Author’s response to reviews

**Title:** A Study Protocol for the Development and Internal Validation of a Multivariable Prognostic Model to Determine Lower Extremity Muscle Injury Risk in Elite Football (Soccer) Players, with Further Exploration of Prognostic Factors

**Authors:**

Tom Hughes (tom.hughes.physio@manutd.co.uk)

Richard Riley (r.riley@keele.ac.uk)

Jamie Sergeant (jamie.sergeant@manchester.ac.uk)

Michael Callaghan (michael.callaghan@mmu.ac.uk)

**Version:** 1  **Date:** 24 May 2019

**Author’s response to reviews:**

Physiotherapy Dept, Manchester United Football Club,

AON Training Complex,

Birch Road,

Off Isherwood Road,

Carrington,

Manchester.

M31 4BH.

21st May 2019

Dear Dr Tzoulaki,

We would like to take the opportunity to extend our thanks to both yourself and the other reviewer for taking the time to consider our manuscript for publication in Diagnostic and Prognostic Research. We have found both reviewer’s comments very helpful and feel that by addressing the issues outlined we now have a stronger paper for your consideration.

Please find our replies (in bold type) to the comments suggested (in italic type), which explain how they have been dealt with in the revised paper. All amended text in the manuscript has been highlighted in yellow. We have copied and pasted excerpts of the text in the replies below where appropriate, to identify the changes made.

Thank you once again for considering this manuscript. We are looking forward to hearing from you in due course.

Yours truly,

Tom Hughes
The purpose of this manuscript is to describe the protocol of a study which aims to develop a prognostic model for using periodic health assessment data to determine individual muscle injury risk profiles of professional soccer players. The authors have developed a very interesting research project and I am looking forward to reading about the findings of this research study in the future. In addition, I do want to take this opportunity to applaud the authors' effort to increase transparency and openness in research by deciding to publish a protocol paper. This manuscript is very high quality and the writing is clear. The development of the prognostic model is well defined and based on solid scientific methods.

Thank you so much for your positive comments and taking the time to consider our manuscript. We feel that increasing the transparency of sports medicine research is vital and this is something we are keen to promote. After conducting the initial systematic review for my PhD project, it is clear that a lot of evidence relating to prognostic (risk) factors in football suffer from issues that could introduce bias. By identifying these issues, it has helped us with planning this study and to attempt to conduct the project according to the gold standard methodology in this field, starting with this protocol.

I do have some comments, which follow.

TITLE
The title is long, which makes it confusing. This being a protocol paper, I am not sure if it of any value to include "The Value of Periodic Health Examination for Injury Prediction in Elite Football (soccer)" in the title. Unless that is the name of the study, which does not come across from the manuscript. To increase clarity, I would advise to either start or end the title with "study protocol." For example: "A Study Protocol for Development and Internal Validation...." or "Determining the Value of Periodic Health Examination for Injury Prediction in Elite Football (soccer): A Study Protocol." Or using the study title registered for ClinicalTrials.gov: "The Role of Periodic Health Examination in Determining Lower Extremity Muscle Injury Risk in Elite Football (Soccer): A Study Protocol." These are just some examples to clarify my point.

Thank you for your opinion regarding the title. We found this quite difficult to develop, given all the elements that are required in a title to conform to the TRIPOD reporting guidelines. We completely agree with your suggestion of changing the title, which has now been amended and changed to:

“A Study Protocol for the Development and Internal Validation of a Multivariable Prognostic Model to Determine Lower Extremity Muscle Injury Risk in Elite Football (Soccer) Players, with Further Exploration of Prognostic Factors” (Page 1)

ABSTRACT

I am not comfortable with the term of "retrospective review" with this methodology as I associate that with determining coverage after treatment. I would suggest rephrasing the first sentence in the "METHODS." For example: "This is a protocol for a retrospective cohort study. The injury and PHE data were collected…"

Thank you for your suggestion regarding this point. As requested the text has now been amended to:
“This is a protocol for a retrospective cohort study. PHE and injury data were routinely collected over 5 seasons (1st July 2013 to 19th May 2018), from a population of elite male players aged 16-40 years old.” (page 3, row 63)

BACKGROUND
Page 5, row 106
I would include commas in the numbers: 17,000 and 20,000.

Thank you for identifying this omission. As requested, this has now been changed to:

“the daily cost to a participating team in the UEFA Champions League, is approximately €17,000 to €20,000” (page 5, row 105)

Other than that minor detail, the Background is well written and clearly underlines the importance of this research. Relevant references are used.

METHODS
Study Design
It is unclear to me why the study was limited to players aged 16 to 40 years? I would understand if those were the age limits of the cohort but I assume that is not the case as age is mentioned as an inclusion criteria.

