Methods matter and the 'too much, too soon' theory (part 2): what is the goal of your sports injury research? Are you describing, predicting or drawing a causal inference?

Rasmus Oestergaard Nielsen [1,2] Nina Sjoerup Simonsen,[3] Marti Casals [4,5] Emmanuel Stamatakis [6] Mohammad Ali Mansournia [7]

INTRODUCTION

Imagine a sports injury researcher who gets a bright idea in relation to the 'too much, too soon' theory (too much training, too soon, leads to injury).[1] Then, the researcher proposes the following aim: to examine the association between workload and sports injury. Think twice: Does the word 'association' truly assist you, the reader, in understanding the underlying goal?

Every researcher should be clear on the goal of their research. The goal should be immediately clear to the intended readers. Failure to be clear about goals may lead to inappropriate, misleading and flawed conclusions.[2] Hernán, an expert epidemiologist, declared in 2018 that '…being explicit about the goal of the analysis is a prerequisite for good science',[3] since being explicit about the objective of a study reduces ambiguity in the scientific question, errors in the data analysis, and excesses in the interpretation of the results. Possibly, using the term 'association' (above) is ambiguous as we are unaware of the underlying goal.

Although definitions may vary depending on the source,[4 5] research into sports injury usually has one of three goals: To describe, to predict or to explain. To describe, for example, how athletes train and/or how many athletes sustains sports injury over time (eg, injury risk), without investigating their relationship. In contrast to describing, researchers can also examine relationships, or so-called associations, which can have two vastly different goals: to predict groups at risk of sustaining sport injury (here referred to as predicting) or to explain why some athletes sustain workload-related injury (here referred to as causal inference).

In the first educational piece in this 'too much, too soon' series,[6] we provided readers with examples of causal research questions, described analytical concepts and outlined the main differences between population-based prevention and personalised prevention. This second BJSM 'Methods matter' education review[7] aims to incorporate the three research goals into the 'too much, too soon' theory, with three target groups in mind. First, we aim to help researchers crystallise their thinking about the goal of the research and to interpret their results accordingly. Second, we intend to help editors and peer reviewers evaluate the value of manuscripts. Third, we seek to aid readers make better use of the presented findings. Table 1 shows the main characteristics of these goals.

DESCRIBING: *HOW* DO ATHLETES TRAIN?

In workload research, describing can be the task of using data to provide a quantitative summary of how athletes perform their training. Here, the researchers consider a single workload-related variable (eg, distance run, balls bowled or minutes played) at a time. This variable can be measured repeatedly over time, without using sports injury as an outcome in the analysis. For instance, to describe the workload of bowlers in elite cricket, Dennis *et al*[8] collected data about the frequency of bowling, the type of bowling performed (match or training), and the time frame within which the bowling was completed. What exactly constitutes 'workload', and which workload-related variables are relevant to describe in each type of sport remains open to debate.

From an injury perspective, researchers may investigate the changes in injury incidence over time (eg, using Poisson regression). This is an example of an analysis with multiple variables in the model. Still, the goal is to describe as no association is investigated (table 1).

PREDICTING: *WHO* SUSTAINS WORKLOAD-RELATED SPORT INJURY?

Predicting is fundamentally different from describing. Here, the researchers aim to predict who is at risk of sustaining workload-related sport injury in the future; they aim to evaluate whether such risk is greater among some athletes than among others. A sport injury study found that daily smoking was associated with lower odds of sustaining a lower-extremity sports injury compared with those who did not smoke (OR 0.23, 95% CI 0.05 to 1.00, p=0.05).[9] Those who smoked daily could be identified as having less risk of a lower-extremity sports injury. The authors acknowledged that a causal association between smoking and lower extremity injury was not plausible. In other words, the authors were unable to explain why smokers were at decreased risk of injury, but they were able to identify who was less likely to sustain lower extremity injury in terms of smoking habits. Some have advocated that injury prevention should be guided by injury aetiology,[10] but it is extremely unlikely that an intervention encouraging athletes to smoke daily would reduce the risk of a low-extremity sports-related injury.

