Received: 20 April 2022 | Revised: 27 June 2022 | Accepted: 21 July 2022

DOI: 10.1111/jep.13748

ORIGINAL PAPER

Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

WILEY

Mechanistic evidence and exercise interventions: Causal claims, extrapolation, and implementation

William Levack-Payne PhD

Department of Philosophy, the University of Kent, Canterbury, UK

Correspondence
William Levack-Payne, PhD, Department of Philosophy, the University of Kent, Canterbury, UK.
Email: william.levackpayne@gmail.com

Funding information
The University of Kent

Abstract

Rationale: Exercise interventions and policies are widely prescribed in both sport and healthcare. Research investigating exercise interventions and policies is generally conducted using an Evidence-Based framework, placing an emphasis on evidence gathered from randomised controlled trials (RCTs).

Aims and objectives: To explore the idea that, in addition to the assessment of evidence from RCTs when investigating exercise interventions, mechanistic studies ought to also be assessed and considered.

Methods: This article assesses the rationale supporting the use of RCTs as evidence for exercise interventions, and the use of evidence of mechanisms in establishing efficacy, determining external validity, and tailoring interventions.

Results and conclusions: The article argues that evidence from mechanistic studies ought to be considered alongside evidence from RCTs because: as RCTs investigating exercise interventions tend to be of low quality, mechanistic studies ought to be used to reinforce the evidence base; further, evidence from mechanistic studies is highly useful for both questions of extrapolation and implementation. This article argues for this on theoretical grounds, and also draws on a number of case studies.

KEYWORDS
evidence-based medicine, exercise, mechanisms, mechanistic reasoning, sports

1 | INTRODUCTION

The aim of this paper is to argue for the importance of assessing evidence from mechanistic studies in the evaluation and development of exercise related interventions and policies. Exercise interventions, and policies promoting them, are researched in both Evidence-Based Medicine (EBM), and the Evidence-Based movement in Sport and Exercise Science (SES). The goals of EBM and SES are to improve relevant outcomes, be they physical and mental health related, or sport related, by providing a framework for the rigorous analysis of evidence, aiming to use the best available evidence to make decisions about intervention and policy design and implementation. The stance typically given in EBM literature, and literature and practice in SES, emphasizes the importance of evidence from clinical studies like randomised controlled trials (RCTs). This often means privileging it over other sources of evidence.[1] It also, often, dismisses the importance of evidence from mechanistic studies. This can clearly be seen in evidence quality hierarchies proposed for SES, such as that given by Knudson et al.,[2] which claims that evidence from mechanistic studies can be considered to be 'hypothesised evidence'

where RCTs, and reviews of them, produce the best evidence. As is shown by Clarke et al.,[3] the GRADE evidence rating scheme utilised by NICE also de-emphasizes the importance of mechanistic studies.

The value of assessing mechanistic studies argued for in this paper is supported by two key premises. The first is that, due to the low quality of evidence that may be produced by association studies like RCTs in much exercise related research, evidence from mechanistic studies is often necessary to establish causal claims, which is, then, necessary to make Evidence-Based decisions. The second is that it is fruitful to understand mechanisms in exercise related research, rather than simply observe correlations, as it aids the extrapolation of claims, and the implementation and delivery of interventions. This notion of assessing mechanistic evidence alongside evidence from association studies is gaining traction in a number of Evidence-Based fields. For instance, Aronson et al.[4] argue for a similar stance to the one expressed in this article, but for medicine and drug approval, rather than exercise interventions. They argue that mechanistic evidence ought to be used in medicine and drug approval, in addition to the evidence from clinical studies typically relied upon. The goal of their argument is that, by clarifying the role of mechanistic evidence, and arguing for its explicit evaluation in addition to evidence from clinical studies, this can lead to improved health outcomes over a reliance on evidence from clinical studies alone, furthering the goals of the programme of EBM. In recent years, a similar stance on the use of mechanistic evidence has been adopted for cancer research by IARC,[5] which employs mechanistic evidence to help grade the likelihood that exposures or chemicals are carcinogenic. Research is also emerging advocating for the use of evidence of mechanism in the social sciences to develop Evidence-Based policies.[6] The goal of this paper is aligned with the goals of Aronson et al.[4]: by clarifying the role of mechanistic evidence as it relates to exercise related interventions, and arguing for its explicit evaluation in addition to evidence from association studies, sport and health related outcomes can be improved, furthering the goals of sports medicine, and SES.

