BMJ Open Sport & Exercise Medicine

Development of an algorithm-based approach using neuromuscular test results to indicate an increased risk for non-contact lower limb injuries in elite football players

Jente Wagemans ,[1] Arie-Willem De Leeuw,[2] Peter Catteeuw,[3] Dirk Vissers[1]

To cite: Wagemans J, De Leeuw A-W, Catteeuw P, *et al*. Development of an algorithm-based approach using neuromuscular test results to indicate an increased risk for non-contact lower limb injuries in elite football players. *BMJ Open Sport & Exercise Medicine* 2023;**9**:e001614. doi:10.1136/bmjsem-2023-001614

Accepted 9 June 2023

Check for updates

© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

[1]Department of Rehabilitation Science and Physiotherapy, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium
[2]IMEC, University of Antwerp, Antwerpen, Belgium
[3]Royal Antwerp Football Club, Antwerp, Belgium

Correspondence to
Jente Wagemans; jente.wagemans@uantwerpen.be

ABSTRACT

Objectives This retrospective cohort study explored an algorithm-based approach using neuromuscular test results to indicate an increased risk for non-contact lower limb injuries in elite football players.

Methods Neuromuscular data (eccentric hamstring strength, isometric adduction and abduction strength and countermovement jump) of 77 professional male football players were assessed at the start of the season (baseline) and, respectively, at 4, 3, 2 and 1 weeks before the injury. We included 278 cases (92 injuries; 186 healthy) and applied a subgroup discovery algorithm.

Results More injuries occurred when between-limb abduction imbalance 3 weeks before injury neared or exceeded baseline values (threshold≥0.97), or adduction muscle strength of the right leg 1 week before injury remained the same or decreased compared with baseline values (threshold≤1.01). Moreover, in 50% of the cases, an injury occurred if abduction strength imbalance before the injury is over 97% of the baseline values and peak landing force in the left leg 4 weeks before the injury is lower than 124% compared with baseline.

Conclusions This exploratory analysis provides a proof of concept demonstrating that a subgroup discovery algorithm using neuromuscular tests has potential use for injury prevention in football.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ No machine learning models to prevent injuries exist in football.

WHAT THIS STUDY ADDS

⇒ A proof of concept of an algorithm that shows that neuromuscular tests are relevant for injury prevention in football.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This algorithm can be the basis for further developing an injury prevention algorithm in professional football.

INTRODUCTION

In recent years, injuries and injury prevention have become major topics within football. Pooled data from a recent systematic review with meta-analysis showed an injury incidence in professional football players of 8.1 injuries per 1000 hours of exposure.[1] When we look at the location of the injury, the authors ascertained that lower limb injuries have the highest incidence (6.8/1000 hours of exposure).[1] Also, traumatic injuries (5.9/1000 hours of exposure) occur more than overuse (non-contact) injuries (2.4/1000 hours of exposure).[1] Injuries have shown significant physical, psychological and financial consequences.[2]

In football, a number of intrinsic risk factors (older age, kicking/preferred leg) and extrinsic risk factors (lower limb strength imbalances) are known as significant risk factors for developing a lower limb injury.[3 4] Within this topic, multiple articles have been published researching risk factors of these injuries and the effectiveness of injury prevention programmes.[5 6] As sport injuries are mostly a consequence of complex interactions of multiple risk factors,[7] plenty of data has already been collected—with the objective of injury prevention[8] Some examples of the most used and reliable data collection methods for load monitoring in football are: electronic performance tracking system (distance, speed, accelerations,…), physiological monitoring (maximal heart rate, heart rate variability, etc), force plate testing (jump height, power output, etc), blood sampling (cortisol, testosterone, etc), questionnaires (wellness, rate of perceived exertion, etc) and tensiomyography (muscle contraction time, etc).[9] In this abundance of data, many applications of artificial intelligence (AI) and machine learning (ML) are relevant to sports practitioners in this domain.[10]

In sports medicine, AI methods have already been proven beneficial to tackle this challenging multifaceted task of identifying injury risk factors.[11 12] For example, ML methods can be used to improve injury prediction and allow proper approaches to injury prevention. Regarding the popularity of the different ML techniques, tree-based models are currently the most popular. They provide easy visualisation and interpretation of data and aim to be the most cost-effective method.[7] Although these methods signal some risk factors, there is still a need for further exploratory research involving injury prediction to obtain predicting values, better known as ORs ultimately.[7] The goal of this explorative retrospective cohort study is to examine whether neuromuscular test variables might be able to predict a lower limb non-contact injury for elite football players.

