Hamstring strain injury risk factors in Australian Football change over the course of the season

Sim, Aylwin, Timmins, Ryan G., Ruddy, Joshua D., Shen, Haifeng, Liao, Kewen, Maniar, Nirav, Hickey, Jack T., Williams, Morgan D. and Opar, David A.

. . . Published ahead of Print

Hamstring Strain Injury Risk Factors in Australian Football Change over the Course of the Season

Aylwin Sim[1,2], Ryan G. Timmins[3,4], Joshua D. Ruddy[3,4], Haifeng Shen[1,2], Kewen Liao[1,2], Nirav Maniar[3,4], Jack T. Hickey[3,4,5], Morgan D. Williams[6], and David A. Opar[3,4]

[1]Peter Faber Business School, Australian Catholic University, Sydney, AUSTRALIA; [2]Human-Centred Intelligent Learning and Software Technologies Research Lab (HilstLab), Sydney, AUSTRALIA; [3]School of Behavioural and Health Sciences, Australian Catholic University, Fitzroy, Victoria, AUSTRALIA; [4]Sports Performance, Recovery, Injury and New Technologies (SPRINT) Research Centre, Australian Catholic University, Fitzroy, Victoria, AUSTRALIA; [5]Department of Sport Science and Nutrition, Maynooth University, County Kildare, IRELAND; [6]Faculty of Life Sciences and education, University of South Wales, Pontypridd, UNITED KINGDOM

Accepted for Publication: 29 August 2023

Medicine & Science in Sports & Exercise, Publish Ahead of Print
DOI: 10.1249/MSS.0000000000003297

Hamstring Strain Injury Risk Factors in Australian Football Change over the Course of the Season

Aylwin Sim[1,2], Ryan G. Timmins[3,4], Joshua D. Ruddy[3,4], Haifeng Shen[1,2], Kewen Liao[1,2], Nirav Maniar[3,4], Jack T. Hickey[3,4,5], Morgan D. Williams[6], and David A. Opar[3,4]

[1]Peter Faber Business School, Australian Catholic University, Sydney, AUSTRALIA; [2]Human-Centred Intelligent Learning and Software Technologies Research Lab (HilstLab), Sydney, AUSTRALIA; [3]School of Behavioural and Health Sciences, Australian Catholic University, Fitzroy, Victoria, AUSTRALIA; [4]Sports Performance, Recovery, Injury and New Technologies (SPRINT) Research Centre, Australian Catholic University, Fitzroy, Victoria, AUSTRALIA; [5]Department of Sport Science and Nutrition, Maynooth University, County Kildare, IRELAND; [6]Faculty of Life Sciences and education, University of South Wales, Pontypridd, UNITED KINGDOM

Address for Correspondence: Aylwin Sim, Australian Catholic University – North Sydney Campus, 40 Edward St., North Sydney, Sydney, New South Wales, 20260, Australia; Phone: +61 484 108 046; E-mail: aylwin.sim@myacu.adu.au

Conflict of Interest and Funding Source:

No external funding sources were required for this study. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation. The results of this study do not constitute endorsement by the American College of Sports Medicine. Dr David Opar is listed as a co-inventor on a patent filed for a field-testing device of eccentric hamstring strength (PCT/AU2012/001041.2012) and is a

minority shareholders in Vald Performance Pty Ltd, the company responsible for commercialising the device. The association between measures derived from the device and future hamstring strain injury is directly examined in this manuscript. Dr Opar has received research funding from Vald Performance, for work unrelatd to the current manuscript. Dr Opar is also formerly the Chair of the Vald Performance Research Committee, a role that was unpaid. Dr Opar's brother is an employee and minor shareholder in Vald Performance. Dr Opar's brother-in-law is an employee of Vald Performance. Dr Joshua Ruddy is a former employee of Vald Performance Pty Ltd, the company responsible for commercialising the field testing device of eccentric hamstring strength used in the current study. Dr Ruddy's primary contribution to this manuscript occurred post his employment at Vald Performance. Dr Morgan Williams is a member of the Vald Research Committee. Dr Morgan Williams has been provided donations of equipment, and funds for travel and subsidence by Vald Performance to conduct research unrelated to this project. Dr Morgan Williams has received payment for reports for Vald Performance unrelated to this and any research study. No other authors declared any conflicts.

ABSTRACT

Background/aim: To determine which factors were most predictive of hamstring strain injury (HSI) during different stages of the competition in professional Australian Football.

Methods: Across two competitive seasons, eccentric knee flexor strength and biceps femoris long head (BFlh) architecture of 311 Australian Football players (455 player seasons) were assessed at the start and end of pre-season and in the middle of the competitive season. Details of any prospective HSIs were collated by medical staff of participating teams. Multiple logistic regression models were built to identify important risk factors for HSI at the different time points across the season. **Results:** There were 16, 33 and 21 new HSIs reported in preseason, early in-season, and late in-season, respectively across two competitive seasons. Multivariate logistic regression and recursive feature selection revealed that risk factors were different for pre-season, early in-season and late in-season HSIs. A combination of prior HSI, age, height and muscle thickness were most associated with pre-season injuries (median AUC, 0.83). Pennation angle and fascicle length had the strongest association with early in-season injuries (median AUC, 0.86). None of the input variables were associated with late in-season injuries (median AUC, 0.46). Identification of early in-season HSIs and late in-season HSIs was not improved by the magnitude of change of data across pre-season (median AUC, 0.67).

Conclusions: Risk factors associated with prospective HSIs were different across the season in Australian Rules Football, with non-modifiable factors (prior HSI, age, and height) mostly associated with pre-season injuries. Early in-season HSIs were associated with modifiable factors, notably BFlh architectural measures. The prediction of in-season HSIs was not improved by assessing the magnitude of change in data across pre-season.

Key Words: HAMSTRING, INJURY, AUSTRALIAN FOOTBALL SEASON

INTRODUCTION

Hamstring strain injuries (HSIs) are a common injury across many sports (1) including Australian Football (2) and significant work has been conducted to identify factors associated with an increased risk of future injury (3). These risk factors are often categorized as either modifiable or non-modifiable. Older age and a history of HSI are the two most commonly reported non-modifiable risk factors (3). Modifiable risk factors, which can be addressed through intervention, are extensive, but biceps femoris long head (BFlh) fascicle length (4) and eccentric hamstring strength (5) are current prominent variables.

Most prospective cohort studies assess risk factors at a single time-point, typically the beginning of a season (e.g., in the pre-season) (6)). This approach has limitations, as any changes in the measured variables leading up to injury, which could be up to nine months after the pre-season assessment, are not accounted for. Few have explored if more frequent assessments of risk factors improve the association with and/or prediction of future HSI risk. It was recently reported that more frequent assessments of eccentric knee flexor strength and biceps femoris long head (BFlh) architecture did not improve the ability to predict new HSIs in Australian Football (7). However, that study did not consider the possibility that risk factors may vary depending on the time of season, nor did it examine if changes in possible risk factors across time (e.g., an increase in eccentric strength from the start to the end of pre-season) at an individual level altered the ability to predict HSI.

Therefore, the primary aim of this study was to determine which factors were most predictive of the risk of HSI during pre-season, early in-season and late in-season in professional Australian Football. The secondary aim was to determine if the magnitude of change in possible risk factors across pre-season was predictive of future in-season HSIs.

METHODS

The methods used in this study pertaining to the study design, participants, and data collection have been described previously (7) but are included in detail here for the ease of the reader.

Study design and participants

This study was approved by the ACU Human Research Ethics Committee (approval number: 2017-208H) and was conducted across two Australian Football League seasons (November 2017 to August 2018 and November 2018 to August 2019, including pre-season but not including finals). Each player from six teams competing in the Australian Football League provided written informed consent prior to their participation.