Before continuing, some clarificatory work must be done. First, this paper is concerned with how evidence is used in evaluating exercise interventions, and interventions with an exercise component, largely from the viewpoint of exercise science. Exercise science is characterised by the use of principles from SES with a view to improving physical and mental health.[7] Given this, despite discussing some sports related interventions, the primary concern of this paper is how healthcare outcomes can be improved by using mechanistic evidence when assessing intervention effect claims, and investigating intervention implementation and delivery.

Second, this article uses the *minimal mechanism* account given by Glennan and Illari[8] as a sufficient working definition of mechanisms:

> "A mechanism for a phenomenon consists of entities (or parts) whose activities and interactions are organised so as to be responsible for the phenomenon." (p.2)

This definition of mechanisms is useful as it accounts for the varied composition of mechanisms that may arise in exercise research. The entities, activities, and interactions involved in a mechanism may include, for instance: social, psychological, physical, and biological elements.[9] Commonly, particularly when considering the whole system relevant to an exercise intervention or policy, the relevant mechanisms will involve a complex interplay of factors of all these types. For instance, as will be seen later in this paper, the mechanism linking the prescription of an exercise-based injury prevention programme to reduced injury rates will have social components, such how team dynamics relate to intervention enthusiasm and adherence; and biological components such as whether the exercises used in the intervention are sufficient to increase the tolerance of the body to forces exerted on it.

It is also important to highlight what is meant by *mechanistic study*, and to explain what is meant by *understanding a mechanism*. First, a mechanistic study is a study that provides evidence for how some cause gives rise to some effect.[9] This can include: autopsy, medical imaging, in vitro experiments, and established theory. It also includes association studies which provide evidence of some mediating variable X between proposed cause A, and proposed outcome B. For instance, association studies finding associations between A and X, and X and B, provide evidence that there is a mechanism between A and B, by identifying a mediating variable, and are thus mechanistic studies for the claim that A causes B. Second, in this paper, I discuss understanding mechanisms, details of mechanisms, and causal factors, seemingly interchangeably. What is important to the argument of this paper is that, in terms of furthering the goals of exercise research, we ought to have some understanding of what goes on between a proposed cause and effect, be that a deep knowledge of the mechanism, or simply the ability to point out details or causal factors. This is because, as I will argue, it is far more fruitful to be able to explain mechanisms on some level than it is to simply observe that causal relationships exist, leaving the explanation as a *black box* where causes go in, and outcomes come out.

2 | ESTABLISHING EFFICACY

The approach of EBM and SES is to emphasise the importance and reliability of evidence derived from association studies, like RCTs. These are studies which provide evidence that a putative cause, A, is correlated with some outcome, B. The quality of evidence provided by association studies is conditional on the quality of those studies. The lower the quality of an association study, the more likely it is that any correlation observed between A and B is not causal, and is in fact the result of confounding, chance, bias, or some other noncausal relationship. However, if, as well as establishing a correlation, we can establish the existence of a mechanism complex that can account for a correlation, we are justified in making a causal claim. The need to establish the existence of a mechanism and correlation to establish a causal claim is called the Russo-Williamson Thesis.[10] The Russo-Williamson Thesis has received some criticisms, however, as strong of defence of the Thesis,[10] and a thorough treatment of these criticisms,[11] have recently been published, I will proceed without

addressing them here. In general, RCTs are viewed as providing high quality evidence because, through thorough controlling, it is unlikely that bias and confounding can explain observed outcomes. Where this is the case, this allows us to indirectly rule in the existence of a mechanism responsible for an observed correlation, justifying a causal claim. This leads to the privileging of evidence from RCTs and, as is easily apparent through experience of the practice of some sport and exercise practitioners, the use of evidence from RCTs to justify the use of some interventions, without the assessment of the quality of that RCT, or regardless of its quality.

