Invited Review

Does increasing an athletes' strength improve sports performance? A critical review with suggestions to help answer this, and other, causal questions in sport science

James Steele, James Fisher, Derek Crawford

Objectives: Researchers and practitioners in sports science aim to generate, and apply, knowledge to improve sports performance. One area of interest is the role that muscular strength, and thus approaches to improve this (i.e. resistance training), has upon sports performance. In this review we briefly consider the evidence regarding an answer to the causal question "*Does increasing an athletes' strength improve sports performance?*".

Design & Methods: We first consider the Applied Research Model for the Sport Sciences (ARMSS) to frame the problem and answer this. We then highlight barriers to answering it (and other causal questions) before offering suggestions to address these.

Results: Muscular strength typically differentiates elite and non-elite athletes, and is correlated with proxy measures of sports performance. However, there is insufficient evidence to make a definitive statement regarding the causal effect of muscular strength upon sports performance.

Conclusions: Considering the ARMSS, evidence is lacking whether improving muscular strength is causally related to sports performance. Present evidence is primarily observational and cross-sectional, experimental evidence is limited and focused upon *proxy* measures of sports performance, primarily conducted in small samples, and with little consideration regarding meaningfulness of effects. Suggestions to help improve research in this area and better answer this question include: larger sample sizes, determination of smallest effect sizes of interest for outcomes including muscular strength and *proxy* measures of sports performance (using both anchoring and/or expert opinion), and use of causal inference methods for observational data (*actual* sports performance, performance indicators, and fitness measures) including graphical causal diagrams and mediation analysis.

(*Journal of Trainology* 2020;9:20-32)

Key words: sports performance ■ muscular strength

INTRODUCTION

Sports science as a discipline including its researchers and practitioners generally aims to (and its existence is perhaps predicated upon) generate evidence and apply knowledge to improve sports performance. Indeed, this is generally agreed upon by both academics and practitioners[1] fuelling collaboration between these two domains[2] as both believe the discipline genuinely contributes towards sports performance[1,3]. Sports coaches generally perceive their domain of expertise to be in tactical and technical areas of sport, and that research in these areas to be important; however, applied sport science practitioners perceive research in the area of physical fitness and conditioning to be of greater value perhaps due to their occupational focus upon this.[4] Regardless, a component of physical fitness which continues to receive considerable interest from both domains is muscular strength, and thus applications of approaches (i.e. resistance training) to improve this variable, in the belief that it might be causally implicated in improving sports performance[5-7]. However, the role of increased strength and resistance training in improving sports performance has been questioned[8-13] and debate continues regarding its importance[14].

Bishop[15] proposed an Applied Research Model for the Sport Sciences (ARMSS) to aid in the identification of a problem (i.e. how to improve sports performance) and its solution through: observation and description to generate testable hypotheses, predictive modelling, experimental testing, optimisation of performance predictors, testing of efficacy, identification of barriers to uptake, followed finally by implementation research. Considering the generally accepted role of sport science in improving sports performance, and the interests of researchers and practitioners in the role of muscular strength and its improvement in achieving this goal, this critical review considers the evidence base regarding this. Using the ARMSS, it seeks to consider an answer to the question "Does increasing an athletes' strength improve sports performance?" whilst attempting to identify current barriers to understanding and gaps in the evidence base. Finally, it offers some suggestions to help researchers and practitioners contribute better evidence towards answering this, and other,

Received February 8, 2020; accepted April 28, 2020

From the[1] School of Sport, Health, and Social Sciences, Solent University, Southampton, UK (J.S., J.P.F.), ukactive Research Institute, ukactive, London, UK (J.S.), and School of Nutrition, Kinesiology, and Psychological Science, University of Central Missouri, Warrensburg, USA (D.C.)

Communicated by Jeremy P. Loenneke, PhD (Guest Editor)

Correspondence to: Dr. James Steele, Solent University, East Park Terrace, Southampton, UK

Email: james.steele@solent.ac.uk

causal questions in sport science.

Summarising the problem and the evidence

The ARMSS consists of eight stages, though within this review only the first four stages are considered. These first four stages (1. Defining the problem; 2. Descriptive research; 3. Predictors of performance; 4. Experimental testing of predictors) are sufficient to consider, in the context of the question posed, which is causal in nature. Later stages (i.e. 5. Determinants of key performance predictors; 6. Efficacy studies; 7. Barriers to uptake; 8. Implementation studies) are reliant upon an answer to this question first. There is little sense in seeking to identify how an intervention should be manipulated, and which variables of an exposure (i.e. resistance training) determine the optimisation of a mediating variable (i.e. muscular strength) to improve a desired outcome (i.e. sports performance) if we first have no knowledge of whether that proposed mediator is even causally associated with the outcome. Unfortunately, the field of sport science appears to have pursued this line of inquiry solely on the assumption that muscular strength and sports performance are causally related. Similarly, consideration of the latter stages of the ARMSS regarding implementation into practice appear redundant if we are stopped at the earlier hurdle of understanding the causal relationships between muscular strength and sports performance. Implementation research in general is lacking in both sport science and sport and exercise medicine[15,16], but again it makes little sense to even consider this if we first have no evidence of a causal effect.

Considering the first 4 stages of the ARMSS we could frame the following questions:

1. Defining the problem – Should I commit time, athlete energy, and recovery in to trying to make my athlete stronger (vs. more skill specific training, more rest, mental skills training, etc.)?
2. Descriptive research – Does strength differentiate elite and non-elite performers?
3. Predictors of performance – Are stronger athletes better at performing their sport?
4. Experimental testing of predictors – Does increasing an athlete's strength improve their sports performance?

From the perspective of the coach and/or sport science practitioner working with an athlete or team of athletes there is only so much time to dedicate to the different components of their preparation for sports performance. Thus any choice to implement an intervention presents an opportunity cost as it means that something else cannot be implemented. As noted, though coaches see technical and tactical components as more important, applied sport science practitioners place more value upon development of physical fitness.[4] However, strength and conditioning practices (i.e. resistance training) to improve components of physical fitness (i.e. muscular strength) must be considered of importance within high level sport from the fact that numerous surveys document their implementation in this domain.[17-23] Though implementation is already occurring, perhaps suggesting evidence must support the causal role of increasing muscular strength in improving sports performance, it is worth considering whether this existing practice is indeed based upon robust evidence.

Focusing on review level evidence, the most recent and comprehensive review of descriptive studies has suggested that, at the group level, muscular strength differs between non-elite and elite performers (including starters and non-starters in elite sporting samples), and that muscular strength is a *predictor* of (i.e. correlated with) a range of sports performance related outcomes[6]. Group level comparisons between non-elite and elite performers (where the variable differentiating the groups is in fact their *actual* sports performance e.g. amateur vs elite) and certain outcome measures for certain athletes notwithstanding (e.g. for powerlifters their *actual* sports performance *is* strength i.e. a one repetition maximum [1RM], and time to complete a distance for time trial athletes is in fact their *actual* sports performance[a]), one issue with this evidence is that the studies conducted typically examine *proxies* of sports performance (e.g. rate of force development, 'power', sprint speed, change of direction, jump performance etc.) rather than *actual* sports performance. Indeed, it has been argued that many of these proxy measures may be poor surrogates for actual sports performance particularly considering the role of specific practice of tasks in skill acquisition.[24] Experimental research employing interventions such as resistance training to increase muscular strength does suggest that many of these proxy measures improve.[9,10] Yet this is not a consistent finding even for performance outcomes that could be considered as *actual* sports performance such as time trial performance.[5,8,11,13]

There may be some contextual considerations for the causal effect of increasing muscular strength. Indeed, certain sports involve performances/movements that are closer in nature to the tasks typically performed to improve strength (e.g. powerlifting and other strength sports, and resistance training). However, the lack of consistency in changes in outcomes across the studies noted may be due to the lack of a "generality of strength adaptation" [b](the notion that making a muscle stronger through one movement will result in transference to strength in other movements). In fact, the further removed a performance task/movement is from the specific training task/movement performed to improve strength, the less likely the strength improvements from training will transfer to improve its performance.[24] Even where the *proxies* of sports performance mentioned involve similar musculature, they differ considerably compared to the typical resistance training that might be performed to improve strength in a measure such as a one repetition maximum. *Actual* sports performance then represents a possible further step away. This highlights a large gap in our understanding of the causal pathways (e.g.