At the start of pre-season, team medical staff were responsible for providing details of individual players' history of HSI in the past 12 months and if they had ever sustained an anterior cruciate ligament (ACL) injury. Eccentric knee flexor strength and BFlh architecture were assessed at the start of pre-season (November/December), end of pre-season (February/March), and middle of the competitive season (May/June), respectively. Due to scheduling constraints, the actual dates for assessments were not identical across the six different teams involved in this study. A standardised injury report form was completed by the medical staff if any player sustained a HSI during the study period.

Eccentric knee flexor strength

The assessment of eccentric knee flexor strength was performed during the execution of the Nordic hamstring exercise similar to previous studies using an instrument device (NordBord, VALD, Queensland, Australia) (2, 4, 5, 8). Players knelt on a cushioned board with their ankles secured immediately superior to the lateral malleolus by individual ankle

hooks attached to uniaxial load cells. All players completing this assessment were familiar with the Nordic hamstring exercise. Participants were told to slowly lean forward forcefully contracting their knee flexors to control their descent. All players maintained their trunk and hips in a neutral position while holding their hands across the chest throughout the exercise. Players performed a single set of 1-3 maximal repetitions as determined by each team's practices after a self-selected warm-up. The highest peak force produced by each leg throughout the test were recorded as eccentric knee flexor strength. Relative eccentric knee flexor strength was obtained by scaling its value relative to the mass of the player (N/kg) (7).

Biceps femoris long head architecture

The assessment of BFlh architecture has been reported previously (9-12). The measurements of muscle thickness, pennation angle, and fascicle length of the BFlh were obtained from ultrasound images taken along the longitudinal axis of the muscle belly utilising a two-dimensional, B-mode ultrasound (frequency, 12 MHz; depth, 8cm; field of view, 14 x 47mm) (GE Healthcare Vivid-i, Wauwatosa, U.S.A). Along the line of BFlh, the halfway point between the ischial tuberosity and the knee joint fold was determined as the scanning site. The architecture assessments were conducted on players lying on a massage plinth after at least 5 minutes of inactivity. The assessor (RGT) adjusted the orientation of the probe accordingly. The reliability of the assessor has been previously established with an intraclass correlation > 0.90 reported for BFlh fascicle length.

Offline analysis was undertaken after the images were collected (MicroDicom, Version 0.7.8, Bulgaria). Muscle thickness was determined by the distance between the superficial and intermediate aponeuroses of the BFlh. Pennation angle was determined by the angle between the intermediate aponeurosis and a fascicle of interest. The angles of superficial and intermediate aponeurosis were defined as the angle between the line marked

as the aponeurosis and an intersecting horizontal reference line across the capture image (13). Due to part of the fascicle not being visible in the ultrasound probe's field of view, the following equation from Blazevich and colleagues was used for estimation (13):

$$FL = sin(AA + 90°) \times MT/sin(180° - (AA + 180° - PA))$$

where FL=fascicle length, AA=aponeurosis angle, MT=muscle thickness and PA=pennation angle. Fascicle length was reported in absolute terms (cm) and relative to muscle thickness from a single image. The same assessor (RGT) collected and analysed all scans. The assessor has evidenced reliability in determining measures of BFlh muscle architecture at rest with ICCs >0.95 and %TE <5.0% across the measurement of all architectural variables.

Prospective hamstring strain injury reporting

An HSI was defined as posterior thigh pain that prevented a player from performing subsequent exercise and was confirmed by physical examination by the team physiotherapist or doctor (14, 15). The team medical staff filled out a standard injury report form for each HSI, which requested details about the injured limb, injured muscle, activity type performed when the injury occurred, and the number of days required for the player to return to full participation in training and competition.

Statistical Analysis

Statistical analyses were performed using the python 3.9.2 programming language (Python Software Foundation, https://www.python.org/) and the following packages: scikit-learn, statsmodel, panda, numpy, matplotlib and seaborn.

General Modelling Approach

The general modeling approach applied to this study can be found in Figure 1a.

Data pre-processing. For all analyses, an observation was removed if it consisted of at least one missing value. Additionally, players who sustained an HSI in previous time points within a season were censored from building models to predict HSIs that occurred in later timepoints. For example, a player who sustained an HSI in pre-season was excluded from training models to predict HSIs that occurred in early in-season and later in-season. Likewise, players who sustained an HSI in early in-season were excluded from training models to predict HSIs occurring in late in-season.

Correlation analysis was conducted on input predictor variables to identify redundant predictors. A Pearson's correlation coefficient threshold of > 0.8 was applied and if the pairwise correlation between two predictors exceeded the threshold of 0.8, the predictor with the higher mean pairwise correlation across all other predictor variables was removed.

Following this, the remaining input predictor variables were normalized (16) into the range of 0 and 1, using the following equation:

$$x_{\text{norm}} = \frac{x - min(x)}{max(x) - min(x)}$$

where x is the value to scale, $min(x)$ is the smallest value of the predictor, and $max(x)$ is the largest value of the predictor.

Predictor selection. The aim of the predictor selection process in the current study was to eliminate redundant predictors and identify which subset of risk factors achieved the highest predictive performance across the different time points. A wrapper feature selection method,

specifically recursive feature elimination, was used to search through different subsets of risk factors associated with HSI (17). Recursive feature elimination, which is robust to overfitting (18), was conducted in this study by fitting a logistic regression with all input predictor variables and recursively eliminating predictors that were less important based on the coefficients. Once the predictor with the lowest coefficient was removed, the model was fitted with the remaining predictors to repeat the process. This process was repeated until there was only one remaining predictor, after which the importance of individual predictors was ranked. Preliminary analyses using this dataset showed that models built using recursive feature elimination outperformed models built using all predictors. Recursive feature elimination, however, does not identify the optimal number of predictors.

Finding the optimal number of predictors. Stratified k-fold cross validation was utilised to determine the optimal number of predictors. In this study, k = 5 was applied to divide data into 5 stratified folds. For each split, 1-fold of the data was used as testing data and the remaining folds (k-1) were used as training data. The number of selected predictors resulting in the highest AUC averaged across 5 folds was chosen as the optimal number of predictors.

Performance evaluation. Once the optimal number of risk factors was determined, the final step was to evaluate the performance of logistic regression with selected risk factors. It is a common practice to allocate 20-30% of data for testing and 70%-80% of data for training (6). Any split within this threshold has been shown to have an accurate estimation of the model's performance (19). A 20%/80% train-test split was used in this study. Stratified cross validation was utilised to preserve the percentage of injured and uninjured athletes for all iterations. Since the given dataset was relatively small (<455 observations), 1000 iterations of evaluation were performed. The metric used to evaluate predictive performance was area under the curve (AUC) (20). AUC measures the ability of the models to correctly predict

prospectively injured and uninjured players. An AUC of 0.5 indicates prediction no better than random chance whereas an AUC of 1.0 indicates perfect prediction.

Analysis 1

The aim of Analysis 1 was to determine which risk factors best predicted HSIs at different time points throughout the season.

The general modelling approach was applied to Analysis 1. The subset of data utilised for Analysis 1 has been illustrated in Figure 1b, where *d1* are data assessed at the start of pre-season, *d2* are data assessed at the end of pre-season, and *d3* are data assessed in the middle of in-season. *i1* is the window following *d1* during which prospective HSIs could have occurred throughout pre-season, *i2* is the window following *d2* during which prospective HSIs could have occurred early in-season, and *i3* is the window following *d3* during which prospective HSIs could have occurred during late in-season.

Analysis 1 utilised all non-modifiable risk factors assessed at the start of pre-season and modifiable risk factors assessed at multiple time points (*d1* or start of pre-season, *d2* or end of pre-season, and *d3* or middle of in-season) as predictor variables. Prospective HSIs that occurred between individual assessment time frames (*i1* or between the start and end of pre-season, *i2* or between the end of pre-season and the middle of in-season, and *i3* or between the middle of in-season and the end of the in-season prior to the commencement of finals) were the target of the prediction models. (Refer to Table 1 for types of input predictor variables and target variables included in each of the individual models).

Analysis 2

Analysis 2 aimed to determine whether the magnitude of change in data between the start and end of pre-season, as well as more frequent assessment during pre-season, improved the ability to predict in-season HSIs, beyond the data collected at the start and end of pre-season alone.