Key features an RCT requires in order that it can provide high quality evidence and effectively rule out bias, confounding, and chance, are: large sample sizes, adequate placebo controlling, and adequate blinding of patients and those involved in conducting trials. Due to the nature of research relating to many exercise interventions, it is often difficult, and sometimes impossible, to adequately fulfil the requirements of all these features. This means that the evidence produced by RCTs for exercise interventions, is often insufficient to justify claims about intervention or policy efficacy. I will discuss each of these features, and why it is necessary to provide high quality evidence in turn.

Large sample sizes are necessary both to rule out chance as an explanation for observed correlations, and to allow us to observe small-to-medium effects. Just as we may need to flip a fair coin many times to balance out the impact of chance on the number of heads versus tails, we also need a large sample size in an association study to rule out the impact of chance on observed outcomes. Similarly, if an intervention has a small-to-medium effect size, a sufficiently large sample size is needed to observe a statistically significant difference in outcomes between trial groups to make inferences about intervention effects.

The difficulty of obtaining large sample sizes for research on sport and exercise interventions is well known.[12–14] In fact, in a systematic review of intervention studies in SES that considered all 'articles providing information on the recruitment of adults into interventions involving sport and reporting physical activity or participation outcomes', Cooke and Jones found that, of those studies examined, only half reached their recruitment goals.[15] Further, an analysis of four leading sports science journals found that, of all studies published between 2009 and 2013, in a large proportion of these experimental studies, sample sizes were too small to detect 'small-to-medium effects'.[16] In addition to this, whilst relating to sports medicine rather than exercise specifically, the reality of the problem whereby chance influences real world research, and its impact, is demonstrated by the fact that, in a review of sports surgery trials published between 2005 and 2015, it was possible to reverse the statistical significance of most studies by changing the outcome of only a few patients.[17]

Moving on from sample size, placebo controls are needed to correctly estimate the effect size of interventions under investigation, and to rule out psychological effects on observed outcomes. They are used, in part, to allow us to isolate the effects of an intervention under investigation on an outcome of interest. Where a placebo is too similar to an intervention under investigation, we may underestimate its effects. Conversely, where it is too dissimilar, we may overestimate its effects. The Grünbaum criteria for a placebo which, whilst not uncontested, remains operationally the most useful, is that a placebo must[18]:

- Have no features of the treatment it is being compared to that may cause recovery, or must have none of the features proposed to cause recovery that are under investigation. These are called the characteristic features.
- Have every feature that is in the true treatment being tested, but that would not cause a recovery, or that is not under investigation. These are called the incidental features.
- Have no more features.

Accordingly, a placebo should be indistinguishable from an intervention being tested, except that it should not include the active part of the intervention under investigation. Adequate placebos are relatively easy to develop for many medicines. For instance, one may simply need a similar appearing and tasting pill without the active ingredient under investigation. It is even possible, in some cases, to improve placebos by ensuring that they produce similar side effects to the real intervention. When we are concerned with exercise and sports interventions, however, placebo controlling is very difficult. Jeremy Howick lists several things an exercise placebo would need to control for without introducing any of the outcome-influencing features of the intervention under investigation. This includes[19]: the belief that one is being treated with exercise; the psychological benefits of exercise; participant/investigator interaction; and increased metabolic rate, heart rate and temperature. Imagine trying to develop a placebo control for an intervention investigating the efficacy of bike riding as a treatment for high blood pressure. A participant would need to undergo all the incidental features of bike riding, without the characteristic ones. For instance, a participant would need to get sweaty, out of breath, hungry, tired, and feel like they were riding a bike, all without improving their fitness in a way that could affect blood pressure to properly isolate the effects of bike riding. This example illustrates that, in many instances, it may be incredibly difficult, or almost impossible, to develop adequate placebos to test sports medicine and exercise interventions. There is also increasing concern that things we deploy as placebos when testing sports medicine and exercise interventions may, in some instances, have their effect by some mechanism other than a placebo mechanism.[20] Where we cannot develop and deploy adequate placebos, evidence produced by trials will be of low quality because we cannot be sure that we have sufficiently isolated the effects of the intervention under investigation. Potential solutions to this problem, such as conducting active controlled trials and dose response trials also have problems associated with them, including those posed by assay sensitivity and the difficulty of finding sample sizes sufficient to conduct dose response trials.

