Methods Matter: Clinical Prediction Models Will Benefit Sports Medicine Practice, But Only if They Are Properly Developed and Validated

*Garrett S. Bullock, PT, DPT,[1,2] *Tom Hughes, PT, PhD[3,4] Jamie C. Sergeant, MSci, DPhil[5,6] Michael J. Callaghan, MCSP, MPhil, DPhil[3,4,5] Richard D. Riley, PhD[7] Gary S. Collins, PhD[8,9]

1. Centre for Sport, Exercise and Osteoarthritis Research *Versus Arthritis,* University of Oxford, United Kingdom
2. Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom
3. Manchester United Football Club, AON Training Complex, Birch Road, Off Isherwood Road, Carrington, Manchester M31 4BH, UK
4. Department of Health Professions, Manchester Metropolitan University, Manchester, UK.
5. Centre for Epidemiology Versus Arthritis, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK
6. Centre for Biostatistics, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK.
7. Centre for Prognosis Research, School of Medicine, Keele University, Staffordshire, UK.
8. Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford UK
9. Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

*Joint first authors

Corresponding author:

Garrett S. Bullock PT, DPT
Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences
University of Oxford
Oxford, United Kingdom
OX3 7LD
(865) 227-374
garrett.bullock@ndorms.ox.ac.uk

Introduction

Sports medicine clinicians are expected to make accurate diagnoses, estimate prognoses, and identify athletes at risk of sustaining an injury [1]. These complex decisions are dependent on clinical reasoning, which is informed by, and often biased toward, a practitioner's scientific knowledge and experience. Clinical prediction models are developed by researchers to help facilitate such decisions in practice [2]; data for multiple predictor variables are combined to estimate an individual's risk of a health outcome either being present (diagnosis) or whether it will occur in future (prognosis) [3]. Despite being employed widely in clinical medicine, clinical prediction models are uncommon in sports medicine. Clinical prediction models can offer benefits to both practitioners and athletes, but only if they are developed and validated using rigorous methods and transparently reported so that potential users can judge their accuracy and usefulness.

Therefore, the purpose of this editorial is to describe the recommended steps for clinical prediction development and validation and to guide practitioners using and interpreting prediction models in sports medicine.

Model Development

The first step in developing a prediction model is to identify its clinical need, the target population, and how and when it would fit into the clinical workflow. Models should predict outcomes that are relevant to sport stakeholders, and be clearly defined, including how and when assessed [4].

Next is to identify any existing models that could be evaluated or updated. If not, then before developing a new model, a publicly accessible protocol should be developed [4]. A summary of the recommended steps is in Table 1 [3].

The natural design for developing a prediction model is a cross-sectional study for developing a diagnostic model and a longitudinal study for a prognostic model [3]. For the latter, follow-up periods should be of sufficient duration to measure the desired outcome. Datasets used to develop prediction models are rarely complete. Omitting individuals with incomplete data should be avoided, as it reduces the sample size and may lead to bias. Multiple imputation should be considered for handling missing data [2,3].

Typically, many predictors are available for potential inclusion in a prediction model and reduction is often needed. Omitting predictors based on univariable association with the outcome should be avoided. Instead, predictors considered for inclusion should be identified based upon existing evidence, and clinical reasoning to determine their importance, relevance and plausibility related to the outcome [3]. Model fitting is typically done using regression methods, such as logistic (for binary outcomes), Cox (time-to-event outcomes) and linear regression (for continuous outcomes), although machine learning methods are gaining interest [5].

During model fitting, many predictors will be continuous (e.g., age). Categorising continuous variables should be avoided because it assumes that risk suddenly changes when measurements fall either side of a cut point, which is implausible. Categorising also discards information and

decreases predictive accuracy [6]. It is important to determine whether a predictor has a linear or non-linear association with the outcome.

When modelling, it is important to control model complexity to circumvent overfitting - i.e., where a model performs well in the development dataset, but performs poorly in new data (termed optimism) [7]. Consequently, an appropriate sample size calculation is imperative, e.g., using *pmsampsize* available in R and Stata [8], to establish the number of predictor parameters that can be considered while mitigating the risk of overfitting and improving targeted precise outcome risk estimation.