[a] Though it could be argued here that these performances under laboratory or non-competition conditions is not strictly speaking the same as under competitive conditions due to the lack of direct competition, crowds, potentially changing environment etc.

[b] Interestingly this lack of general adaptation is not limited to physical task performance in relation to strength increases. Sala and Gobet[25] highlight the lack of a generality of cognitive adaptation resultant from cognitive training.

does increased strength lead to improved sprint speed and thus improved sports performance?). Further, this is the case even without considering the potential for the intervention with which muscular strength is increased (resistance training, dietary supplementation, other manipulations) to have different indirect *moderating* effects on the *mediating* direct potential role of muscular strength to cause improved sports related skills performance and for this to transfer to actual sports performance. This complex area will be touched upon again later in this review directing the reader to additional materials in causal inference.

Returning to our four questions encompassing the first four stages of the ARMSS we might then offer the following answers. With respect to the evidence:

2. Does strength differentiate elite and non-elite performers? *Yes, elite performers are typically stronger than non-elite performers in their sport*.
3. Are stronger athletes better at performing their sport? *Maybe. For primarily proxy measures of sports performance (e.g., vertical jump), there is some evidence that stronger athletes tend to perform better*.
4. Does increasing an athlete's strength improve their sports performance? *Unclear. For* ***actual*** *sports performance we currently have no clear evidence*.[c]

Then our answer to the coach or practitioner to help solve their problem is:

1. Should I make my athlete stronger (vs doing other things)? *Maybe. At present we don't really know whether it will or won't improve actual sports performance*.

The reason for this answer to the defined problem is that, at present, we only really have evidence to questions two and three of the ARMSS; both of which are *correlational* questions. Our problem however is a *causal* one (i.e. should I do x if I want to cause y) and needs to be solved by answering the *causal* question[d] of whether doing *x* (increasing muscular strength) causes *y* (improved sports performance). This problem is not unique to sport science however, as clinical sciences have similar issues within their respective fields.[27] **Currently, we do not believe there is sufficient evidence to make a definitive statement regarding the causal effect of muscular strength upon sports performance**. Why that is the case, and potential solutions for this issue, will be the focus of the remainder of this review.

Barriers to answering this question

(Between person) correlation does not imply (within person) causation: One issue with the use of correlational data pertains to what is known as the 'ecological fallacy' wherein the results from between person analyses may not correspond to those from within person analyses.[28,29] The following example may help to conceptualise this in our present context[e]. Suppose we are interested in the relationship between 1RM squat strength and 10m sprint time. A cross sectional analysis between persons (figure 1a) may indeed show that those who are stronger (high 1RM squat) tend to be faster (low 10m sprint speed). In fact, we know this from the body of literature reviewed by Suchomel et al.[6] If we took that relationship and generalised it to the within person level, we might conclude that if a particular person becomes stronger then they will become faster. But, as noted above, we have a finite amount of time and resource to commit to different training activities. Let's assume for the purpose of this example we decide to spend more time training our athletes to increase 1RM squat strength (i.e. heavy squatting), at the expense of spending time training to improve 10m sprint speed specifically (i.e. less sprint training). Again, for the purpose of the example, let's assume we decide to measure both regularly during this period thus tracking the *change* in each variable within individuals. In this case, if we were to plot all of these measurements, we might still see a group level relationship between 1RM squat strength and 10m sprint time. But if we consider this within individual athletes we see that the within person correlations suggest that increasing strength is associated with getting slower (figure 1b). This example is of course purely hypothetical and assumes that the changes in either are proportional to the time spent training them. In reality increasing 1RM squat strength may in fact contribute to improved 10m sprint time within person if we condition upon the time spent training either specifically. However, the point of the example is to highlight the issue with drawing conclusions from between person cross sectional correlations and assuming that they apply to within person causal effects. Indeed, this has recently been applied to the related question of whether changes in muscular hypertrophy are causally related to changes in muscular strength with further argument regarding the need for hierarchical models allowing individual person intercepts and slopes to vary also.[30]

Small sample sizes: One issue which many in the field of sport science will be familiar with, particularly with respect to experimental intervention research, is that of small samples and subsequently underpowered research designs. This is certainly not something new[31] but is something that, in combination with other practices, may result in a misleading view of the role of a particular variable upon sports performance. Prior work has shown that publications with small sample sizes are inherently more likely to show inflated effects (as small samples require large effects to achieve traditional thresholds for statistical significance in null hypothesis testing) and publication bias favours the publication of significant effects.[32-34] Though this has not specifically been investigated

[c] Powerlifting strength sports athletes notwithstanding as their *actual* sports performance is in fact strength i.e. 1RM

[d] We direct the reader to the recently published popular press text *"The Book of Why: The New Science of Cause and Effect"*[26] by Judea Pearl and Dana MacKenzie in which the difference between levels of causal inference, and associational or correlational and causal questions, is discussed.

[e] Note, the example provided here is the same to that presented by Solomon Kurz – see https://solomonkurz.netlify.com/post/individuals-are-not-small-groups-ii-the-ecological-fallacy/. The data and figures are reproduced from the code provided but are repurposed here for the present context i.e. the relationship between strength upon sports performance.

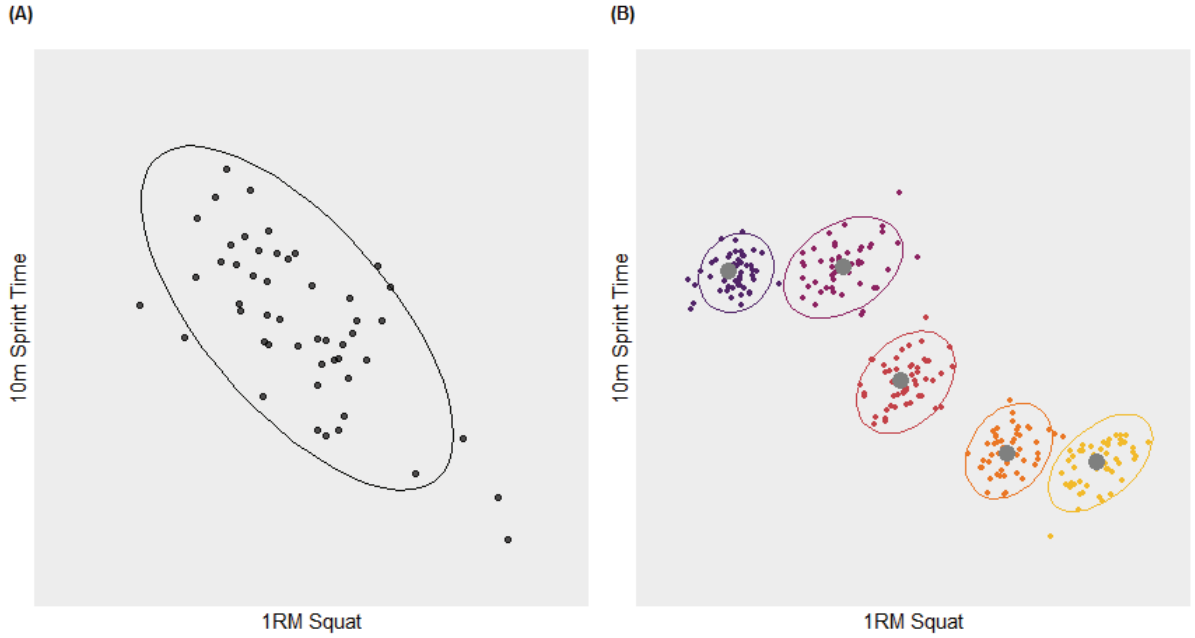

Figure 1 Example highlighting the ecological fallacy showing (A) between person correlations may not reflect (B) within person correlations (Five individual athletes shown as an example).

within sport science, there is no reason to believe that the same is not occurring within this discipline and the field may be rife with false positive findings.[35] Ignoring for the moment that we still lack data on actual sports performance outcomes, it seems possible that many within the field, including researchers and practitioners, may be pontificating upon false positives and inflated effect sizes regarding *proxy* outcomes for sports performance and thus, overestimating any benefits of increasing muscular strength through interventions such as resistance training.