Finally, the importance of blinding a trial must be discussed. Ensuring that neither participants, nor those conducting a trial, know who is receiving a treatment and who is receiving a placebo is also an important part of running high quality trials. It is important so that

intuitions and hunches about treatment and placebo allocation do not have an impact on outcomes, or interpretation of outcomes.[21] If blinding is inadequate, the resulting bias and psychological confounding can influence observed outcomes, providing an explanation other than that the intervention under investigation is effective, meaning that the quality of evidence the trial produces will be of low quality. Just as was seen in the case for placebo controlling exercise related trials, blinding adequately will also be difficult, or impossible in some cases. This is due to the difficulty in creating adequate placebos, but also the involvement of those running trials in interventions. For instance, a physiotherapist will know if they are providing sham manipulations and treatments to patients, and an experienced coach will know that they are providing someone with a placebo exercise intervention. This compares unfavourably, for instance, with many trials on medicines, which may be blinded for both participant and experimenter relatively easily with a sugar pill.

As has been seen, a small sample size, inadequate placebo controlling, and inadequate blinding, can all lower the quality of evidence a trial can produce. This is because, where these trial features are inadequate, any correlation observed between putative cause and effect is likely to have other possible explanations, and thus cannot be ruled in as a causal relationship. Further, the less well these features are implemented, the more likely it is that an observed correlation is not causal, and has some explanation other than that the intervention under investigation has caused it. In addition to this, in the case of many exercise related interventions, it will be difficult or impossible to conduct a trial which does all of these things well, due to the nature of the type of research being conducted. As such, the privileging of evidence from RCTs and other types of association study is likely to mean the adoption or use of interventions without well justified efficacy.

So, the problem arises because, although we identify a correlation, we are unable to rule it in as causal because the quality of the trial is insufficient to rule out non-causal explanations. If, however, there is a way to rule in causal explanations, and rule out non-causal ones, we will, then, be justified in making a causal claim about the efficacy of an intervention. It is here that evidence from mechanistic studies becomes useful. Where a trial, or set of trials, is insufficient to rule out non-causal explanations for observed outcomes, if we can establish that a mechanism exists which can account for the observed outcome, we can then rule in a causal claim. Mechanistic studies allow us to do this by providing evidence for details of the mechanism between putative cause and effect. What this means is that, where the quality of evidence of efficacy is low, because it is difficult or impossible to conduct high quality association studies, we ought to assess evidence from mechanistic studies, in addition to association studies, to have a higher quality evidence base with which to justify the use of an intervention. Accordingly, when investigating exercise related interventions, where it is likely that the quality of evidence produced by any trial will be low, we ought to assess evidence from trials and mechanistic studies together in most cases.

3 | EXTERNAL VALIDITY

When we have some established understanding of the mechanism by which an intervention works, this can aid greatly in problems of external validity and extrapolating knowledge about intervention effects between populations. This is often called *mechanisms-based extrapolation*. This is particularly important when investigating exercise interventions as, given the difficulties associated with conducting them, if it is possible to understand and compare intervention relevant mechanisms between populations, this can reduce the number of high-quality studies that need to be conducted before we can roll out an intervention with justified efficacy. This, in turn, can save both money and time. Notably, understanding how mechanisms differ between populations can also help us to ensure that we do not roll out an intervention in a population where it is unlikely to be effective.

Several authors have commented on how one may perform mechanisms-based extrapolation.[22,23] A compelling account of it is given by Steel,[22] who claims that, by comparing the mechanism by which something has its effect in a test group, and comparing that mechanism to a target group at the last stage where it is most likely to differ, if the mechanisms are sufficiently similar, we may use this as justification for extrapolation. For example, suppose there is a mechanism:

$$A \rightarrow X \rightarrow Y \rightarrow B \rightarrow Z$$

where X, Y, and Z are the points where a target and a model organism are likely to have differences in how the mechanism acts, and A and B are places where they're very similar. In this instance, changes in X and Y will result in changes in Z. As such, if the result of the mechanism is sufficiently similar at stage Z, this provides evidence in favour of justifiably extrapolating intervention effects.