Prediction-related research should not guide interventions, not even when training load is included in the analysis. In a basketball study,[11] athletes with three or fewer decelerations per game (IRR 4.36, 95% CI 1.78 to 10.6) and those running 1.3 or less miles per game (lower workload) (IRR 6.42, 95% CI 2.52 to 16.3) had a higher risk of sustaining injury during the games (p<0.01 in both cases). Based on this, the authors seem to make a causal conclusion: 'Increasing external workload may likely reduce the risk of injury

[1]Department of Public Health, Section for Sports Science, Aarhus University, Aarhus, Denmark
[2]Research Unit for General Practice, Aarhus, Denmark
[3]Department of Public Health, Aarhus University, Aarhus, Denmark
[4]Sport and Physical Activity Studies Centre (CEEAF), Faculty of Medicine, University of Vic-Central University of Catalonia (UVic-UCC), Barcelona, Spain
[5]Medical Department, Futbol Club Barcelona, Barça Innovation Hub, Barcelona, Spain
[6]School of Public Health, University of Sydney, Sydney, New South Wales, Australia
[7]Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran, Islamic Republic of

Correspondence to Dr Rasmus Oestergaard Nielsen, Department of Public Health, Section for Sports Science, Aarhus University, 8000 Aarhus, Denmark; roen@ph.au.dkProfessor Mohammad Ali Mansournia; mansournia_ma@yahoo.com

Table 1 Three research goals and their main characteristics within workload-related sports injury research

	Three goals of research		
	To describe	**To predict**	**To draw causal inference**
Research objective can be	To describe **how** athletes perform their training or **how** many athletes that sustain injury over a given time span	To predict **who** is at risk of injury (by comparing athlete workloads and characteristics)	To understand **why** injury occurs (by comparing athlete workloads and characteristics)
Researchers aim to	Describe, not to investigate associations	Investigate an association which is non-causal	Investigate a causal association
Researchers can use	At least one variable in the analysis*	At least two variables in the analysis	At least two variables in the analysis
Role of additional variables	No additional variables are included in the analysis	Subgroup stratification (optional)	Confounders, mediators and/or effect-measure modifiers
Should the analysis be based on frameworks for sports injury occurrence?	No	No	Yes
Some examples of statistical techniques	Regression Cluster analysis …	Regression Classification …	Regression IPW IV-estimation G-estimation G-formula Standardisation …
Type of study	Exploratory or confirmatory	Exploratory or confirmatory	Exploratory or confirmatory
Goal allows researchers	To describe a workload-related or an injury-related variable	To predict which athletes are most vulnerable to injury. These athletes may be target population for interventions.	To identify potential targets for sports injury prevention. Here, authors assess the role of causal factors which if targeted for interventions, are likely to reduce sports injury risk.

Classification includes, but is not limited to, decision trees, random forests, neural networks.
*If the aim is to describe injury development over time, two variables are needed (injury and time). Please note that the term 'how', which is used in relation to describing in the table, has previously been used in the context of causal inference in relation to the mechanisms underpinning sports injury occurrence in other educational pieces.[20]
IPW, inverse probability weighting; IV estimation, Instrumental variable estimation.

in professional basketball. More studies are needed to confirm these findings so that adequate prevention programs can be implemented to decrease the number of injuries in professional basketball and other sports'. Such conclusion is counter-intuitive as exposing basketball players to more training load in a game is unlikely to reduce the risk of injury. Therefore, we question whether those players with three or less decelerations per game and those running 1.3 or less miles per game have increased risk of injury because they had a lower dose of training. These basketball players might also be those with reduced load capacity or other characteristics that made them more vulnerable to injury. If this is true, and they follow the advice of adding more training load during a game, they may face an even higher injury risk. Hence, the 'preventive' programme would fail because it would introduce, not prevent, sports injury.