(Standardised) Effect size mania!: The reliance on null hypothesis significance testing, without consideration of the magnitude of effects under examination, has been considered an issue for many fields.[36] In this context, sport science over recent years has seemingly began to focus more upon the reporting of effect estimates using effect sizes alongside the results of null hypothesis significance testing (i.e. *p* values). Some have even called for setting aside falsification and hypothesis testing in favour of estimation approaches within sport science.[37] It is common to see effect sizes reported including Cohen's *d* and Hedge's *g* in papers examining the effects of sport science interventions upon outcomes. Yet, many seem to think that to report an 'effect size' means that what should be reported is a *standardised* effect size such as the ones mentioned above.

Returning to the problem we are trying to solve for the coach or practitioner (i.e. should I make my athlete stronger (vs doing other things)?), results such as "sprint speed increased ($d = 0.47$)", or "Hedge's *g* indicated the effect was large" we suspect are not easily interpreted by coaches, practitioners, or athletes. In fact, some research has shown that though non-scientists find certain standardised effect sizes more 'informative' than others,[38] most are relatively unimpressed by even 'large' standardised effects[39]. The use of standardised effect sizes likely makes decision making difficult regarding the problem that a coach is looking to solve and ultimately is counterintuitive to the goal of sports scientists. Coaches may wish to consider the cost-benefit ratio with respect to the investment (time and resources required to increase muscular strength through resistance training displacing other activities) and the pay-off (how much of an improvement in sports performance will occur and is it meaningful). Further, it has been noted that across fields there is lack of consistency in exactly how standardised effect sizes are even calculated with many not reporting the calculations used.[36] Within sports science, varying denominators appear to be used without specification and also the varying use of either within- or between-group calculations and comparisons of effects inappropriate for the study designs are often employed.[40]

The family of effect sizes typically used are commonly referred to as Cohen's *d* with subscripts recommended to specify the exact effect sizes computed;[41] though this practice is rarely used in the sport sciences. Laken's[42] provides a useful tutorial and overview of this family including their formulae. Most commonly used within sport science are Cohen's d_s, Hedge's g_s, and Glass' Δ for between-group study designs (e.g. comparing the pre-post change scores [Δ = post – pre] between two intervention groups typically using the pooled baseline SD for intervention studies as the denominator), or Cohen's d_z for within-group study designs (e.g. comparing pre-post change score [Δ = post – pre] within a single group and typically using the SD of the change score as the denominator). However, it is not uncommon for all manner of variety of combinations to be used in sport science studies, including the mixing of both within- and between-group effect sizes in

Table 1 Comparison of standardised effect sizes (d) for outcomes from Griffiths et al.[43] for both within- and between-group using different denominators.

	TRAD			EXP			Between Group (TRAD-EXP)		
Denominator	CMJ	10m Sprint	1RM	CMJ	10m Sprint	1RM	CMJ	10m Sprint	1RM
Independent Group Change SD	0.75	-1.84	2.25	2.08	-1.42	2.59	-	-	-
Pooled Change SD	-	-	-	-	-	-	-0.78	0.07	0.48
Pooled Baseline SD	0.23	-0.42	0.63	0.58	-0.52	0.57	-0.21	-0.21	0.02
Pooled Pre/Post SD	0.24	-0.45	0.64	0.45	-0.47	0.53	-0.21	-0.21	0.02

Within-group Δ calculated as: post – pre

Between-group Δ calculated as: $TRAD_{\text{within-group }\Delta}$ - $EXP_{\text{within-group }\Delta}$

All pooled SDs calculated as: $SD_{pooled} = \sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_2^2}{n_1+n_2-2}}$

and for Pre/Post: $SD_{pooled} = \sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_2^2+(n_3-1)SD_3^2+(n_4-1)SD_4^2}{n_1+n_2+n_3+n_4-4}}$

between-groups pre- and post-test randomised controlled study designs, and applications of varying denominators across these[f].

By way of example, in a recent randomised trial[43] conducted by our group two resistance training interventions (TRAD and EXP) were examined for their comparative effectiveness in amateur soccer players upon a range of outcomes including muscular strength measured as one-repetition maximum (1RM), 10 m sprint time, and counter movement jump (CMJ) height. To highlight the issue of interpretability when reporting standardised effect sizes in sport science research (particularly without detail of how they were calculated) table 1 presents the standardised effects sizes calculated using different denominators both for within-groups (i.e. mean change score [Δ = post – pre] for that group divided by chosen denominator) as well as for the between-groups difference (mean difference in change scores [Δ = $TRAD_{\text{within-group }\Delta}$ - $EXP_{\text{within-group }\Delta}$] between groups divided by the chosen denominator). The considerable variance in the reported standardised effect sizes dependent upon the method used is clearly evident, none of which we contend are easily interpretable for a coach, practitioner, or athlete as standalone statistics.

Contrastingly to the confusing results in table 1, in the study we also reported the results as point estimates (unadjusted and adjusted estimated marginal means from an analysis of covariance model with baseline scores as covariates) along with their precision (95% confidence intervals) for each outcome and each group in their raw units (i.e. kg, seconds, and cm). Few seem to realise that simply reporting a point estimate (i.e. a mean difference in change scores between groups, and/or mean values for change for each group independently) with precision of that estimate (i.e. confidence or credible intervals) **in the raw units of measurement** is in fact reporting an effect size.[36] It seems likely that a coach is more easily able to interpret, contrast, and make decisions about whether either intervention is worthwhile based upon reporting of independent group changes of 0.7 cm [0.3 cm to 1.1 cm] compared with 1.3 cm [0.9 cm to 1.7 cm] in jump height for example as compared to $d_z = 0.75$ compared to $d_z = 2.08$ (or between group effect size of $d_s = -0.21$). However, this still begs the question of whether a change in any of these variables (muscular strength, or proxies of sports performance) believed to be potential mediators of actual sports performance are in fact *meaningful*.

What changes are meaningful?: Is a decrease in 10m sprint speed of 0.05s meaningful? This is how much the soccer players in the study by Griffiths et al.[43] improved on average. This amounts to an improvement in sprint speed of 2.2%. Answering whether this change is meaningful or not is more difficult than it seems. One might ask by way of clarification in response to the question "meaningful for what?". Well, in the case of our coach/practitioner's problem, we likely want to know whether this results in a 'big enough' change in *actual* sports performance for us to justify spending the time to acquire it in our athletes. Herein lies a further issue with the use of *proxy* measures of sports performance. In the case of a soccer player it might sound 'intuitive' and 'logical' that a faster athlete is likely to perform better in an actual soccer match. Indeed, all other things being equal we might expect a faster athlete would be more likely to win more loose balls, outrun opposition players etc. Yet, it is a long causal pathway from resistance training → increased muscular strength → improved sprint speed → winning more loose balls → scoring more goals than the opposition team. The final part in that pathway is, in this example, the *actual* measure of sports per-

[f] Note, we are not claiming to be innocent of this ourselves and have inadvertently contributed to the confusion by employing such practices in prior work.

formance (assuming performance in one match, not necessarily the average performance over a season or multiple seasons). Indeed, in Griffiths et al.[43] it was argued that this change of 0.05s might not be deemed particularly meaningful even though others have interpreted similar magnitudes of change as being meaningful[5,44].