An example of an exercise programme intended to reduce injuries in football provides a useful example of how understanding intervention relevant mechanisms can help us to see where it may and may not be appropriate to employ an exercise intervention. The FIFA 11+ is a football injury prevention exercise programme that is intended to reduce incidences of hamstring and other lower limb injuries by improving hamstring strength and training athletes' neuromuscular skill, proprioceptive ability, and functional balance (their ability to use non-deleterious movement patterns). The intervention employed a range of leg strengthning exercises and running and jumping drills to improve these outcomes[24]. It was originally tested with a large scale RCT in adolescent Scandinavian female football players.[24] Without changing the exercise programme, it was also rolled out to a variety of other populations worldwide, and has been effective in many other populations.[25] However, it has also been ineffective in a number of populations including veteran aged men, and also men of intermediate and above skill levels.[25–28]

Interestingly, in early research into the 11+, it was noted that:

> Our prevention programme is multifaceted and addresses many factors that could be related to the

> risk of injury [...] it is not possible to determine exactly which exercises or factors might have been responsible for the observed effects.[24]

This argument of this section and Section 4 highlight how avoiding this type of *black box* approach to research would have been more fruitful.

The disparity in effectiveness between populations in which the FIFA 11+ was employed highlights how important mechanisms-based extrapolation would have been in this instance to determine what populations it was likely to be effective in. The similarity between different groups of adolescent female football players means that little work would need to be done to suggest that the mechanisms of action are sufficiently similar that the 11+ will be effective between groups. However, the social and physical differences between veteran men and men of intermediate and above skill level, and adolescent women are vast, and mechanisms-based extrapolation could have been employed fruitfully here to not employ an ineffective exercise programme. Consider the physical mechanisms at play, first. Pre-intervention injury rates will be conditional on athletes' pre-intervention strength and movement patterns. Thus, to reduce injury rates, the 11+ has to be able to improve strength and movement patterns. Given that adolescent females will likely be at an earlier stage in their football lifespan than veteran or intermediate and above skilled men, they will likely begin the 11+ weaker, and with less refined movement patterns. The stronger the athlete, or the more refined the movement patterns, the more intensity and difficulty is required to improve those attributes through training. Accordingly, as is suggested by Beijsterveldt et al.,[26] whilst the 11+ was sufficiently challenging to improve movement patterns and strength in adolescent females, it is unlikely to have been sufficient to do so in some older male populations.

Social mechanisms should also be considered when conducting mechanisms-based extrapolation. Hammes et al.[28] suggest that another difference between adolescent females and some male groups that explains differences in effectiveness is adherence to the intervention. Adolescent female groups performed the 11+ two to three times a week,[28] during 77% of their training or match sessions.[24] Some male groups with low intervention effectiveness employed the 11+ comparatively less, around once per week.[28] Clearly, if a group does not engage in an injury prevention programme, they cannot benefit from it. As was done in the case of Hammes et al.[28] mechanistic reasoning examining differences in social structure between effective and ineffective populations could have been done to provide insight that the intervention was unlikely to be effective in older male groups. Hammes et al. suggest that job-related commitments and the perceived importance of the 11+ likely lead to low adherence. Conversely, in young female groups where coaches may be regarded with more respect, it is unlikely that a coach will allow athletes to skip the 11+.

The rollout of the FIFA 11+ is similar to the rollout of the Bangladeshi Integrated Nutrition Programme (BINP) discussed by Cartwright and Hardie, who assess policy extrapolation decisions more widely.[29] With the FIFA 11+, its success in adolescent female populations was used as justification for its effectiveness in other populations, some of which it ended up being ineffective in. In the case of the BINP, the success of a highly effective nutrition policy aimed at reducing rates of childhood malnourishment in Tamil Nadu was used as justification for the development of a similar nutrition policy in Bangladesh. Unfortunately, the policy being effective in one population was insufficient evidence that it would be effective in another and because of differences in the mechanisms by which food is distributed in Tamil Nadu and Bangladesh. Whilst the policy reduced rates of child malnourishment in Tamil Nadu, it had no effect in Bangladesh. Using their case study, Cartwright and Hardie motivate the importance of assessing differences in social factors between populations before extrapolating. Similarly, I use the case of the 11+ to highlight the importance of assessing biological and social mechanisms, and their differences between populations that may affect intervention effectiveness, to highlight populations where effectiveness is likely, and also populations where it is unlikely. It is clear that this example speaks for much exercise related research, demonstrating the value of conducting and assessing mechanistic studies.

