injury risk where there are many other existing factors, i.e. sleep, recovery and nutrition, that can be contributors to injury risk.[31] The ACWR has striking limitations for injury prediction that is vulnerability to sparse data bias, unmeasured confounding, time-dependent confounding and recurrent injuries.[27] Therefore, the statistical properties of ACWR, are in its current form, an arguably rudimentary and inaccurate metric for injury prediction.[27,32]

Limitations

The limitations of the included studies are also limitations of this study as per the inherent study design of a systematic review. The scope of the literature search was limited to PubMed and Embase databases published exclusively in English. Therefore, it was possible that studies that satisfied the eligibility criteria were not identified. The risk of bias appraisal of studies in this systematic review was conducted using the Newcastle-Ottawa Scale, an assessment tool used for nonrandomised clinical studies. As this systematic review focused on studies that trialled machine learning techniques for injury prediction instead of interventions, diagnostic or prognostic variables typically trialled in clinical studies, the Newcastle-Ottawa Scale can be called into question on its suitability to appraise the included studies. To the authors' knowledge, no assessment tool has been developed to appraise studies that trials machine learning techniques for injury prediction to date. The prediction models developed were for a range of sports and therefore, mechanics of injury can differ, with various injuries more susceptible than others in certain sports, questioning the generalisability and transferability of the concept of the prediction models overall. The included studies were also predominantly of a confirmed male population and therefore, the constructed models reviewed may not be completely applicable to females. Although a 'time-loss' injury definition was most commonly confirmed, the 3 studies that did not report a specific injury definition may have used self-reported injuries that may have yielded much larger cases of injury, skewing the comparisons.[18,22,23] There can also be many other factors (both known and unknown) beyond the included studies that contribute to a causal injury path.

PERSPECTIVES

The negative results of the studies that employed the baseline testing approach can be explained by variation in training and game exposure between players. The performance of prediction models in the literature has been poor, caused by a fundamental difficulty in discovering real effects in small sample sizes with low injury rates. A better understanding of how training and game exposure is associated with the data components for data preprocessing and subsequently associated with injury is warranted. The incorporation of these considerations is critical for the development of robust injury prediction models. In the real world, a clinician can be biased towards generating negative results for injury prediction. If a clinician felt a player was at high risk of injury, they would intervene to prevent that injury. The development of a robust prediction model could allow clinicians to better understand the limits of a player prior to the occurrence of an injury but to also provide estimated probabilities of which informed decisions could occur (accepting known risk given the context of game/training and its importance), ultimately providing the advantage in player health and sport management.

REFERENCES

1. Brooks JH, Fuller CW, Kemp SP, Reddin DB. Incidence, risk, and prevention of hamstring muscle injuries in professional rugby union. *Am J Sports Med*. 2006;34:1297–1306.
2. Ekstrand J, Hägglund M, Waldén M. Epidemiology of muscle injuries in professional football (soccer). *Am J Sports Med*. 2011;39:1226–1232.
3. Langevoort G, Myklebust G, Dvorak J, Junge A. Handball injuries during major international tournaments. *Scand J Med Sci Sports*. 2007;17:400-407.
4. Hägglund M, Waldén M, Magnusson H, Kristenson K, Bengtsson H, Ekstrand J. Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA Champions League injury study. *Br J Sports Med*. 2013;47:738-742.
5. Hurley OA. Impact of player injuries on teams' mental states, and subsequent performances, at the Rugby World Cup 2015. *Front Psychol*. 2016;7:807.
6. Gabbett TJ. Changes in physiological and anthropometric characteristics of rugby league players during a competitive season. *J Strength Cond Res*. 2005;19:400-408.
7. Fernández-Cuevas I, Gomez-Carmona P, Sillero-Quintana M, Noya-Solces J, Arnaiz-Lastras J. Economic costs estimation of soccer injuries in first and second Spanish division professional teams. *15th Annual Congress of the European College of Sport Sciences*. 2010.
8. Knowles SB, Marshall SW, Miller T, et al. Cost of injuries from a prospective cohort study of North Carolina high school athletes. *Inj Prev*. 2007;13:416-421.
9. Alanazi HO, Abdullah AH., Qureshi KN. A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care. *J Med Syst.* 2017;41:69.
10. Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Trans Syst Man Cybern Part C (Appl Rev)*. 2012;42:463-484.
11. Fontana MA, Lyman S, Sarker GK, Padgett DE, MacLean CH. Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? *Clin Orthop Relat Res*. 2019;477:1267-1279.