A similar example can be made with respect to other 'sports related performance' outcomes. Bimson et al.[45] performed a randomised trial comparing regular soccer practice to the addition of an isometric knee extension resistance training intervention alongside regular soccer practice in collegiate level female soccer players. In response to the isometric resistance training intervention there were no changes in sprint speed (with or without a soccer ball) and additionally no changes in change of direction speed. Bimson et al.[45] did however find improvements in CMJ and the distance a soccer ball could be kicked. The latter improved by 2.73 ± 0.65 metres and, similarly to the example of sprint speed, it would be possible to come up with a plausible sounding explanation for why this improvement in kick distance could be deemed meaningful. Yet, the fact remains that we don't know if being able to kick a soccer ball further, or run faster, means that actual soccer performance improves.

Measuring changes in actual sports performance is difficult: An argument could be made that any change in *actual* sports performance that results in a 'win' is meaningful. Yet, it is worth considering whether it is even possible to measure such changes. It seems obvious to all in sport science that measuring changes in *actual* sports performance (win/loss ratio, pts/goals scored for a given match, etc.) and trying to determine what has caused them is not an easy feat. Even within the field of performance analysis it not always easy to link specific performance indicators during actual sports performance to outcomes such as winning/losing.[46] As an outcome, actual sports performance is inherently noisy with an incredible number of degrees of freedom and confounding variables. This applies to almost any sport (though some such as time trial and powerlifting/weightlifting athletes may be notable exceptions with fewer degrees of freedom and confounding variables affecting actual performance – in fact it has been argued that where such measurement of actual performance is possible it should of course be preferred[47]) but notably sports with opposition, and team sports, magnify this exponentially. Indeed, within soccer the match to match variation in just time spent in different speed zones[48] in addition to other technical elements,[49] let alone *actual* performance, can be considerable. Given this noisy outcome, it seems incredibly unlikely that any intervention is going to induce a detectable effect of a large, meaningful magnitude. Most effects from any kind of intervention seem more than likely to be trivial or at best small with the effect likely decreasing in magnitude in higher levels of sports performance. Indeed, much of the variance from sources other than the inclusion of any intervention may be such that it is impossible to even detect these trivial effects outside of such measurement error. The inclusion of time matched control groups of athletes would aid in understanding the magnitude of this variation and the magnitude of effects we *could* detect. But, it is difficult enough to persuade both coaches and athletes to participate in research that involves intervention, let alone control periods, particularly when it requires them to stop doing something which they may presently believe to be effective (irrespective of whether it actually is or not).

A hypothetical study…

Considering the barriers highlighted, and making reasonable assumptions that the effects of any intervention are likely to be both small and the data noisy, let's consider briefly planning a hypothetical study that might help provide an answer to our problem. Let's say we are a coach wanting to know whether we should let our athletes spend time with the sport science practitioner to do resistance training instead of time practicing other components of the sport. We speak to the sports science practitioner and decide to consider a simple between-groups, pre- and post-test randomised controlled design of the effects of resistance training (intended to increase muscular strength which we will assume occurs for this example) upon *actual* sports performance (the actual outcome measure is not necessarily relevant here). We want to randomise athletes to either receive resistance training in addition to their current training, or not. Considering how noisy our outcome measure is, and that we anticipate a small effect (though we might still consider this to be important if it is *actual* sports performance we are considering, particularly if this has implications for performance-related income as an example), we decide that we want a study that can detect small effects based on traditional standardised effect size thresholds (standardised effect size thresholds are used in this example as we have not specified the actual outcome measure). So, we opt to try and power our study to be able to detect small effects[g] ($f = 0.1$) with traditionally accepted level of $\alpha = 0.05$ and $\beta = 0.20$. Our dependent variable here is the change score (Δ = post – pre) and our independent variable is the group (either receiving the resistance training intervention, or not). We know that regression to the mean might impact our analysis comparing the change in performance in the intervention group to the change in performance in the control group and so we opt to analyse our data using an analysis of covariance with the baseline performance values as a covariate. Further, we know that this model is statistically more powerful than others typically used for pre-post trials.[51] Based upon this we plug our numbers into a commercially available software package (G*Power, v3.1.9.2, Universitat Kiel, German) to see how many athletes we need. Lo' and behold we only need a total of 787 athletes to take part in our study. This number is not dissimilar from if we wanted to instead plan our study based upon the estimation of the magnitude of the effect with a certain level of precision. For example, using available software (ESCI, La Trobe University, Australia) for a two-group design would require 833 athletes if we wanted to have confidence intervals as precise as $f = -0.1$ to 0.1 around our effect estimate with assurance of $\gamma = 99$ (note, this does not consider

[g] Which has been argued to be what sport scientists should be aiming for if they want to make an impact[50]

adjustment for the baseline covariate). However, we can consider what the expected (or known) pre- to post-test correlation (r) in our dependent variable might be which would affect our sample estimate[52]. We can adjust our sample estimate using $N_{Ancova}(1-r^2)$, where N_{Ancova} is the sample size from our unadjusted estimate above. For example, and considering a range of plausible correlations, with a correlation of 0.5 we would need 590 athletes, with 0.7 we would need 401 athletes, and for 0.9 we would need 150 athletes. The implications of these estimates of course need to be considered in terms of their cost and benefit. Running a study with 590 athletes is obviously costlier in terms of time and money compared to running one with 150. The benefit of improving *actual* sports performance by a 'small' magnitude could be considered in terms of its economic pay off and this considered relative to the cost of running the study. But, there are also the issues of whether there are even enough athletes available and willing to sample. Though not entirely unfeasible, even the lower estimate above of 150 athletes would exceed considerably the typical sample sizes achieved in most interventional research in sport science. Further, whether or not the smaller or larger estimates are needed depends on some knowledge of the dependent variable of interest also. In all cases, the practicality of performing the study may be questionable. So, what can we do to try and provide some solution to our coaches' problem?

What can we do?

The example of a hypothetical study above was deliberately given within the context of a Frequentist statistical framework wherein the aim is to draw inferences that aid in our decisions of how we should act (i.e. as though the null or alternative hypothesis were true). However, given the very real practical problems of conducting the type of experimental research normally considered as necessary to draw causal conclusions, it seems unworkable to be able to provide answers regarding action in relation to a particular hypothesis. Methods such as estimation and the use of magnitude-based inference[37] have been argued to be a potential solution to the problems sport science faces. However, this approach has been heavily criticised recently for presenting itself as being Bayesian in nature when in fact it is not (see Sainani et al.[53]). Yet, fully Bayesian approaches, wherein the aim is to quantify the degree of belief we should hold in a particular hypothesis, have been suggested as a potential solution to the small effects and sample sizes that exist within sport science[54] and estimation approaches within Bayesian statistical frameworks may be useful in this regard[55]. Indeed, the example given above is deliberately simple whereas in reality more complex factorial designs may be desirable including for example both intervention doses, implementation strategies, and any potential interactions. In this case, Bayesian approaches using hierarchical priors may make such studies far more feasible by reducing sample sizes required and improving precision of effect estimates.[56] However, instead of focusing on the debates regarding the statistical approaches which might be used, here we focus on other possible solutions to help with answering the question "Does increasing an athlete's strength improve their sports performance?"

The 'so what' factor: It appears at present sport scientists are practically limited to looking at *proxy* measures of sports performance. However, as noted above, whether changes in these outcomes are meaningful is often overlooked. One thing that we could be doing is at least trying to first answer the question "so what?"[57]. Traditionally, benchmarks for standardised effect sizes are used to interpret the meaningfulness of changes in outcomes. Varying benchmarks have been proposed and applied in sport science[58] yet most commonly used are Cohen's[41] original 'small', 'medium', and 'large' effects. Most however do not realise that Cohen's[59] definition of these conventional thresholds was to a large extent arbitrary with a 'medium' effect size intended to be *"likely visible to the naked eye of a careful observer"*, and a 'small' effect size to be *"noticeably smaller than medium but not so small as to be trivial."* Further, these were first designed to be used within the field of psychological research wherein measurements are often taken using scales for which there is no clear intuition as to what a meaningful magnitude of change is in their raw, often ordinal, units of measurement. As we have noted, we suspect most coaches and practitioners are unlikely to be able to interpret effects reported as standardised effect sizes and often the measurements taken are more easily understood when expressed in their raw units. In this sense, some have argued for use of the smallest worthwhile difference/change.[60] For individual athletes this has been proposed based upon simulation as 0.3 × "within participant race to race σ" being equal to an additional podium placement per 10 races/events. For team athletes this has been based upon Cohen's arbitrary 'small' threshold as 0.2 × "between participant σ". In both cases however, it is difficult to detect whether these smallest worthwhile changes have actually occurred or not as noise in most proxy performance measures typically exceeds these values. So how can we determine what the smallest effect size of interest should be?