4 | TAILORING INTERVENTIONS

Understanding details of mechanisms relevant to an intervention and its proposed pathway is also highly beneficial when it comes to tailoring interventions to different groups. The FIFA 11+, again, provides a useful example of this. As was previously discussed, the 11+ was ineffective in some male populations, and relevant research suggested that this was likely due to it being insufficiently intense and challenging to promote improvements in strength and movement patterns in those male populations. Identifying this detail of the intervention mechanism allows us to suggest adaptations to the 11+ for some male populations to increase its intensity and, therefore, its effectiveness. For instance, an exercise trial that used the same exercises as the 11+ to reduce injury rates, and which was effective in skilled adult men, used a load to stress the hamstrings that greater than that used in the 11+, and the number of exercise repetitions was also greater, thus better promoting an adaptive response in already strong athletes.[30] Similarly, identifying that problems with adherence lead to the 11+ not being effective allows us to suggest realistic changes to the intervention that may improve adherence. Hammes et al.[28] highlight this by suggesting that adapting the 11+ to have at home components may improve adherence by lowering the time burden, and adding ball handling skills may improve adherence by making the intervention more engaging. If the work had been done to identify and address differences between populations before the 11+ was employed in these ineffective male populations, perhaps FIFA would have rolled out an intervention that was effective in more of the populations that took it up. A key benefit of developing an understanding of the details of intervention relevant mechanisms to tailor and adapt interventions is that, in doing so, we may reduce the

number of high-quality trials that need to be conducted to find effective treatments. When conducting trials, it is highly beneficial to have some idea of why a treatment might be effective, this can reduce the cost and time necessary to developing and utilising effective interventions.[31] In the case of the 11+, if we had examined the mechanistic differences between populations, conducting a trial on an intervention where the likelihood of effectiveness was low could have been pre-empted and avoided.

Whilst the example of changes that may be made to the FIFA 11+ to improve its effectiveness in different populations provides us with a theoretical example of the fruitfulness of understanding details of intervention relevant mechanisms, the case of treatment for Relative Energy Deficiency in Sport (RED-S) provides us with a real-world, concrete, example. RED-S arises when an athlete's energy intake is chronically lower than their energy output, generally as a result of sport or exercise. RED-S symptoms include reduced metabolic rate, menstrual function, bone health, immunity, protein synthesis, and cardiovascular health.[32] The general recommendation for REDS treatment is to encourage athletes to eat more and exercise less, thus tipping the energy balance in favour of energy intake. This treatment, however, does not take into account the nexus of mechanisms by which an athlete gets RED-S. Highlighting the importance of taking account of these mechanisms, a recent review[33] discussed how we ought to treat RED-S in light of our understanding of these mechanisms. The review notes that athletes likely get RED-S as a result of high energy output from a commitment to their sport and training, which they are unwilling to change. Further, the review notes that RED-S treatments are associated with weight gain, which may negatively affect athlete performance. Finally, athletes may simply find it too difficult to eat sufficient calories to combat RED-S as the volume may be too high.

Noting these key factors, the review noted that changing athlete mealtimes to reduce the time over which muscle breakdown occurs, increasing carbohydrate intake and reducing fibre intake can increase calorie intake and reduce volume, and encouraging weight training can combat bone deterioration. All these methods reduce or improve RED-S symptoms without needing an athlete to eat more volume, or drastically reduce their training load. Without examining the intervention relevant mechanisms, this much more viable approach to treatment is unlikely to have been developed. These two cases both provide examples of the fruitfulness that can be afforded exercise research by considering mechanisms relevant to interventions and their success.

5 | CONCLUSION

I have presented several areas in which, for exercise related research, it is beneficial to systematically assess evidence from mechanistic studies. This was particularly important in the case of assessing the efficacy of interventions in light of the quality of evidence produced by association studies in exercise related research. The examples given are intended to be illustrative of the importance of assessing mechanistic evidence, rather than exhaustive.