The general modelling approach was applied to Analysis 2. The subset of data utilised for Analysis 2 has been illustrated in Figure 1c, where *d1* are data assessed at the start of pre-season, *d2* are data assessed at the end of pre-season, *d2-d1* is the magnitude of change in the risk factors across pre-season. *i2* is the window during which prospective HSIs could have occurred early in-season and *i3* is the window during which prospective HSIs could have occurred during late in-season.

Analysis 2 utilised all non-modifiable risk factors assessed at the start of pre-season and modifiable risk factors assessed at the start and end of pre-season as predictor variables. Prospective HSIs that occurred during the in-season periods (*i2* and *i3*) were the target of the prediction models. Additionally, the magnitude of change in modifiable risk factors was determined as the absolute difference between values captured at the end of pre-season and values captured at the start of pre-season. (Refer to Table 1 for types of input predictor variables and target variables included in individual modelling approaches).

RESULTS

Three-hundred and eleven male Australian Football players, with a total number of 455 player seasons (23.7 ± 3.8yrs, 188.1 ± 7.6cm, 86.5 ± 8.8kg) across the 2018 and 2019 seasons were evaluated on at least one occasion. Of these player seasons, 381 (83.7%) did not sustain an HSI and 74 (16.3%) did.

After the removal of missing values for Analysis 1, the total number of injured and uninjured player seasons during *i1* was 14 and 339 respectively (*d1->i1*; Table 2). For *i2*, the total number of injured and uninjured player seasons with complete datasets assessed at *d2* was 24 and 259 respectively (*d2->i2*; Table 2). For *i3*, the total number of injured and uninjured player seasons (with complete datasets assessed at *d3*) was 11 and 225 respectively (*d3->i3*; Table 2).

For Analysis 2, the total number of injured and uninjured player seasons with complete datasets during early in-season (i.e. *i2*) was 23 and 219 respectively (*i2*; Table 3). For late in-season (i.e. *i3*), the total number of injured and uninjured player seasons with complete datasets was 9 and 210 respectively.

Analysis 1

The performance of the individual models in Analysis 1 can be found in Figure 2. Data that were assessed at the end of pre-season and used to predict HSIs that occurred early in-season displayed the best predictive performance (median AUC = 0.86, interquartile range (IQR) = 16; Table 2) (*d2->i2*, Figure 2). The prediction of pre-season HSIs utilising data assessed at the start of pre-season (*d1->i1*; Figure 2) resulted in a median AUC of 0.83 and an IQR of 0.16. In contrast, data assessed at the middle of the in-season period and used to predict HSIs that occurred late in-season (*d3->i3*, Figure 2) resulted in the poorest predictive performance (median AUC = 0.46, interquartile range (IQR) = 0.25; Table 2).

Pre-season HIS. Players with history of HSI are more likely to sustain an HSI in pre-season (Figure 3a-c, $p < 0.01$). Shorter players displayed a higher risk of sustaining HSI in pre-season (Figure 3a). Results also showed that older athletes were associated with an increased risk of HSI in pre-season (Figure 3b, $p < 0.05$; Supplemental Table 1, Supplemental Digital

Content, http://links.lww.com/MSS/C912, The p-value of individual risk factors determined by multivariate logistic regression models in Analysis 1) and players who had thicker BFlh muscles were also more susceptible to HSI in pre-season (Figure 3c).

Early in-season HIS. Players with a greater BFlh pennation angle and shorter fascicle length were at significantly increased risk of sustaining HSI during the early in-season period (Figure 3d, 3e; $p < 0.05$; Supplemental Table 1, Supplemental Digital Content, http://links.lww.com/MSS/C912).

Late in-season HIS. Although height, age, history of ACL injury, BFlh pennation angle, fascicle length, relative eccentric knee flexor strength, as well as relative eccentric knee flexor strength imbalance were selected as predictive predictors (Figure 3f-k), the overall predictive performance of AUC was below 0.5 (median AUC = 0.46, interquartile range (IQR) = 0.25; Table 2).

Analysis 2

The performance of the individual models in Analysis 2 can be found in Figure 4a and 4b. Neither the predictions of early in-season HSIs (median AUC = 0.67, interquartile range (IQR) = 0.15; Table 3) nor late in-season HSIs (median AUC = 0.67, interquartile range (IQR) = 0.26; Table 4) were improved by assessing the magnitude of change in data across preseason. For HSIs occurring early in-season, the model with the best predictive performance utilised BFlh fascicle length and pennation angle, which were assessed at the end of pre-season. The resulting median AUC was 0.84 and the IQR was 0.16 (Table 3). Predicting late in-season injuries utilising the absolute change in BFlh pennation angle and fascicle length across pre-season, as well as history of ACL displayed the best predictive performance (median AUC = 0.67, interquartile range (IQR) = 0.26; Table 4). However, the

predictive performance was not significantly improved when compared to relative BFlh fascicle length and fascicle length, which were assessed at the start of pre-season only (median AUC = 0.65, interquartile range (IQR) = 0.25; Table 4).

DISCUSSION

This study aimed to assess whether the factors associated with HSI in professional Australian Football changed across the season. The current study found that the subset of risk factors that best predicted the occurrence of HSI was different between the pre-season and in-season periods. This study also aimed to assess whether the magnitude of change in HSI risk factors across the pre-season period improved the prediction of HSIs sustained in-season beyond using measures taken at the start or end of pre-season alone. The magnitude of change in eccentric knee flexor strength and BFlh muscle architecture variables across the pre-season period generally displayed poorer predictive performance than the absolute measures themselves (particularly those taken at the end of pre-season).

Did more frequent assessment of risk factors improve the prediction of future HSI?

The best performing model in the current study was built using BFlh fascicle length and pennation angle assessed at the end of pre-season and aimed to predict only HSIs that occurred during the first half of the in-season period. This model predicted prospective HSIs with a median AUC of 0.86. A previous study attempted to predict HSIs in elite Australian Footballers using age, previous HSI and eccentric knee flexor strength data, collected across two AFL seasons (6). When predicting HSIs that occurred within the same season, the median AUC values for the 2013 and 2015 AFL seasons were 0.58 and 0.57, respectively (6). In this previous study, when data from the 2013 AFL season were used to predict HSIs that occurred during the 2015 AFL season, the median AUC was 0.52 (6). It was suggested that more frequent measures of the risk factors examined may have improved predictive

performance. However, another study reported that more frequent measurements of modifiable risk factors did not improve the ability to identify athletes at an increased risk of HSI beyond data collected at a single timepoint (7). In support of these previous findings (7), we observed that more frequent measurements did not improve the ability to predict the occurrence of HSI. However, the assessment of different risk factors at different timepoints did improve predictive performance. In addition, we utilised recursive feature elimination to optimise predictive performance and improve the interpretability of built models. Results from preliminary analyses suggest that the selected predictors are likely to deliver better predictive performance than utilising all predictors. The findings of our study suggest that a subset of risk factors, as opposed to all risk factors, used in previous studies may have been more effective in predicting prospective HSIs.

Does the magnitude of change in risk factor data across pre-season improve the ability to predict HSI throughout the season beyond the absolute values?

In addition to suggesting that more frequent measures of the risk factors examined may improve predictive performance (6), prior work has also noted that assessing risk factors at the start of pre-season alone assumes that these factors will remain constant throughout the season (or up to the point of HSI). It is suggested that changes in HSI risk factors may influence the risk of injury to a greater extent than the absolute values of those risk factors measured at a single timepoint (6, 21). AUC values of 0.7 and above are regarded as having significant impacts in sport science domains (22). In the current study, models built with the magnitude of change in risk factors across pre-season were less optimal when attempting to predict HSIs during early in-season, or *i2* (median AUC of 0.66), as well as HSIs during late in-season, or *i3* (median AUC of 0.63). Conversely, models built using the absolute values measured at the end of pre-season, or *d2*, performed better when predicting HSIs during early

in-season, or *i2* (median AUC of 0.83). However, the performance of all models attempting to predict late in-season HSIs, or *i3*, were the poorest.