Thank you for highlighting this and sorry that this is unclear. Indeed, the age range of 16-40 represented the actual age range for full-time elite players at this club during the data collection process (and this is also representative of the typical age range at other clubs). To aid clarity, we
have explained that these were the age limits of the cohort and also removed the age range from the inclusion criteria. Changes have been made to the following:

A) Study Design subsection in the Methods section (page 6, row 134):

“This study will be of retrospective cohort design, using a population of male elite football players aged 16-40 years old, who were employed on a full-time basis at an English Premier League club.”

B) Participants and eligibility subsection in the Methods section (page 9, row 193):

“During any season, participants were eligible for inclusion into the analysis if they: 1) had an outfield position (i.e. not a goalkeeper); 2) participated in PHE testing for the relevant season. Participants were excluded from the analysis for any season if they were a triallist player or not contracted to the club at the time of PHE. (Page 9, row 193)”

C) Abstract (page 3, row 63):

“PHE and injury data were routinely collected over 5 seasons (1st July 2013 to 19th May 2018), from a population of elite male players aged 16-40 years old.”

Data sources

This part mentions "both studies" but it is unclear what the two studies are? Are the two objectives considered as separate studies?

Again, sorry this is unclear. We consider this project as one study with two parts: (1) the development and internal validation of a prognostic model and (2) exploration of prognostic factors.
Preseason PHE data collection

On row 151 it is mentioned that "Typically, the musculoskeletal and ..." Does this mean that the protocol for PHE varied between the five years? If that is the case, what were the changes in protocol?

Thank you for this comment. Most tests were consistently performed across all five seasons, although some were not because of changes to the PHE procedure. We have acted upon your comments and the second reviewer’s comments regarding missing data. Thus, we have now incorporated a new subsection in the methods specifically covering missing data and any tests that varied across seasons have been outlined with reasons for this. This section is now located on page 13, row 286 and states:

“As presented in Table 1, all medical history and age factors were complete (23 factors). Of the 37 remaining candidate PFs, the proportion of missingness ranged from 5.68% (for height and weight) to 76.34% (for body fat). Eleven of these had >15% missing observations (which included body fat, toe touch in standing, sacroiliac kinematic function, all Y Balance Test and upper body peak power variables). For these factors, the large degree of missingness was because of procedural changes in the PHE process, which meant that these tests were not conducted across all seasons.

For candidate PFs with < 15% missing observations, all tests were conducted consistently across all 5 seasons. For these factors, the sample characteristics of cases with complete PF data were compared to incomplete cases which had at least one missing observation (Table 2).”
With the PHE being mandatory, were the players free from injury at the time of PHE?

Thank you for highlighting this omission. This was included on the original manuscript, but for some reason was deleted during the editing process. This has now been reintroduced and it located on page 8, row 162:

“If a participant was injured at the time of PHE, a risk assessment was completed by medical staff. In such instances, participants completed only tests that were deemed appropriate and safe for the participant’s condition; examiners were therefore not blinded to the injury status of participants.”

Detailed descriptions of all the measurements and tests included in the PHE need to be included in this section, including the manufacturers of the measurement tools. This is important for replication. Please also reference previous research if the test protocols are based on published studies.

Thank you for highlighting this. We originally included this as a section in the draft manuscript, although we elected to remove it in the first instance due to the length. We agree that for increased transparency and for potential future replication, this is an important element to include. As such, we have included this section in its entirety. However, because of the level of detail included it is around 3500 words. We would value the editor’s input and advice, but we have decided that rather than include this in the main manuscript, it should be included as an Additional File (see attached ‘Additional File 1- Detailed descriptions of all PHE tests.’). We have also elected not to cut and paste the excerpt section into this reply letter, but instead have included a line in the main text referring the reader to this file, and can be located at page 7, row 151. This now states:

“Typically, the musculoskeletal and performance components of the PHE included: 1) anthropometric measurements; 2) medical history (i.e. previous injury history); 3) musculoskeletal examination tests; 4) functional movement and balance tests; 5) strength and power tests. Detailed descriptions of all tests are provided in Additional File 1.”
Instead of "club doctors" I would use "club medical doctors" throughout the manuscript. 

Thank you for identifying this. We have amended the text accordingly and now feel this aids clarity. The text has been changed in several locations:

“Each component of the PHE test battery was standardised according to a written protocol and examined by physiotherapists, sports scientists or club medical doctors.” page 7, row 158.

“For every player in the squad, any injuries that occurred during the season were assessed and electronically documented within 24 hours by a club medical doctor or physiotherapist in accordance with the Consensus Statement on Injury Definitions and Data Collection Procedures in Studies of Football Injuries” page 8, row 172.