Unfortunately, given the quality of evidence produced by association studies in exercise related research, and the difficulty of generating very high quality mechanistic evidence (particularly when social and psychological mechanisms are considered), whilst advocating the assessment of evidence from both mechanistic studies and association studies together is a step forward in the Evidence-Based programme of SES, it does not mean that we will always have sufficient evidence to justify a given intervention's effectiveness. So, whilst we must guard against the use of an intervention justified on the grounds that it seemed to be effective in an association study, regardless of that study's quality, we must also guard against the wheeling out of low-quality mechanistic evidence to support claims about interventions. Accordingly, we must systematically assess the quality of evidence from mechanistic and association studies together when investigating interventions.

ACKNOWLEDGEMENTS

The research necessary to write this article was possible thanks to a Graduate Teaching Assistant scholarship at the University of Kent and the feedback of two very helpful anonymous reviewers. Funding provided by the University of Kent Graduate Teaching Award.

CONFLICT OF INTEREST

The author declares no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

William Levack-Payne http://orcid.org/0000-0003-2543-7906

REFERENCES

1. Ivarsson I, Andersen MB. What counts as "evidence" in evidence-based practice? Searching for some fire behind all the smoke. *Journal of Sport Psychology in Action*. 2016;7(1):11-22. doi:10.1080/21520704.2015.1123206
2. Knudson D, Elliott B, Hamil JI. Proposing application of results in sport and exercise research reports. *Sports Biomechanics*. 2014;13(3):195-203.
3. Clarke B, Gillies D, Illari P, Russo F, Williamson J. Mechanisms and the evidence hierarchy. *Topoi*. 2014;33(2):339-360. doi:10.1007/s11245-013-9220-9
4. Aronson K, La Caze A, Kelly MP, Parkkinen V-P, Williamson J. The use of mechanistic evidence in drug approval. *J Eval Clin Pract*. 2018;24(5):1166-1176. doi:10.1111/jep.12960.
5. IARC. *Preamble to the IARC Monographs on the Identification of Carcinogenic Hazards to Humans*. 2019.
6. Shan Y, Williamson J. Applying evidential pluralism to the social science. *European Journal for Philosophy of Science*. 2022;11:96. doi:10.1007/s13194-021-00415-z
7. BASES. *About Sport and Exercise Science*. https://www.bases.org.uk/spage-about_us-about_sport___execise_science.html.
8. Glennan S, Illari P. Varieties of mechanisms. *The Routledge handbook of mechanisms and mechanical philosophy*. Routledge; 2017:91-103.