Model performance

Once developed, model performance should be assessed through calibration and discrimination. Calibration is the agreement between the predictions from the model against what was observed, and is best visualised using a calibration plot [2,3]. Discrimination is the ability of the model to differentiate between individuals with and without the outcome, usually quantified by the c-index [3,4].

It is important to internally validate the model, using resampling methods such as bootstrapping or cross validation. These approaches quantify the model optimism which can then be used to adjust performance measures and regression coefficients to form a final model more reliable for use in practice. Randomly splitting data for development and internal validation should be

avoided because it increases the risk of overfitting in smaller datasets and will not sufficiently test model performance in larger datasets [3,4]. Before using the model in clinical practice, it should be independently externally validated on a separate dataset, representative of the intended population and with an appropriate sample size.

Studies that develop or validate clinical prediction models should follow the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) Statement [9]. The model should be fully reported, so that it can be independently evaluated or used by others.

Brief Example

A prediction model was developed to estimate the risk of sustaining a lower extremity injury over the course of a English Premier League season [10]. 138 events occurred over 5 seasons and an *a priori* sample size of 12 parameters could be included in a logistic regression model. Less than 15% of data were missing and multiple imputation was performed. Bootstrapping was performed for internal validation. Model performance was poor with a calibration slope of 0.718 (95% CI: 0.275-1.161) and c-statistic of 0.589 (95% CI: 0.528-0.651). Poor model performance was hypothesized to be related to the inadequate predictive value of the selected predictors. The authors recommended the model should not be externally validated or used in clinical practice. Poor performance was related to data limitations and not methodology. Importantly, using proper methodology does not necessitate clinical practice integration. Only models that demonstrate high model performance, and after external validation should a model be considered for clinical

implementation.

Conclusion

Clinical prediction models can assist sports medicine practitioners with estimating an athlete's risk of sustaining an adverse health outcome in future, or the probability of a health condition being present. However, such models require careful development and validation if they are to be fit for purpose. By increasing awareness that methods matter in prediction research, we hope this improves future studies and allows clinicians to better appraise and identify models that are beneficial to sports medicine.

Contributions: GSB, TH, JCS, MJC, RDR, GSC conceived the study idea. GSB, TH, JCS, MJC, RDR, GSC were involved in design and planning. GSB, TH, RDR, GSC wrote the first draft of the manuscript. GSB, TH, JCS, MJC, RDR, GSC critically revised the manuscript. GSB, TH, JCS, MJC, RDR, GSC approved the final version of the manuscript.

Funding: GSC was supported by the NIHR Biomedical Research Centre, Oxford, and Cancer Research UK (programme grant: C49297/A27294).

Competing Interests: None

Ethics Approval: Not Applicable

Data Availability: Not Applicable

Patient Public Involvement: A series of symposiums are planned at various sports medicine conferences to help further educate clinicians on this topic

151 **Table 1**. Common Clinical Prediction Model Design Recommendations in Sport, based upon the

152 Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis

153 (TRIPOD) statement

154

Predictive Model Steps	Considerations
Define a health-related outcome that is to be predicted	• Health-related outcomes can be binary (e.g., whether athletes sustained a certain injury or not), time-to-event (e.g., when athletes sustained a certain injury), or continuous (e.g., athlete performance metrics or time to return to play after a certain injury)
Identify a data source that could be used to develop a model	• Data should be representative of the target population and sport. • Prospective data collection is ideal; however, existing or routinely collected data can be used.
Perform a sample size calculation to determine the number of predictor parameters that can be used	• The effective sample size for continuous outcomes is the total number of individuals; whilst it is the total sample size as well as the minimum of the number of events (those with the outcome) and non-events (those without the outcome) for binary outcomes, and the total number of events for time-to-event outcomes. • Sample size (as implemented in pmsampsize) for developing a prediction model (binary or survival outcomes) depends not only on the number of events relative to the number of candidate predictor parameters but also on the total number of participants, the outcome proportion and the expected predictive performance (e.g., R^2 or c-index) that minimize the risk of model overfitting.
Evaluate and manage missing data	• The quantity of and reasons for missing data should be explored. • Complete case analyses (i.e., deleting cases with missing data) reduces sample size, and may lead to biased predictions and poor model performance in new datasets or populations. • Multiple imputation is generally recommended to handle missing values, where missing values are predicted (imputed) to estimate the distribution of the data conditional on other known variables.
Handling of continuous predictors	• Any predictors that are based on continuous data should be retained on their continuous scale, with dichotomisation or categorisation avoided.