The current results suggest that risk factor data assessed at the end of pre-season provides the strongest performance when predicting in-season HSIs. Despite the magnitude of change in modifiable risk factor data performing poorly from a prediction standpoint, it is important to acknowledge that significant adaptations in eccentric knee flexor strength and BFlh muscle architecture can be elicited in as little as two weeks (23). For example, an increase of BFlh fascicle length and a reduction in BFlh pennation angle has been observed following just 14 days of an eccentric strength training intervention (23). Given this, it is likely that athletes saw significant adaptations across the pre-season period and that modifiable risk factor data assessed at the end of pre-season provided a better indication of athletes' physical status during the in-season period compared to data collected at the start of pre-season. In contrast to this, data collected at the midpoint of the in-season period displayed the worst predictive performance when used to predict injuries that occurred during the second half of the in-season period. This suggests that despite this data being more aetiologically relevant, there may exist other factors that influence the risk of HSIs occurring during the latter half of the season to a greater extent than those examined in this study.

In which phase of the season was the predictive performance for HSI best?

The best performing model aimed to predict HSIs during the first half of the in-season period and was built using data collected at the end of pre-season (median AUC of 0.86; Table 2). In contrast, the poorest performing model was built using data collected at the midpoint of the in-season period and aimed to predict HSIs in the second half of the in-season period (median AUC of 0.46; Table 2). It has previously been reported that an increase in BFlh fascicle length during early in-season can be observed across all players (9).

However, it was observed that players with a history of HSI saw greater decreases in BFlh fascicle length during the latter part of the in-season period when compared to players without a history (9). This may, to an extent, explain why BFlh fascicle length assessed at the end of pre-season did not present strong association with late in-season HSIs, when compared to early in-season HSIs (9).

Absolute risk factor data assessed at the end of pre-season may provide practitioners with the most insight regarding HSI risk, and that additional assessments of the studied variables throughout the in-season period may not add further value. The relatively poor performance of the models built to predict late in-season HSIs suggests that there may be additional factors that influence the risk of injury to a greater extent in the latter stages of the season.

Limitations

Since BFlh fascicles were longer than the ultrasound field of view (14x4.7mm), extrapolation methods were used to calculate BFlh fascicles (24). Although the extrapolation method was proven to be highly reliable in an earlier study (ICC>0.97) when validated against cadaveric data (13), the drawback is that it may overestimate BFlh fascicle length (25). Due to the lack of a standardised classification system (26), not all HSIs reported in this study offer details pertaining to the muscle that was injured. Further subgroup analysis may be conducted if more injury data of the injured muscle were recorded. Since player exposure data were not presented in this study, reported HSI incidence did not consider the amount of time spent training and competing. In addition, the use of athlete tracking technologies to account for high-speed running and strength training exposure may offer more insights regarding HSI risk. Warm up procedures were not standardised for strength assessments. Future studies should consider standardising warm-up practices to limit the impact it may

have on the strength outcomes. The data utilised in the current study were collected across multiple clubs at differing time points as part of their routine practices. Accordingly, it was not possible to standardise the warm-up and testing protocols for the assessment of eccentric knee flexor strength. The impact that a standardised warm-up protocol across all participating teams could have had on the peak eccentric knee flexor force values observed is unknown. Additionally, whether identifying peak eccentric knee flexor force from one versus three repetitions of the NHE would have impacted the current results is also unknown. It is possible that there is a learning/feedback effect that may influence the results observed after three repetitions compared to one repetition, however, participants in this study were all very familiar with the assessment of eccentric knee flexor force during the NHE, which likely reduced the possibility of this. The use of logistic regression in this study assumes linearity between target variable and risk factors. Complex non-linear models may be utilised with proper hyperparameter tuning practice. Although previous studies showed the use of non-linear models outperformed logistic regression in injury prediction (22, 27), these studies were conducted on a larger dataset. Earlier work showed no improvements in predictive performance when complex modelling approach was used (6). The absence of a standardised fine-tuning process on small and imbalanced dataset may be the cause, which result in overfitting. Despite this study recording a high number of prospective HSIs in comparison to previous research (28), the relatively low injury rates and the class imbalance problem that this presents remains a limitation of this study and as well as most prospective sports injury studies in general. It is unclear whether predictive performance would be improved if class imbalance was addressed. Furthermore, the presence of missing data results in reduced numbers of player seasons used for the analysis in this study. Although AUC is used in many studies (6, 7, 29), other metrics should be considered thoroughly when evaluating the generalisation of binary classifiers. In addition, future studies should utilise interpretability methods in machine learning to help experts better understand the decisions of trained models

beyond predictive performance. Finally, previous work suggests that HSI risk factors are not transferable to different sporting populations (30) so applications of the current findings to other sports (e.g., soccer, rugby) should be done with caution.

CONCLUSIONS

This study has demonstrated that the risk factors most associated with prospective HSIs change throughout an Australian Football season. Non-modifiable risk factors (history of HSI, age and height) demonstrated a strong association with pre-season HSIs, whereas early in-season HSIs were better explained by modifiable risk factors. Conversely, late in-season injuries did not present any strong associations with either modifiable or non-modifiable risk factors examined in this study. The magnitude of change in modifiable risk factors across pre-season did not improve the prediction of in-season HSIs. The results of this study suggest that assessing the same risk factors at multiple time points throughout the season may not be the best approach when identifying athletes at an increased risk of HSI. Instead, assessing different risk factors at specific time points may provide practitioners with more insight, however, the practical relevance of this is questionable. Future research is warranted to investigate the effectiveness of assessing risk factors at varying time points to improve HSI risk mitigation efforts.

Acknowledgements

No external funding sources were required for this study. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation. The results of this study do not constitute endorsement by the American College of Sports Medicine.

Conflict of interest

Human-Centred Intelligent Learning and Software Technologies Research Lab (HilstLab), Peter Faber Business School, Australian Catholic University, Sydney, Australia.

Sports Performance, Recovery, Injury and New Technologies (SPRINT) Research Centre, Australian Catholic University, Fitzroy, Victoria, Australia & School of Behavioural and Health Sciences, Australian Catholic University, Fitzroy, Victoria, Australia.

Dr David Opar is listed as a co-inventor on a patent filed for a field-testing device of eccentric hamstring strength (PCT/AU2012/001041.2012) and is a minority shareholders in Vald Performance Pty Ltd, the company responsible for commercialising the device. The association between measures derived from the device and future hamstring strain injury is directly examined in this manuscript. Dr Opar has received research funding from Vald Performance, for work unrelatd to the current manuscript. Dr Opar is also formerly the Chair of the Vald Performance Research Committee, a role that was unpaid. Dr Opar's brother is an employee and minor shareholder in Vald Performance. Dr Opar's brother-in-law is an employee of Vald Performance.

Dr Joshua Ruddy is a former employee of Vald Performance Pty Ltd, the company responsible for commercialising the field testing device of eccentric hamstring strength used

in the current study. Dr Ruddy's primary contribution to this manuscript occurred post his employment at Vald Performance.

Dr Morgan Williams is a member of the Vald Research Committee. Dr Morgan Williams has been provided donations of equipment, and funds for travel and subsidence by Vald Performance to conduct research unrelated to this project. Dr Morgan Williams has received payment for reports for Vald Performance unrelated to this and any research study.

No other authors declared any conflicts.