“Ultrasound scans were performed by the club medical doctor using a Toshiba Aplio 500 or 1900 machine (Toshiba Corporation, Tokyo, Japan). Magnetic Resonance Imaging (MRI) was performed as appropriate, using a Canon Vantage Titan 3T Scanner (Canon Medical Systems, Otowara, Japan) according to sequences determined by the club medical doctor. Images were evaluated by a club medical doctor and an independent musculoskeletal radiologist.” page 8, row 178.

“An IMI was confirmed during the injury assessment procedure outlined above, and graded by the club medical doctor or physiotherapist according to the Munich Consensus Statement for the Classification of Muscle Injuries in Sport.” page 10, row 226.

Ethics and Data Use

Page 9, row 197
If an ethics approval number has been assigned to this by the University of Manchester, please add.

Because this study will utilise data that has already been recorded as a mandatory condition of the player’s employment, The Research Ethics Service at the University of Manchester requested only a detailed letter of permission to use data from the Football Club in a research study. This letter sufficiently covered all aspects required by the Research Ethics Service, although an ethics approval number was not provided (see attached files in submission inventory).

Data extraction

It is unclear why all this work with the data has already been done, before the development of the protocol.

Thank you for identifying this. Because a retrospective dataset was used, we had to audit the data prior to writing the protocol to establish whether we had enough outcomes and candidate prognostic factors for model development. After reading the section again, we can see that this was not explained to the reader. As such, we have now amended the text in two locations, which now states:

“The extracted data were audited in parallel with the development of this protocol to determine the number of available index IMI events in the dataset. This was essential to allow calculation of the maximum number of candidate PFs that could be included in model development in order to limit the effects of statistical overfitting. (33)” page 11, row 249.

“The extracted PHE data were audited as per current methodological recommendations,(23) to establish data quality and quantify missing values. This process was also conducted in parallel with development of this protocol, to assist selection of candidate PFs to be included in either model development or exploration a priori and to inform strategies for handling missing data in the final analysis.” page 12, row 274.
Outcome measures

Please clarify at some point what is meant by "both studies." See my previous comment on Data sources.

Please refer to our previous reply to your comment in Data Sources above. The text has now been amended accordingly to:

“For this study, the primary outcome measure will be the occurrence of an initial (index) lower extremity IMI sustained by a participant during a season.” page 10, row 221.

Sample Size

The fact that sample size is "fixed" as you state does not mean you do not need to complete power calculations. Yes, the size of a team is fixed but you need to determine if that is enough to conduct the planned analysis without type II error. This piece is missing.

Thank you for your comment. For a prediction modelling study such as this, our primary concern is recognizing and adjusting for statistical overfitting (which will provide stable model estimates) rather than consideration of a type II error. As such, we have determined the number of events and this has been used to guide the number of candidate PFs for inclusion within the model, by not exceeding the traditional 10 Events Per Variable (EPV) rule of thumb for model development (Peduzzi et, 1996). We have also used a newer method (Riley et al, 2018) to determine the predicted amount of shrinkage of the model to adjust for overfitting.
Model development and internal validation

It is unclear why >15% of data is missing from so many tests and measurements of a mandatory PHE. This can be included in Discussion.

Thanks for highlighting this. As explained in an earlier point, we have incorporated all information into the ‘Missing Data’ subsection in the Methods section.

Statistical analysis

This section is clear.

Discussion

Discussion is well written and clearly highlights the relevance of this study while stating the exploratory nature of this work. Some clarification for the high percentage of missing data could be included here (see my previous comment) as well as the potential influence of the missing data on results. For example, the Y-balance has been studied a lot in younger cohorts with contradictory results. It would have been of interest to see if a functional balance test has any value among professional soccer players but due to large share of data missing, that is not possible. It is important to clarify the reasons behind the missing data, especially with a professional soccer club being a highly controlled environment.

Thank you for this comment, which is related to the above point regarding the missing data. In conjunction with adding the subsection on missing data, we have also included a line in the discussion to highlight that potentially useful prognostic factors were excluded (page 19, row 438):

“It is also possible that some potentially useful factors have been excluded on the basis of having > 15% of missing values. As such, only modest performance of this initial model is expected.”
This is a well written protocol on the development and internal validation of risk prediction model for injuries in elite athletes. The protocol followed well the TRIPOD recommendations for reporting. I have minor comments.

Thank you also for taking the time to review our submission. We feel that the comments and feedback you have provided have helped improve the manuscript.

Throughout the article there is no information on sample size and number of events. I assume that the measurements have taken place, can the authors show some quantitative information? They state which measurements have been deleted as predictors due to QC and missing data but there is no quantitative information. How many athletes have been measured? If this is not completed, what is the estimated sample sizes and number of events?