MATERIALS AND METHODS

Participants

We carried out a retrospective cohort study following the ethical principles for medical research involving human subjects. We used data from 77 professional players (26.16±4.63 years) of a team competing in the highest Belgian league from 2020–2021 season to November 2022, containing both injured and healthy cases. All players volunteered to participate and gave written consent.

We included players based on the following eligibility criteria: (1) the injury must be located in the lower limb or the lower back; (2) the injury must be due to a non-contact mechanism; (3) the player had to be unfit to play/train for at least 1 day, and he should have received medical attention. Healthy players were included based on the following criterium: no present injury; no data can be used from 2 months prior and after a previous injury occurred—to prevent the influence of recent injuries.

Data acquisition

We extracted results of various neuromuscular assessments from Smartabase (product of Teamworks, Durham, North Carolina, USA), a data-driven platform used by the team to store their gathered data. The baseline was measured for each player at the start of the season. When an injury occurred, four measurements in time were noted: T4, test result 4 weeks before the injury; T3, 3 weeks before the injury; T2, 2 weeks before the injury; T1, 1 week before the injury. This sequence of the injury event as starting point and the four preceding weeks were defined as a 'case'. We used the same principle with healthy cases; one value of the included neuromuscular tests for each week of the 4 weeks before a specified time.

The following neuromuscular tests were performed for each player weekly: eccentric hamstring strength test, isometric adduction/abduction strength test and countermovement jump test. The neuromuscular tests were evaluated with the following testing systems: NordBord (NB), ForceFrame (FF) and ForceDecks (FD) (VALD Performance, Newstead, Australia). Table 1 depicts the measured outcome variables. For each measured variable, an average of three measurements were automatically calculated by the Smartabase software. Additionally, we calculated between limbs and the relative value of each variable concerning the baseline.

Test protocols

Eccentric hamstring strength

The eccentric hamstring strength test was performed using a VALD Performance NB testing system. The player takes a kneeling position with both knees on the board flexed 90°, with the ankles attached and the arms crossed in front of the chest. The player extends the hips and looks forwards, holding a straight posture. The player moves forward by extending the knees and descending slowly while maintaining a straight posture. During this motion, the arms stay crossed on the upper body.

Table 1 Neuromuscular tests, test systems and their respective outcome variables

Neuromuscular tests	Test system (VALD)	Measured variables	Added variable
Eccentric hamstring strength	NordBord	Left eccentric hamstring force (N)	Left/right imbalance (%)
		Right eccentric hamstring force (N)	
Isometric hip adductor and abductor strength	ForceFrame	Left adduction force (N)	Left/right adduction imbalance (%) Left adduction/abduction imbalance (%) Left/right abduction imbalance (%) Right adduction/abduction imbalance (%)
		Right adduction force (N)	
		Left abduction force (N)	
		Right abduction force (N)	
Countermovement jump	ForceDecks	Jump height (cm)	n/a
		Left peak landing force (N)	Left/right peak landing force imbalance (%)
		Right peak landing force (N)	
		Left take-off peak force (N)	Left/right take-off peak force imbalance (%)
		Right take-off peak force (N)	

n/a, not applicable; N, Newton.

Wagemans J, *et al. BMJ Open Sp Ex Med* 2023;**9**:e001614. doi:10.1136/bmjsem-2023-001614

Isometric hip adductor and abductor strength

The isometric hip adductor and abductor strength tests were performed using the VALD Performance FF strength testing system. This test is performed in a standing position where the player performs isometric abduction and adduction of the leg by pushing the distal part of the lower leg against the force detectors.

Countermovement jump

The countermovement jump test was conducted using the VALD performance FD force platform. Players were instructed to initiate a rapid downward motion, rapidly followed by an explosive upward motion, jumping as high as possible. After each jump, land with both feet on the FD force platforms.

Data management

After collecting all data, we calculated the missing data percentages for the different neuromuscular testing variables. Hereafter, we imputed the missing values using three different methods.