REFERENCES

1. Maniar N, Carmichael DS, Hickey JT, et al. Incidence and prevalence of hamstring injuries in field-based team sports: a systematic review and meta-analysis of 5952 injuries from over 7 million exposure hours. *Br J Sports Med.* 2023;57(2):109-16.
2. Orchard JW, Seward H, Orchard JJ. Results of 2 decades of injury surveillance and public release of data in the Australian Football League. *Am J Sports Med.* 2013;41(4):734-41.
3. Green B, Bourne MN, van Dyk N, Pizzari T. Recalibrating the risk of hamstring strain injury (HSI): A 2020 systematic review and meta-analysis of risk factors for index and recurrent hamstring strain injury in sport. *Br J Sports Med.* 2020;54(18):1081-8.
4. Maniar N, Shield AJ, Williams MD, Timmins RG, Opar DA. Hamstring strength and flexibility after hamstring strain injury: a systematic review and meta-analysis. *Br J Sports Med.* 2016;50(15):909-20.
5. Opar DA, Williams MD, Timmins RG, Hickey J, Duhig SJ, Shield AJ. Eccentric hamstring strength and hamstring injury risk in Australian footballers. *Med Sci Sports Exerc.* 2015;47(4):857-65.
6. Ruddy JD, Shield AJ, Maniar N, et al. Predictive modeling of hamstring strain injuries in elite Australian footballers. *Med Sci Sports Exerc.* 2018;50(5):906-14.
7. Opar DA, Ruddy JD, Williams MD, et al. Screening hamstring injury risk factors multiple times in a season does not improve the identification of future injury risk. *Med Sci Sports Exerc.* 2022;54(2):321-9.
8. Opar DA, Piatkowski T, Williams MD, Shield AJ. A novel device using the Nordic hamstring exercise to assess eccentric knee flexor strength: a reliability and retrospective injury study. *J Orthop Sports Phys Ther.* 2013;43(9):636-40.

9. Timmins RG, Bourne MN, Hickey JT, et al. Effect of prior injury on changes to biceps femoris architecture across an Australian Football League season. *Med Sci Sports Exerc.* 2017;49(10):2102-9.

10. McGrath TM, Hulin BT, Pickworth N, Clarke A, Timmins RG. Determinants of hamstring fascicle length in professional rugby league athletes. *J Sci Med Sport.* 2020;23(5):524-8.

11. Pollard CW, Opar DA, Williams MD, Bourne MN, Timmins RG. Razor hamstring curl and Nordic hamstring exercise architectural adaptations: impact of exercise selection and intensity. *Scand J Med Sci Sports.* 2019;29(5):706-15.

12. Hickey JT, Timmins RG, Maniar N, et al. Pain-free versus pain-threshold rehabilitation following acute hamstring strain injury: a randomized controlled trial. *J Orthop Sports Phys Ther.* 2020;50(2):91-103.

13. Kellis E, Galanis N, Natsis K, Kapetanos G. Validity of architectural properties of the hamstring muscles: correlation of ultrasound findings with cadaveric dissection. *J Biomech.* 2009;42(15):2549-54.

14. van Dyk N, Bahr R, Burnett AF, et al. A comprehensive strength testing protocol offers no clinical value in predicting risk of hamstring injury: a prospective cohort study of 413 professional football players. *Br J Sports Med.* 2017;51(23):1695-702.

15. van Dyk N, Bahr R, Whiteley R et al. Hamstring and quadriceps isokinetic strength deficits are weak risk factors for hamstring strain injuries: a 4-year cohort study. *Am J Sports Med.* 2016;44(7):1789-95.

16. Majumdar A, Bakirov R, Hodges D, Scott S, Rees T. Machine learning for understanding and predicting injuries in football. *Sports Med Open.* 2022;8(1):73.

17. Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell.* 1997;97(1–2):273–324.

18. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn.* 2002;46(1):389-422.

19. Gholamy A, Kreinovich V, Kosheleva O, editors. Why 70/30 or 80/20 relation between training and testing sets: a pedagogical explanation. 2018.

20. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med.* 2013;4(2):627-35.

21. Meeuwisse WH, Tyreman H, Hagel B, Emery C. A dynamic model of etiology in sport injury: the recursive nature of risk and causation. *Clin J Sport Med.* 2007;17(3):215-9.

22. Lövdal SS, Den Hartigh RJR, Azzopardi G. Injury prediction in competitive runners with machine learning. *Int J Sports Physiol Perform.* 2021;16(10):1522-31.

23. Timmins RG, Ruddy JD, Presland J, et al. Architectural changes of the biceps femoris long head after concentric or eccentric training. *Med Sci Sports Exerc.* 2016;48(3):499-508.

24. Franchi MV, Raiteri BJ, Longo S, Sinha S, Narici MV, Csapo R. Muscle architecture assessment: strengths, shortcomings and new frontiers of in vivo imaging techniques. *Ultrasound Med Biol.* 2018;44(12):2492-504.

25. Franchi MV, Fitze DP, Raiteri BJ, Hahn D, Spörri J. Ultrasound-derived biceps femoris long head fascicle length: extrapolation pitfalls. *Med Sci Sports Exerc.* 2020;52(1):233-43.

26. Pollock N, James SL, Lee JC, Chakraverty R. British athletics muscle injury classification: a new grading system. *Br J Sports Med.* 2014;48(18):1347-51.

27. Lu Y, Pareek A, Lavoie-Gagne OZ, et al. Machine learning for predicting lower extremity muscle strain in National Basketball Association athletes. *Orthop J Sports Med.* 2022;10(7):23259671221111742. Epub 20220726. doi: 10.1177/23259671221111742.

28. Opar DA, Timmins RG, Behan FP, et al. Is pre-season eccentric strength testing during the Nordic hamstring exercise associated with future hamstring strain injury? A systematic review and meta-analysis. *Sports Med.* 2021;51(9):1935-45.

29. Luu BC, Wright AL, Haeberle HS, et al. Machine learning outperforms logistic regression analysis to predict next-season NHL player injury: an analysis of 2322 players from 2007 to 2017. *Orthop J Sports Med.* 2020;8(9):2325967120953404.

30. Lee Dow C, Timmins RG, Ruddy JD, et al. Prediction of hamstring injuries in Australian football using biceps femoris architectural risk factors derived from soccer. *Am J Sports Med.* 2021;49(13):3687-95.

FIGURE LEGENDS

Figure 1a. Adopted workflow process to identify important risk factors and build optimal models for performance evaluation.

Figure 1b. The modelling approach for Analysis 1. *d* represents data assessment at different time points where *d1* are data assessed at the start of pre-season; *d2* are data assessed at the end of pre-season; *d3* are data assessed in the middle of in-season. *i* represents hamstring strain injuries (HSIs) that occurred within individual assessment time frames where *i1* are prospective HSIs that occurred in pre-season; *i2* are prospective HSIs that occurred in early in-season; *i3* are prospective HSIs that occurred in late in-season.

Figure 1c. The modelling approach for the Analysis 2. *d* represents data assessment in different time points where *d1* are data assessed at the start of pre-season; *d2* are data assessed at the end of pre-season, *d2-d1* are magnitude of change of data in preseason. *i* represents hamstring strain injuries (HSIs) that occurred within individual assessment time frames where *i2* are prospective HSIs that occurred in early in-season; *i3* are prospective HSIs that occurred in late in-season; *i2+i3* are prospective HSIs that occurred throughout in-season.

Figure 2. The results of Analysis 1. The performance of models built with selected predictors assessed and evaluated at the start of pre-season and hamstring strain injuries (HSIs) that occurred in pre-season (*d1->i1*), end of preseason and HSIs that occurred in early in-season (*d2->i2*), in the middle of preseason and HSIs that occurred in late in-season (*d3->i3*).

Figure 3a. The impact of change in height on hamstring strain injury (HSI) probability in pre-season with other factors set as mean constants (*Age* = 23.54 years, *Muscle thickness* = 2.64 cm).

Figure 3b. The impact of change in age on hamstring strain injury (HSI) probability in pre-season with other factors set as mean constants (*Height* = 188.07 cm, *Muscle thickness* = 2.64 cm).

Figure 3c. The impact of change in muscle thickness on hamstring strain injury (HSI) probability in pre-season with other factors set as mean constants (*Height* = 188.07 cm, *Age* = 23.54 years).

Figure 3d. The impact of change in pennation angle on hamstring strain injury (HSI) probability in early in-season with fascicle length set as mean constant (*Fascicle length* = 10.72 cm).

