Sports Medicine
https://doi.org/10.1007/s40279-022-01655-6

CURRENT OPINION

# Black Box Prediction Methods in Sports Medicine Deserve a Red Card for Reckless Practice: A Change of Tactics is Needed to Advance Athlete Care

**Garrett S. Bullock[1,2,3] · Tom Hughes[4,5] · Amelia H. Arundale[6,7] · Patrick Ward[8] · Gary S. Collins[3,9,10] · Stefan Kluzek[2,3,11]**

Accepted: 26 January 2022


## Abstract

There is growing interest in the role of predictive analytics in sport, where such extensive data collection provides an exciting opportunity for the development and utilisation of prediction models for medical and performance purposes. Clinical prediction models have traditionally been developed using regression-based approaches, although newer machine learning methods are becoming increasingly popular. Machine learning models are considered 'black box'. In parallel with the increase in machine learning, there is also an emergence of proprietary prediction models that have been developed by researchers with the aim of becoming commercially available. Consequently, because of the profitable nature of proprietary systems, developers are often reluctant to transparently report (or make freely available) the development and validation of their prediction algorithms; the term 'black box' also applies to these systems. The lack of transparency and unavailability of algorithms to allow implementation by others of 'black box' approaches is concerning as it prevents independent evaluation of model performance, interpretability, utility, and generalisability prior to implementation within a sports medicine and performance environment. Therefore, in this Current Opinion article, we: (1) critically examine the use of black box prediction methodology and discuss its limited applicability in sport, and (2) argue that black box methods may pose a threat to delivery and development of effective athlete care and, instead, highlight why transparency and collaboration in prediction research and product development are essential to improve the integration of prediction models into sports medicine and performance.

✉ Garrett S. Bullock
garrettbullock@gmail.com

1 Department of Orthopaedic Surgery & Rehabilitation, Wake Forest School of Medicine, Winston-Salem, NC, USA

2 Centre for Sport, Exercise and Osteoarthritis Research Versus Arthritis, University of Oxford, Oxford, UK

3 Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford, UK

4 Department of Health Professions, Manchester Metropolitan University, Manchester, UK

5 Manchester United Football Club, Manchester, UK

6 Red Bull Athlete Performance Center, Thalgua, Austria

7 Icahn School of Medicine, Mount Sinai Health System, New York, NY, USA

8 Seattle Seahawks, Seattle, WA, USA

9 Centre for Statistics in Medicine, University of Oxford, Oxford, UK

10 Oxford University Hospitals NHS Foundation Trust, Oxford, UK

11 University of Nottingham, Nottingham, UK

**Key Points**

Transparent reporting of prediction models through full equations or compete code is of vital importance.

Without transparency, black box models cannot be externally validated, which is essential to understanding model performance and generalisability.

Black box prediction models may pose a threat to delivery and development of effective athlete care.

## 1 Introduction

There is growing interest in the role of data analytics for prediction purposes in sport [1] and sports medicine. Indeed, being able to accurately predict the risk of incurring future injuries or changes in performance as early as possible has recently been viewed as the 'holy grail' of sports medicine research [2].

Published online: 17 February 2022

The extensive volume of medical, training and performance data collected in elite sport provides an exciting opportunity for the development and utilisation of clinical prediction models for medical and performance purposes. Clinical prediction models can be used to assist practitioners with clinical decision making; they incorporate data from multiple predictor variables (termed predictors herein) measured at a point in time, to estimate an individual's probability of a health- or performance-related outcome being present at the time of measurement (diagnosis) or whether it will occur in the future (prognosis) [2, 3] (for list of key terms please refer to Table 1) [4, 5].

It is imperative to understand that predictors can be either causal or non-causal, as long as they have an association with the health outcome of interest. Causal factors have predictive value because they contribute to the cause of an event (such as an injury or change in performance) through direct or indirect mechanisms [6]. Non-causal factors are simply associated with an outcome so have predictive value, but they do not have a direct or indirect influence on whether or not an event happens [5]. Individual predictors usually have poor predictive value if used in isolation. However, when multiple predictors are used in combination, the ability of producing more nuanced individualised predictions can be realized [6, 7].

Clinical prediction models have traditionally been developed using regression-based approaches, although newer machine learning (ML) methods are becoming increasingly popular [8, 9]. ML methods are often viewed as opaque as the underlying architecture is typically too complex to disentangle all the predictor-outcome relationships and the availability of implemented software to obtain individualised predictions is rarely seen [8, 10–12]. For this reason, ML models are considered as 'black boxes' [10]. Specifically, 'black box' is defined as when a model or algorithm is not interpretable by humans or the reasons underlying a model risk score or choice are not available [13]. In parallel with the increase in ML, there is also an emergence of proprietary prediction models (and associated tools to facilitate implementation, e.g., software, online calculators, smartphone apps) that have been developed by researchers with the aim of becoming commercially available [1, 14]. Consequently, because of the profitable nature of proprietary systems [15], developers are often reluctant to transparently report, or make freely available, the development and validation of their prediction algorithms [8, 16]; the term 'black box' therefore also applies to these systems [9, 11].

While the number of prediction model development studies conducted in elite sport is currently modest overall [17], there is evidence that regardless of the statistical approach, the majority are poorly developed, opaque in their reporting (through not reporting the full complete code, algorithm description or model equations), and are not externally validated [17, 18].

The lack of transparency of 'black box' approaches is concerning as it prevents independent evaluation of model performance, interpretability, utility and generalisability prior to implementation within a sports medicine and performance environment [3]. Further, this opaqueness hinders model uptake and clinic implementation as medical professionals may not understand or correctly interpret how different predictors relate to the outcome [19]. Therefore, in this Current Opinion article, we: (1) critically examine the use of black box prediction methodology and discuss its limited applicability in sport, and (2) argue that black box methods may pose a threat to effective athlete care and, instead, highlight why transparency and collaboration in prediction research and product development is essential to improve the integration of prediction models into sports medicine and performance.

## 2 Sophisticated Black Boxes Have Limited Real-World Use

Hernán categorizes data analytics into three types [20]. Firstly, descriptive tasks involve using data to quantitatively summarise certain features of interest (e.g., a

**Table 1** Key terms

| Term | Definition |
|---|---|
| Predictors | Any variable that is predictive of the outcome (i.e., injury) |
| Causality | The relation of cause to effect |
| Clinical prediction models | Are multivariable mathematical models combining multiple predictors to estimate the risk or probability of an event. Both causal and non-causal factors can be used to aid risk prediction. They are developed to aid clinicians in determining the risk of a patient developing an outcome |
| Overfitting | An overly complex model that fits well to the idiosyncrasies in the data used to develop the model (capturing spurious unimportant predictor-outcome relationships), but fails to predict for new individuals |
| Internal validation | An evaluation of the model in the underlying population where the data used to develop the model originated from (using bootstrapping or cross-validation) to quantify overfitting of the developed model and estimate the optimism in model performance |
| External validation | Evaluating model performance in new data |

descriptive task could be a clinical audit to summarise the proportion of hamstring muscle injuries observed within a football team). Secondly, prediction tasks, also known as predictive modelling [21], select particular data and map these to outcomes of interest [20]. These tasks can therefore determine associations between predictors and a health or performance outcome, or can utilise multiple predictors in a model to compute the probability of a future event occurring for an individual [22]. For example, as a prediction task, one could calculate the risk of a hamstring injury occurring to an athlete during a season using a model that uses any combination of causal and non-causal predictors. Thirdly, counterfactual prediction tasks, also known as explanatory modelling [21], determine the probability of an outcome as if circumstances were different in some way because of an intervention, and are underpinned by a causal inference framework [20]. As an example of counterfactual prediction, one could potentially calculate the risk of a hamstring injury by incorporating or not incorporating eccentric hamstring exercises into an athlete's injury mitigation program, on the basis that reduced hamstring strength was a causal predictor for hamstring injury [23]. Therefore, any interventions that are aimed at modifying causal factors will consequently change their predictive relationship, thus modifying an individual's probability of an event occurring [24].

Determining predictors that are causal in relation to an outcome can improve a practitioner's ability to create more precise and impactful interventions [25], and this can be achieved through careful evaluation of the literature. However, a limitation of the current evidence—especially from studies that investigate the association between training load and injury risk – is that results are often incorrectly framed as though causality has been evaluated, even when their methodology does not allow such inferences; there is a clear need for further investigation of injury aetiology in elite sport using a counterfactual, explanatory framework [26–28].

Traditionally, prediction models have been developed using regression-based statistical methods (such as logistic, linear and survival analyses) to calculate a risk score for an individual [29]. If the aim of a model is to predict an outcome, then these traditional approaches can be used to develop models with any predictors that have an association with the outcome of interest [20, 21]. But if the aim is to develop a model that has the potential to identify causal relationships, an additional benefit of these traditional approaches is that models can be built around a counterfactual, explanatory framework where existing evidence, expert knowledge and clinical reasoning can be used to select predictors considered important both in terms of clinical relevance and to adjust for the effect of confounding factors [20, 21].

In contrast, ML (e.g., tree-based methods, gradient boosting machines, support vector machines) and deep learning (artificial neural networks) methods use bespoke algorithms that identify data patterns and determine associations within a dataset, using minimal assumptions about the data being used [30]. These approaches are perceived to offer increased flexibility to capture nonlinear associations and higher order interactions that exist between a set of predictors and outcome [31, 32]. However, similar to proprietary systems, if developed using ML, full models [3, 33] and the methodologic assumptions used in a model's development are unknown [5, 22, 34], thus contributing to their opacity.

In addition, while these methods typically employ extremely complex processes to determine associations between potential predictors and outcomes, they are not built using expert knowledge and clinical reasoning, so it is likely that during model development, predictors are selected from a dataset that have no causal relevance [20, 23]. This means that they cannot provide insights into: (1) what interventions could be implemented or (2) the magnitude of those interventions needed to reduce athlete risk.

Because such models have little or no causal relevance, and the underlying models and their assumptions are not reported, the black box methods are helpful for prediction tasks only. In other words, models developed from such processes in elite sport should only be used to communicate risks or probabilities of an event to practitioners, athletes and coaching staff, rather than assisting with selecting interventions that are designed to change the probability of an outcome. However, a danger is that, in practice, a lack of awareness regarding the conceptual limitations of black box models means that practitioners could erroneously use them for this latter function, thus inadvertently assuming a causal relationship between a predictor and an outcome, even where there is no evidence of one. The assumption of a causal relationship is problematic because an intervention is therefore unlikely to be effective, or, worse, may influence an unknown or different causal pathway. This may result in a possible cost to the athlete, such as increasing the risk of a different injury for instance [26, 27].

## 3 Black Boxes Cause Problems for Validation and Implementation

The absence of reporting transparency associated with black box models is problematic because it inhibits the understanding of how they have been internally or externally validated, which are essential processes to understand model performance and generalisability [18, 35].

Prediction models usually demonstrate improved (or optimistic) performance on the original development data compared to their performance if they are used on new individuals or future data [36]. This is often due to overfitting, which occurs when fitted models are too complex (with respect to the available data) and are overly adapted to all the idiosyncrasies within the data [34]. In order to counteract or minimise the effect of overfitting during development, internal validation should be performed to obtain an optimism-corrected (unbiased) assessment of model performance [37]. Splitting data into development ('training') and validation ('test') data sets should be avoided as this approach will decrease the sample size for both model development and performance assessment, actually increasing the risk of overfitting [37]. Ideally, all available data should be used to develop and internally validate the model using bootstrapping or cross-validation to counteract the risk of overfitting [38, 39].

Crucially, because black box models do not report full equations (for regression-based models) or full code, hyperparameters or provide algorithms on a repository (for ML models) [40, 41], they are at high risk for overfitting. In particular, while ML methods may have great potential for making accurate predictions [42], these approaches typically result in overfitted models that have frequently demonstrated poor prospective validation [43].

It is recommended that prior to using any model in a clinical setting, they should be externally validated in the target population for which they are intended, as models may not perform well in different sport organizations or athletic settings [44]. However, the lack of transparency of black box models is also problematic for external validation processes [18, 35]. Indeed, in a recent systematic review, prediction model external validation quality and reporting transparency was very poor, with 57% not reporting the full model or updated model [18].

External validation can consist of temporal- (same setting or population), geographic- (e.g., different sport club) or domain- (e.g., college athletes) based methods [44]. It should be noted that merely performing an external validation does not necessarily mean that the model is useful—external validation is assessing the performance of the model in different data. For example, a model developed in one organization to predict the risk of professional baseball pitcher arm injuries could be externally validated in another organization. The model may show poor external validation performance, which could imply a lack of clinical usefulness and potential harm, and so would not be recommended for clinical implementation.