1. Mean method: the average of the available data of that period per test was taken, and this outcome was entered into the missing values.
2. Forward–backward method: depending on which week was missing, the data from 1 week before or 1 week after the given date was duplicated. When T1 and T3 were given, the backward method was used for both missing values. So, T2 is completed by taking T1 and T4 by using the values from T3. When T2 and T3 are known, the forward method is used for T1 and the backward method is used for T4. When periods T1 and T4 were given, T2 was completed using the backward method and T3 was completed using the forward method. When only one period was not given, we used the backward method for periods T2 and T4 and the forward method for T1 and T3.
3. Linear regression method: when data are missing between 2 weeks, the average of the two surrounding periods is taken, and this average is calculated. This value is used for the remaining 3 weeks when only 1 week's data are available. If only T4 is missing, a backward method from T3 is performed. If T1 is missing, a forward method is performed from T2. If T1 and T2 are missing, a forward method from T3 is performed and if T3 and T4 are missing, a backward method from T2 is performed. An exception can be found when missing T2 and T3, and then a linear evolution is determined using these formulas:
   - $T3=T4+(T1-T4)/3$
   - $T2=T1-(T1-T4)/3$

Each method resulted in a complete dataset; therefore, we have three separate datasets used for further analysis.

Data analysis

We used the data mining technique Subgroup Discovery, a descriptive, exploratory technique that can handle relatively small datasets, as considered here. With this technique, we can extract easy-interpretable findings that have proven useful in sport-specific settings.[13–17]

This technique aims to find subsets with a different distribution of the outcome variable than the distribution in the entire data collection. In this case, the outcome variable is binary and equals 1 or 0, corresponding to an injured or healthy person, respectively. The obtained subgroups are subsets of the entire dataset that contain the instances that satisfy single or multiple conditions on the independent variables or so-called features, for example, FF abduction strength imbalance or FD peak landing force at T4. The results are considered significant if there is less than a 5% probability that the result is a consequence of considering too many features.[18]

We used the subgroup discovery tool Cortana for our analysis (https://datamining.liacs.nl/cortana.html). We present all significant subgroups described by a restriction on a single feature. Moreover, we also performed a more in-depth analysis by allowing a combination of restrictions on two features.

RESULTS

A total of 92 injuries were included in the study, representing 56 players (mean age: 26.05±4.80 years, weight: 79.28±8.53 kg, height: 184.45±7.72 cm, body mass index (BMI): 23.26±1.49). Most cases represent upper leg injuries (30/92), followed by groin (20/92) and knee injuries (19/92). Only 12 ankle injuries occurred, while lower limb injuries (7/92) and lower back injuries (4/92) are the least represented. Healthy cases were added to the dataset to allow for comparative analysis. This resulted in a total of 186 healthy cases, representing 21 players (mean age: 26.46±4.29 years, weight: 79.13±8.34 kg, height: 183.00±7.31 cm, BMI: 25.59±1.63) (table 2).

Subgroup discovery with one restriction

Table 3 depicts the statistically significant results for the three datasets. Datasets are distinguished based on the three missing data-filling techniques. The largest proportion of missing data was found in the FD assessments (34.67%), while 18.06% of data was missing in FF variables and 17.90% in NB variables, respectively. The dataset of missing data filled by the mean method contains six statistically significant results, while the two other datasets have three statistically significant findings. Only restrictions on two variables have a statistically significant injury risk score across all three datasets: increased abduction strength imbalance between limbs 3 weeks before injury compared with baseline (threshold: 0.97); decreased right leg adduction strength 1 week before injury compared with baseline (threshold: 1.01). There was one feature with a statistically significant injury risk score across two datasets—mean method and linear regression: Decreased left leg adduction strength 4 weeks before injury compared with baseline (threshold: 1.08).

Table 2 Characteristics of the included players and type of injuries

	Injured (n=56)	Healthy (n=21)
Age (years)	26.05±4.79 (17–37)	26.46±4.29 (19–37)
Height (cm)	184.45±7.72 (166–197)	183.00±7.31 (166.5–196)
Weight (kg)	79.28±8.53 (61.4–96.4)	79.13±8.34 (63.2–95.8)
Body mass index	23.26±1.49 (20.21–25.60)	25.59±1.63 (20.60–28.41)
Injury	N=92	
Lower back/pelvis		
Non-specific low back pain	2	
Lumbar disc injury	2	
Groin pain		
Adductor related	16	
Psoas related	3	
Rectus femoris related	1	
Upper leg		
Quadriceps injury	8	
Hamstring injury	22	
Knee		
ACL	2	
Patella tendinopathy	4	
Patellofemoral pain syndrome	6	
Medial collateral ligament	3	
Meniscus injury	2	
Osteochondral lesion	2	
Lower leg		
Soleus injury	5	
Gastrocnemius injury	1	
Achilles tendinopathy	1	
Ankle		
Tendon injury	4	
Medial ankle sprain	1	
Lateral ankle sprain	3	
Syndesmosis injury	2	
Osteochondral lesion	1	
Fracture	1	