Figure 3e. The impact of change in fascicle length on hamstring strain injury (HSI) probability in early in-season with pennation angle set as mean constant (*Pennation angle* = 15.37 degrees).

Figure 3f. The impact of change in height on hamstring strain injury (HSI) probability in late in-season with other factors set as mean constants (*Age* = 23.13 years, *Pennation angle* = 15.39 degrees, *Fascicle length* = 10.74 cm, *Relative eccentric knee flexor force* = 5.45 N/kg, *Eccentric knee flexor force imbalance* = 9.33%).

Figure 3g. The impact of change in age on hamstring strain injury (HSI) probability in late in-season with other factors set as mean constants (*Height* = 188.05 cm, *Pennation angle* = 15.39 degrees, *Fascicle length* = 10.74 cm, *Relative eccentric knee flexor force* = 5.45 N/kg, *Eccentric knee flexor force imbalance* = 9.33%).

Figure 3h. The impact of change in pennation angle on hamstring strain injury (HSI) probability in late in-season with other factors set as mean constants (*Height* = 188.05 cm, *Age* = 23.13 years, *Fascicle length* = 10.74 cm, *Relative eccentric knee flexor force* = 5.45 N/kg, *Eccentric knee flexor force imbalance* = 9.33%).

Figure 3i. The impact of change in fascicle length on hamstring strain injury (HSI) probability in late in-season with other factors set as mean constants (*Height* = 188.05 cm, *Age* = 23.13 years, *Pennation angle* = 15.39 degrees, *Relative eccentric knee flexor force* = 5.45 N/kg, *Eccentric knee flexor force imbalance* = 9.33%).

Figure 3j. The impact of change in relative eccentric knee flexor force on hamstring strain injury (HSI) probability in late in-season with other factors set as mean constants (*Height* = 188.05 cm, *Age* = 23.13 years, *Pennation angle* = 15.39 degrees, *Fascicle length* = 10.74 cm, *Eccentric knee flexor force imbalance* = 9.33%).

Figure 3k. The impact of change in eccentric knee flexor force imbalance on hamstring strain injury (HSI) probability in late in-season with other factors set as mean constants (*Height* = 188.05 cm, *Age* = 23.13 years, *Pennation angle* = 15.39 degrees, *Fascicle length* = 10.74 cm, *Relative eccentric knee flexor force* = 5.45 N/kg).

Figure 4a. The performance of models built with selected predictors assessed at start of pre-season (*d1*), end of pre-season (*d2*), start and end of pre-season (*d1, d2*), the magnitude of change of data in pre-season (*d2-d1*), data assessed at the start and end of pre-season and the magnitude of change of data in pre-season (*d1, d2, (d2-d1)*) as predictor variables, and hamstring strain injuries (HSIs) that occurred in early in-season (*i2*) as target variable. AUC = area under the curve

Figure 4b. The performance of models built with selected predictors assessed at start of pre-season (*d1*), end of pre-season (*d2*), start and end of pre-season (*d1, d2*), the magnitude of change of data in pre-season (*d2-d1*), data assessed at the start and end of pre-season and the magnitude of change of data in pre-season (*d1, d2, (d2-d1)*) as predictor variables, and hamstring strain injuries (HSIs) that occurred in late in-season (*i3*) as target variable. AUC = area under the curve

SUPPLEMENTAL DIGITAL CONTENT

SDC 1: Supplemental Material 1.docx

Table S1: The p-value of individual risk factors determined by multivariate logistic regression models in Analysis 1.