INVESTIGATING CAUSAL INFERENCE: *WHY* DO ATHLETES SUSTAIN WORKLOAD-RELATED INJURY?

In the basketball study above, the authors considered confounders when analysing their data. As seen in table 1, considering variables as confounders changes the goal of the research into investigating causal inference. Importantly, a variety of exposures must be included in the analysis if the goal is to investigate causal inference in observational studies. Consequently, sports injury researchers must be cautious about their assumptions as to how these variables are interrelated, which can be supported by using causal frameworks and/or diagrams.[12 13]

Hernán and Robins have published a freely available book that helps readers understand how to visualise their causal assumptions.[14] To closer examine why-related questions, the data analyses should be aligned with these causal frameworks. In our field, sports injury researchers are unlikely to shed light on why-questions if they do not account for exposure to sport participation, training load and/or workload (however defined) because these are major components across all causal sports injury frameworks.[15–17] This has been referred to as the 'too much, too soon' theory. Unfortunately, we believe there are, as yet, no published sports injury papers, within the causal inference domain, that demonstrates a clear alignment between the goal of the research, formulation of research question, statistical analyses, interpretations of results and conclusions (in the 'too much, too soon' research area).

In causal theory, it is widely accepted that every causal factor can be viewed as a predictor, although sometimes a weak one, but not every predictor is a cause.[18 19] In the present education review, we used smoking as an example to illustrate a predictor that is not a causal factor. However, in sports science, many causal factors include internal factors (weight, kinematics, kinetics) and external factors (change in training load). Importantly, a crude association between a causal factor and a sports injury does not mean that a why-question is explored (eg, the basketball example). To move closer to a causal association, we must include (change in) training load in the analysis and consider the capacity of the body to withstand load.[17] In addition, exploring causal inference require the researcher to carefully consider the applied analytical strategy, as described in table 1.

CONCLUSION

A clearly stated research goal (to describe, to predict or to investigate causal inference) will help researchers crystallise the rationale behind their work. This requires a deep understanding of the methods differences outlined in table 1 that are prompted by the chosen type of the research goal. In addition, it requires researchers to acknowledge that the commonly used word 'association' becomes ambiguous when used without proper explanation. A deeper understanding of methods can help researchers to design more robust studies with low risk of bias, to perform better statistical analyses, and to interpret their results in a more useful manner. A clear research goal (describe, predict, investigate cause) is an essential first step in the process of writing a good research publication

without flawed, ambiguous or even outright misleading conclusions.

Key points

- The goal of sports injury research must be clear; do the researchers aim to *describe* how athletes train, to *predict* who is likely to sustain injury, or to investigate *causal inference*?
- If the goal is to describe, the data may summarise how athletes perform their training, and the analysis requires only one workload variable at a time (measured repeatedly over time).
- If the goal is to identify/predict who sustains injury, target populations for interventions can be identified (in which population(s) is action needed?). Still, no conclusions regarding targets of intervention can be drawn as the injury prevention is based on knowledge of injury aetiology and injury mechanisms.[10]
- If the goal is to investigate why athletes/players sustain injury, researchers aim to explain by exploring causal inference. Research into sports injury aetiology should be guided by causal frameworks or directed acyclic graphs to inform population-based or personalised sports injury prevention.[6 12 13 21]
- The commonly used word 'association' becomes ambiguous when used without explanation as it can refer to both prediction and causal inference.
- If confounders are included in the analysis, the goal of the research automatically becomes to draw a causal inference. In this case, the sports injury researchers must disclose their causal assumptions in a framework or a diagram.