Subgroup discovery with two restrictions

Additionally, to a single restriction, we also investigated subgroups described by restrictions on two features. If we compare the results of the three datasets, we found that the subgroup with the largest deviation from the distribution of the injury score in the entire data collection is the combination of increased abduction imbalance with respect to the baseline 3 weeks before injury (threshold: 0.97) and decreased peak landing force in the left leg 4 weeks before injury (threshold: 1.24). The subgroup contained, on average, 148 cases over the three datasets. When these conditions were met, we found roughly 50% injured players compared with only 33% in the data collection.

Table 3 Significant results for subgroup discovery with one restriction

Feature	Time point	Threshold	Subgroup (%)
Filling missing data: mean method			
Take off-peak force right	T2	≥0.98	76/176 (43%)
Abduction strength imbalance*[1]	T3	≥0.97	78/184 (42%)
Take off-peak force right	T3	≥0.96	81/195 (42%)
Take off-peak force left	T2	≥0.98	76/181 (42%)
Adduction strength left†	T4	≤1.08	77/185 (42%)
Adduction strength right*[2]	T1	≤1.01	61/138 (44%)
Filling missing data: forward–backward method			
Abduction strength imbalance*[1]	T3	≥0.97	80/187 (43%)
Adduction strength right*[2]	T1	≤1.01	80/131 (61%)
Abduction strength imbalance	T2	≥0.97	80/198 (40%)
Filling missing data: linear regression method			
Abduction strength imbalance*[1]	T3	≥0.97	80/189 (42%)
Adduction strength right*[2]	T1	≤1.01	60/131 (46%)
Adduction strength left†	T4	≤1.08	75/182 (41%)

Subgroup represents the number of injuries in relation to the total amount of cases.
*Significant across all three missing data-filling techniques.
†Significant across two missing data-filling techniques.

DISCUSSION

This study examined if neuromuscular test variables can predict a lower limb non-contact injury for elite football players. We performed an ML technique—subgroup discovery—on retrospective data regarding occurred injuries and neuromuscular assessments performed throughout the 2020–2021 and 2021–2022 seasons, and the 2022–2023 season until December 2022 from professional football players, to identify predictors that could indicate an increased injury risk. The key result is that we found significant predictors, based on neuromuscular tests, that could signal an increased risk of injury.

Results and clinical implications

Hitherto, published research regarding ML to predict injuries in professional football is scarce.[19] Hecksteden *et al* ascertained that screenings and data collection yields promising results for predicting non-contact lower limb injuries.[20] Our findings align with this study—we prove that ML can be used for the above-mentioned purposes. Bahr criticised preseason screening as unfit for injury prediction due to lack of substantiation.[21] Nevertheless, screening can still highlight key risk factors for sports injuries.[22] The current literature comprises similar proofs of concept in the NBA[11 12] or applied to a specific injury—ACL injury.[23] The study of Cohan *et al* predicted injury based on the injury mechanism, player's characteristics and game statistics,[11] while Lu *et al* focused mostly on injury history and past concussions.[12] Jauhiainen *et al* included an extensive screening protocol comprising neuromuscular and functional tests. Their results show that they could not predict ACL injuries in clinical practice.[23]

We found that increased abduction imbalance for the baseline 3 weeks before the injury and decreased right adduction for the baseline 1 week before injury were significant predictors of risk of injury. The analysis with two restrictions showed significant predictive value for the combination of increased abduction imbalance for the baseline 3 weeks before the injury and decreased peak landing force in the left leg 4 weeks before the injury for risk of injury. Concretely, if the test results of a player would highlight one of the thresholds mentioned above, a red flag should be raised that the player is more prone to injury. This would enable the technical and medical staff to alter training and rehabilitation programmes following the respective red flag.

The increase between-leg abduction imbalance and decreased adduction muscle strength follow the proportion of groin injuries in the current sample. Lower adductor strength shows to yield a higher risk of a groin injury.[24] Considering the number of hamstring injuries in the current sample, one would expect a significant finding regarding the eccentric hamstring strength test. In the recent literature, the Nordic hamstring test has come out as one of the best predictors of a hamstring strain injury.[25] Another article executed by various Premier League teams found that more than 15% asymmetry is a significant predictor for a hamstring injury.[26] Both articles found that decreased eccentric hamstring strength is a predictor for an injury as opposed to our results.