12. Mazurowski MA, Habas PA, Zurada JM, Lo JY, Baker JA, Tourassi GD. Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance. *Neural Netw*. 2008;21:427–436.
13. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med*. 2009;151:W65-94.
14. Fuller CW, Ekstrand J, Junge A, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Br J Sports Med*. 2006;40:193-201.
15. Howick J, Chalmers I, Glasziou P, et al. The Oxford 2011 levels of evidence. *Oxford Centre for Evidence-Based Medicine* 2011. 2011. http://www.cebm.net/index.aspx?o=5653.
16. Wells GA, Shea B, O'Connell D, et al. Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. *Ottawa Hospital Research Institute, Ottawa, Canada*. 2012.
17. Gabbett TJ. The development and application of an injury prediction model for noncontact, soft-tissue injuries in elite collision sport athletes. *J Strength Cond Res*. 2010;24:2593-2603.
18. Talukder, H., Vincent, T., Foster, G., et al. Preventing in-game injuries for NBA players. *MIT Sloan Analytics Conference. Boston* 2016. 2016.
19. López-Valenciano A, Ayala F, Puerta JM, et al. A preventive model for muscle injuries: a novel approach based on learning algorithms. *Med Sci Sports Exerc*. 2018;50:915-927.
20. Ruddy JD, Shield AJ, Maniar N, et al. Predictive modeling of hamstring strain injuries in elite Australian footballers. *Med Sci Sports Exerc*. 2018;50:906-914.
21. Rossi A, Pappalardo L, Cintia P, Iaia FM, Fernàndez J, Medina D. Effective injury forecasting in soccer with GPS training data and machine learning. *PLoS One*. 2018;13:e0201264.
22. Carey DL, Ong K, Whiteley R, Crossley KM, Crow J, Morris ME. Predictive modelling of training loads and injury in Australian football. *Int J Comput Sci*. 2018;17:49-66.
23. Carbuhn AF, Sanchez Z, Fry AC, Reynolds MR, Magee LM. A simplified prediction model for lower extremity long bone stress injuries in male endurance running athletes. *Clin J Sport Med*. 2018.
24. Colby MJ, Dawson B, Peeling P, et al. Improvement of prediction of noncontact injury in elite

Australian footballers with repeated exposure to established high-risk workload scenarios. *Int J Sports Physiol Perform*. 2018;13:1130-1135.

25. Connaboy C, Eagle SR, Johnson C, Flanagan S, Mi Q, Nindl BC. Employing machine learning to predict lower extremity injury in U.S. Special Forces. *Med Sci Sports Exerc*. 2018.
26. Ayala F, López-Valenciano A, Gámez Martín JA, et al. A preventive model for hamstring injuries in professional soccer: learning algorithms. *Int J Sports Med*. 2019;40:344-353.
27. Wang C, Vargas JT, Stokes T, Steele R, Shrier I. Analyzing activity and injury: lessons learned from the acute:chronic workload ratio. *Sports Med*. 2020.
28. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. *Springer*; 2009.
29. Murray NB, Gabbett TJ, Townshend AD, Blanch P. Calculation acute:chronic workload ratios using exponential weighted moving averages provides a more sensitive indicator of injury likelihood than rolling averages. *Br J Sports Med*. 2017;51:749-754.
30. Bornn L, Ward P, Norman D. Training schedule confounds the relationship between acute:chronic workload ratio and injury. *MIT Sloan Analytics Conference. Boston* 2016. 2016.
31. Maupin D, Schram B, Canetti E, Orr R. The relationship between acute: chronic workload ratios and injury risk in sports: a systematic review. *Open Access J Sports Med*. 2020;11:51-75.
32. Impellizzeri FM, Tenan MS, Kempton T, Novak A, Coutts AJ. Acute:chronic workload ratio: conceptual issues and fundamental pitfalls. *Int J Sports Physiol Perform*. 2020;1-7.