Thank you for your comment. The protocol was written at a similar time to when the final season’s data was being cleaned. We did have a section in the original protocol that was submitted that stated the anticipated number of events, number of prognostic factors for inclusion and anticipated shrinkage. However, since the protocol has been reviewed the data cleaning and extraction process is complete and so we have the quantitative information available. As such, the sample size section has been updated accordingly, and now states (page 12, row 261):

“Following the audit, the number of independent participant-seasons that will be included for analysis is 317, with 138 index IMI events recorded during the 5-season period. Therefore, we
have chosen to restrict the number of parameters for the candidate PFs (variables) for inclusion in model development to 12, which corresponds to having >10 EPV and thus above the minimum recommendation of 10. We also checked if this met the criteria to minimise overfitting recently proposed by Riley et al. (33) Assuming the model will have a modest Nagelkerke R-squared of 25%, then with an outcome proportion of 0.435 our 12 candidate predictor parameters corresponds to targeting an approximate shrinkage factor of 0.85, and thus a relatively small amount of overfitting (15%) (33). We deemed this a suitable compromise between increasing the number of predictor parameters and minimising the overfitting.”

The authors state the data will be assumed to be missing at random. Maybe it is better to check that this is missing at random and make informed choices?

Thank you for your comment. In line with the other reviewer’s comments, we have now included a subsection entitled ‘Missing Data” in the methods section. This quantifies the proportion of missing data. A table has also been created to highlight the mean values of age, height, weight and body mass index in participants with complete data and those with at least one missing observation. This has helped to assist our justification for a missing at random mechanism. The text can be located on page 13, row 286 and now states:

“As presented in Table 1, all medical history and age factors were complete (23 factors). Of the 37 remaining candidate PFs, the proportion of missingness ranged from 5.68% (for height and weight) to 76.34% (for body fat). Eleven of these had >15% missing observations (which included body fat, toe touch in standing, sacroiliac kinematic function, all Y Balance Test and upper body peak power variables). For these factors, the large degree of missingness was because of procedural changes in the PHE process, which meant that these tests were not conducted across all seasons.

For candidate PFs with < 15% missing observations, all tests were conducted consistently across all 5 seasons. For these factors, the sample characteristics of cases with complete PF data were compared to incomplete cases which had at least one missing observation (Table 2).
For complete cases, the mean values of all characteristics were less than incomplete cases, with the largest differences observed in age (20.83 and 23.55 years respectively) and weight (74.15 and 77.86 kg respectively). Therefore a complete case only analysis was not appropriate, and we will rather assume that the mechanism of missingness can be considered as missing at random (MAR), where the distribution of missing values are related to values of observed variables,(26) to allow imputation and so inclusion of individuals with missing data.”

Are the assessors glided to the status of the athletes when measurements are taking place?

Thank you for highlighting this omission from the manuscript. This point has now been addressed and can be located on page 8, row 162:

“If a participant was injured at the time of PHE, a risk assessment was completed by medical staff. In such instances, participants completed only tests that were deemed appropriate and safe for the participant’s condition; examiners were therefore not blinded to the injury status of participants.”

The authors chose backward selection as their variable selection method. Could they also examine other variable selection methodologies that do not require the elimination of predictors to 10 but start with a higher number?

Thanks for your suggestion. After careful consideration and discussion, although other approaches such as the Lasso might allow more predictors to be considered than our restricted set of 12, this also increases the potential for instability in results as larger adjustments would be needed for overfitting. With larger overfitting, the uncertainty in the penalisation needed increases, leading to more instability in the final model (Riley et al, 2018). Additionally,
backward selection will also allow us to develop a more robust, parsimonious and potentially clinically deployable model.

Additional information:

We felt the section for the exploratory statistical analysis was not particularly clear. For the exploratory analysis, we have now clarified that we intend to conduct multivariable logistic regression analyses for the remaining prognostic factors while adjusting for age, height and body weight, rather than constructing multivariable models using all prognostic factors of each type. This was because these factors are considered as the most important potential confounding factors. This has now been amended in both the abstract and the main text as follows:

“The remaining 35 candidate PFs are eligible for further exploration, using univariable logistic regression to obtain unadjusted risk estimates. Exploratory PFs will be incorporated into multivariable logistic regression models to determine risk estimates whilst adjusting for age, height and body weight.” (Abstract: page 4, row 71)

“Prognostic factor exploration

All remaining candidate factors that are eligible for exploration (Table 5) will undergo univariable logistic regression analyses to determine unadjusted associations with IMIs. Candidate PFs will also be incorporated into multivariable logistic regression models to determine odds ratios after adjustment for age, height and body weight. Note that because age was included as a candidate in the original model and will also be used for adjustment purposes in the exploratory multivariable models, the total number of candidate PFs eligible for exploration is 36. Exploration of non-linear associations between candidate factors and index IMI outcomes will also be evaluated using a fractional polynomial approach (49).”

(Main text: page 18, row 395)