Subgroup discovery

Our data mining approach obtained results that can be applied clinically. More specifically, this analysis provides specific thresholds that can be used as a red flag for increased risk of injury. The medical and technical staff can include these values in their daily programmes. Second, using a subgroup discovery algorithm, we obtained that a combination of features is even more relevant. This follows previous studies that demonstrated that injuries result from complex interactions of features that interrelate.[7] Determining combinations of features, and recognising patterns, can help with injury prevention.

Limitations

Our study is not without limitations. First, we had a considerable amount of missing data and recommendations regarding which imputation method should be used ambiguously. We chose three single imputation methods within our study: mean value substitution, last observation carried forward/backwards and linear regression method.[27] After our research, neither one of those methods proved to be superior or inferior. The main statistical finding remained the same over the three datasets. It has to be noted that single imputation methods could implement biased estimates even if data are missing completely at random. In case of fewer missing data, we recommend using multiple imputation procedures. These will create more reliable results and a better understanding of the impact of missing data.[27]

Second, the included non-contact injuries represented various injuries—from muscle strains to fractures. Even though groin injuries and hamstring injuries account for almost half of the included injuries, such equivocality can underpin confounding factors. Furthermore, it would be interesting to distinguish injuries based on anatomical area and perform the subgroup discovery analysis for each type of injury separately. Unfortunately, our dataset was too small for this kind of analysis.

Third, we included only a selection of potential predictors in agreement with the medical and technical staff of the club. More salient outcome variables, such as rate of force development, reactive strength index, power and stiffness, were not included. Considering the aim of this study—to develop a proof of concept—this is only a minor limitation. However, further development of this model should include evidence-based outcome variables.

Lastly, these results should be considered cautiously: our analysis was performed on a small dataset, our ascertained thresholds are based on average data over all participants instead of individually, we included only a small number of variables and we limited our time frame to 4 weeks before the injury; as such, the current results are not clinically applicable. Nevertheless, this dataset served our purposes, since we could determine that our model can identify red flags that indicate an increased risk of injury. We need a more robust dataset with more variables for further development to a clinically applicable model.

CONCLUSION

The use of AI for injury prediction and, in the end, prevention is a new but fast-growing research area. The lack of consensus regarding which variables to use for data collection, the imputation method for missing data and the data mining technique make it challenging to create a prediction model. We used a subgroup discovery algorithm, an explorative data mining technique, to examine possible predictive values in the data set. This exploratory analysis provides insights into important parameters that might indicate increased injury risk. Possible avenues for future research are more in-depth analysis with bigger sample sizes, including more features in the model, determining causality and constructing data-driven injury risk scores.

Twitter Jente Wagemans @jentewagemans

Acknowledgements We gratefully acknowledge the contribution of Professor Dr Tim Verdonck, Professor at the University of Antwerp, for proofreading this manuscript. We also acknowledge the contribution of Arno Lamberts, Frederik Goosens and Aline Delbaen, Master students in Rehabilitation Sciences and Physiotherapy at the University of Antwerp, for acquiring data.

Contributors All authors—JW, A-WDL, PC and DV—were involved in developing the concept and design of this working, drafting and revising the manuscript and gave final approval for publishing this version; all authors accept responsibilty for the overall content as the guarantors of this manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by University of Antwerp: UZA—ID: 5048. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

Author note Our research study is conducted in a professional football population with players from all continents, young and old, and from various socioeconomic backgrounds. Our research team comprises different disciplines: physiotherapy, engineering sciences and sports sciences. One author is in an early career stage, one in a mid-career stage, one working in the field and one a senior lecturer.