Figure 1

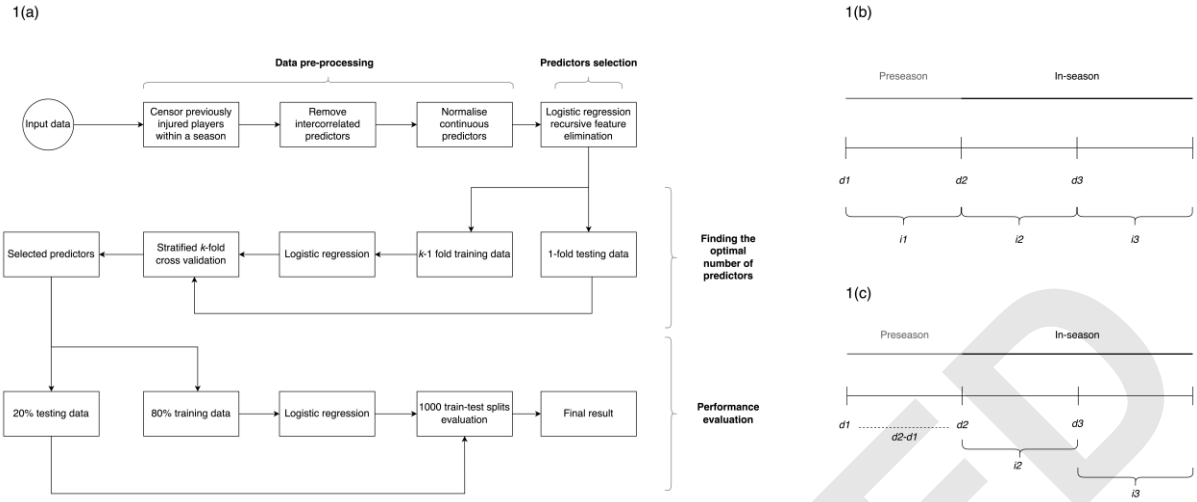

Figure 2

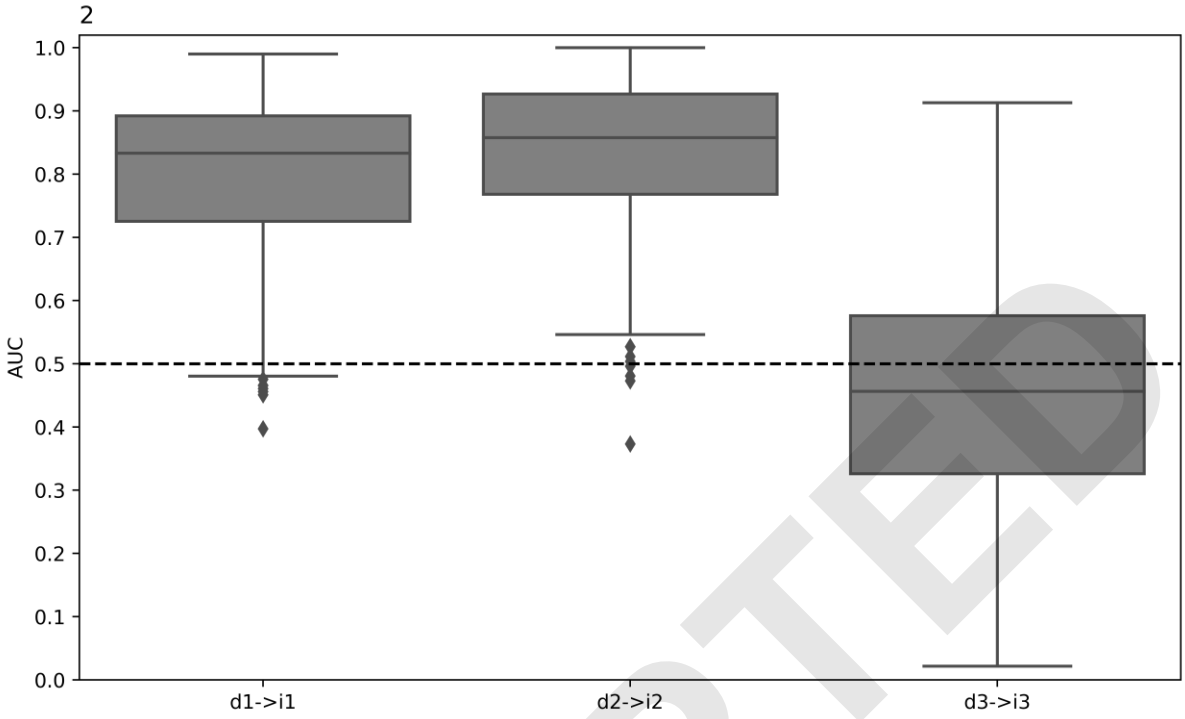

Figure 3

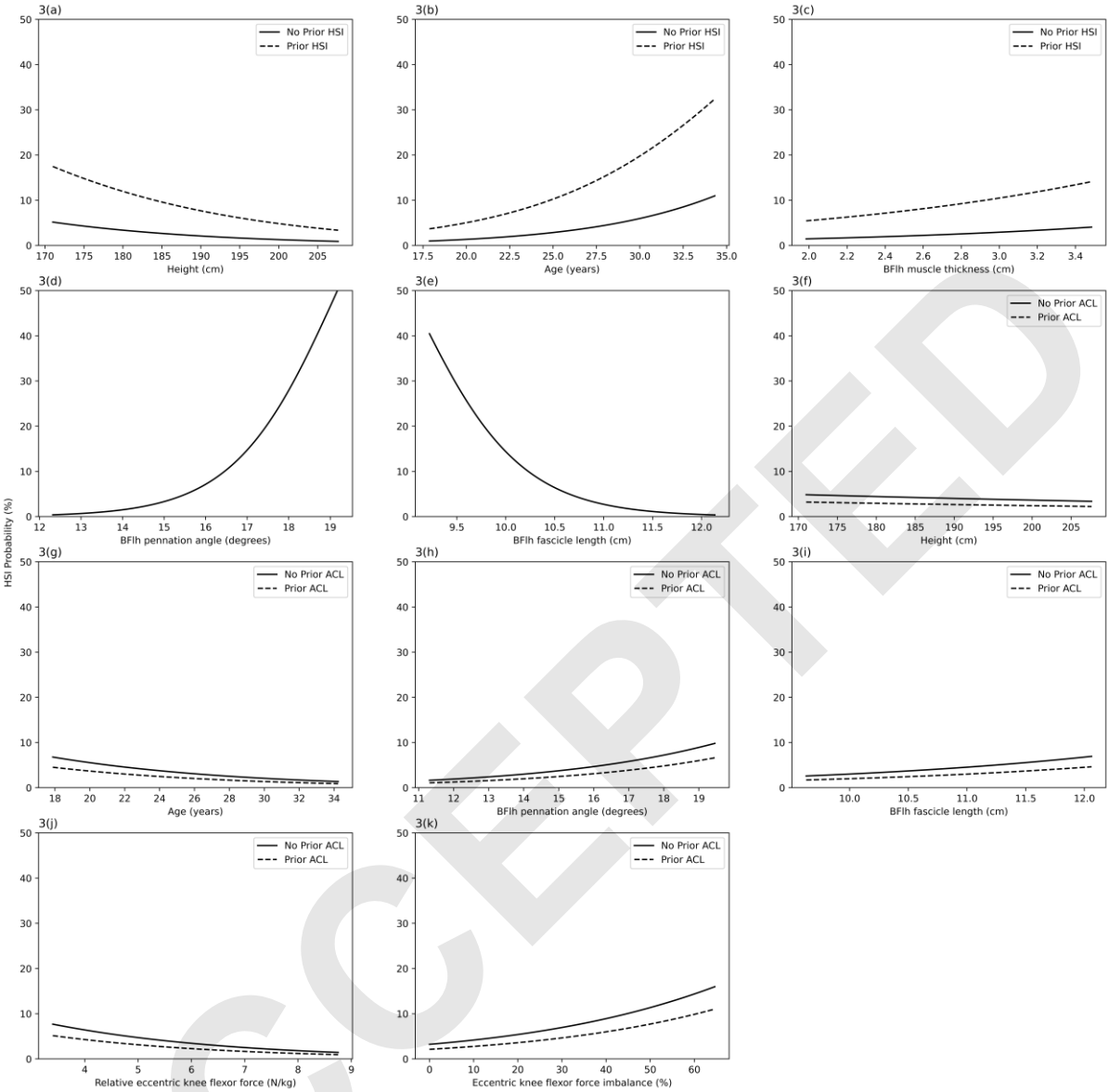

Figure 4

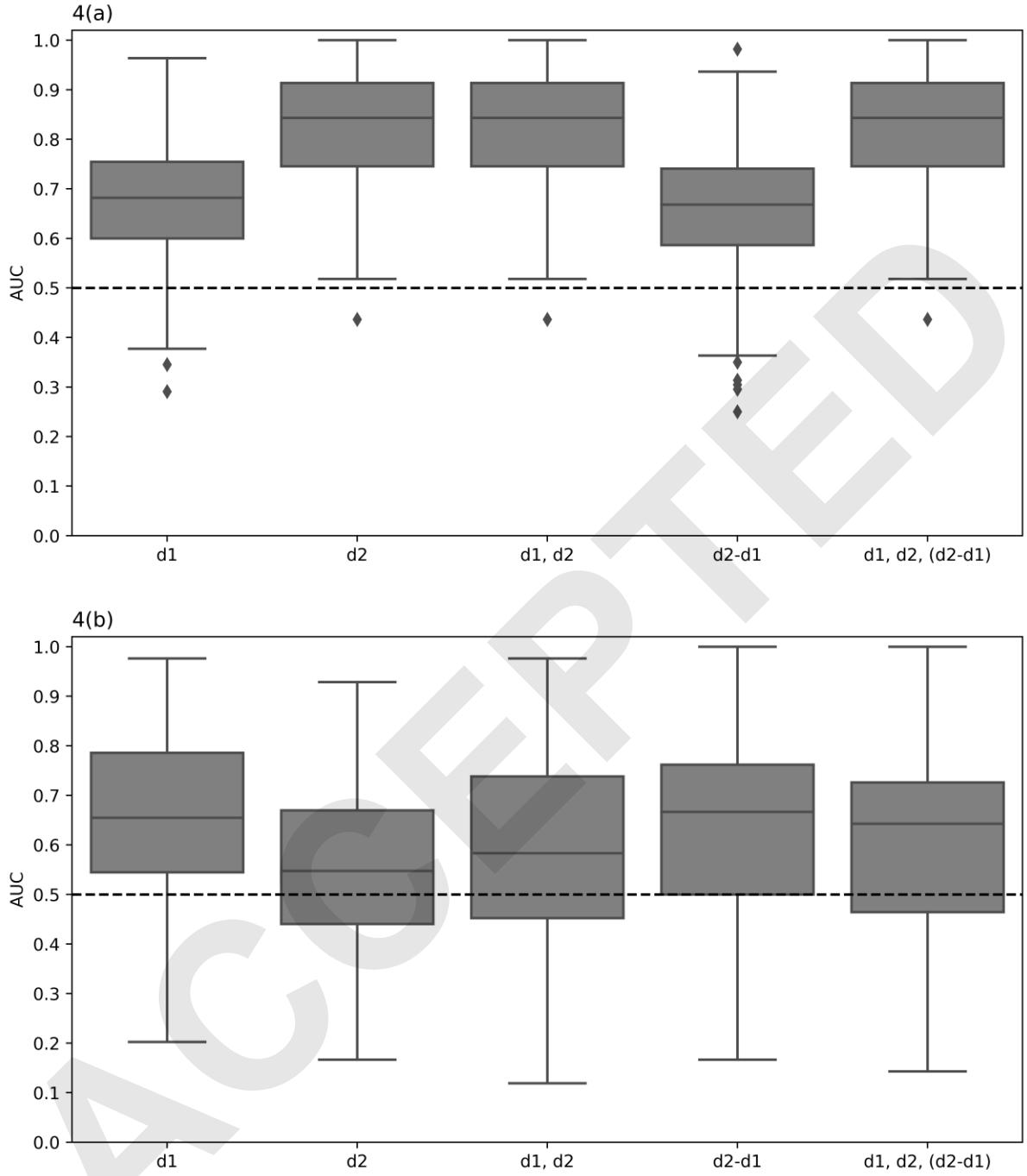

Table 1. Types of predictor variables and target variables included in individual models for Analysis 1 and Analysis 2.