Table 1. Study and player characteristics

Study	Study design	LoE	QoE	Prediction model focus	Sport	Participation level	Population (n)	No. of players injured	No. of injuries	Mean age (years)	Mean height (cm)	Mean body mass (kg)	Follow-up (months)
Gabbett. *J Strength Cond Res*. 2010[17]	Prospective cohort study	2	7	Non-contact injuries	Rugby league	Professional	91	NR	159	23.7	183.2	94	2 sporting seasons
Talukder et al. *MIT Sloan Analytics Conference*. 2016[18]	Retrospective cohort study	3	7	All injuries	Basketball	Professional	NR	NR	500	NR	NR	NR	2 sporting seasons
López-Valenciano et al. *Med Sci Sports Exerc*. 2018[19]	Prospective cohort study	2	7	Lower body muscular injuries	Soccer (n = 98) handball (n = 34)	Professional	132	NR	32	NR	NR	NR	9
Ruddy et al. *Med Sci Sports Exerc*. 2018[20]	Prospective cohort study	2	7	Hamstring injuries	Australian football	Professional	S_1 = 186 S_2 = 176	S_1 = 27 S_2 = 26	NR	S_1 = 23.2 S_2 = 25	S_1 = 188 S_2 = 187.6	S_1 = 87.6 S_2 = 87	22
Rossi et al. *PLoS One*. 2018[21]	Prospective cohort study	2	7	Non-contact injuries	Soccer	Professional	26	13	23	26	179	78	5.29
Carey et al. *Int J Comp Sci*. 2018[22]	Prospective cohort study	2	7	Hamstring injuries	Australian football	Professional	75	NR	NR	NR	NR	NR	24
Carbuhn et al. *Clin J Sport Med*. 2018[23]	Retrospective cohort study	3	7	Long bone injuries	Endurance sport	Professional	27	NR	NR	18 to 23	NR	NR	3
Colby et al. *Int J Sports Physiol Perform*. 2018[24]	Retrospective cohort study	3	7	Lower body non-contact injuries	Australian football	Professional	60	NR	58	23.3	188.9	88.1	3 sporting seasons
Connaboy et al. *Med Sci Sports Exerc*. 2018[25]	Prospective cohort study	2	7	Lower body injuries	Military	Special Forces	140	38	NR	27.0	177.6	83.8	12
Ayala et al. *Int J Sports Med*. 2019[26]	Prospective cohort study	2	7	Hamstring injuries	Soccer	Professional	96	NR	18	NR	NR	NR	9

LoE, level of evidence; NR, not reported; QoE, quality of evidence; S_1, season 1; S_2, season 2

Table 2. Data preprocessing

Study	Data component collection timing	Data components 1	2	3	4	5
Gabbett. *J Strength Cond Res*. 2010[17]	30 minutes post-training	Workload data, training load: 1. Duration of training session 2. Modified rating of perceived exertion scale	NA	NA	NA	NA
Talukder et al. *MIT Sloan Analytics Conference*. 2016[18]	NR	Player profile 1. Age 2. Salaries	Body composition data 1. Body mass 2. Height	Player statistics (play-by-play game data)	Workload data, SportsVU: 1. Average distance covered 2. Average minutes of play 3. Average number of field goals attempted 4. Average speed during games 5. Number of games played	NA
López-Valenciano et al. *Med Sci Sports Exerc*. 2018[19]	Baseline: all data	Player profile 1. Age 2. Dominant leg 3. Position 4. Sport 5. Sport level	Body composition data 1. Body mass 2. Height	Injury history in previous season	Psychological assessments: 1. Athlete burnout questionnaire 2. Pittsburgh sleep diary	Neuromuscular assessments: 1. Core stability 2. Dynamic postural control 3. Lower extremity joints ROM 4. Isokinetic knee flexion and extension strength 5. Isometric hip abduction and adduction strength
Ruddy et al. *Med Sci Sports Exerc*. 2018[20]	Baseline: all data	Player profile 1. Age 2. Position	Body composition data 1. Body mass 2. Height	Injury history: 1. Any ACL injury history 2. Hamstring injury in previous 12 months	Eccentric hamstring strength 1. Between-limb imbalance 2. Knee flexor strength during Nordic hamstring exercise	NA