Twitter Rasmus Oestergaard Nielsen @RUNSAFE_Rasmus, Marti Casals @CasalsTMarti and Emmanuel Stamatakis @M_Stamatakis

Contributors All authors contributed equally to drafting the content of the educational editorial. RON drafted the educational editorial. All authors revised the article for important intellectual content and approved the submission to British Journal of Sports Medicine.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

To cite Nielsen RO, Simonsen NS, Casals M, *et al*. *Br J Sports Med* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/bjsports-2020-102144

Accepted 14 July 2020

Br J Sports Med 2020;**0**:1–3. doi:10.1136/bjsports-2020-102144

ORCID iDs
Rasmus Oestergaard Nielsen http://orcid.org/0000-0001-5757-1806
Marti Casals http://orcid.org/0000-0002-1775-8331
Emmanuel Stamatakis http://orcid.org/0000-0001-7323-3225
Mohammad Ali Mansournia http://orcid.org/0000-0003-3343-2718

REFERENCES

1 Vandenbroucke JP, Pearce N. From ideas to studies: how to get ideas and sharpen them into research questions. *Clin Epidemiol* 2018;10:253–64.
2 Smoliga JM, Zavorsky GS. Team Logo predicts concussion risk: lessons in protecting a vulnerable sports community from Misconceived, but highly Publicized epidemiologic research. *Epidemiology* 2017;28:753–7.
3 Hernán MA. The C-Word: scientific Euphemisms do not improve causal inference from observational data. *Am J Public Health* 2018;108:616–9.
4 Pearl J, MacKenzie D. *The book of why: the new science into cause and effect*. 1st ed. New York: Basic Books, Hachette Book Group, 2018.
5 Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. *Chance* 2019;32:42–9.
6 Nielsen RO, Bertelsen ML, Møller M, *et al*. Methods matter: exploring the 'too much, too soon' theory, part 1: causal questions in sports injury research. *Br J Sports Med* 2020. doi:10.1136/bjsports-2018-100245. [Epub ahead of print: 05 Mar 2020].
7 Verhagen E, Stovitz SD, Mansournia MA, *et al*. *BJSM* educational editorials: methods matter. *Br J Sports Med* 2018;52:1159–60.
8 Dennis R, Farhart P, Goumas C, *et al*. Bowling workload and the risk of injury in elite cricket fast bowlers. *J Sci Med Sport* 2003;6:359–67.
9 Van Middelkoop M, Kolkman J, Van Ochten J, *et al*. Risk factors for lower extremity injuries among male marathon runners. *Scand J Med Sci Sports* 2008;18:691–7.
10 Finch C. A new framework for research leading to sports injury prevention. *J Sci Med Sport* 2006;9:3–9.
11 Caparrós T, Casals M, Solana Álvaro, Solana A, *et al*. Low external workloads are related to higher injury risk in professional male Basketball games. *J Sports Sci Med* 2018;17:289–97.
12 Stovitz SD, Verhagen E, Shrier I. Distinguishing between causal and non-causal associations: implications for sports medicine clinicians. *Br J Sports Med* 2019;53:398–9.
13 Shrier I. Understanding causal inference: the future direction in sports injury prevention. *Clin J Sport Med* 2007;17:220–4.
14 Hernán M, Robins J. *Causal inference: what if*. Boca Raton: Chapman & Hall/CRC, 2020.
15 Meeuwisse WH, Tyreman H, Hagel B, *et al*. A dynamic model of etiology in sport injury: the recursive nature of risk and causation. *Clin J Sport Med* 2007;17:215–9.
16 Windt J, Gabbett TJ. How do training and competition workloads relate to injury? the workload-injury aetiology model. *Br J Sports Med* 2017;51:428–35.
17 Nielsen RO, Bertelsen ML, Møller M, *et al*. Training load and structure-specific load: applications for sport injury causality and data analyses. *Br J Sports Med* 2018;52:1016–7.
18 Moons KGM, Royston P, Vergouwe Y, *et al*. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;338:b375.
19 Shmueli G. To explain or to predict? *Statist. Sci.* 2010;25:289–310.
20 Bahr R, Krosshaug T. Understanding injury mechanisms: a key component of preventing injuries in sport. *Br J Sports Med* 2005;39:324–9.
21 Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol* 2008;8:70-2288-8-70.