One approach which is commonly used within the clinical sciences is that of anchoring. Anchor-based approaches have been used to set minimal clinically important changes for outcomes such as pain and disability in rehabilitation for example[61] and have recently been proposed as a better-informed method to determine the smallest effect size that can be deemed meaningful[62]. This involves considering the measurement outcome that you intend to use and examining it in relation to a global perceived outcome measurement. For example, within a clinical setting you might measure change in pain using a 0-100 visual analogue scale pre- and post- an intervention and also ask post-intervention for a global rating of perceived condition using an ordinal scale such as:

1. Very much improved
2. Much improved
3. A little improved
4. No change
5. A little deterioration
6. Much deterioration
7. Very much deterioration

In this case it is determined *a priori* how much of a perceived change will be the cut off on the global perceived outcome scale (e.g. 'Much improved' or higher considered as a meaningful change, and vice versa in the opposite direction of effect) and also what range of responses will be deemed, for all intents and purposes, to be considered as a stable effect (e.g. 'A little improvement' and 'A little deterioration' are considered to not differ practically from 'No change'). The change score for pain associated with the *a priori* cut offs for perceived outcome are then used as estimates for the smallest effect size of interest (e.g., we may see that a 20 ± 3.4% reduction in pain score is needed before patients report "much improvement").

There is no reason why this approach couldn't be used with *proxy* measurements of sports performance and Anvari and Lakens[62] provide useful guidance on the use of anchor-based methods. For example, what is the mean sprint speed change required to elicit a global perceived improvement in sprint speed of 'Much improved' or higher? There should also be consideration of whether it might be possible to anchor changes in proxy outcomes based upon actual sports performance. A research program aimed at specifically determining this (i.e. anchoring changes in proxy outcome measures to win/loss ratio, pts/goals scored for a given match or matches over a period of time) may then permit researchers, coaches, and practitioners to utilise *proxy* measures with greater confidence in their ability to infer from them whether or not they lead to changes in actual sports performance. However, in the case of our current question (i.e. Does increasing athletes' strength improve sports performance?) it may be that changes in strength need to be anchored to *actual* sports performance directly, or to *proxy* measures of sports performance after these have been anchored to *actual* sports performance. Such is the nature of the causal pathway relating to our coach/practitioners problem. Yet, both of these may still be difficult approaches as they require the measurement of *actual* sports performance. Coaches however may already have some expertise and insight based upon their experience that might aid in the interpretation of changes in *proxy* outcomes and whether they are meaningful.

Integrating expert opinion: As noted, areas such as rehabilitation have used anchor-based methods for some time in the determination of minimal clinically important changes. However, fields such as these have not limited themselves to only using such approaches and have combined this with expert clinical and academic expertise.[61] Surprisingly, this seems to be relatively uncommon within sports science (though there is some use of conceptual analysis of expert coach opinion on development of competence/efficacy self-report measures[63]). A recent study however demonstrates its application. Kyprianou et al.[64] elicited the experience and opinion of 49 elite soccer practitioners regarding what they would consider to be an acceptable level of measurement error when measuring sprint speed. They then used the median response from their survey respondents to set equivalence bounds for the smallest effect size of interest when comparing the agreement of two measures of maximal sprint speed. This novel approach could easily be applied in the planning of research to determine whether or not an intervention intended to increase muscular strength (i.e. resistance training) can produce a meaningful change in sports performance. The researcher could work with coaches and practitioners, perhaps even athletes themselves, to determine *a priori* the smallest change in a *proxy* measure of sports performance that they would deem to be meaningful. Then after the implementation of the intervention they can interpret the magnitude of change observed in light of this smallest effect size of interest. If the intervention is unable to improve the performance outcome measure at least as much as the smallest effect size of interest then from the coaches, practitioners, and athlete's perspective it, by definition, does not produce a meaningful effect. From a decision-making process this information could be incredibly useful in deciding whether to commit time to completing such interventions in the future and for other coaches and practitioners considering whether or not to do so. Further, within the Bayesian statistical framework there exists tools developed specifically for the integration of expert opinion into determining probability distributions for effects (e.g. MATCH[65]). This could be used to both elicit prior probability distributions regarding intervention effects, but also distributions reflecting the uncertainty in outcome measures to help with determination of the smallest effect size of interest.

All of this is however, to some extent, reliant upon the ability to conduct experimentation with randomisation. That is to say that some athletes are randomised to receive an intervention and others not in order to provide an unbiased estimate of the causal effect of the intervention (i.e. resistance training) upon performance. This is prior to even considering whether that causal effect is meaningful or not. However, as noted experimental research is difficult to conduct in athletic settings for a range of reasons. Observational research in the form of correlations however are fairly easy to come by particularly considering the volume of such research regarding muscular strength[6]. As we have noted there are limitations to the use of observational research, but perhaps there should be consideration of how to better use this observational data if it is currently the best we have.

Causal inference from observational data: In some circumstances it is either ill advised, or not possible, to conduct randomised experiments in order to draw causal inferences.[66] Most of the problems that both scientists and applied practitioners arguably seek to solve are causal in nature; much like the one that is the focus of our discussion – "Should I make my athlete stronger (vs doing other things)?" – and which requires an answer to the causal question – "Does increasing an athlete's strength improve their sports performance?". What can be done however in the case where we cannot utilise true experimentation? The classic example here is the causal effect of smoking upon lung cancer. Ultimately the causal role of smoking was elucidated through careful use of observation and epidemiological research. But the ability to yield unbiased (or minimally biased) causal estimates from observational research requires the use of approaches which have, until recently, remained relatively unused in sport sci-

ence. Techniques such as graphical causal diagrams and potential outcomes framework approaches including propensity scores, *G* methods, and mediation analysis have recently begun to be introduced to the field of sport and exercise and may hold promise for causal inference from observational data.[67,68]

For example, causal diagrams are essentially maps of probable causal pathways between variables that are drawn *a priori* and in essence encode the model that is to be tested. They have existed since the 1920s when Sewall Wright developed what he referred to as path diagrams. In the past 20 or so years they have gained attention from health, social, and behavioural scientists[26] and their use with large observational datasets (alongside other techniques to 'imagine' counterfactuals) permit consideration of things such as comparative effectiveness even in the absence of experimentation[66]. Under the assumptions of the model, causal diagrams can be combined with propensity scores, *G* methods, and causal mediation analysis to provide unbiased estimates of comparative interventions effects, including time-varying longitudinal effects, mediated through specific pathways or variables[26,68-70]. A commonly employed causal map is the directed acyclic graph (DAG) wherein the assumptions of the model are explicit and drawn a *priori*. DAGs make it clear as to (a) what variables are being proposed as related; (b) the direction of causation proposed; (c) whether direct or indirect pathways are being proposed; and (d) can handle both linear and non-linear time-varying estimates. It is beyond the scope of this article to fully explain their use and application. However, rules for developing DAGs are intended to reduce confounding and aid in identifying causal effects; both guidance[26,71-74] and freely available software (e.g. www.dagitty.net) exist to aid in their construction. Despite this, it is recommended that they are developed with those experienced in causal inference from observational data using such methods.

