Reviewer’s report

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

Version: 0 Date: 23 Feb 2019

Reviewer: Anu Räisänen

Reviewer's report:

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.

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 tittle 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.

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...."
BACKGROUND

Page 5, row 106

I would include commas in the numbers: 17,000 and 20,000.

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.

Data sources

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

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? With the PHE being mandatory, were the players free from injury at the time of PHE?

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.

Page 7, row 158

Instead of "club doctors" I would use "club medical doctors" throughout the manuscript.
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.

Data extraction

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

Outcome measures

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

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.

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.

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.

**Level of interest**
Please indicate how interesting you found the manuscript:

An article of importance in its field

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons
CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal