Regression models are not inherently predictive. An educational review on why sports exercise and medicine research should be more careful with the term "predictors" *(Version 1)*

Supplementary materials: Not applicable

For correspondence: jneves@fade.up.pt

José Afonso[1*], Chris Bishop[2], Rodrigo Ramirez-Campillo[3], Filipe Manuel Clemente[4,5], Renato Andrade[6,7,8], Tania Pizzari[9], Anthony Turner[2], Richard Inman[10]

[1]Centre of Research, Education, Innovation, and Intervention in Sport (CIFI$_2$D), Faculty of Sport, University of Porto, Porto, Portugal. [2]Faculty of Science and Technology, London Sports Institute, Middlesex University, London, UK. [3]Exercise and Rehabilitation Sciences Institute. Faculty of Rehabilitation Sciences. Universidad Andres Bello. Santiago, 7591538, Chile. [4]Escola Superior Desporto e Lazer, Instituto Politécnico de Viana do Castelo, Rua Escola Industrial e Comercial de Nun'Álvares, 4900-347 Viana do Castelo, Portugal. [5]Gdansk University of Physical Education and Sport, 80-336 Gdańsk, Poland. [6]Porto Biomechanics Laboratory (LABIOMEP), University of Porto, Portugal. [7]Clínica Espregueira – FIFA Medical Centre of Excellence, Porto, Portugal. [8]Dom Henrique Research Centre, Porto, Portugal. [9]School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, Australia. [10]Centro de Investigação em Psicologia para o Desenvolvimento (CIPD), Universidade Lusíada, Porto, Portugal.

* Corresponding author: José Afonso (jneves@fade.up.pt)

Please cite as: Afonso et al. (2024). Regression models are not inherently predictive. An educational review on why sports exercise and medicine research should be more careful with the term "predictors". *SportRχiv*.

All authors have read and approved this version of the manuscript. This article was last modified on Month XX, YEAR.

Authors X @*xtwitter* and Y @*ytwitter* can be reached on Twitter.

ABSTRACT

The term “predictor” is commonly used in regression modelling in substitution of the more accurate “independent variables”, suggesting a predictive capacity that regression inherently lacks. The goal of this educational review is to raise awareness of the misuse of the term “predictor” when associated with regression models, with a focus on sports exercise and medicine. We start by elucidating the fundamentals of regression modelling and explain its descriptive rather than predictive nature. We then address key conceptual pre-requisites for predictive modelling: sample representativeness, context, expected consistency of relationships over time, trustworthiness of measurements, problems with multiple testing, and confounders. Next, we establish why external validation is warranted before deeming a model “predictive” and present a conceptual model for progressive extrapolation. While these steps apply to other statistical models, regression modelling is particularly prone to the use of the term “predictors” as a default terminology, fostering the misconception that regression models are inherently predictive. While regression models provide relevant insights into the relationships between a chosen set of variables, they are not inherently predictive, and their extrapolation is contingent upon rigorous validation and contextual appropriateness. Lastly, we provide an algorithm and checklist to guide researchers when the terms “predictor” or “predictive” may be applicable. We advocate that research using regression modelling should eschew from the default use of “predictive” terminology to avoid inaccurate interpretations and scientifically misleading the

audience. Awareness of these nuances is crucial to strive for scientific integrity and to appropriately interpret findings from research that uses regression models.

1. Introduction

Prediction

"A statement about what you think will happen in the future."

(https://dictionary.cambridge.org/dictionary/english/prediction)

Statistical models can be used to describe, infer, or predict features of the world around us [1-3]. In most cases, it is too difficult, too expensive, or even outright impossible to analyse the entire target population, and therefore statistical models are built upon subsets called samples [4, 5]. Limited resources mean that statistical models are normally not continuously updated with new data, but instead rely on data collected at a single or small number of assessments. Consequently, statistical models can be great for describing the features of a sample (e.g., handgrip strength) during the assessed time windows, but researchers may aim to extend the models' applications beyond this original data set [6-8]. Prediction, in particular, is a goal of many such modelling endeavours [9-11], involving the construction of a statistical model based on current data aiming to forecast unknown future observations [2, 3].

Prediction is a common objective in sports exercise and medicine research into screening, injury risk factors, diagnosis, and therapeutic interventions [1, 12]. As an example, statistical models have been fit to data to describe how a change in one or more independent variables (e.g., score in the Adductor Performance Test) affect hip adductor injury risk (the dependent variable) [9]. However, as in the aforementioned case [9], instead of interpreting this relationship as a general pattern of association, specific to a particular sample within a given context and timeframe, researchers often infer (erroneously) a functional relationship where an individual's score for a 'predictor' (independent variable), can be used to accurately forecast a future 'outcome' (dependent variable) for that individual [9, 13]. For example, if athlete 1 scores *X* on the Brazilian Adductor Performance Test it is highly expected he will get a hip adductor injury in the future [9].

However, using models to make inferences about the future is at odds with the nature of statistical models, which are always based on collected (i.e., retrospective) data [3]. Strictly speaking, statistical models are *retrodictive*, not predictive: they are mathematical models (equations) that summarize available data, enabling researchers to make inferences about underlying relationships. Even if these models provide an excellent fit to the dataset upon which they were built, they may not necessarily fit well to different datasets (different samples), or even to the same sample at future time points [1-3, 14, 15]. For a statistical model to assume a predictive role, additional conceptual leaps must be undertaken (which will be discussed in section 3). Although some scientific realms (e.g., regression modelling) have started to use "prediction" in a much looser manner, this detracts from its general understanding and may impair the proper communication of scientific findings.

When studying relationships among variables, researchers often adopt relatively neutral terminology, such as association (e.g., when referring to chi-square tests of independence for categorical variables [16, 17]) or correlation (e.g., Pearson correlation for continuous variables [18, 19]). Regression models are a notable exception. When applying regression modelling, the common statistical concepts of independent and dependent variables are frequently replaced with the terms "predictor(s)" and "outcome" [20-25] and the resulting models are often termed "predictive" [26-30]. This is apparent in statistical textbooks [1, 31, 32], and extends to empirical research in sports exercise and medicine [33], including topics related to injury risk [9, 11, 34-38] and performance enhancement [7, 8, 39-41]. This language is commonly applied regardless of the specific type of regression model used.

Regression models do not present special features that warrant a default use of the prediction terminology. Some research using regression modelling uses more moderate language (e.g., association) and avoids applying the terms "predictors" or "predictive" [42-46]. Indeed, regression models are often used to report associations applied to observational data [2]. A recent systematic review heavily criticized regression-based models for making unreliable predictions regarding risk of sports injuries due to several methodological shortcomings [47] (more information in section 4)[1]. Perhaps this terminology is a convention that caught-on, despite lacking the foundations for that purpose. This may have far-reaching implications, as the

[1] We further advise readers to consult the methodological review of Windt et al. [48], which pinpointed several concerns in how statistical analyses are being performed and interpreted in longitudinal studies examining the relationship between workload and injury risk in sports.

concept of a predictor may be misleading and result in inappropriate and spurious interpretation of data [2].

Statistical relationships between dependent and independent variables do not necessarily imply causation [1, 14, 49, 50] and, even if a causal or explanatory relationship is reasonably well-established, it may not suffice to predict the outcome [1, 2, 51]. The choice of variables to be assessed and the selection of potential considered predictors requires choices by the researchers, meaning the models may suffer from omitted variable bias [52, 53]. Those choices may exclude relevant variables and limit the choices to variables that the researchers can measure (in their specific context, facilities, and financial and temporal budget). These shortcomings are known to affect regression modelling [53].

Prediction is a natural goal of (at least part of) the scientific endeavour. Likewise, there is nothing inherently wrong with using regression modelling. However, both depend on often forgotten and bypassed premises, with terms such as “predictors” or “predictive” becoming a default language mode in the realm of regression modelling, without ensuring that all the relevant foundations are properly implemented. Therefore, the goals of this work are to: (i) establish what regression modelling consists of and what can be inferred from regression models based exclusively on their mathematical properties; (ii) what additional (often non-mathematical steps) must be ensured to classify such models as predictive; and (iii) warn against the automatic coupling of regression modelling and prediction language.

2. What is regression? A quick overview

2.1. Regression 101

Regression is an umbrella term describing a statistical approach to estimating the relationship between a dependent variable (*Y*) and one or more independent variables by fitting a regression model to available data [1, 6, 14, 49, 52, 54-58]. Regression models attempt to maximize the proportion of variance in the dependent variable accounted for by an independent variable (or a set of independent variables), i.e., they estimate coefficients that describe the relationship between dependent and independent variables by finding a model with the smallest amount of error between the observed and expected values. Different types of regression models are appropriate depending on the number of independent variables, the type of dependent variable (i.e., discrete vs. continuous), and assumed relationship (i.e., linear vs. non-linear) [1, 14, 59].

At its core, regression is used to summarize data, specifically variance in a dependent variable, by fitting some type of statistical model to it. Thus, researchers can use regression to describe relations between independent and dependent variables, often inferring that such relationships may generalize to a broader target population (a topic that will be developed in section 3). All remaining variance not accounted for by the model, which will not fit the data perfectly, can be referred to as error. Models that do a good job at summarizing data include a smaller amount of error, while models that underperform are paired with a larger error term:

$$\textbf{Equation 1}: Y = (model) + error$$

When using a simple linear regression (Equation 2, General Linear Model [GLM]), researchers are essentially summarizing data with a straight line. In this case, the term 'model' is replaced with the equation of a straight line, where the parameter β_0 is the intercept and β_1 is the gradient (or slope). The intercept β_0 is the estimated value of the dependent variable *Y* when the independent variable X_1 is zero. The gradient β_1 is the estimated value that describes the magnitude of the linear association between independent variable X_1 and dependent variable *Y*: in more simple terms, by which rate *Y* changes by each unit of X_1.

$$\textbf{Equation 2}: Y_i = (\beta_0 + \beta_1 X_1) + error_i$$

Multiple linear regression extends the logic of simple linear regression to include scenarios where there is more than one independent variable. In this model, β_1 describes the magnitude of the linear association between independent variable X_1 and dependent variable *Y,* β_2 describes the magnitude of the linear association between independent variable X_2 and dependent variable *Y,* and so forth.

$$\textbf{Equation 3}: Y_i = (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \ + \cdots \ \beta_p X_{pi}) + error_i$$

Independent variables can be added to a regression model *ad infinitum* and other types of research questions, data, and assumptions can demand models based on different and often more complex equations[2]. Nonetheless, all regression equations include unknown regression coefficients (e.g., β_0 and $\beta_1$etc.) that are estimated using known values from the sample dataset (for example, in the case of simple linear regression, the values of *Y* and X_1). In the case of linear regression, particularly in the domain of sports exercise and medicine, the Ordinary Least Squares (OLS) method is ubiquitous (although not the only option) for this process of estimation. Regardless of estimation method used and complexity of the model, the common feature is that the regression coefficients obtained are derived directly from, and are therefore descriptive of, a specific set of data points.

2.2. Assumptions

To instil confidence in an estimated regression model, researchers must be certain that relevant underlying statistical assumptions are met. Otherwise, the regression coefficients obtained from a model may be biased, providing a poor summary of the data upon which the model is based (let alone other datasets one may wish to extrapolate to). Critically, even when these assumptions are met, it does not follow that regression coefficients estimated from a specific sample will be accurate for a target population or other samples (albeit the chance of it

[2] We shall not consider the equations underpinning logistic regression, ridge regression, multilevel regression and so on, as they fall beyond the scope of this work.

being accurate does increase). Given their relevance, it is worth briefly considering some of the primary assumptions, focusing specifically on those that when not met can bias the regression coefficients [1, 31, 59].

Misjudgement of assumptions can be illustrated by three simple examples on: (i) influential outliers, (ii) incorrect beliefs about the nature of the relationship between the variables, and (iii) highly correlated independent variables. First, a regression model may do a poor job of summarizing observed data – that is, demonstrate poor fit and present biased regression coefficients – if the observed data contains outliers that exert a disruptive influence on the estimation process [14, 60, 61]. A real example of an outlier affecting sports exercise and medicine-related data can be reviewed elsewhere [61]. Second, a regression model may also underperform when summarizing observed data if it assumes and then tries to fit a mathematical function that is uncharacteristic of the true relationship between the variables. For example, simple linear regression assumes that the independent and dependent variable share a linear relationship, and as such, summarizes data using the equation of a straight line (section 2.1). However, reality is often far more complex and nonlinear, and fitting straight lines will do a poor job if the relationship between two variables is, for example, curvilinear. Indeed, the linear assumptions that often underly the regression models may not hold true at more extreme data points, rendering the model ineffective [59]. Third, regression models with multiple independent variables can have unstable regression coefficients – that is, estimates can change dramatically with small changes in data - if the independent variables are highly correlated [62, 63].

Other assumptions are linked to the standard error and thus, the reliability of significance tests (but not necessarily point estimates of regression coefficients), are the normality of distribution (for continuous variables), homoscedasticity (i.e., the variance of residuals is constant across observations; valid for linear regression models), and the independence of error terms [1, 31, 59]. An additional assumption is that the relationship between variables will behave relatively consistently over time [14] (more on the topic in section 3). Other specific assumptions may be applicable to specific regression models. If these assumptions are met, the model may do a reasonable job at capturing reality, *at least within the narrow confines of available data within a specific study*.

2.3. Is regression inherently predictive?

In the context of regression, researchers often refer to independent variables as "predictors" (regardless of whether they are using it to describe relations in available data or in an applied sense) [1, 6, 8, 14, 38, 47, 49, 55, 60], although perhaps a more fitting term would be "regressors" [31, 53, 55]. Equation 2 means that once the model has been fit to a specific dataset, the regression coefficients (*ß*) and error term have been estimated. At this point, by inserting values for the independent variable(s) (e.g., X_1, X_2) we can solve the equation and calculate the value for the dependent variable (Y): the Y is said to be "predicted" from X, hence the independent variables being termed "predictors". Strictly speaking though, if the value for Y already exists and is simply unknown, then this is a process of retrodiction. For example, the main author of this work (JA) has previously stated that "quality of reception and attack tempo were shown to be predictors of attack type" (p. 244) in volleyball [13]. However, the model was entirely built upon past data and there was no attempt to forecast future events (e.g., different sample or same sample at a later time), and therefore there was no prediction involved. Likewise, it is possible to rearrange the equation and calculate X if a value of Y is known, so what constitutes a "predictor" would largely depend on research design and *a priori* theoretical assumptions. Especially in cross-sectional designs, the dependent and independent variables are mostly interchangeable (e.g., in the previous case [13], it could also be stated that attack type predicted attack tempo). These relations are not specific to regression modelling and are not inherently predictive.

It is also worth mentioning that Equation 2 (the GLM) forms the basis of many statistical models. For example, an analysis of variance (ANOVA) to test differences in means between three groups, for example, is multiple regression but with a categorical independent variable. As with regression, it is entirely plausible to fit the model, estimate β values, and then estimate an individual's score for the dependent variable by replacing the values of X (in this case, with zeros or ones that represent group membership); however, this approach is not used with the same predictive intent as regression. Similarly, it is possible to estimate a Pearson's correlation coefficient (r) for observed data, take a value of X, rearrange the associated equation, then solve it to calculate a value of Y (or vice versa). Once again, mathematics allows for these processes, and yet we do not so readily use the term "prediction" in these contexts. Prediction is not a determined property of these statistics: it also requires an intention to predict, and a design that allows for it (e.g., establishing a causal relationship between dependent and independent variables).

The fact is that regression models estimate mathematical functions that closely fit the specific data set that was used to create those models [1-3, 6, 14, 38, 47, 49]. These models are therefore built upon, and fit to, a specific data set from a specific sample at a specific time point. As such, these models may become too specific to the available data (i.e., overfit), modelling both the underlying relations and the noise that is idiosyncratic to the specific data. This is particularly problematic if the model is not validated with new data (a subset of data that has not been used to build the model may still be considered new data). The ability to predict how that sample will behave at future timepoints is a challenging task. The confidence we have in applying a model from one sample to another, or for the same sample at a later time, depends on the extent to which we believe it is generalizable (which depends on multiple assumptions that will be explored in section 3). Therefore, a relevant challenge involves out-of-sample predictions: i.e., understanding how generalizable the model is and/or using it to predict how other samples will behave.

3. Generalizing regression and making predictions

As previously mentioned, regression is often used for out-of-sample prediction (i.e., extrapolation), whereby the relations modelled with a specific data set are used to estimate outcomes outside that data's range [1, 14, 47, 49, 64-66]. However, such extrapolation requires relevant assumptions to be met, which should be explicitly stated (and tested) [1, 14, 49, 58]. No statistical regression model is predictive *per se* [58, 67, 68][3], and often regression models underperform when applied to new samples (i.e., not included in the original data set used to create the model) [1, 6, 14, 49]. Again, although this is not exclusive to regression models, it is this class of models that applies "predictors" as the default terminology. Extrapolating from a (typically) small sample to the broader target population involves a sizeable leap of faith [1, 6, 14, 38, 49, 68], which often goes unstated. Although this problem affects every statistical model, most do not adopt the terminology of prediction so readily as regression modelling.

The mathematical assumptions beyond regression modelling are insufficient to extrapolate and require additional criteria before applying the term "predictors", criteria which involve human judgment and decision-making [1, 14, 49, 58]. To illustrate this point, we refer to a study with soccer players [69]. Using binary logistic regression to identify the variables

---

[3] The predictive ability of any given proposal will rely not only on statistical modelling, but also on overall study design. But even if the study design is appropriate for generating predictions, the statistical model's accuracy should still be considered, as no model will be 100% accurate.

associated with team success (i.e., match victory vs. draw/loss), the goal was "(...) to identify those variables which would best *predict future* success for the team" (p.164) [69]. First, the term prediction is misused because the models were applied exclusively to past data and not to forecast future success (or to test whether the model performed well with a subset of data not used to build the model). Second, even for the data upon which the model was built, it correctly "predicted" 71.7% of match outcomes – if predicting the past is difficult, predicting the future may be just a fancy guess. For an illustrative example of how humans make poor predictions, we suggest reading An et al. [70].

A different (but likely instructive example) comes from a study with rock climbers: "Linear regression, adjusted for age and experience (years), revealed that forearm oxidative capacity index, treadwall maximal deoxygenation (Δ), and treadwall VO_2peak all significantly predicted self-reported red-point sport climbing ability" (p. 3534) [41]. Given the cross-sectional nature of this study, self-reported performance might have predicted how well those rock climbers performed in the standardized tests (e.g., through the boost of self-confidence), or an untested variable could have explained both, i.e., the authors tested associations and there was no attempt to study how well the model performed with a new sample. Similar cases of potentially misleading application of the term "predictors" when using regression models can be found in sports such as basketball [71], golf [72], swimming [73], tennis [74], track and field [75], and volleyball [13] (incidentally, the main author of the current work (JA) co-authored the volleyball study [13], so the critique is extensible to our own previous research).

A more complex example derives from a study with 187 college students aged 18 to 22 years, where the Functional Movement Screen (FMS^{TM}) was tested and then these individuals were followed for 12 months to monitor injury incidence [35]. The authors used binary logistic regression to conclude that a composite FMS^{TM} value below 17.5 was predictive of sports injuries [35]. However, even applying a longitudinal design, the model was calculated after all data had been collected, with no attempt at forecasting. For now, let's ignore whether the sample was representative or not. The FMS^{TM} is a paradigmatic example, whereby the literature has shown that each sample has its own cut-off values [12, 76, 77], and the test battery presents low values of specificity and sensitivity, and overall reduced predictive ability [10, 15, 76, 78]. Therefore, the relationship between the FMS^{TM} scores and injury risk is idiosyncratic and lacks predictive power, as it fits only the sample that was analysed and can only be calculated retrospectively. Although this may be common in scientific research, it suggests caution before using (the authors) and interpreting (the readers) expressions such as "predictors".

Prediction relies on factors and decisions that are not exclusively based on the statistical properties of a model, and this includes regression models. Features related to the sample, context, assessment procedures, among others, should be considered before assuming that a model can be generalized beyond its specific data.

3.1. Sample representativeness

Small sample sizes are typically used in sports exercise and medicine-related research [79-82], and are more likely to be unrepresentative of the target population relative to larger samples [1, 14, 83]. Small sample sizes may also result in regression models incorrectly failing to reject the null hypothesis (incorrectly suggesting that independent and dependent variables are unrelated when, in fact, they are related in the target population) due to a lack of statistical power (type II error) [1, 14, 79, 80, 83]. There is, of course, the inverse problem: a regression model (particularly one based on a very large sample) may provide parameter estimates that incorrectly lead to the rejection of the null hypothesis, suggesting that independent and dependent variables are related when, in fact, they are unrelated in the target population [1, 14, 59, 84][4]. In line with the central limit theorem(s), the larger the sample, the more likely its properties will converge on target population properties [86, 87].

Beyond sample size, sample characteristics should also be representative of, or relevant to, the wider target population [1, 4, 14, 51, 64, 68, 88-90]. Representativeness is dependent on the aims of the models [64, 66, 90]. For example, studies investigating the response of stretching interventions in sedentary populations [91] should not be generalized to elite-level athletes [92].

[4] Any sufficiently large sample may result in even very minimal effects becoming "statistically significant", and so researchers should try to determine what is the minimal clinically important difference (MCID) or the smallest worthwhile effect (SWE) [84, 85].

Sample specificity can limit the generalizability of findings to broader target populations, leading to potential inaccuracies or biases in the equations derived from those samples. It underscores the importance of diverse and representative samples in research to capture the variability across different ethnicities and target populations accurately.

Merely stating that a sample is representative is insufficient and must be substantiated with adequate evidence [64, 88, 89]. While true random samples *on average* tend to converge on population properties, individual samples can differ from the population, most probably by a small amount but potentially to a large degree. Moreover, samples are often not obtained randomly, and instead constructed with the aim to be representative according to a few selected variables deemed relevant (e.g., sex, age, economic status, geographical location, competitive level) [1, 38, 88, 89], again potentially falling into the omitted variable bias trap [52, 53]. A more detailed discussion of what representative sampling is and how it can be recruited falls outside the scope of our work (cf., [64, 88, 89]). However, pragmatic considerations (e.g., high cost, difficulty of access) often inhibit recruiting a representative sample [66, 89]. Frequently, samples in sports exercise and medicine are chosen by convenience [93-96], even in studies using regression models to establish predictors of performance, injury risk, or other outcomes [37, 93, 97-99]. Convenience sampling usually presents certain pragmatic advantages, such as easier access to the sample and reduced temporal and financial costs [65, 66]. Nevertheless, without representativeness of the sample being established, findings should not be promptly generalized to the target population [1, 64-66, 90].[5]

3.2. Context matters

The conditions or environment under which a study was implemented should be controlled to reduce extrinsic sources of variation [14, 64, 101] (e.g., randomized trials) or described with sufficient detail to provide rich context (e.g., observational studies). In conflict with this, the narrower the set of conditions giving rise to the model's data, the less likely that model is applicable outside the original data set [1, 2, 14, 38, 58]. For example, are data collected during pre-season generalizable to the in-season? And if a data set was collected during the Spring, is it generalizable to Autumn? In sports, performance may be affected by circannual rhythms and/or by the structuring of the training periods [102, 103], although this effect is not universal [104]. If the model was based on a club setting, does it work well when applied to clubs with distinct settings (e.g., different facilities or equipment available)? Even the historical context

[5] Due to temporal, financial, and technical demands, most studies are likely to recruit relatively small samples and be statistically underpowered. Perhaps we should be pushing statistical analysis at the individual level instead of relying solely on group-based calculations, as has been suggested elsewhere [100]. This approach could provide meaningful insights for practitioners and afford greater understanding on inter- and intraindividual variability. A discussion of this scenario falls outside the scope of this work.

and local specificities may generate a distinct set of relations and require different interventions [101, 105]. While some findings may translate well between contexts, that is by no means guaranteed [1, 101], even when the contexts are very similar [1]. Even within the same trial, manipulating a single variable (e.g., supervision) may result in very different outcomes [106]. Therefore, any predictive model should ensure that the context where its data were extracted from is translatable to other contexts before considering any extrapolation of their findings.

3.3. Consistency of relationships over time

For a cause to produce a consistent, reproducible effect, the same combination of circumstances must be repeated [1, 54]. Regression models rest on the assumption that relationships between variables are relatively consistent over time [14]: the relationships between variables are expected to remain similar, within a certain (inherent) degree of variability. However, such close relationships do not always hold true [1, 6] (e.g., secular trends in children's physical activity [107]). Relationships between variables may change depending on numerous contextual elements (e.g., politics, culture, economy, environment [1, 14, 101, 105]), or due to the introduction of novel measures such as new pain management protocols implemented over time in healthcare settings [6].

More specific to sports, this may include changes to game rules (e.g., increased number of substitutions in soccer [108]) or the introduction of new technologies (e.g., full body swimsuits

[109]). Motor learning may follow diverse paths (in form and rate), and occurs under dynamic and unpredictable contexts, making generalization inappropriate [110]. Expected long-term stability is paramount for establishing predictions [1, 14], and when the relationships between variables are expected to change in the near future, speaking of predictors is a risky venture [1, 68, 110]. Otherwise, prediction becomes retrodiction or postdiction [1, 67, 68, 82], whereby we are just "predicting" the past (regardless of studies being retrospective or prospective).

3.4. Trustworthiness of the measurements

Trustworthiness could be interpreted as encompassing validity (i.e., whether the results of a test represent what they are supposed to measure) and reliability (i.e., whether the test delivers consistent measurements). As an example, we had previously discussed that the FMS™ was not *valid* for predicting injury risk, despite research showing it may show moderate-to-excellent intra- and inter-rater reliability [111-113]. However, for the purposes of this section, we will focus on issues of reliability.

When analysing the results from any statistical model, it is common to assume that measurements or data points were accurately recorded [1, 14]. However, measurement instruments often present different degrees of accuracy depending on the range of data being assessed (e.g., the reliability of blood lactate analysers varies depending on lactate concentrations [114]), so the measurements are not equally trustworthy across the spectrum of possible values [1, 14]. Moreover, some outcomes are inherently unreliable due to innate biological variability (e.g., blood pressure), requiring multiple measurements to increase reliability [1]. Consequently, a discussion on the degree of confidence in the data points (e.g., measurements) should occur [14], and assume (when relevant) the limited accuracy of measurement tools [1, 110]. This should go beyond calculation of random error and/or reliability in measurements (even if acceptable consistency was reported in prior research) and extend into how data are interpreted.

Reliability plays a major role in predictive statements [1, 8, 49]. For example, Cronbach's alpha (α) [115] is commonly used to determine reliability of measurement of continuous outcomes (e.g., assessing intra- and/or interobserver reliability [1]) and its *arbitrary* thresholds for acceptable values may vary from 0.70 to 0.95 [116, 117]. However, the α values are rarely given proper background or interpretation [116, 117], i.e., their magnitude is not considered when discussing the confidence in the results of a statistical model. Values of 0.70 probably have different statistical implications than values of 0.95, and low values naturally challenge a model's generalizability [1], given they imply reduced levels of reliability in the analysed dataset. We are not aware of regression model studies using such values to moderate their analyses. For example, different reliability values might affect (e.g., increase) the width of confidence intervals, therefore reducing the precision of estimations and potentially affect the interpretation of regression models.

3.5. Predictors by chance: the problem of multiple testing

By assessing multiple outcomes, researchers may attempt to dodge or minimize omitted variable bias (which can never be fully achieved). However, performing multiple tests increases the likelihood of finding statistically significant results by chance (i.e., type I error or false positive) [1, 14, 49, 118, 119][6]. This is doubly problematic for regression models because: (i) there may be many outcomes assessed with a small sample size, and (ii) authors may indefinitely experiment with many regression models until finding one that fits the data. The latter limitation could be partially resolved through well targeted, theory-based selection of the relevant variables [1, 8, 14, 38, 58, 60, 64, 88, 90]: choosing variables that have shown to be pertinent in previous research or whose known mechanisms suggest they may be important [50]. Otherwise, we might be throwing darts into a dark room.

When multiple testing is implemented, the traditional (arbitrary and conventional) *p*-value threshold of 0.05 should be adjusted to more conservative (i.e., lower) cut-off values, decreasing the likelihood of finding significant results by chance [118, 119]. Although there is an ongoing

[6] Another risk involved with multiple testing is reporting bias, whereby the authors undertake questionable research practices and report only the outcomes that delivered significant results (e.g., through p-hacking, HARKing [hypothesizing after the results are known] or cherry-picking [47, 120]), hence providing an incomplete perspective on the phenomenon under analysis [118, 121]. The risk of reporting bias falls outside the scope of this work, but one of its flavors (bias towards statistical significance) has recently been well documented in sports exercise and medicine-related research [122].

debate regarding the best statistical procedures, some form of family-wise error rate control should be implemented [118, 119]. For example, a season-long study with soccer players correlated interlimb asymmetry with unilateral countermovement jump, unilateral drop jump, sprint time, and change of direction speed [123]. The authors applied Bonferroni corrections to account for the family-wise error rate (type I), dividing the usual p-value (0.05) by four (the number of outcomes), having established a significance level at $p < 0.012$. There are additional examples of family-wise error rate corrections in sports sciences [124-126]. Because regression models are often used to generate predictions, the issues involving multiple testing should be carefully considered.

3.6. Confounders

There is always the risk of the relationship between two variables being mediated or even totally explained by a third, unaccounted variable [1, 14, 50]. This requires that predictive models start from a sound theoretical understanding of a phenomenon and demands careful controlling of potential confounders [1, 8, 14, 38, 58, 60, 64, 88, 90]. The variables to assess should be chosen based on an explicit rationale, instead of being added to the study design just because the researchers have the tools to assess it [50]. An illustrative example of the influence of confounders: in multi-arm interventions, different baseline levels on the different arms of the study will confound any pre-post analyses [90]. I.e., if the two study arms differ considerably – at baseline – in a key outcome, any eventual differences in the effectiveness of the interventions of that outcome may potentially be explained away due to baseline imbalances.

To reiterate, although these barriers to prediction are not exclusive to regression, regression represents a class of statistical models where the predictor terminology has been adopted more readily as a default. Even when – quite optimistically – all conceptual assumptions for prediction were appropriately established, would that suffice to venture into the prediction realm? Not quite. Predictive models are built for predicting *new observations* (i.e., they are prospective), unlike explanatory models (which are retrospective) [2]. Therefore, external validation is required by re-testing the model with new data [59].

4. External validation

Can regression models be extrapolated? A typical measure that accompanies regression models is goodness-of-fit [14, 55, 57], but this only serves to internally validate the model and should not be used for predictions outside the sample that was analysed [1, 2, 14, 57]. In fact, several regression models may fit the initial data, while a single regression model may fit multiple, unrelated data sets [3, 14, 49]. Goodness-of-fit should not be used to infer causation [14, 57] nor to infer out-of-sample predictive power. Indeed, statistical models (including regression) risk overfitting the original data set and therefore its "predictions" fail when applied to different samples [1-3, 6, 14, 38, 47, 49]. Even the usual procedure of splitting the original dataset into a training set and a testing set [1, 2, 14] still faces the problem that all the data results from a very specific sample analysed at very specific timepoints and under specific contexts.

For purposes of extrapolation (the basis for out-of-sample predictions), the models should be tested in data sets not used to develop said models or in subsets of the original data that were not used to create the original model [1, 6, 8, 14, 49, 51, 58, 68]. However, validation with different data sets (or sub-sets) is often left unperformed [14, 47, 49]. The so-called predictive analysis in sports is often performed in cross-sectional studies, which may reflect nothing more than idiosyncratic associations found for very specific samples, under very specific contexts (e.g., match analysis in team ball sports [82])[7]. Since the strength of predictive models relies in delivering accurate predictions, this can and should be empirically tested [2, 3].

When external validation is performed, the original prediction models are often abandoned in favour of new models [1, 6]: the models were not predictive at all, but retrodictive or postdictive [67, 82]. This links to the wider problem of reproducibility and replicability in science [1, 127-130]. At a minimum, we should successfully reproduce our results twice before delivering more definitive statements [14]; but preferably, independent verifications of the

[7] A discussion of predictive models based on machine learning and other methods is outside the scope of this work, even if some of those methods may start from regression models (although often performing external validation). Here, the focus lies entirely on the legitimacy of regression models *per se* to establish predictors.

models should be performed [1, 14]. Every model should be considered provisional, and an ongoing cycle of testing, refining, and updating the model is needed [1, 6, 14, 38, 47]. Therefore, any regression models built upon a given sample should not be readily referred to as predictive; first, such models should be replicated with different samples, and preferably across similar (but not necessarily equal) contexts [1, 14, 47, 49]. Expanding circles of extrapolation can be thought out, with the initial regression models being tested using progressively more dissimilar samples (within a target population) and extending to ever more diverse contexts (Figure 1).

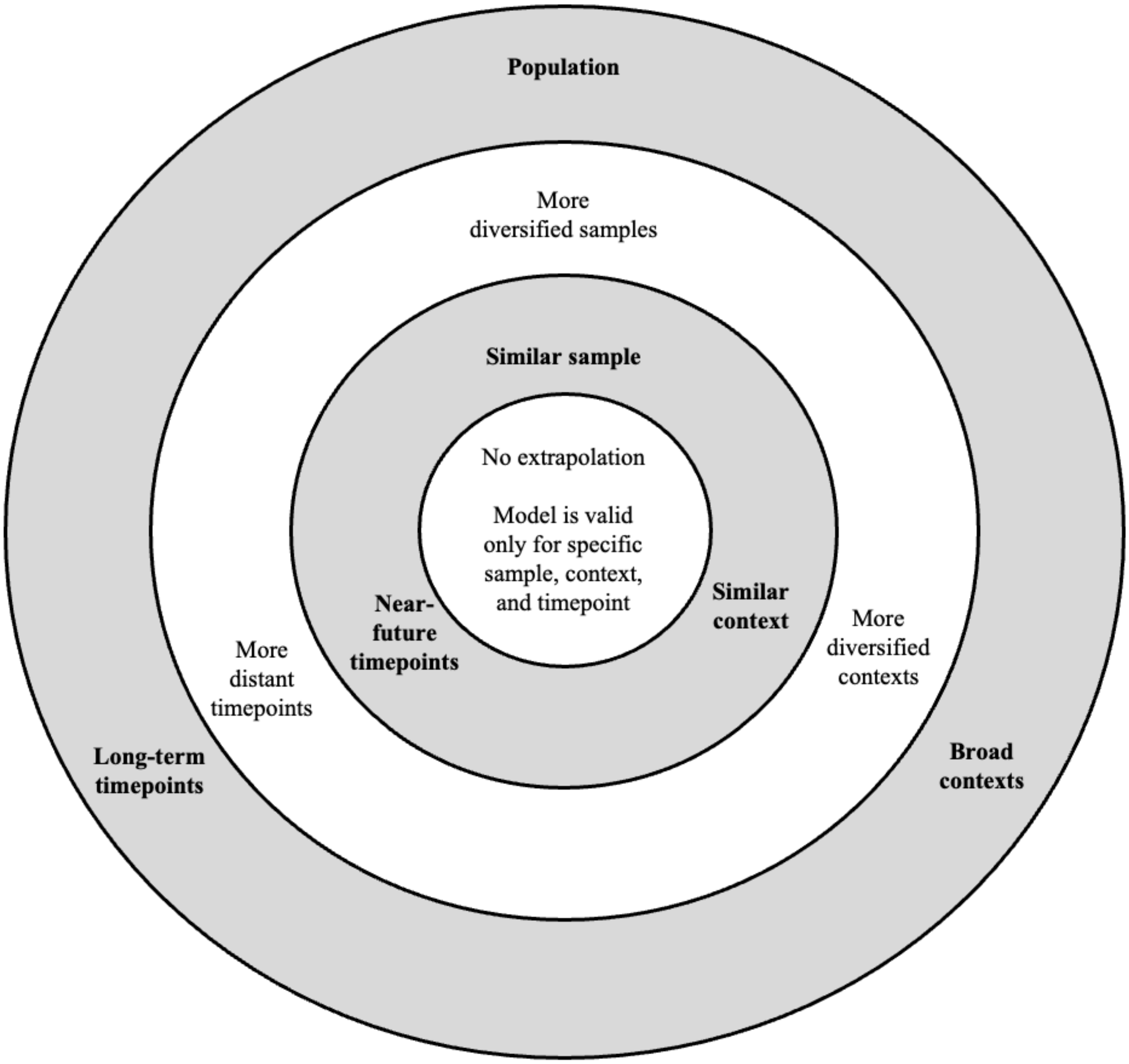

Figure 1. Expanding circles of extrapolation.

An illustrative case can be provided for sports exercise and medicine: Bullock et al. [38] systematically reviewed injury prediction models in sports, and their conclusions largely converge with the concerns we have outlined throughout this work. The authors showed that 60% of the included studies (*n* = 18) exclusively used regression to define a prediction model. None of these studies, however, provided any external validation for the model; ergo, no extrapolation should have been made. Most models had also been developed using small sample sizes and were judged at high risk of bias [38].

5. When using regression models, when can we apply the term "predictors"?

Ultimately, regression models are calculated based on *past* data relative to a specific sample in a specific context [6, 14, 49, 64, 88]. Even if their explanatory power is sound, it does not mean they will have good predictive power [2]. Therefore, applying the term "predictors" based solely on using a regression model represents an unjustified, untested, and perilous concept, as well as a leap of faith [1, 6, 14, 38, 49, 68]. In Figure 2, we propose an algorithm of criteria that should be considered before venturing into predictive-like language. A formal checklist is provided as supplementary table 1, after the references.

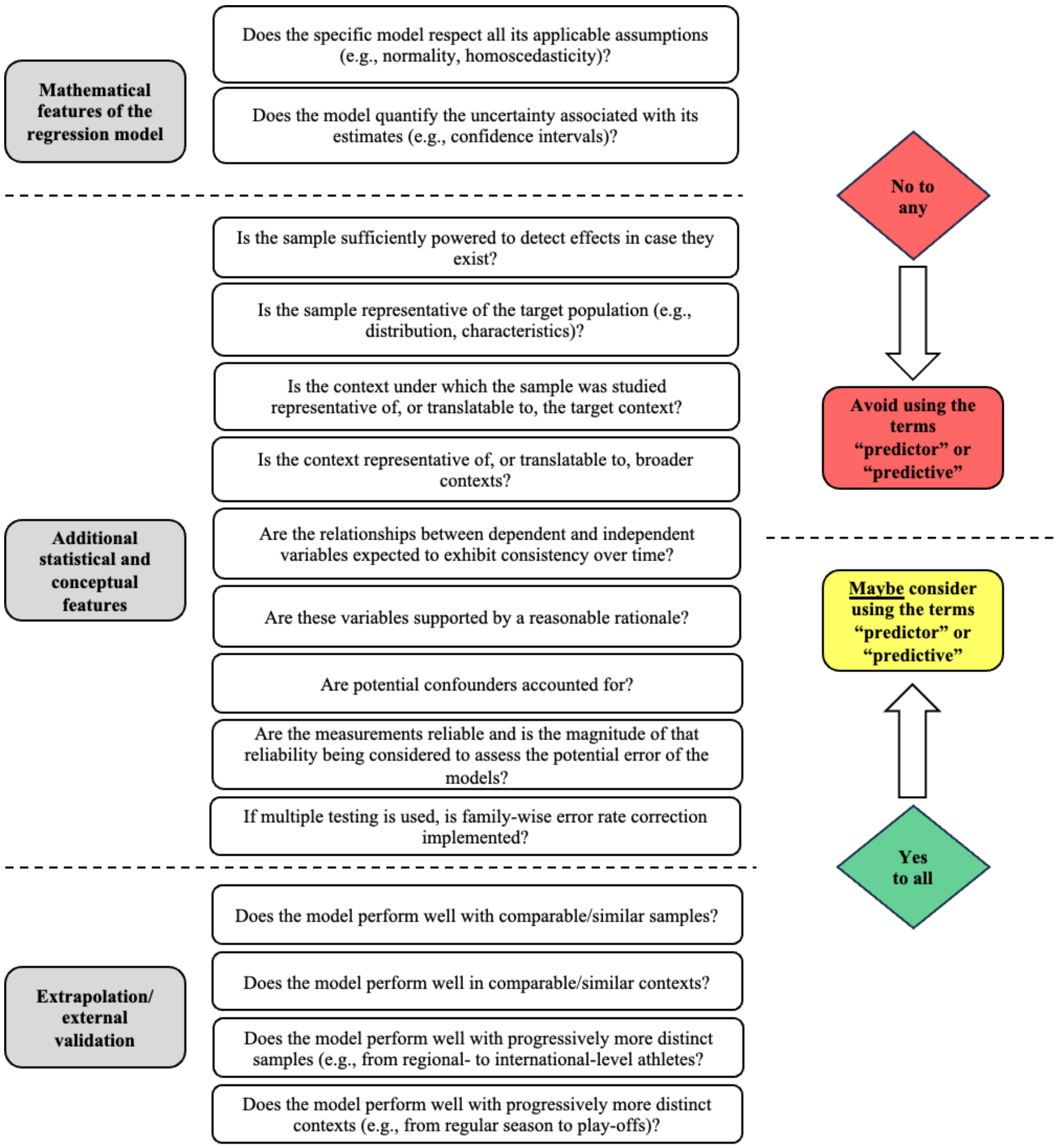

Figure 2. Algorithm to guide regression-based research use of the terms “predictor” or “predictive”.

In cases where all the above-mentioned criteria are met, researchers may wish to ponder mentioning predictors and predictive modelling when using regression models. In the absence of such assurance, researchers would perhaps provide a better service to the wider community by eschewing prediction terminology. Regression models, even when not used for prediction, can still provide valuable insights into relationships between different sets of variables [32], and therefore their value does not rely on a putative predictive ability. We reiterate and underpin previous calls to be more careful when using predictor-related terminology [2].

6. Concluding remarks

Regression models provide valuable insights into the relations between chosen sets of variables. However, they are limited in scope, and can hardly be extrapolated to other samples or to the target population, especially when relying on small and/or non-representative samples. The results of exploratory regression models, created without a strong underlying rationale, should be analysed with caution. Mathematically, regression models have no special features (in comparison with other models, including ANOVA and correlation) that imbue them with predictive power. Applying the term “predictors” generalizes what (potentially) should not be generalized. Ensuring that both sample and context are representative, that sound methodological procedures were followed, and that the relations are expected to remain relatively stable in time are necessary, but not sufficient conditions for prediction: external validation is required for extrapolation to be implemented. Even so, regression models should best be viewed as ongoing processes, subject to continuous monitoring. This pitfall is prevalent across most scientific fields, but also within sports exercise and medicine, where the term "predictors" should probably not be applied to research using regression models, nor should the titles include words such as "predictor" or "predictive".

Contributions

JA conceived the initial idea. All authors contributed significantly to conceive, write, and revise the initial drafts, as well as the current version of the article. All authors read and reviewed the manuscript critically for important intellectual content and approved the final version to be submitted. All authors agreed to upload to *SportRxiv*.

Acknowledgments

The authors are deeply grateful to Patrick Ward for his valuable critical comments and discussion, which helped steering the organization and writing of the manuscript. This does not imply that Patrick Ward endorses the views expressed in the manuscript, for which the authors take full responsibility.

Funding information and competing interests

CIFI$_2$D is financed by the Portuguese Foundation for Science and Technology, under the DOI https://doi.org/10.54499/UIDB/05913/2020.

The authors have no competing interests.

Data and Supplementary Material Accessibility

Supplementary table 1 available after the references.

REFERENCES

1. Steyerberg EW. Clinical prediction models. 2nd ed. Berlin: Springer; 2019.

2. Shmueli G. To Explain or to Predict? Statistical Science. 2010;25(3):289-310. doi: https://doi.org/10.1214/10-sts330.

3. Breiman L. Statistical Modeling: The Two Cultures. Statistical Science. 2001;16(3):199-215.

4. Mujere N. Sampling in Research. In: Baran ML, Jones JE, editors. Mixed Methods Research for Improved Scientific Study. Hershey, PA, USA: IGI Global; 2016. p. 107-21.

5. Smith G. Chapter 5 - Sampling. In: Smith G, editor. Essential Statistics, Regression, and Econometrics. Boston: Academic Press; 2012. p. 141-62.

6. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol. 2008;61(1):76-86. doi: https://doi.org/10.1016/j.jclinepi.2007.04.018.

7. Haydon DS, Pinder RA, Grimshaw PN, Robertson WSP, Holdback CJM. Prediction of Propulsion Kinematics and Performance in Wheelchair Rugby. Front Sports Act Living. 2022;4:Article 856934. doi: https://doi.org/10.3389/fspor.2022.856934.

8. Palmer PB, O'Connell DG. Regression analysis for prediction: understanding the process. Cardiopulm Phys Ther J. 2009;20(3):23-6.

9. de Queiroz JHM, Frota JP, Dos Reis FA, de Oliveira RR. Development and Predictive Validation of the Brazilian Adductor Performance Test for Estimating the Chance of Hip Adductor Injuries in Elite Soccer Athletes. Int J Sports Physiol Perform. 2023;18(6):653-9. doi: https://doi.org/10.1123/ijspp.2022-0306.

10. Hoover DL, Killian CB, Tinius RA, Bellar DM, Wilkinson SG, Esslinger FT, Judge LW. Predictive Validity of a Functional Movement Screen in Professional Basketball Players. Medicina. 2020;56(12):724. doi: https://doi.org/10.3390/medicina56120724.

11. Tears C, Rae G, Hide G, Sinha R, Franklin J, Brand P, et al. The British Athletics Muscle Injury Classification grading system as a predictor of return to play following hamstrings injury in

professional football players. Phys Ther Sport. 2022;58:46-51. doi: https://doi.org/10.1016/j.ptsp.2022.08.002.

12. Coogan SM, Schock CS, Hansen-Honeycutt J, Caswell S, Cortes N, Ambegaonkar JP. Functional Movement Screen™ (FMS™) scores do not predict overall or lower extremity injury risk in collegiate dances. Int J Sports Phys Ther. 2020;15(6):1029-35. doi: https://doi.org/10.26603/ijspt20201029.

13. Costa G, Afonso J, Barbosa R, Coutinho P, Mesquita I. Predictors of attack efficacy and attack type in high-level Brazilian women's volleyball. Kinesiology. 2014;46(2):242-8.

14. Good PI, Hardin JW. Common errors in statistics (and how to avoid them). 4th ed. New Jersey: Wiley; 2012.

15. Fraser P, Dimitra B, Edward KC, Caroline S, Claire S, Kim M, Anand DP. Study of the measurement and predictive validity of the Functional Movement Screen. BMJ Open Sport Exerc Med. 2018;4(1):Article e000357. doi: https://doi.org/10.1136/bmjsem-2018-000357.

16. Lau R, Mukherjee S. Prevalence of Shoulder and Elbow Overuse Injuries Among Competitive Overhead Youth Athletes in Singapore. Orthop J Sports Med. 2023;11(3):Article 23259671231156199. doi: https://doi.org/10.1177/23259671231156199.

17. Biese KM, McGuine TA, Kliethermes SA, Bell DR, Post EG, Watson AM, et al. Sport specialization and sport participation opportunities and their association with injury history in female high school volleyball athletes. Phys Ther Sport. 2020;45:86-92. doi: https://doi.org/10.1016/j.ptsp.2020.06.005.

18. Gou Q, Li S. Study on the correlation between basketball players' multiple-object tracking ability and sports decision-making. PLoS One. 2023;18(4):Article e0283965. doi: https://doi.org/10.1371/journal.pone.0283965.

19. Loturco I, Barbosa AC, Nocentini RK, Pereira LA, Kobal R, Kitamura K, et al. A Correlational Analysis of Tethered Swimming, Swim Sprint Performance and Dry-land Power Assessments. Int J Sports Med. 2016;37(3):211-8. doi: https://doi.org/10.1055/s-0035-1559694.

20. Csuka SI, Rohánszky M, Konkolÿ Thege B. Gender differences in the predictors of quality of life in patients with cancer: A cross sectional study. Eur J Oncol Nurs. 2023;68:Article 102492. doi: https://doi.org/10.1016/j.ejon.2023.102492.

21. Karabay E, Tosun Ç, Karsiyakali N, Kayar K, Aykan S, Aslan AR, et al. Identifying the predictors of estimated glomerular filtration rate after partial nephrectomy with a nonlinear regression model. Int J Clin Pract. 2021;75(3):Article e13763. doi: https://doi.org/10.1111/ijcp.13763.

22. Kasimanickam MR, Kerr DCR, Killion RE, Naimi TS, Lira MC, Bae H. State Alcohol Policy Environments of U.S. Colleges: Predictors of Sexual Assault and Alcohol-Related Arrest and Disciplinary Action. Am J Prev Med. 2024;66(1):1-9. doi: https://doi.org/10.1016/j.amepre.2023.09.015.

23. Lowenstein NA, Chang Y, Mass H, Mercurio AM, Ukogu C, Katz JN, Matzkin EG. Preoperative Predictors of Arthroscopic Partial Meniscectomy Outcomes: The APM Index Score. Am J Sports Med. 2024;52(1):116-23. doi: https://doi.org/10.1177/03635465231210303.

24. Mohamed NF, Jiun Ting T, Ab Manan N, Mohd Senari IF, Muhammad Firdaus Chan MF, Rahmatullah B, et al. Prevalence and predictors of social anxiety disorders among Malaysian secondary school students during the COVID-19 pandemic: Exploring the influence of internet gaming disorder and impulsivity. Clin Child Psychol Psychiatry. In press. doi: https://doi.org/10.1177/13591045231206967.

25. Sarfika R, Wenny BP, Muliantino MR, Novrianda D, Saifudin I. Exploring predictors of perceived stress: a cross-sectional study among nursing students during their clinical practice. J Res Nurs. 2023;28(6-7):469-82. doi: https://doi.org/10.1177/17449871231198770.

26. Azad MS, Khan SS, Hossain R, Rahman R, Momen S. Predictive modeling of consumer purchase behavior on social media: Integrating theory of planned behavior and machine learning for actionable insights. PLoS One. 2023;18(12):Article e0296336. doi: https://doi.org/10.1371/journal.pone.0296336.

27. Huang R, Cai Y, He Y, Yu Z, Zhao L, Wang T, et al. Predictive Model of Oxaliplatin-induced Liver Injury Based on Artificial Neural Network and Logistic Regression. J Clin Transl Hepatol. 2023;11(7):1455-64. doi: https://doi.org/10.14218/jcth.2023.00399.

28. Liu M, Li Q, Zhang J, Chen Y. Development and Validation of a Predictive Model Based on LASSO Regression: Predicting the Risk of Early Recurrence of Atrial Fibrillation after Radiofrequency Catheter Ablation. Diagnostics (Basel). 2023;13(22):Article 3403. doi: https://doi.org/10.3390/diagnostics13223403.

29. Loria A, Cai X, Gao S, Zhao T, Juviler P, Li Y, et al. Development and validation of multivariable predictive models for recurrence and mortality following nonoperative management of sigmoid volvulus. Colorectal Dis. 2024;26(2):356-63. doi: https://doi.org/10.1111/codi.16849.

30. Mukherjee S, Martínez-González J, Dowling DP, Gowen AA. Predictive modelling of the water contact angle of surfaces using attenuated total reflection - Fourier transform infrared (ATR-FTIR) chemical imaging and partial least squares regression (PLSR). Analyst. 2018;143(15):3729-40. doi: https://doi.org/10.1039/c8an00320c.

31. Bingham NH, Fry JM. Linear Regression. In: Bingham NH, Fry JM, editors. Regression: Linear Models in Statistics. London: Springer London; 2010. p. 1-32.

32. Smith G. Chapter 8 - Simple Regression. In: Smith G, editor. Essential Statistics, Regression, and Econometrics. Boston: Academic Press; 2012. p. 219-58.

33. Gu X, Shen X, Chu JH, Fang TT, Jiang L. Frailty, Illness Perception and Lung Functional Exercise Adherence in Lung Cancer Patients After Thoracoscopic Surgery. Patient Prefer Adherence. 2023;17:2773-87. doi: https://doi.org/10.2147/ppa.S435944.

34. Hewett TE, Myer GD, Ford KR, Heidt RS, Jr., Colosimo AJ, McLean SG, et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. Am J Sports Med. 2005;33(4):492-501. doi: https://doi.org/10.1177/0363546504269591.

35. Liu H, Ding H, Xuan J, Gao X, Huang X. The functional movement screen predicts sports injuries in Chinese college students at different levels of physical activity and sports

performance. Heliyon. 2023;9(6):Article e16454. doi: https://doi.org/10.1016/j.heliyon.2023.e16454.

36. Plisky PJ, Rauh MJ, Kaminski TW, Underwood FB. Star Excursion Balance Test as a predictor of lower extremity injury in high school basketball players. J Orthop Sports Phys Ther. 2006;36(12):911-9. doi: https://doi.org/10.2519/jospt.2006.2244.

37. Molina-Cárdenas Á, Álvarez-Yates T, García-García O. Predicting Hamstring Strains in Soccer Players Based on ROM: An Analysis From a Gender Perspective. Research Quarterly for Exercise and Sport. 2023;94(2):493-9. doi: https://doi.org/10.1080/02701367.2021.2011091.

38. Bullock GS, Mylott J, Hughes T, Nicholson KF, Riley RD, Collins GS. Just How Confident Can We Be in Predicting Sports Injuries? A Systematic Review of the Methodological Conduct and Performance of Existing Musculoskeletal Injury Prediction Models in Sport. Sports Med. 2022;52(10):2469-82. doi: https://doi.org/10.1007/s40279-022-01698-9.

39. Maszczyk A, Roczniok R, Waśkiewicz Z, Czuba M, Mikołajec K, Zajac A, Stanula A. Application of regression and neural models to predict competitive swimming performance. Percept Mot Skills. 2012;114(2):610-26. doi: https://doi.org/10.2466/05.10.Pms.114.2.610-626.

40. Murray NP, Lawton J, Rider P, Harris N, Hunfalvay M. Oculomotor Behavior Predict Professional Cricket Batting and Bowling Performance. Front Hum Neurosci. 2021;15:Article 768585. doi: https://doi.org/10.3389/fnhum.2021.768585.

41. Fryer SM, Giles D, Palomino IG, de la OPA, España-Romero V. Hemodynamic and Cardiorespiratory Predictors of Sport Rock Climbing Performance. J Strength Cond Res. 2018;32(12):3534-41. doi: https://doi.org/10.1519/jsc.0000000000001860.

42. M Y, Dave AK, Patel SS, Parbat R, Shah V, Gandhi R. Association Between Sarcopenia and Chronic Renal Failure (Overt and Concealed) in Chronic Obstructive Pulmonary Disease (COPD) Patients: A Cross-Sectional Study. Cureus. 2023;15(10):Article e46870. doi: https://doi.org/10.7759/cureus.46870.

43. Song L, Zhang Z, Wang C, Yu J, Liu L, Wang M. Binary Logistic Regression Analysis on the Association of Occlusion with the Temporomandibular Joint Arthralgia in the Partial Denture Wearers-A Retrospective Study. Int J Prosthodont. 2023;36(6):154-67. doi: https://doi.org/10.11607/ijp.8118.

44. Aktar MF, Chowdhury MH, Rahman MS. A quantile regression approach to identify risk factors for high blood glucose levels among Bangladeshi individuals. Health Sci Rep. 2023;6(12):Article e1772. doi: https://doi.org/10.1002/hsr2.1772.

45. Tobías A, Casals M, Saez M, Kamada M, Kim Y. Impacts of ambient temperature and seasonal changes on sports injuries in Madrid, Spain: a time-series regression analysis. BMJ Open Sport Exerc Med. 2021;7(4):Article e001205. doi: https://doi.org/10.1136/bmjsem-2021-001205.

46. Legget KT, Wylie KP, Cornier MA, Melanson EL, Paschall CJ, Tregellas JR. Exercise-related changes in between-network connectivity in overweight/obese adults. Physiol Behav. 2016;158:60-7. doi: https://doi.org/10.1016/j.physbeh.2016.02.031.

47. Bullock GS, Ward P, Impellizzeri FM, Kluzek S, Hughes T, Dhiman P, et al. The Trade Secret Taboo: Open Science Methods are Required to Improve Prediction Models in Sports Medicine and Performance. Sports Med. 2023;53(10):1841-9. doi: https://doi.org/10.1007/s40279-023-01849-6.

48. Windt J, Ardern CL, Gabbett TJ, Khan KM, Cook CE, Sporer BC, Zumbo BD. Getting the most out of intensive longitudinal data: a methodological review of workload-injury studies. BMJ Open. 2018;8(10):Article e022626. doi: https://doi.org/10.1136/bmjopen-2018-022626.

49. Harrell FE, Jr., Lee KL, Matchar DB, Reichert TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. Cancer Treat Rep. 1985;69(10):1071-77.

50. Impellizzeri FM, Shrier I, McLaren SJ, Coutts AJ, McCall A, Slattery K, et al. Understanding Training Load as Exposure and Dose. Sports Med. 2023;53(9):1667-79. doi: https://doi.org/10.1007/s40279-023-01833-0.

51. Bahr R. Why screening tests to predict injury do not work – and probably never will…: a critical review. Br J Sports Med. 2017;50:776-80. doi: https://doi.org/10.1136/bjsports-2016-096256.

52. Wilms R, Mäthner E, Winnen L, Lanwehr R. Omitted variable bias: A threat to estimating causal relationships. Methods in Psychology. 2021;5:Article 100075. doi: https://doi.org/10.1016/j.metip.2021.100075.

53. Hirukawa M, Murtazashvili I, Prokhorov A. Yet another look at the omitted variable bias. Econometric Reviews. 2023;42(1):1-27. doi: https://doi.org/10.1080/07474938.2022.2157965.

54. Francis G. Kinship and Correlation. Statistical Science. 1989;4(2):81-6. doi: https://doi.org/10.1214/ss/1177012581.

55. John A. Fisher and Regression. Statistical Science. 2005;20(4):401-17. doi: https://doi.org/10.1214/088342305000000331.

56. Yule GU. On the Theory of Correlation. Journal of the Royal Statistical Society. 1897;60(4):812-54. doi: https://doi.org/10.2307/2979746.

57. Fisher RA. The Goodness of Fit of Regression Formulae, and the Distribution of Regression Coefficients. Journal of the Royal Statistical Society. 1922;85(4):597-612. doi: https://doi.org/10.2307/2341124.

58. Palmer J, Carlson P. Problems with the Use of Regression Analysis in Prediction Studies. Journal of Research in Crime and Delinquency. 1976;13(1):64-81. doi: https://doi.org/10.1177/002242787601300106.

59. Smith G. Chapter 9 - The Art of Regression Analysis. In: Smith G, editor. Essential Statistics, Regression, and Econometrics. Boston: Academic Press; 2012. p. 259-95.

60. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc. 2009;41(1):3-13. doi: https://doi.org/10.1249/MSS.0b013e31818cb278.

61. Kadlec D, Sainani KL, Nimphius S. With Great Power Comes Great Responsibility: Common Errors in Meta-Analyses and Meta-Regressions in Strength & Conditioning Research. Sports Med. 2023;53(2):313-25. doi: https://doi.org/10.1007/s40279-022-01766-0.

62. Yoo W, Mayberry R, Bae S, Singh K, Peter He Q, Lillard JW, Jr. A Study of Effects of MultiCollinearity in the Multivariable Analysis. Int J Appl Sci Technol. 2014;4(5):9-19.

63. Gregorich M, Strohmaier S, Dunkler D, Heinze G. Regression with Highly Correlated Predictors: Variable Omission Is Not the Solution. Int J Environ Res Public Health. 2021;18(8):Article 4259. doi: https://doi.org/10.3390/ijerph18084259.

64. Jacqueline ER, Yongqi Z, Priya D, Shruti HM, Bryan L. Defining representativeness of study samples in medical and population health research. BMJ Medicine. 2023;2(1):Article e000399. doi: https://doi.org/10.1136/bmjmed-2022-000399.

65. Jager J, Putnick DL, Bornstein MH. II. More than just convenient: The scientific merits of homogeneous convenience samples. Monogr Soc Res Child Dev. 2017;82(2):13-30. doi: https://doi.org/10.1111/mono.12296.

66. Andrade C. The Inconvenient Truth About Convenience and Purposive Samples. Indian Journal of Psychological Medicine. 2021;43(1):86-8. doi: https://doi.org/10.1177/0253717620977000.

67. Taagepera R. Predictive versus postdictive models. European Political Science. 2007;6(2):114-23. doi: https://doi.org/10.1057/palgrave.eps.2210120.

68. Afonso J, Fonseca H, Ramirez-Campillo R, Olivares-Jabalera J, Rocha-Rodrigues S. Prevention Strategies of Lower Limb Muscle Injuries. Orthopaedic Sports Medicine. 2023. p. 1-31.

69. Kite CS, Nevill A. The Predictors and Determinants of Inter-Seasonal Success in a Professional Soccer Team. J Hum Kinet. 2017;58:157-67. doi: https://doi.org/10.1515/hukin-2017-0084.

70. An Z, Jalles JT, Loungani P. How well do economists forecast recessions? International Finance. 2018;21(2):100-21. doi: https://doi.org/10.1111/infi.12130.

71. Reiter CR, Killelea C, Faherty MS, Zerega RJ, Westwood C, Sell TC. Force-plate derived predictors of lateral jump performance in NCAA Division-I men's basketball players. PLoS One. 2023;18(4):Article e0284883. doi: https://doi.org/10.1371/journal.pone.0284883.

72. Chamberlain ST, Hale BD. Competitive state anxiety and self-confidence: intensity and direction as relative predictors of performance on a golf putting task. Anxiety Stress Coping. 2007;20(2):197-207. doi: https://doi.org/10.1080/10615800701288572.

73. Pérez-Olea JI, Valenzuela PL, Aponte C, Izquierdo M. Relationship Between Dryland Strength and Swimming Performance: Pull-Up Mechanics as a Predictor of Swimming Speed. J Strength Cond Res. 2018;32(6):1637-42. doi: https://doi.org/10.1519/jsc.0000000000002037.

74. Cowden RG, Fuller DK, Anshel MH. Psychological predictors of mental toughness in elite tennis: an exploratory study in learned resourcefulness and competitive trait anxiety. Percept Mot Skills. 2014;119(3):661-78. doi: https://doi.org/10.2466/30.PMS.119c27z0.

75. Markström JL, Olsson CJ. Countermovement jump peak force relative to body weight and jump height as predictors for sprint running performances: (in)homogeneity of track and field athletes? J Strength Cond Res. 2013;27(4):944-53. doi: https://doi.org/10.1519/JSC.0b013e318260edad.

76. Dorrel B, Long T, Shaffer S, Myer GD. The Functional Movement Screen as a Predictor of Injury in National Collegiate Athletic Association Division II Athletes. J Athl Train. 2018;53(1):29-34. doi: https://doi.org/10.4085/1062-6050-528-15.

77. Zarei M, Soltanirad S, Kazemi A, Hoogenboom BJ, Hosseinzadeh M. Composite functional movement screen score predicts injuries in youth volleyball players: a prospective cohort study. Sci Rep. 2022;12(1):Article 20207. doi: https://doi.org/10.1038/s41598-022-24508-8.

78. Newton F, McCall A, Ryan D, Blackburne C, aus der Fünten K, Meyer T, et al. Functional Movement Screen (FMS™) score does not predict injury in English Premier League youth academy football players. Science and Medicine in Football. 2017;1(2):102-6. doi: https://doi.org/10.1080/24733938.2017.1283436.

79. Skorski S, Hecksteden A. Coping With the "Small Sample–Small Relevant Effects" Dilemma in Elite Sport Research. Int J Sports Physiol Perform. 2021;16(11):1559-60. doi: https://doi.org/10.1123/ijspp.2021-0467.

80. Abt G, Boreham C, Davison G, Jackson R, Nevill A, Wallace E, Williams M. Power, precision, and sample size estimation in sport and exercise science research. J Sports Sci. 2020;38(17):1933-5. doi: https://doi.org/10.1080/02640414.2020.1776002.

81. Afonso J, Olivares-Jabalera J, Fernandes RJ, Clemente FM, Rocha-Rodrigues S, Claudino JG, et al. Effectiveness of Conservative Interventions After Acute Hamstrings Injuries in Athletes: A Living Systematic Review. Sports Med. 2023;53(3):615-35. doi: https://doi.org/10.1007/s40279-022-01783-z.

82. Sarmento H, Clemente FM, Afonso J, Araújo D, Fachada M, Nobre P, Davids K. Match Analysis in Team Ball Sports: An Umbrella Review of Systematic Reviews and Meta-Analyses. Sports Med Open. 2022;8(1):Article 66. doi: https://doi.org/10.1186/s40798-022-00454-7.

83. Beck TW. The importance of a priori sample size estimation in strength and conditioning research. J Strength Cond Res. 2013;27(8):2323-37. doi: https://doi.org/10.1519/JSC.0b013e318278eea0.

84. Willigenburg NW, Poolman RW. The difference between statistical significance and clinical relevance. The case of minimal important change, non-inferiority trials, and smallest worthwhile effect. Injury. 2023;54:Article 110764. doi: https://doi.org/10.1016/j.injury.2023.04.051.

85. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. J Clin Epidemiol. 2017;82:128-36. doi: https://doi.org/10.1016/j.jclinepi.2016.11.016.

86. Kwak SG, Kim JH. Central limit theorem: the cornerstone of modern statistics. Korean J Anesthesiol. 2017;70(2):144-56. doi: https://doi.org/10.4097/kjae.2017.70.2.144.

87. Chow YS, Teicher H. Central Limit Theorems. In: Chow YS, Teicher H, editors. Probability Theory: Independence, Interchangeability, Martingales. New York, NY: Springer New York; 1997. p. 313-53.

88. Kruskal W, Mosteller F. Representative Sampling, III: The Current Statistical Literature. International Statistical Review / Revue Internationale de Statistique. 1979;47(3):245-65. doi: https://doi.org/10.2307/1402647.

89. van Hoeven LR, Janssen MP, Roes KCB, Koffijberg H. Aiming for a representative sample: Simulating random versus purposive strategies for hospital selection. BMC Medical Research Methodology. 2015;15(1):90. doi: https://doi.org/10.1186/s12874-015-0089-8.

90. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch V. Cochrane Handbook for Systematic Reviews of Interventions. 2nd ed. Chichester (UK): John Wiley & Sons; 2019.

91. Moreno MA, Catai AM, Teodori RM, Borges BL, Cesar Mde C, Silva E. Effect of a muscle stretching program using the Global Postural Reeducation method on respiratory muscle strength and thoracoabdominal mobility of sedentary young males. J Bras Pneumol. 2007;33(6):679-86. doi: https://doi.org/10.1590/s1806-37132007000600011.

92. Moreno-Perez V, Hernandez-Davo JL, Nakamura F, Lopez-Samanes A, Jimenez-Reyes P, Fernandez-Fernandez J, Behm DG. Post-activation performance enhancement of dynamic stretching and heavy load warm-up strategies in elite tennis players. Journal of Back and Musculoskeletal Rehabilitation. 2021;34(3):413-23. doi: https://doi.org/10.3233/bmr-191710.

93. Payne KA, Berg K, Latin RW. Ankle injuries and ankle strength, flexibility, and proprioception in college basketball players. J Athl Train. 1997;32(3):221-5.

94. Zarei H, Bervis S, Piroozi S, Motealleh A. Added Value of Gluteus Medius and Quadratus Lumborum Dry Needling in Improving Knee Pain and Function in Female Athletes With

Patellofemoral Pain Syndrome: A Randomized Clinical Trial. Arch Phys Med Rehabil. 2020;101(2):265-74. doi: https://doi.org/10.1016/j.apmr.2019.07.009.

95. Field A, Harper LD, Chrismas BCR, Fowler PM, McCall A, Paul DJ, et al. The Use of Recovery Strategies in Professional Soccer: A Worldwide Survey. Int J Sports Physiol Perform. 2021;16(12):1804-15. doi: https://doi.org/10.1123/ijspp.2020-0799.

96. Lee TI, Wang MY, Huang BR, Hsu CY, Chien CY. Effects of Psychological Capital and Sport Anxiety on Sport Performance in Collegiate Judo Athletes. Am J Health Behav. 2022;46(2):197-208. doi: https://doi.org/10.5993/ajhb.46.2.9.

97. Wendtlandt M, Wicker P. The Effects of Sport Activities and Environmentally Sustainable Behaviors on Subjective Well-Being: A Comparison Before and During COVID-19. Front Sports Act Living. 2021;3:Article 659837. doi: https://doi.org/10.3389/fspor.2021.659837.

98. Elsahoryi NA, Trakman G, Al Kilani A. General and sports nutrition knowledge among Jordanian adult coaches and athletes: A cross-sectional survey. PLoS One. 2021;16(11):Article e0258123. doi: https://doi.org/10.1371/journal.pone.0258123.

99. Mendez-Rebolledo G, Ager AL, Ledezma D, Montanez J, Guerrero-Henriquez J, Cruz-Montecinos C. Role of active joint position sense on the upper extremity functional performance tests in college volleyball players. PeerJ. 2022;10:Article e13564. doi: https://doi.org/10.7717/peerj.13564.

100. Neumann ND, Van Yperen NW, Brauers JJ, Frencken W, Brink MS, Lemmink K, et al. Nonergodicity in Load and Recovery: Group Results Do Not Generalize to Individuals. Int J Sports Physiol Perform. 2022;17(3):391-9. doi: https://doi.org/10.1123/ijspp.2021-0126.

101. Rhiannon EE, Peter C, Pat H, Hannah L, Laurence M, Simon M, et al. When and how do 'effective' interventions need to be adapted and/or re-evaluated in new contexts? The need for guidance. Journal of Epidemiology and Community Health. 2019;73(6):481. doi: https://doi.org/10.1136/jech-2018-210840.

102. Krzywański J, Pokrywka A, Młyńczak M, Mikulski T. Is vitamin D status reflected by testosterone concentration in elite athletes? Biol Sport. 2020;37(3):229-37. doi: https://doi.org/10.5114/biolsport.2020.95633.

103. Wang H, Theall BM, Early KS, Vincellette C, Robelot L, Sharp RL, et al. Seasonal changes in physiological and psychological parameters of stress in collegiate swimmers. Sci Rep. 2023;13(1):Article 10995. doi: https://doi.org/10.1038/s41598-023-37124-x.

104. Fort-Vanmeerhaeghe A, Bishop C, Buscà B, Vicens-Bordas J, Arboix-Alió J. Seasonal variation of inter-limb jumping asymmetries in youth team-sport athletes. J Sports Sci. 2021;39(24):2850-8. doi: https://doi.org/10.1080/02640414.2021.1968123.

105. Villeval M, Gaborit E, Berault F, Lang T, Kelly-Irving M. Do the key functions of an intervention designed from the same specifications vary according to context? Investigating the transferability of a public health intervention in France. Implementation Science. 2019;14(1):35. doi: https://doi.org/10.1186/s13012-019-0880-8.

106. Baumann FT, Reimer N, Gockeln T, Reike A, Hallek M, Ricci C, et al. Supervised pelvic floor muscle exercise is more effective than unsupervised pelvic floor muscle exercise at improving urinary incontinence in prostate cancer patients following radical prostatectomy - a systematic review and meta-analysis. Disabil Rehabil. 2022;44(19):5374-85. doi: https://doi.org/10.1080/09638288.2021.1937717.

107. Dollman J, Norton K, Norton L. Evidence for secular trends in children's physical activity behaviour. Br J Sports Med. 2005;39(12):892-7. doi: https://doi.org/10.1136/bjsm.2004.016675.

108. García-Aliaga A, Martín-Castellanos A, Marquina Nieto M, Muriarte Solana D, Resta R, López del Campo R, et al. Effect of Increasing the Number of Substitutions on Physical Performance during Periods of Congested Fixtures in Football. Sports. 2023;11(2):25. doi: https://doi.org/10.3390/sports11020025.

109. Foster L, James D, Haake S. Influence of full body swimsuits on competitive performance. Procedia Engineering. 2012;34:712-7. doi: https://doi.org/10.1016/j.proeng.2012.04.121.

110. Schollhorn WI, Rizzi N, Slapsinskaite-Dackeviciene A, Leite N. Always Pay Attention to Which Model of Motor Learning You Are Using. Int J Environ Res Public Health. 2022;19(2):Article 711. doi: https://doi.org/10.3390/ijerph19020711.

111. Alkhathami K, Alshehre Y, Wang-Price S, Brizzolara K. Reliability and Validity of the Functional Movement Screen™ with a Modified Scoring System for Young Adults with Low Back Pain. Int J Sports Phys Ther. 2021;16(3):620-7. doi: https://doi.org/10.26603/001c.23427.

112. Teyhen DS, Shaffer SW, Lorenson CL, Halfpap JP, Donofry DF, Walker MJ, et al. The Functional Movement Screen: A Reliability Study. J Orthop Sports Phys Ther. 2012;42(6):530-40. doi: https://doi.org/10.2519/jospt.2012.3838.

113. Sanchez-Lastra MA, Moldes JM, Diz JC, Martínez-Lemos RI, Ayán C. Feasibility and reliability of the Functional Movement Screen battery in adults with intellectual disability. Journal of Intellectual Disability Research. 2022;66(6):568-75. doi: https://doi.org/10.1111/jir.12916.

114. Bonaventura JM, Sharpe K, Knight E, Fuller KL, Tanner RK, Gore CJ. Reliability and accuracy of six hand-held blood lactate analysers. J Sports Sci Med. 2015;14(1):203-14.

115. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16(3):297-334. doi: https://doi.org/10.1007/BF02310555.

116. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ. 2011;2:53-5. doi: https://doi.org/10.5116/ijme.4dfb.8dfd.

117. Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Research in Science Education. 2018;48(6):1273-96. doi: https://doi.org/10.1007/s11165-016-9602-2.

118. Jafari M, Ansari-Pour N. Why, When and How to Adjust Your P Values? Cell J. 2019;20(4):604-7. doi: https://doi.org/10.22074/cellj.2019.5992.

119. Leday GGR, Hemerik J, Engel J, van der Voet H. Improved family-wise error rate control in multiple equivalence testing. Food Chem Toxicol. 2023;178:Article 113928. doi: https://doi.org/10.1016/j.fct.2023.113928.

120. Büttner F, Toomey E, McClean S, Roe M, Delahunt E. Are questionable research practices facilitating new discoveries in sport and exercise medicine? The proportion of supported hypotheses is implausibly high. Br J Sports Med. 2020;54(22):1365-71. doi: https://doi.org/10.1136/bjsports-2019-101863.

121. Afonso J, Ramirez-Campillo R, Clemente FM, Buttner FC, Andrade R. The Perils of Misinterpreting and Misusing "Publication Bias" in Meta-analyses: An Education Review on Funnel Plot-Based Methods. Sports Med. 2024;54(2):257-69. doi: https://doi.org/10.1007/s40279-023-01927-9.

122. Borg DN, Barnett AG, Caldwell AR, White NM, Stewart IB. The bias for statistical significance in sport and exercise medicine. Journal of Science and Medicine in Sport. 2023;26(3):164-8. doi: https://doi.org/10.1016/j.jsams.2023.03.002.

123. Bishop C, Read P, Bromley T, Brazier J, Jarvis P, Chavda S, Turner A. The Association Between Interlimb Asymmetry and Athletic Performance Tasks: A Season-Long Study in Elite Academy Soccer Players. J Strength Cond Res. 2022;36(3):787-95. doi: https://doi.org/10.1519/jsc.0000000000003526.

124. Zhang K, Jan YK, Liu Y, Zhao T, Zhang L, Liu R, et al. Exercise Intensity and Brain Plasticity: What's the Difference of Brain Structural and Functional Plasticity Characteristics Between Elite Aerobic and Anaerobic Athletes? Front Hum Neurosci. 2022;16:Article 757522. doi: https://doi.org/10.3389/fnhum.2022.757522.

125. Wallace A, Briggs MS, Onate J, DeWitt J, Rinehart-Thompson L. Perceived Management of Acute Sports Injuries and Medical Conditions by Athletic Trainers and Physical Therapists. Int J Sports Phys Ther. 2021;16(6):1548-65. doi: https://doi.org/10.26603/001c.29850.

126. Terry DP, Mewborn CM, Miller LS. Repeated Sport-Related Concussion Shows Only Minimal White Matter Differences Many Years After Playing High School Football. J Int Neuropsychol Soc. 2019;25(9):950-60. doi: https://doi.org/10.1017/s1355617719000754.

127. Ioannidis JPA. Why Most Published Research Findings Are False. PLoS Medicine. 2005;2(8):Article e124. doi: https://doi.org/10.1371/journal.pmed.0020124.

128. Moonesinghe R, Khoury MJ, Janssens ACJW. Most Published Research Findings Are False—But a Little Replication Goes a Long Way. PLoS Medicine. 2007;4(2):Article e28. doi: https://doi.org/10.1371/journal.pmed.0040028.

129. Hillary FG, Medaglia JD. What the replication crisis means for intervention science. International Journal of Psychophysiology. 2020;154:3-5. doi: https://doi.org/10.1016/j.ijpsycho.2019.05.006.

130. Mesquida C, Murphy J, Lakens D, Warne J. Replication concerns in sports and exercise science: a narrative review of selected methodological issues in the field. R Soc Open Sci. 2022;9(12):Article 220946. doi: https://doi.org/10.1098/rsos.220946.

Supplementary table 1. Checklist of features to consider predictors in regression modeling.

Mathematical features of the regression model	**Yes/No**
Does the specific model respect all its applicable assumptions (e.g., normality, homoscedasticity)?	
Does the model quantify the uncertainty associated with its estimates (e.g., confidence intervals, prediction intervals)?	
Additional statistical and conceptual features	**Yes/No**
Is the sample sufficiently powered to detect effects in case they exist?	
Is the sample representative of the target population (e.g., distribution, characteristics)?	
Is the context under which the sample was studied representative of, or translatable to, the target context?	
Is that context representative of, or translatable to, broader contexts?	
Are the relationships between dependent and independent variables expected to exhibit consistency over time?	
Are these variables supported by a reasonable rationale?	
Are potential confounders accounted for?	
Are the measurements reliable and is the magnitude of that reliability being considered to assess the potential error of the models?	
If multiple testing is used, is family-wise error rate correction implemented?	
Extrapolation/external validation	
Does the model perform well with comparable/similar samples?	

Does the model perform well in comparable/similar contexts?	
Does the model perform well with progressively more distinct samples (e.g., if the model was built upon data from regional level athletes, does it generalize well to international level athletes)?	
Does the model perform well with progressively more distinct contexts (e.g., are predictive models of performance for the regular season matches applicable to performance in the play-offs)?	