The use of causal diagrams and mediation analysis might aid in determining the causal effects of increasing strength, as a result of interventions such as resistance training, upon sports performance. It has recently been argued that a similar area of debate – the causal role of muscular hypertrophy, resultant from resistance exercise, upon muscular strength – would benefit from the application of this approach[67]. Further, Nuzzo et al.[67] note that *"causal mediation analysis has the potential to be used to clarify a host of other questions in exercise science, such as whether training induced changes in muscle strength cause improvements in functional performance"* and present a simple DAG depicting this (figure 1d in their paper). DAGs could be used to map out carefully, and *a priori*, the proposed relationships between variables to allow careful consideration of these and how they need to be handled in analyses to yield a minimally biased causal estimate. As noted they also permit consideration of both direct and indirect causal pathways. Many sports teams, coaches, practitioners, and athletes now capture data on a range of variables including *actual* sports performance, performance indicators such as those focused upon in performance analysis,[46] in addition to measures of fitness such as strength. Applications of causal thinking, and techniques such as those noted above, to the types of large datasets that are becoming more widely available in sport may aid in answering many causal questions (e.g. see Binney[75]).

For example, in figure 2a a simple DAG proposes that resistance training may have both a direct effect upon *actual* sports performance, but also an effect which is indirect and mediated by the effect this exposure has upon muscular strength. Considering the discussion relating to the use of *proxy* measures of sports performance we might further construct a DAG adding this variable and mapping the proposed relationship (figure 2b). DAGs can get complicated quite quickly and though we have merely labelled 'confounders' in our figures careful consideration of potential confounders is required such that they can be carefully measured with minimal error if it is required that they are adjusted for to allow minimally biased estimates to be produced. These approaches may be useful in permitting clearer conceptualisation of the proposed causal pathways and thus their investigation with observational data. Indeed, muscular strength represents an indirect causal pathway for the effects of resistance training. However, recently it has been argued[12] that another indirect pathway may exert a stronger causal effect and be the primary pathway for the causal effects of exposure to resistance training interventions: injury risk (figure 2c).

Are the effects of resistance training and/or strength due to injury risk?

An injured athlete is normally unable to perform to the best of their ability, and sometimes is unable to perform at all, so injury risk presents a plausible mechanism through which resistance training and improved muscular strength might improve sports performance. Recent evidence suggests that higher levels of strength may be predictive of injury risk[76]; though, this does not imply *improving* strength will reduce injury risk. There is good evidence from meta-analyses that resistance training can reduce sports injury risk.[77,78] In 2013, Lauersen et al.[77] performed a meta-analysis which compared resistance training, proprioception, stretching, and multicomponent interventions finding the greatest risk reductions from resistance training. Further, in an updated meta-analysis including 6 studies of 5 interventions in 7738 participants aged 12-40 years old and experiencing 177 acute or overuse injuries, Lauersen et al.[78] reported a reduction in risk of 66% (with 95% certainty of a halving of risk [95%CI 52% to 76%]) as a result of resistance training interventions. The proposed mechanisms for the effects of resistance training upon injury risk include improved coordination, enhanced technique in training/match situations, strengthening of adjacent tissues reducing critical joint loads and better psychological perception of high-risk situations, but also strengthening of the musculature itself[78,79]. However, similarly as with the issues of causally linking the effects of a resistance training intervention through increased strength as a mediator, it is not entirely clear whether it is *strength* increases per se that are responsible.

Lauersen et al.[78] found that a 10% increase in resistance

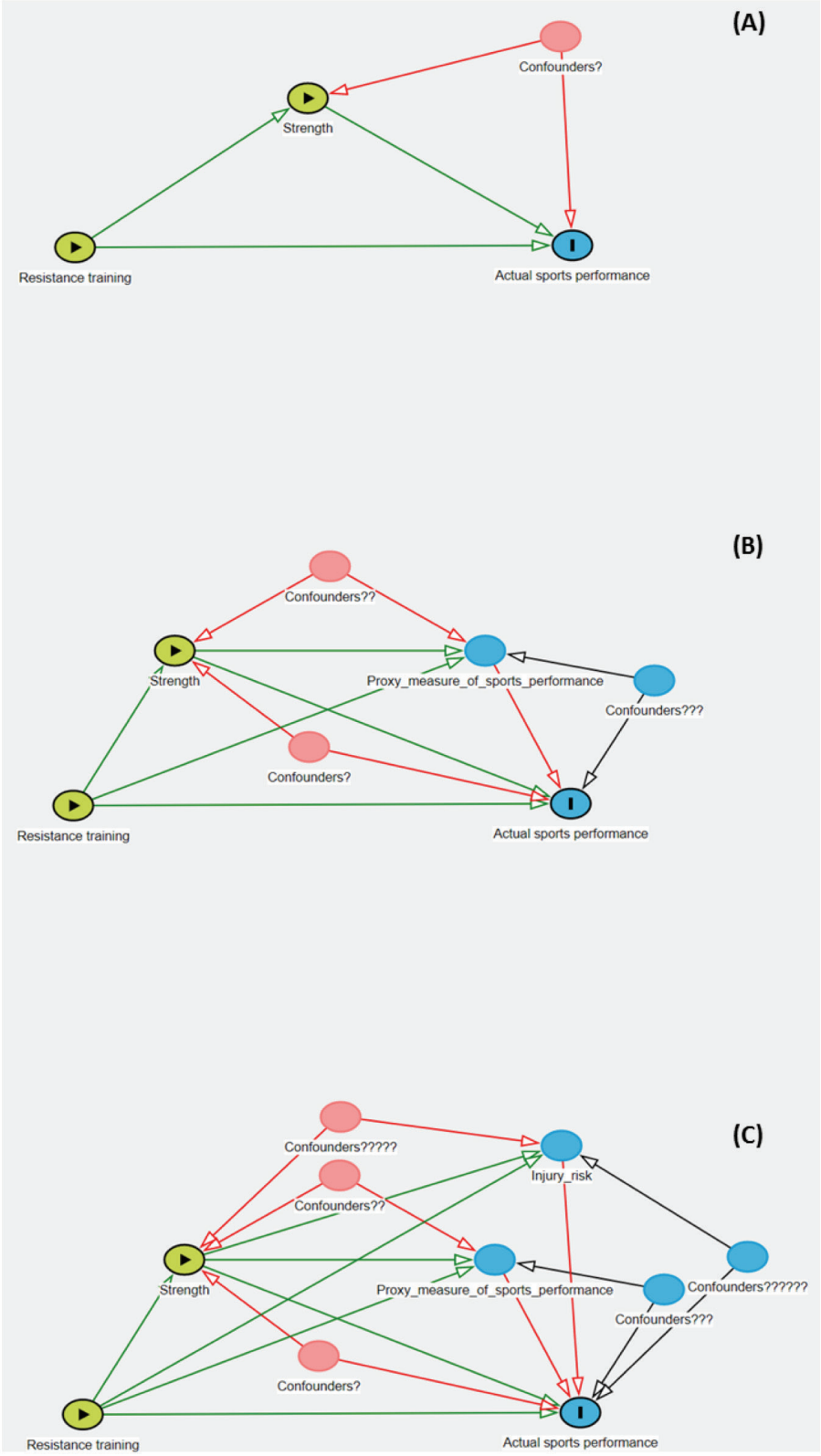

Figure 2 Directed Acyclic Graphs (DAGs) depicting the effect of resistance training upon *actual* sports performance mediated by (a) muscular strength, (b) adding the further mediator of *proxy* measures of sports performance, and (c) adding the further mediator of injury risk.

training volume was associated with a 4.3% additional injury risk reduction, though there is limited data directly comparing different manipulations of resistance training volumes upon injury risk. As such, it has been argued that resistance training to improve muscular strength should perhaps be applied in the least amount possible to optimise injury risk reduction.[12,80]