Rossi et al. *PLoS One*. 2018[21]	Baseline: 1 and 2; per training session: 3, 4 and 5	Player profile 1. Age 2. Position	Body composition data 1. Body mass 2. Height	Injury history	Workload data, GPS data: 1. Dynamic stress load 2. Explosive distance 3. Fatigue index 4. High metabolic load distance 5. High metabolic load distance per minute 6. High speed running 7. Metabolic distance 8. Number of accelerations above 2m/s^2	GPS data, continued: 9. Number of accelerations above 3m/s^2 10. Number of decelerations above 2m/s^2 11. Number of decelerations above 3m/s^2 12. Total distance Other workload data: 1. Minutes of play 2. Number of games played
Carey et al. *Int J Comp Sci*. 2018[22]	Baseline: 1; per training session: 2 and 3	Player profile: 1. Age	Workload data, training load: 1. Modified rating of perceived exertion scale	Workload data, GPS data: 1. Total distance 2. Moderate speed running 3. High speed running 4. Player load	NA	NA
Carbuhn et al. *Clin J Sport Med*. 2018[23]	Baseline: all data	Body composition data: 1. Body mass 2. Height 3. DEXA variables	NA	NA	NA	NA
Colby et al. *Int J Sports Physiol Perform*. 2018[24]	Per training	Workload data, training load: 1. Modified rating of perceived exertion scale	Workload data, GPS data: 1. Total distance 2. Sprint distance 3. Maximal velocity	NA	NA	NA
Connaboy et al. *Med Sci Sports Exerc*. 2018[25]	Baseline: all data	Body composition data	Injury history	Neuromuscular assessments: 1. Isokinetic lower body strength 2. Flexibility 3. Single-leg standing mechanics	Cardiopulmonary assessments: 1. Aerobic capacity 2. Anaerobic capacity	NA

Ayala et al. *Int J Sports Med*. 2019[26]	Baseline: all data	Player profile: 1. Age 2. Dominant leg 3. Position 4. Sport 5. Sport level	Body composition data: 1. Body mass 2. Height	Injury history in previous season	Psychological assessments: 1. Athlete burnout questionnaire 2. Pittsburgh sleep diary	Neuromuscular assessments: 1. Core stability 2. Dynamic postural control 3. Lower extremity joints ROM 4. Isokinetic knee flexion and extension strength 5. Isometric hip abduction and adduction strength

ACL, anterior cruciate ligament; DEXA, dual-energy x-ray absorptiometry; GPS, global positioning system; NA, not applicable; NR, not reported; ROM, range of motion