ORCID iD

Jente Wagemans http://orcid.org/0000-0001-7856-463X

REFERENCES

1 López-Valenciano A, Ruiz-Pérez I, Garcia-Gómez A, *et al*. Epidemiology of injuries in professional football: a systematic review and meta-analysis. *Br J Sports Med* 2020;54:711–8.
2 Ekstrand J, Hägglund M, Waldén M. Injury incidence and injury patterns in professional football: the UEFA injury study. *Br J Sports Med* 2011;45:553–8.
3 Hägglund M, Waldén M, Ekstrand J. Risk factors for lower extremity muscle injury in professional soccer: the UEFA injury study. *Am J Sports Med* 2013;41:327–35.
4 Croisier J-L, Ganteaume S, Binet J, *et al*. Strength imbalances and prevention of hamstring injury in professional soccer players: a prospective study. *Am J Sports Med* 2008;36:1469–75.
5 Nuhu A, Jelsma J, Dunleavy K, *et al*. Effect of the FIFA 11+ soccer specific warm up programme on the incidence of injuries: a cluster-randomised controlled trial. *PLoS One* 2021;16:e0251839.
6 Al Attar WSA, Soomro N, Sinclair PJ, *et al*. Effect of injury prevention programs that include the nordic hamstring exercise on hamstring injury rates in soccer players: a systematic review and meta-analysis. *Sports Med* 2017;47:907–16.
7 Bittencourt NFN, Meeuwisse WH, Mendonça LD, *et al*. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition-narrative review and new concept. *Br J Sports Med* 2016;50:1309–14.
8 Van Eetvelde H, Mendonça LD, Ley C, *et al*. Machine learning methods in sport injury prediction and prevention: a systematic review. *J Exp Orthop* 2021;8:27.
9 Claudino JG, Cardoso Filho CA, Boullosa D, *et al*. The role of veracity on the load monitoring of professional soccer players: a systematic review in the face of the big data era. *Appl Sci* 2021;11:6479.
10 Rein R, Memmert D. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *Springerplus* 2016;5:1410.
11 Cohan A, Schuster J, Fernandez J. A deep learning approach to injury forecasting in NBA basketball. *JSA* 2021;7:277–89.
12 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;10:23259671221111742.
13 de Leeuw A-W, van der Zwaard S, van Baar R, *et al*. Personalized machine learning approach to injury monitoring in elite volleyball players. *Eur J Sport Sci* 2022;22:511–20.
14 Knobbe A, Orie J, Hofman N, *et al*. Sports analytics for professional speed skating. *Data Min Knowl Disc* 2017;31:1872–902.
15 de Leeuw AW, Hoekstra A, Meerhoff L, *et al*. Machine learning and knowledge discovery in databases 2020. In: *Tactical Analyses in Professional Tennis*. Cham: Springer International Publishing, 2020.
16 de Leeuw A-W, Meerhoff LA, Knobbe A. Effects of pacing properties on performance in long-distance running. *Big Data* 2018;6:248–61.
17 de Leeuw A-W, van Baar R, Knobbe A, *et al*. Modeling match performance in elite volleyball players: importance of jump load and strength training characteristics. *Sensors (Basel)* 2022;22:7996.
18 Duivesteijn W, Knobbe A. Exploiting false discoveries -- statistical validation of patterns and quality measures in subgroup discovery. 2011 IEEE 11th International Conference on Data Mining;
19 Majumdar A, Bakirov R, Hodges D, *et al*. Machine learning for understanding and predicting injuries in football. *Sports Med Open* 2022;8:73.
20 Hecksteden A, Schmartz GP, Egyptien Y, *et al*. Forecasting football injuries by combining screening, monitoring and machine learning. *Science & Medicine in Football* 2022:1–15.
21 Bahr R. Why screening tests to predict injury do not work-and probably never will…: a critical review. *Br J Sports Med* 2016;50:776–80.
22 Verhagen E, van Dyk N, Clark N, *et al*. Do not throw the baby out with the bathwater; screening can identify meaningful risk factors for sports injuries. *Br J Sports Med* 2018;52:1223–4.
23 Jauhiainen S, Kauppi J-P, Krosshaug T, *et al*. Predicting ACL injury using machine learning on data from an extensive screening test battery of 880 female elite athletes. *Am J Sports Med* 2022;50:2917–24.
24 Bakken A, Targett S, Bere T, *et al*. Muscle strength is a poor screening test for predicting lower extremity injuries in professional male soccer players: a 2-year prospective cohort study. *Am J Sports Med* 2018;46:1481–91.
25 Shalaj I, Gjaka M, Bachl N, *et al*. Potential prognostic factors for hamstring muscle injury in elite male soccer players: a prospective study. *PLoS One* 2020;15:e0241127.
26 McCall A, Carling C, Davison M, *et al*. Injury risk factors, screening tests and preventative strategies: a systematic review of the evidence that underpins the perceptions and practices of 44 football (soccer) teams from various premier leagues. *Br J Sports Med* 2015;49:583–9.
27 Jamshidian M, Mata M. 2 - advances in analysis of mean and covariance structure when data are Incomplete**This research was supported in part by the National Science Foundation grant DMS-0437258. In: Lee S-Y, ed. *Handbook of Latent Variable and Related Models*. Amsterdam: North-Holland, 2007: 21–44.