Using black box models that do not have transparent validation processes in practice can result in a range of consequences. At the very least, prediction performance may be unreliable, and so would not necessarily be generalizable to new individuals or populations, thus resulting in models with little clinical usefulness [3]. However, more seriously, if models are used where their performance is questionable, this can have significant adverse implications for the health of the individuals for which the models are intended [7]. For example, in addition to the issues highlighted earlier that surround the use of black box models for making erroneous causal assumptions, if a poorly performing model is used to decide whether athletes require an intervention, this could result in the delivery of inappropriate or unsuitable interventions, which could be harmful or detrimental to the health and wellbeing of the athlete. To use a further clinical example, Obermeyer et al. [45] independently evaluated a commercial, proprietary black box prediction model that was used widely within American healthcare systems to identify patients with complex health needs and provide additional healthcare resource interventions. The authors used a dataset that contained all of the algorithm's predictions, as well as all of the data variables needed to establish the mechanisms responsible for differences in predicted risks. They found that predictions-derived models suffered from racial bias, which resulted in unequal access to healthcare among different groups, and potentially directly affected patient health.

Therefore, it should be strongly emphasised that if practitioners are considering whether to use a clinical prediction model or buy costly proprietary prediction software systems that have been developed using black box methods, if it is unclear whether a system has been validated accordingly, they should question whether it is safe and appropriate for use in practice.

## 4 A Change in Tactics to Advance Athlete Care: From Black Boxes to Transparency and Collaboration

Irrespective of the methods used to develop and validate prediction models, the importance of transparent development and validation processes cannot be underestimated, and are vital to allow clinicians to make informed choices on the potential risks and benefits of implementing such black box models into practice. While enhancing reproducibility, transparent reporting also allows researchers and practitioners to interpret the validity, performance and, ultimately, the clinical utility of such models in practice [3]. In particular, reporting of complete equations (for regression-based models, e.g., all the regression coefficients including the intercept) [8] or the complete code and tuned hyperparameters (e.g., for machine learning and deep learning models) [3, 46] can allow the creation of easy to use applications, where sports medicine practitioners can input relevant data to calculate an individual athlete's overall risk score [8]. In the case of proprietary risk-prediction models, steps can be taken to keep intellectual property guarded while still performing transparent validation and performance assessments such as reporting likelihood or odds ratios for all predictors, signing non-disclosure agreements, or sharing the algorithm logic, to name a few [47]. If a company does not want to disclose its model, it is possible to reverse engineer prediction outcomes in a separate sample. However, this is untenable as a long-term validation solution [8]. For further information on reporting model transparency, please refer to Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) [3] and TRIPOD: Artificial Intelligence [3].

Increasing the transparency of black box models provides an opportunity for collaboration between sport organizations or leagues. Increasing collaborative opportunities can help circumvent a specific issue that affects the development of prediction models in sport, which is limited sample size [26, 27]. For example, prediction model sample size calculations are based upon the number of events (in this case, injuries), not on the overall sample size [48–50]. Further, external prediction model validation is grounded on the number of events, distribution of the sample risk, and calibration [51, 52]. Most sport organizations do not sustain enough injuries within a given season, or multiple seasons, to develop or externally validate a model accurately. Only by collaborating with multiple sport organizations or leagues could proper prediction model development and external validation be performed. Without model transparency, these collaborations are impossible and, thus, accurate and useful prediction models cannot be developed or validated [53].

## 5 Conclusion

The improvements in technology [54–56] and the rise of data-driven methods [57–60] provide an exciting opportunity for the development, validation and incorporation of prediction models within sports medicine practice. The application of prediction model methodology can aid sports medicine clinicians and performance professionals in identifying predictors that most influence injury risk or changes in performance. However, without full transparency of reporting and the complete presentation of all model equations (regression-based approaches) [3] or code (machine learning) [33], these methods become black boxes that cannot aid in interpretation, or guide intervention. Further, these methods may potentially waste scarce resources including athlete and clinician time. While prediction models in sports medicine are developed with the intention of facilitating athlete care, we argue that, paradoxically, the creation of opaque models significantly inhibits their influence. Furthermore, opaque black box models may indeed hinder or threaten delivery of effective healthcare or training programmes if they are applied in practice. In this Current Opinion article, we strongly advocate the use of transparency and collaboration to enhance the rigour of future model development and validation studies, which may culminate in more accurate and useful future models for implementation within elite sport. Furthermore, it is hoped that this paper will increase practitioner awareness of the issues surrounding black box predictive analytics, and assist with the evaluation of such models to facilitate informed decisions regarding clinical implementation.

**Authors' contributions** GSB, TH, GSC and SK conceived the study idea. GSB, TH, GSC and SK were involved in design and planning. GSB, TH and SK wrote the first draft. GSB, TH, AHA, PW, GSC and SK critically appraised the manuscript. GSB, TH and SK wrote the first draft. GSB, TH, AHA, PW, GSC and SK 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). No other sources of funding were used to assist in the preparation of this article.

## Declarations

**Conflicts of interest/Competing interests** Garrett S. Bullock, Tom Hughes, Amelia H. Arundale, Patrick Ward, Gary S. Collins and Stefan Kluzek declare that they have no conflicts of interest relevant to the content of this article.

**Availability of data and material** Not applicable.

**Code availability** Not applicable.

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

## References

1. Horvat T, Job J. The use of machine learning in sport outcome prediction: a review. Wiley Interdiscipl Rev Data Min Knowl Discov. 2020;10(5):e1380.
2. McCall A, Fanchini M, Coutts AJ. Prediction: the modern-day sport-science and sports-medicine "quest for the holy grail." Int J Sports Physiol Perform. 2017;12(5):704–6.
3. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Int Med. 2015;162(1):W1–73.
4. Riley RD, van der Windt D, Croft P, Moons KG. Prognosis research in healthcare: concepts, methods, and impact. Oxford University Press; 2019.
5. Hughes T, Sergeant JC, van der Windt DA, Riley R, Callaghan MJ. Periodic health examination and injury prediction in professional football (Soccer): theoretically, the prognosis is good. Sport Med. 2018;48(11):2443–8.

6. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med. 2013;10(2):e1001380.
7. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013;10(2):e1001381.
8. Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. Predictive analytics in health care: how can we know it works? J Am Med Inf Assoc. 2019;26(12):1651–4.
9. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.
10. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access. 2018;6:52138–60.
11. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206–15.
12. Dhiman P, Ma J, Navarro CA, Speich B, Bullock G, Damen JA, et al. Reporting of prognostic clinical prediction models based on machine learning methods in oncology needs to be improved. J Clin Epidemiol. 2021;138:60–72.
13. Da Cruz HF, Pfahringer B, Martensen T, Schneider F, Meyer A, Böttinger E, et al. Using interpretability approaches to update "black-box" clinical prediction models: an external validation study in nephrology. ArtifIntell Med. 2021;111:101982.
14. Cook C. Predicting future physical injury in sports: it's a complicated dynamic system. Br J Sport Med. 2016;50(22):1356–7.
15. Shah ND, Steyerberg EW, Kent DM. Big data and predictive analytics: recalibrating expectations. JAMA. 2018;320(1):27–8.
16. Van Calster B, Steyerberg EW, Collins GS. Artificial intelligence algorithms for medical prediction should be nonproprietary and readily available. JAMA Int Med. 2019;179(5):731.
17. Seow D, Graham I, Massey A. Prediction models for musculoskeletal injuries in professional sporting activities: A systematic review. Trans Sports Med. 2020;3(6):505–17.
18. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol. 2014;14(1):40.
19. Watson DS, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR, et al. Clinical applications of machine learning algorithms: beyond the black box. Bmj. 2019;12;364.
20. Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. Chance. 2019;32(1):42–9.
21. Shmueli G. To explain or to predict? Stat Sci. 2010;25(3):289–310.
22. Prosperi M, Guo Y, Sperrin M, Koopman JS, Min JS, He X, et al. Causal inference and counterfactual prediction in machine learning for actionable healthcare. Nat Mach Intell. 2020;2(7):369–75.
23. Hernán MA, Robins JM. Causal inference: what if. Boca Raton: Chapman & Hall/CRC; 2020.
24. Sperrin M, Jenkins D, Martin GP, Peek N. Explicit causal reasoning is needed to prevent prognostic models being victims of their own success. J Am Med Informatic Assoc. 2019;26(12):1675–6.
25. Hingorani AD, van der Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ. 2013;346:e5793.
26. Impellizzeri FM, McCall A, Ward P, Bornn L, Coutts AJ. Training load and its role in injury prevention, part 2: conceptual and methodologic pitfalls. J Athl Train. 2020;55(9):893–901.
27. Impellizzeri FM, Menaspà P, Coutts AJ, Kalkhoven J, Menaspa MJ. Training load and its role in injury prevention, part I: back to the future. J Athl Train. 2020;55(9):885–92.
28. Impellizzeri FM, Ward P, Coutts AJ, Bornn L, McCall A. Training load and injury part 1: the devil is in the detail—challenges to applying the current research in the training load and injury field. J Orthop Sport Phys Ther. 2020;50(10):574–6.
29. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009;338:b375.
30. Bzdok D, Altman N, Krzywinski M. Points of significance: statistics versus machine learning. Nature 2018;14(12):1119.
31. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. J Clin Epidemiol. 2016;76:175–82.
32. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med. 2016;35(2):214–26.
33. Collins GS, Moons KG. Reporting of artificial intelligence prediction models. The Lancet. 2019;393(10181):1577–9.
34. Steyerberg EW. Clinical prediction models. Springer; 2019.
35. Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models. BJOG. 2017;124(3):423–32.
36. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med. 2000;19(4):453–73.
37. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans M, Vergouwe Y, Habbema JDF. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774–81.
38. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. J Clin Epidemiol. 2016;69:245–7.
39. Efron B, Tibshirani RJ. An introduction to the bootstrap. CRC Press; 1994.
40. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. Bmj. 2020;25;368.
41. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. bmj. 2020;20;368.
42. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. N Engl J Med. 2016;375(13):1216.
43. D'Amour A, Heller K, Moldovan D, Adlam B, Alipanahi B, Beutel A, et al. Underspecification presents challenges for credibility in modern machine learning. arXiv preprint arXiv:201103395. 2020.
44. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691–8.
45. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–53.
46. Haibe-Kains B, Adam GA, Hosny A, Khodakarami F, Waldron L, Wang B, et al. Transparency and reproducibility in artificial intelligence. Nature. 2020;586(7829):E14–6.
47. Janssens A. Proprietary algorithms for polygenic risk: protecting scientific innovation or hiding the lack of it? Genes. 2019;10(6):448.
48. van Smeden M, de Groot JA, Moons KG, Collins GS, Altman DG, Eijkemans MJ, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. BMC Med Res Methodol. 2016;16(1):163.
49. van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJ, et al. Sample size for binary logistic prediction models: Beyond events per variable criteria. Stat Methods Med Res. 2019;28(8):2455–74.

50. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE Jr, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II—binary and time-to-event outcomes. Stat Med. 2019;38(7):1276–96.
51. Riley RD, Debray TP, Collins GS, Archer L, Ensor J, van Smeden M, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat Med. 2021; 40(19):4230–51.
52. Snell KI, Archer L, Ensor J, Bonnett LJ, Debray TP, Phillips B, et al. External validation of clinical prediction models: simulation-based sample size calculations were more reliable than rules-of-thumb. J Clin Epidemiol. 2021;135:79–89.
53. 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):1–13.
54. Jennings D, Cormack S, Coutts AJ, Boyd LJ, Aughey RJ. Variability of GPS units for measuring distance in team sport movements. Int J Sport Physiol Perform. 2010;5(4):565–9.
55. Plews DJ, Laursen PB, Stanley J, Kilding AE, Buchheit M. Training adaptation and heart rate variability in elite endurance athletes: opening the door to effective monitoring. Sports Med. 2013;43(9):773–81.
56. Wisbey B, Rattray B, Pyne D. Quantifying changes in AFL player game demands using GPS tracking: 2008 AFL season. Florey (ACT): FitSense Australia; 2008.
57. Me E, Unold O. Machine learning approach to model sport training. Comput Hum Behav. 2011;27(5):1499–506.
58. Alderson J. A markerless motion capture technique for sport performance analysis and injury prevention: toward a ‘big data’, machine learning future. J Sci Med Sport. 2015;19:e79.
59. Zelič I, Kononenko I, Lavrač N, Vuga V. Induction of decision trees and Bayesian classification applied to diagnosis of sport injuries. J Med Syst. 1997;21(6):429–44.
60. Robertson S, Bartlett JD, Gastin PB. Red, amber, or green? Athlete monitoring in team sport: the need for decision-support systems. Int J Sport Physiol Perform. 2017;12(s2):S2-73-S2-9.