Table 3. Data processing

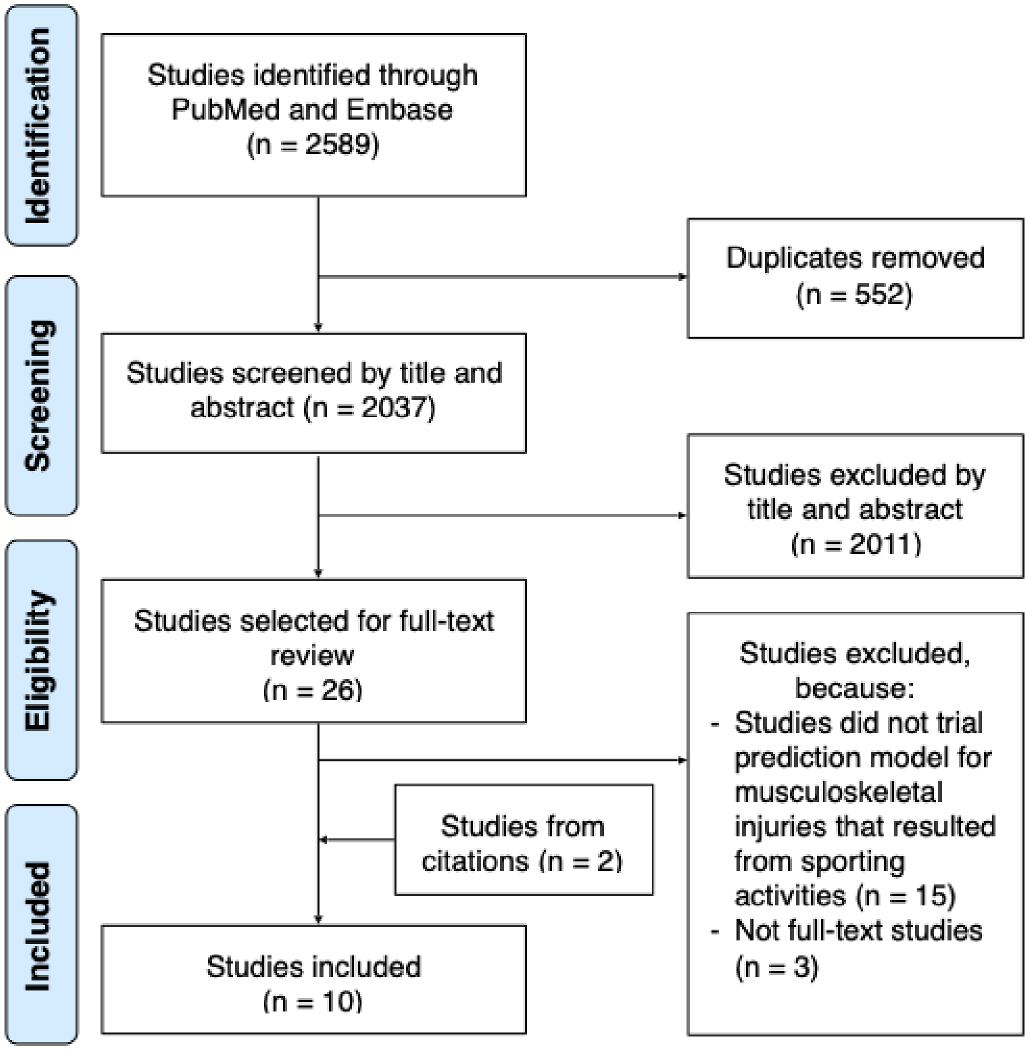
Study	Machine learning technique		Sampling technique	Cross-validation	Prediction window
	Best performing prediction model per study	Other prediction models trailed			
Gabbett. *J Strength Cond Res*. 2010[17]	Logistic regression with binomial distribution and logit link function	NA	NR	Yes (fold not specified)	Season end
Talukder et al. *MIT Sloan Analytics Conference*. 2016[18]	Random forest	Logistic regression	50%-50% random splitting	10-fold	1 week
López-Valenciano et al. *Med Sci Sports Exerc*. 2018[19]	Decision Tree with decision tree algorithm ADTree and ensemble learning algorithm AdaBoost with cost-sensitive classifier	Decision tree with 3 other decision tree algorithms and 7 various ensemble learning algorithms each	Synthetic minority oversampling technique	5-fold	Season end
Ruddy et al. *Med Sci Sports Exerc*. 2018[20]	S_1 = Naïve Bayes S_2 = random forest and support vector machine	S_1 = logistic regression, neural network, random forest, support vector machine S_2 = Naïve Bayes, logistic regression, neural network	Synthetic minority oversampling technique	10-fold	Season end
Rossi et al. *PLoS One*. 2018[21]	Decision tree with decision tree classifier	Decision tree with random forest classifier, decision tree with logit classifier	Oversample minority class with adaptive synthetic sampling approach	2-fold	Season end, next training session/game
Carey et al. *Int J Comp Sci*. 2018[22]	Logistic regression	Random forest, generalised estimating equations, support vector machines	Undersampling for noninjuries and synthetic minority oversampling for injuries	10-fold	Next game
Carbuhn et al. *Clin J Sport Med*. 2018[23]	Binary logistic regression: PLBI = 23.465 - 0.896 BMI + 1.043 TUB - 34.536 leg BMD, where PLBI is the LBI prediction according to the log odds (LBI)	NA	NR	NR	3 months
Colby et al. *Int J Sports Physiol Perform*. 2018[24]	Poisson log-link regression	NA	NR	10-fold	Season end
Connaboy et al. *Med Sci Sports Exerc*. 2018[25]	Decision Tree	NA	NR	K-fold	12 months
Ayala et al. *Int J Sports Med*. 2019[26]	Decision Tree with SmooteBoost ADTree	J48, SimpleCart	Synthetic minority oversampling technique	3-fold	Season end