Maximising deliberate practice of actual sports performance

Indeed, performing 'just enough' resistance training might allow athletes to maximise the time spent practicing their *actual* sports performance under supervision of their coach, and/or recovering. Further, a reduction in injury risk might also increase the time an athlete is able to engage in specific practice of their *actual* sports performance as they are less likely to be unable to do so due to injury. However, the extent to which even this exposure (i.e. deliberate practice) has a causal effect upon sports performance may be questionable. Macnamara et al.[81] conducted a meta-analysis examining the relationship between deliberate practice and sports performance including 33 studies of 52 independent samples yielding 63 effects sizes from a total of 2765 participants. They found deliberate practice only accounted for 18% of the variance in sports performance (note this was composed of both *proxy* and *actual* measures) and accounted for only 1% specifically in elite athletes. There is some debate about the specific definition as to what constitutes 'deliberate practice' and Ericsson[82] argues that "…one-on-one instruction of an athlete by a coach, who assigns practice activities with explicit goals and effective practice activities with immediate feedback and opportunities for repetition…" is the most appropriate definition. The studies included by Macnamara et al.[81] in their meta-analysis however include various elements of 'practice' and at least two of the studies included had estimates of practice time derived from activities including strength and conditioning activities such as resistance training[82]. Further, Macnamara et al.[81] examined the relationships in a between-person manner and thus it is not entirely clear whether within-person relationships exist between deliberate practice and performance (see earlier discussion). Nevertheless, it is somewhat disconcerting for the sport scientist to see it is unclear whether even something as foundational as deliberate practice has much of an impact upon sports performance. If this is unclear, to what extent can we reasonably expect causal effects to be of meaningful magnitudes from interventions such as resistance training and their resultant impact upon muscular strength? With lack of clarity even when it comes to deliberate practice of *actual* sports performance perhaps we should be somewhat more tentative and sceptical of the impact that sport science actually has upon sports performance. Indeed, considering the issues raised and suggestions offered in this brief review, it is interesting to note that Ericsson[82] concluded with respect to deliberate practice that:

> "*Future research should collect objective measures of representative performance with a longitudinal description of all the changes in different aspects of the performance so that any proximal conditions of deliberate practice related to effective improvements can be identified and analyzed experimentally.*"

We would, add to this that future intervention studies should consider the direct comparison of deliberate practice with the inclusion of intervention approaches such as resistance training. Such studies, whether using observational or experimental designs, would be useful in helping coaches and athletes decide how best to allocate their finite time to these two components, and may open up wider consideration of the implementation of other tactical or cognitive elements of preparation in addition to rest and recovery.

CONCLUSIONS

Sport science is intended to improve sports performance and this review considered the question "Does increasing an athletes strength improve sports performance?". Considering the ARMSS, evidence seems lacking regarding whether improvements in muscular strength are causally related to sports performance. Present evidence is primarily observational and cross-sectional in nature, experimental evidence is limited and focused upon *proxy* measures of sports performance, primarily conducted in small samples, and with little consideration to whether the effects reported are even meaningful in terms of their magnitude. As such, it is unknown whether increases in muscular strength resultant from exposure to interventions such as resistance training actually improve sports performance. We have offered some suggestions to help improve research in this area and better answer this question including: larger sample sizes, determination of smallest effect sizes of interest for outcomes including muscular strength and *proxy* measures of sports performance (using both anchoring and/or expert opinion), and use of causal inference methods for observational data including graphical causal diagrams and mediation analysis. Many of these suggestions can further benefit from adherence to principles of open science including detailed a *priori* specification and pre-registration or use of registered reports[35]. It is hoped that this article helps to improve future research not only seeking to answer the question(s) posed here but also for other causal questions in sport science.

REFERENCES

1. Bishop D, Burnett A, Farrow D, et al. Sports-science roundtable: does sports-science research influence practice? *Int J Sport Physiol Perform* 2006;1:161-168.
2. Malone JJ, Harper LD, Jones B, et al. Perspectives of applied collaborative sport science research within professional team sports. *Eur J Sport Sci* 2019;19:147-155.
3. Reade I, Rodgers W, Hall N. Knowledge transfer: How do high performance coaches access the knowledge of sport scientists? *Int J Sports Sci Coaching* 2008;3:319-334.
4. Fullagar HHK, McCall A, Impellizzeri FM, et al. The translation of sport science research to the field: A current opinion and overview on the perceptions of practitioners, researchers and coaches. *Sports Med* 2019:49:1817-1824.
5. Seitz LB, Reyes A, Tran TT, et al. Increases in lower-body strength transfer positively to sprint performance: A systematic review with meta-analysis. *Sports Med* 2014;44:1693-1702.