Model	Input predictor variables					Target variables		
	Non-modifiable risk factors	Modifiable risk factors				HSIs		
	d1	d1	d2	d3	d2-d1	i1	i2	i3
Analysis 1								
d1->i1	✓	✓				✓		
d2->i2	✓		✓				✓	
d3->i3	✓			✓				✓
Analysis 2 **HSI occurred in early in-season (i2)**								
d1	✓	✓					✓	
d2	✓		✓				✓	
d1&d2	✓	✓	✓				✓	
d2-d1	✓				✓		✓	

d1&d2&(d2-d1)	✓	✓	✓		✓		✓	
HSI occurred in late in-season (i3)								
d1	✓	✓						✓
d2	✓		✓					✓
d1&d2	✓	✓	✓					✓
d2-d1	✓				✓			✓
d1&d2&(d2-d1)	✓	✓	✓		✓			✓

Non-modifiable risk factors: Height, weight, age, prior hamstring strain injury (HSI), and prior anterior cruciate ligament (ACL) injury;

Modifiable risk factors: muscle thickness (cm), pennation angle (degrees), fascicle length (cm), relative fascicle length (fascicle length divided by muscle thickness), eccentric knee flexor strength (N), relative eccentric flexor strength (N/kg), and between-limb imbalance (%);

d1: data assessed at the start of pre-season, d2: data assessed at the end of pre-season, d3: data assessed in the middle of in-season, d1&d2: data assessed at start and end of pre-season, d2-d1: magnitude of change in data between start and end of pre-season;

i1: HSIs occurred in pre-season, i2: HSIs occurred in early in-season; i3: HSIs occurred in late in-season.

Table 2. The results of Analysis 1. The performance of models built with selected predictors assessed and evaluated at start of pre-season and hamstring strain injuries (HSIs) that occurred in pre-season (*d1->i1*), end of pre-season and HSIs that occurred in early in-season (*d2->i2*), and middle of in-season and HSIs that occurred in late in-season (*d3->i3*). The descriptive summary is the outcome of 1000 iterations of train-test splits.

Model	Risk factors*	Frequency			AUC						
		HSI	Non-HSI	Total	Interquartile range	Standard Deviation	Minimum	Lower quartile	Median	Upper quartile	Maximum
d1->i1	prior HSI, height, age, muscle thickness	14	339	353	0.16	0.12	0.40	0.73	0.83	0.89	0.99
d2->i2	pennation angle, fascicle length	24	259	283	0.16	0.11	0.37	0.77	0.86	0.93	1.00
d3->i3	prior ACL, height, age, pennation angle, fascicle length, relative eccentric knee	11	225	236	0.25	0.17	0.02	0.33	0.46	0.58	0.91

	flexor force, eccentric knee flexor force imbalance										

Performance is measured as area under the curve (AUC).

*Risk factors were selected by recursive feature elimination and 5-fold cross validation.

Table 3. The results of Analysis 2. The performance of models built with selected predictors assessed at start of pre-season (*d1*), end of pre-season (*d2*), start and end of pre-season (*d1, d2*), the magnitude of change of data in pre-season (*d2-d1*), data assessed at the start and end of pre-season and the magnitude of change of data in pre-season (*d1, d2, (d2-d1)*) as predictor variables, and hamstring strain injuries (HSIs) occurred in early in-season (*i2*) as target variable.

Models	Risk factors*	Frequency			AUC						
		HSI	Non-HSI	Total	Interquartile range	Standard Deviation	Minimum	Lower quartile	Median	Upper quartile	Maximum
d1	fascicle length (d1), relative fascicle length (d1)	23	219	242	0.15	0.11	0.29	0.60	0.68	0.75	0.96
d2	pennation angle (d2), fascicle length (d2)	23	219	242	0.16	0.11	0.44	0.75	0.84	0.91	1.00
d1&d2	pennation angle (d2), fascicle length (d2)	23	219	242	0.16	0.11	0.44	0.75	0.84	0.91	1.00
d2-d1	prior HSI, pennation	23	219	242	0.15	0.11	0.25	0.59	0.67	0.74	0.98

	angle (c1), fascicle length (c1), eccentric knee flexor force imbalance (c1)										
d1&d2&(d2-d1)	pennation angle (d2), fascicle length (d2)	23	219	242	0.16	0.11	0.44	0.75	0.84	0.91	1.00

Performance is measured as area under the curve (AUC).

*Risk factors were selected by recursive feature elimination and 5-fold cross validation.

d1&d2; models built with non-modifiable risk factors assessed at the start of pre-season and modifiable risk factors assessed at the start and end of pre-season.

d2-d1; models built with non-modifiable risk factors assessed at the start of pre-season and magnitude of change of modifiable risk factors between start and end of pre-season.

c1; magnitude of change of specific risk factor between start and end of pre-season.

Table 4. The results of Analysis 2. The performance of models built with selected predictors assessed at start of pre-season (*d1*), end of pre-season (*d2*), start and end of pre-season (*d1, d2*), the magnitude of change of data in pre-season (*d2-d1*), data assessed at the start and end of pre-season and the magnitude of change of data in pre-season (*d1, d2, (d2-d1)*) as predictor variables, and hamstring strain injuries (HSIs) occurred in late in-season (*i3*) as target variable.

Models	Risk factors*	Frequency			AUC						
		HSI	Non-HSI	Total	Interquartile range	Standard Deviation	Minimum	Lower quartile	Median	Upper quartile	Maximum
d1	fascicle length (d1), relative fascicle length (d1)	9	210	219	0.25	0.16	0.20	0.54	0.65	0.79	0.98
d2	eccentric knee flexor force imbalance (d2)	9	210	219	0.23	0.16	0.17	0.44	0.55	0.67	0.93
d1&d2	prior ACL, pennation angle (d1), fascicle length (d1)	9	210	219	0.29	0.18	0.12	0.45	0.58	0.74	0.98

d2-d1	prior ACL, pennation angle (c1), fascicle length (c1)	9	210	219	0.26	0.19	0.17	0.50	0.67	0.76	1.00
d1&d2&(d2-d1)	fascicle length (c1)	9	210	219	0.27	0.20	0.14	0.46	0.64	0.73	1.00

Performance is measured as area under the curve (AUC).

*Risk factors were selected by recursive feature elimination and 5-fold cross validation.

d1&d2; models built with non-modifiable risk factors assessed at the start of pre-season and modifiable risk factors assessed at the start and end of pre-season.

d2-d1; models built with non-modifiable risk factors assessed at the start of pre-season and magnitude of change of modifiable risk factors between start and end of pre-season.

c1; magnitude of change of specific risk factor between start and end of pre-season.

Supplemental Material 1. The p-value of individual risk factors determined by multivariate logistic regression models in Analysis 1.

Model	Risk Factors*	p-value
d1->i1	Prior HSI	< 0.01
	Height	0.112
	Age	0.047
	Muscle thickness	0.267
	Intercept	0.948
d2->i2	Fascicle length	< 0.001
	Pennation angle	< 0.001
	Intercept	0.226
d3->i3	Prior ACL	0.999
	Height	0.809
	Age	0.322
	Pennation angle	0.316
	Fascicle length	0.322
	Relative eccentric knee flexor force	0.348
	Eccentric knee flexor force imbalance	0.293
	Intercept	0.448

*Risk factors were selected by recursive feature elimination and 5-fold cross validation.

Prior HSI: prior hamstring strain injury (HSI),

Prior ACL: prior anterior cruciate ligament (ACL) injury,

d1->i1: data assessed at start of pre-season and hamstring strain injuries (HSIs) that occurred in preseason, d2->i2: data assessed at end of pre-season and hamstring strain injuries (HSIs) that occurred in early in-season, d3->i3: data assessed in the middle of in-season and hamstring strain injuries (HSIs) that occurred in late in-season.