	• Linear associations between continuous data and outcomes should not be assumed; non-linear transformations should be considered. • Fractional polynomials or restricted cubic splines are recommended to assess non-linear relationships
Assessing and handling of predictor interaction	• Interactions occur when the effect of one predictor on the outcome is modified by the value of another predictor • Only biologically plausible interactions are recommended to be considered for inclusion in the model and these should be kept to a minimum
Select predictors for inclusion in a model	• Univariable screening and forward selection are discouraged to select predictors • Penalized methods (ridge regression, lasso, or elastic net) aim to alleviate the problems of overfitting during model estimation, or a global shrinkage factor can be applied post-estimation
Assessing the performance of a model	• Calibration is a measure of agreement between predicted risks (derived from a model) and observed risks in the dataset. • Calibration should be assessed graphically using calibration curves, and can be quantified by the calibration slope and calibration-in-the-large. • Discrimination is a measure of how well predictions from a model differentiate between individuals with the outcome and individuals without the outcome. • Discrimination should be assessed by the c-statistic (which for binary outcomes is equivalent to the area under the receiver operating characteristic curve). • Where a model's predicted risks will be used to change clinical decisions, clinical utility should be assessed using net benefit and decision curves.
Performing internal validation	• Models typically suffer from overfitting during development. This is where they model both the prognostic relationships *and* noise that exist between predictors and outcomes, so are therefore tailored to development datasets. • This means that models have better (or optimistic) apparent performance in the datasets that are used to develop them, but if used in different datasets or populations, performance will usually be worse. • Models should be validated using the entire dataset using bootstrapping or cross-validation, to determine optimism-corrected performance (calibration, discrimination, clinical utility).

	• Regression coefficients in a model can then be adjusted after validation to address overfitting, especially when the sample size is appropriate.
Performing external Validation	• Prior to model implementation in practice, ideally performance should usually be evaluated in an external dataset, called external validation. • This can consist of temporal, geographic, or domain validation, and requires an appropriate sample size to estimate predictive performance precisely
Reporting & Model Presentation	• All model development steps should be fully and transparently reported, following the TRansparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement

155
156

References

1. Emery CA, Roy T-O, Whittaker JL, Nettel-Aguirre A, van Mechelen W. Neuromuscular training injury prevention strategies in youth sport: a systematic review and meta-analysis. *Br J Sports Med*. 2015;49(13):865-870. doi:10.1136/bjsports-2015-094639

2. Riley RD, Van Der Windt DA, Croft P, Moons KG. *Prognosis Research in Healthcare: Concepts, Methods and Impact*. Oxford University Press; 2019.

3. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med*. 2015;162(1):W1-W73.

4. Wynants L, Collins G, Van Calster B. Key steps and common pitfalls in developing and validating risk models. *BJOG: Int J Obstet Gy*. 2017;124(3):423-432. doi:10.1111/1471-0528.14170

5. Seow D, Graham I, Massey A. Prediction models for musculoskeletal injuries in professional sporting activities: A systematic review. *Translational Sports Medicine*. 2020;3(6):505-517. doi:https://doi.org/10.1002/tsm2.181

6. Collins GS, Ogundimu EO, Cook JA, Le Manach Y, Altman D. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med*. 2016;35:4124-4135. doi:10.1002/sim.6986

7. Heinze G, Wallisch C, Dunkler D. Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal*. 2018;60(3):431-449. doi:https://doi.org/10.1002/bimj.201700067

8. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441. doi:10.1136/bmj.m441

9. Collins GS, Reitsma JB, Altman D, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis: The TRIPOD statement. *Ann Intern Med*. 2015;162:55-63.

10. Hughes T, Riley RD, Callaghan MJ, Sergeant JC. The Value of Preseason Screening for Injury Prediction: The Development and Internal Validation of a Multivariable Prognostic Model to Predict Indirect Muscle Injury Risk in Elite Football (Soccer) Players. *Sports Med Open*. 2020;6(1):22. doi:10.1186/s40798-020-00249-8