9. Parkkinen VP, Wallmann C, Wilde M, et al. *Evaluating evidence of mechanisms in medicine: principles and procedures*. Springer; 2018. doi:10.1007/978-3-319-94610-8
10. Williamson J. Establishing causal claims in Medicine. *International Studies in the Philosophy of Science*. 2019;32(1):33-61. doi:10.1080/02698595.2019.1630927
11. Levack-Payne W. *The application of Evidence-Based Medicine methodologies in sports science: problems and solutions*. Doctor of Philosophy (PhD) thesis. University of Kent (2022). doi:10.22024/UniKent/01.02.93560
12. Nevill A, Atkinson G, Hughes M. Twenty-five years of sport performance research in the journal of sports sciences. *J Sports Sci*. 2008;26(4):413-426. doi:10.1080/02640410701714589
13. Pyne D, Hopkins W, Martin D. Inadequate sample sizes in studies of athletic performance at the 2012 ACSM annual meeting. *Sportscience*. 2010;14:1-11.
14. Select Committee on Science and Technology. *Sport and Exercise science and medicine: building on the Olympic legacy to improve the nation's health*. Published by the Authority of the House of Lords, 2012. HL Paper 33, 1st Report of Session 2012-13.
15. Cooke R, Jones A. Recruiting adult participants to physical activity intervention studies using sport: a systematic review. *BMJ Open Sport Exerc Med*. 2017;3(1):1-9. doi:10.1136/bmjsem-2017-000231
16. Schweizer G, Furley P. Reproducible research in sport and exercise psychology: the role of sample sizes. *Psychol Sport Exerc*. 2016;23:114-122. doi:10.1016/j.psychsport.2015.11.005
17. Khan M, Evaniew N, Gichuru M, et al. The fragility of statistically significant findings from randomized trials in sports surgery: a systematic survey. *Am J Sports Med*. 2017;45(9):2164-2170. doi:10.1177/0363546516674469
18. Grünbaum A. The placebo concept. *Behavioural Research and Therapy*. 1981;19(2):157-167.
19. Howick J. *The philosophy of evidence-based medicine*. John Wiley & Sons; 2011. doi:10.1002/9781444342673
20. Beedie C, Benedetti F, Barbiani D, Camerone E, Lindheimer J, Roelands B. Incorporating methods and findings from neuroscience to better understand placebo and nocebo effects in sport. *European Journal of Sport Science*. 2020;20(3):313-325. doi:10.1080/17461391.2019.1675765
21. Straus S, Glasziou P. *Evidence-Based Medicine E-Book: how to Practice and Teach EBM*. Elsevier Health Sciences; 2019.
22. Steel D. *Across the Boundaries: Extrapolation in Biology and Social Science*. Oxford University Press; 2007. doi:10.1017/S0266267110000325
23. Parkkinen VP, Williamson J. Extrapolating from Model Organisms in Pharmacology. In: *Uncertainty in Pharmacology: Epistemology, Methods, and Decisions*. Ed. by Adam LaCaze and Barbara Osimani. Springer International Publishing, 2020, pp. 59–78. doi:10.1007/978-3-030-29179-2_3
24. Soligard T, Myklebust G, Steffen K, et al. Comprehensive warm-up programme to prevent injuries in young female footballers: cluster randomised controlled trial. *BMJ: Online First*. 2008;337:1-9. doi:10.1136/bmj.a2469
25. Bizzini M, Dvorak J. FIFA 11+: an effective programme to prevent football injuries in various player groups worldwide—a narrative review. *Br J Sports Med*. 2015;49(9):577-579.doi:10.1136/bjsports-2015-094765
26. van Beijsterveldt C, van de Port IG, Krist MR, et al. Effectiveness of an injury prevention programme for adult male amateur soccer players: a cluster-randomised controlled trial. *Br J Sports Med*. 2012;46(16):1114-1118. doi:10.1136/bjsports-2012-091277
27. Gatterer H, Ruedl G, Faulhaber M, Regele M, Burtscher M. Effects of the performance level and the FIFA "11" injury prevention program on the injury rate in Italian male amateur soccer players. *J Sports Med Phys Fitness*. 2012;52(1):80-84pmid: 22327090.
28. Hammes D, der Fünten KA, Kaiser S, Frisen E, Bizzini M, Meyer T. Injury prevention in male veteran football players—a randomised controlled trial using "FIFA 11+. *J Sports Sci*. 2015;33(9):873-881. doi:10.1080/02640414.2014.975736
29. Cartwright N, Hardie J. *Evidence-Based policy: a practical guide to doing it better*. Oxford University Press; 2012. doi:10.1017/S0266267114000091
30. Mjølsnes R, Arnason A, Østhagen T, Raastad T, Bahr R. A 10-week randomized trial comparing eccentric vs. concentric hamstring strength training in well-trained soccer players. *Scand J Med Sci Sports*. 2004;14(5):311-317. doi:10.1046/j.1600-0838.2003.367.x
31. Anjum R, Mumford S. *Causation in science and the methods of scientific discovery*. Oxford University Press; 2018. doi:10.1093/oso/9780198733669.001.0001
32. Mountjoy M, Sundgot-Borgen J, Burke L, et al. The IOC consensus statement: beyond the female athlete triad—relative energy deficiency in sport (RED-S). *Br J Sports Med*. 2014;48(7):491-497. doi:10.1136/bjsports-2014-093502
33. Kuikman M, Mountjoy M, Stellingwerff T, Burr JF. A review of nonpharmacological strategies in the treatment of relative energy deficiency in sport. *Int J Sport Nutr Exerc Metab*. 2021;31(3):268-275. doi:10.1123/ijsnem.2020-0211

How to cite this article: Levack-Payne W. Mechanistic evidence and exercise interventions: causal claims, extrapolation, and implementation. *J Eval Clin Pract*. 2022;28:745-751. doi:10.1111/jep.13748