ADTree, alternating decision tree; BMD, bone mineral density; BMI, body mass index; LBI, long bone injuries; NA, not applicable; NR, not reported; PLBI, probability of long bone injuries; S_1, season 1; S_2, season 2; TUB, total upper-body mass

Table 4. Performance of best performing model indicated per study

Study	True positive (n)	True negative (n)	False positive (n)	False negative (n)	Sensitivity/recall	Specificity	Precision	AUC
Gabbett. *J Strength Cond Res*. 2010[17]	121	1589	20	18	0.87	0.99	0.86	NR
Talukder et al. *MIT Sloan Analytics Conference*. 2016[18]	NR	NR	NR	NR	0.19	0.06	NR	0.92
López-Valenciano et al. *Med Sci Sports Exerc*. 2018[19]	19	72	19	10	0.66	0.79	0.50	0.75
Ruddy et al. *Med Sci Sports Exerc*. 2018[20]	NR	NR	NR	NR	NR	NR	NR	S_1 = 0.60 S_2 = 0.57
Rossi et al. *PLoS One*. 2018[21]	NR	NR	NR	NR	0.80	NR	0.50	0.76
Carey et al. *Int J Comp Sci*. 2018[22]	NR	NR	NR	NR	NR	NR	NR	0.72
Carbuhn et al. *Clin J Sport Med*. 2018[23]	4	22	0	1	NR	NR	NR	0.87
Colby et al. *Int J Sports Physiol Perform*. 2018[24]	NR	NR	NR	NR	NR	NR	NR	0.54 to 0.59
Connaboy et al. *Med Sci Sports Exerc*. 2018[25]	NR	NR	NR	NR	NR	NR	0.55	0.83 to 0.92
Ayala et al. *Int J Sports Med*. 2019[26]	NR	NR	NR	NR	0.78	0.84	NR	0.84

AUC, area under curve; NR, not reported; S_1, season 1; S_2, season 2



tsm2_181_f1.tiff

Appendix 1. Item-specific scores of the Newcastle-Ottawa scale

Study	Representativeness of the exposed cohort (1 star)	Selection of the non-exposed cohort (1 star)	Ascertainment of exposure (1 star)	Demonstration that outcome of interest was not present at start of study (1 star)	Comparability of cohorts on the basis of the design or analysis (2 stars available)	Assessment of outcome (1 star)	Was follow-up long enough for outcomes to occur (1 star)	Adequacy of follow up of cohorts (1 star)	Total
Gabbett. *J Strength Cond Res*. 2010[17]	0	1	1	1	1	1	1	1	7
Talukder et al. MIT Sloan Analytics Conference 2016[18]	0	1	1	1	1	1	1	1	7
López-Valenciano et al. *Med Sci Sports Exerc*. 2018[19]	1	0	1	1	1	1	1	1	7
Ruddy et al. *Med Sci Sports Exerc*. 2018[20]	0	1	1	1	1	1	1	1	7
Rossi et al. *PLoS One*. 2018[21]	0	1	1	1	1	1	1	1	7
Carey et al. *Int J Comp Sci*. 2018[22]	0	1	1	1	1	1	1	1	7
Carbuhn et al. *Clin J Sport Med*. 2018[23]	0	1	1	1	1	1	1	1	7
Colby et al. *Int J Sports Physiol Perform*. 2018[24]	0	1	1	1	1	1	1	1	7
Connaboy et al. *Med Sci Sports Exerc*. 2018[25]	0	1	1	1	1	1	1	1	7
Ayala et al. *Int J Sports Med*. 2019[26]	0	1	1	1	1	1	1	1	7