6. Suchomel TJ, Nimphius S, Stone MH. The importance of muscular strength in athletic performance. *Sports Med* 2016;46:1419-1449.
7. Suchomel TJ, Nimphius S, Bellon C, et al. The importance of muscular strength: Training considerations. *Sports Med* 2018;48:765-785.
8. Cronin J, Ogden T, Lawton T, et al. Does increasing macimal strength improve sprint running performance? *Strength Cond J* 2007;29:86-95.
9. Bruce-Low S, Smith D. Explosive exercises in sports training: A critical review. *J Exerc Physiol online* 2007;10:21-33.
10. McGuigan MR, Wright GA, Fleck SJ. Strength training for athletes: Does it really help sports performance? *Int J Sports Physiol Perform* 2012;7:2-5.
11. Dankel SJ, Mattocks KT, Mouser JG, et al. A critical review of the current evidence examining whether resistance training improves time trial performance. *J Sports Sci* 2018;36:1485-1491.
12. Buckner SL, Jessee MB, Dankel SJ, et al. Resistance exercise and sports performance: The minority report. *Med Hypotheses* 2018;113:1-5.
13. Lum D, Barbosa TM. Effects of strength training on Olympic time-based sport performance: A systematic review and meta-analysis of randomized controlled trials. *Int J Sports Physiol Perform* 2019;14:1318-1330.
14. Hornsby WG, Gentles JA, Haff GG, et al. What is the impact of muscle hypertrophy on strength and sport performance? *Strength Cond J* 2018;40:99-111.
15. Bishop D. An applied research model for the sport sciences. *Sports Med* 2008;38(3):253-263
16. Beedie C, Mann S, Jimenez A, et al. Death by effectiveness: exercise as medicine caught in the efficacy trap! *Br J Sports Med* 2016;50:323-324.
17. Sutherland TM, Wiley JP. Survey of strength and conditioning services for professional athletes in four sports. *J Strength Cond Res* 1997;11:266-268.
18. Ebben WP, Blackard DO. Strength and conditioning practices of National Football League strength and conditioning coaches. *J Strength Cond Res* 2001;15:48-58.
19. Ebben WP, Carrol RM, Simenz CJ. Strength and conditioning practices of National Hockey League strength and conditioning coaches. *J Strength Cond Res* 2004;18:889-897.
20. Ebben WP, Hintz MJ, Simenz CJ. Strength and conditioning practices of Major League Baseball strength and conditioning coaches. *J Strength Cond Res* 2005;19:538-546.
21. Simenz CJ, Dugan CA, Ebben WP. Strength and conditioning practices of National Basketball Association strength and conditioning coaches. *J Strength Cond Res* 2005;19:495-504.
22. Dhuering MD, Feldmann CR, Ebben WP. Strength and conditioning practices of United States high school strength and conditioning coaches. *J Strength Cond Res* 2009;23:2188-2203.
23. Ebben WP, Suchomel TJ. Physical demands, injuries, and conditioning practices of stock car drivers. *J Strength Cond Res* 2012;26:1188-1198.
24. Buckner SL, Kuehne TE, Yitzchaki N, et al. The generality of strength adaptation. *J Trainol* 2019;8:5-8.
25. Sala G, Gobet F. Cognitive training does not enhance general cognition. *Trends Cogn Sci* 2019;23:9-20.
26. Pearl J, MacKenzie D. *The Book of Why: The New Science of Cause and Effect*. Penguin Random House, 2018.
27. Beauchamp MK, Lee A, Ward RF, et al. Do Exercise Interventions Improve Participation in Life Roles in Older Adults? A Systematic Review and Meta-Analysis. *Phys Ther* 2017;97:964-974.
28. Thorndike EL. On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them. *Am J Psychol* 1939;52:122-124.
29. Robinson WS. Ecological correlations and the behaviour of individuals. *Am Sociol Rev* 1950;15:351-357.
30. Vigotsky AD, Schoenfeld BJ, Than C, et al. Methods matter: the relationship between strength and hypertrophy depends on methods of measurement and analysis. *PeerJ* 2018;6:e5071.
31. Beck TW. The importance of a priori sample size estimation in strength and conditioning research. *J Strength Cond Res* 2013;27:2323-2337.
32. Ioannidis JPA. Why most discovered true associations are inflated. *Epidemiology* 2008;19:640-648.
33. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013;14:365-376.
34. Zhang Z, Xu X, Ni H. Small studies may overestimate the effect sizes in critical care meta-analyses: a meta-epidemiological study. *Crit Care* 2013;17:R2.
35. Caldwell AR, Vigotsky AD, Tenan MS, et al. Moving sport and exercise science forward: A call for the adoption of more transparent research practices. *Sports Med* 2020 (published online ahead of print).
36. Cumming G. The new statistics: why and how. *Psychol Sci* 2014;25:7-29.
37. Wilkinson M, Winter EM. Estimation versus falsification approaches in sport and exercise science. *J Sport Sci* 2019;37:3-4.
38. Hanel PHP, Mehler DMA. Beyond statistical significance: Which effect sizes are informative to improve scientific communication. *Public Underst Sci* 2019;28:468-485.
39. McPhetres J, Pennycock G. Lay people are unimpressed by the effect sizes typically reported in psychological science. *PsyArXiv.* April 2. doi:10.31234/osf.io/qu9hn.
40. Dankel SJ, Mouser JG, Mattocks KT, et al. The widespread misuse of effects sizes. *J Sci Med Sport* 2017;20:446-450.
41. Cohen J. *Statistical Power Analysis for the Behavioural Sciences*. 2nd ed. Lawrence Erlbaum Associates, 1988.
42. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs. *Front Psychol* 2013;4:863.
43. Griffiths B, Grant J, Langdown L, et al. The effects of in-season traditional and explosive resistance training programs on strength, jump height, and speed in recreational soccer players. *Res Q Exerc Sport* 2019;90:95-102.
44. Styles W, Matthews M, Comfort P. Effects of strength training on squat and sprint performance in soccer players. *J Strength Cond Res* 2016;30:1534-1539.
45. Bimson L, Langdown L, Fisher JP, et al. Six weeks of knee extensor isometric training improves soccer related skills in female soccer players. *J Trainol* 2017;6:52-56.
46. Hughes MD, Bartlett RM. The use of performance indicators in performance analysis. *J Sports Sci* 2002;20:739-754.
47. Jolly S. A critical view of critical velocity testing to predict performance. *J Sports Sci* 2013;31:688-689.
48. Gonçalves B, Coutinho D, Travassos B, et al. Speed synchronization, physical workload and match-to-match performance variation of elite football players. *PLoS One* 2018;13:e0200019.
49. Liu H, Gómez MA, Gonçalves B, et al. Technical performance and match-to match variation in elite football teams. *J Sport Sci* 2016;34:509-518.
50. Bucheit M. Chasing the 0.2. *Int J Sports Physiol Perform* 2016;11:417-418.
51. O'Connell NS, Dai L, Jiang Y, et al. Methods for analysis of pre-post data in clinical research: A comparison of five common methods. *J Biom Biostat* 2017;8:1-8.
52. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow-up measurements. *BMJ* 2001;323:1123.
53. Sainani KL, Lohse KR, Jones PR, et al. Magnitude-based inference is not Bayesian and is not a valid method of inference. *Scand J Med Sci Sports* 2019;29:1428-1436.
54. Mengersen KL, Drovandi CC, Robert CP, et al. Bayesian estimation of small effects in exercise and sports science. *PLoS One* 2016;11:e0147311.
55. Kruschke JK, Liddell TM. The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychon Bull Rev* 2015;25:178-206.
56. Kassler D, Nichols-Barrer I, Finucane M. Beyond "treatment versus control": How Bayesian analysis makes factorial experiments feasible in education research. *Eval Rev* 2019 (published online ahead of print).
57. Stapleton C, Scott MA, Atkinson G. The 'so what' factor: statistical versus clinical significance. *Int J Sports Med* 2009;30:773-774.
58. Bernards JR, Sato K, Haff GG, et al. Current research and statistical practices in sport science and a need for a change. *Sports* 2017;5:87.

59. Cohen J. A power primer. *Psychol Bull* 1992;112:155-159.
60. Hopkins WG. How to interpret changes in an athletic performance test. *Sportscience* 2004;8:1-7.
61. Ostelo RWJG, Deyo RA, Straford P, et al. Interpreting change scores for pain and functional status in low back pain: Towards international consensus regarding minimal important change. *Spine* 2008;33:90-84.
62. Anvari F, Lakens D. Using anchor-based methods to determine the smallest effect size of interest. *PsyArχiv* 2019; https://psyarxiv.com/syp5a/
63. Feltz DL, Lirgg CD. Perceived team and player efficacy in hockey. *J Apply Psychol* 1998;83:557-564.
64. Kyprianou E, Lolli L, Haddad HA, et al. A novel approach to assessing validity in sports performance research: integrating expert practitioner opinion into the statistical analysis. *Sci Med Football* 2019;3:333-338.
65. Morris DE, Oakley JE, Crowe JA. A web-based tool for eliciting probability distributions from experts. *Environ Model Softw* 2014;52:1-4.
66. Hernán M, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758-764.
67. Nuzzo JL, Finn HT, Herbert RD. Causal mediation analysis could resolve whether training-induced increases in muscle strength are mediated by muscle hypertrophy. *Sports Med* 2019;49:1309-1315.
68. Lynch BM, Dixon-Suen SC, Ramirez Varela A, et al. Approaches to improve causal inference in physical activity epidemiology. *J Phys Act Health* 2019 (published online ahead of print).
69. Pearl J. The foundations of causal inference. *Sociol Methodol* 2010;40:75-149.
70. VanderWeele TJ. Mediation analysis: A practitioner's guide. *Annu Rev Public Health* 2016;37:17-32.
71. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol* 2008;8:70.
72. Attia JR, Jones MP, Hure A. Deconfounding confounding part 1: traditional explanations. *Med J Aust* 2017;206:244-245.
73. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37-48.
74. Rohrer JM. Thinking clearly about correlations and causation: Graphical causal models for observational data. *Adv Methods Pract Psychol Sci* 2018;1:27-42.
75. Binney Z. Offensive personnel vs. men in the box: A football causality tutorial. *Football Outsiders* 2019; Available: https://www.footballoutsiders.com/stat-analysis/2019/offensive-personnel-men-box
76. Case MJ, Knudson DV, Downey DL. Barbell squat relative strength as an identifier for lower extremity injury in collegiate athletes. *J Strength Cond Res* 2020 (published online ahead of print).
77. Lauersen JB, Bertelsen DM, Andersen LB. The effectiveness of exercise interventions to prevent sports injuries: a systematic review and meta-analysis of randomised controlled trials. *Br J Sports Med* 2014;48:871-877.
78. Lauersen JB, Andersen TE, Andersen LB. Strength training as superior, dose-dependent and safe prevention of acute and overuse sports injuries: a systematic review, qualitative analysis and meta-analysis. *Br J Sports Med* 2018;52:1557-1563.
79. Fleck SJ, Falkel JE. Value of resistance training for the reduction of sports injuries. *Sports Med* 1986;3:61-68.
80. Fisher J, Steele J, Bruce-Low S, Smith D. Evidence-based resistance training recommendations. *Med Sport* 2011;15:47-62.
81. Macnamara BN, Moreau D, Hambrick DZ. The relationship between deliberate practice and performance in sports: A meta-analysis. *Perspect Psychol* Sci 2016;11:333-350.
82. Ericsson KA. Summing up hours of any type of practice versus identifying optimal practice activities: Commentary on Macnamara, Moreau. & Hambrick (2016). *Perspect Psychol Sci* 2016;11:351-354.