International Journal of Sports Physiology and Performance, 2018, 13, 127-134
https://doi.org/10.1123/ijspp.2017-0322

Human Kinetics
BRIEF REVIEW

Strengthening the Practice of Exercise and Sport-Science Research

Israel Halperin, Andrew D. Vigotsky, Carl Foster, and David B. Pyne

Exercise and sport sciences continue to grow as a collective set of disciplines investigating a broad array of basic and applied research questions. Despite the progress, there is room for improvement. A number of problems pertaining to reliability and validity of research practices hinder advancement and the potential impact of the field. These problems include inadequate validation of surrogate outcomes, too few longitudinal and replication studies, limited reporting of null or trivial results, and insufficient scientific transparency. The purpose of this review is to discuss these problems as they pertain to exercise and sport sciences based on their treatment in other disciplines, namely psychology and medicine, and to propose a number of solutions and recommendations.

Keywords: methodology, replication, null results

Over the passing years, exercise and sport sciences have developed into a large field of study consisting of several disciplines including physiology, biomechanics, psychology, nutrition, performance analysis, motor learning and control, strength and conditioning, and sports medicine. Much like biomedical sciences, exercise and sport sciences serve to inform practitioners. This parallel approach allows exercise scientists to learn from the medical model of research and application. Many of the mistakes made by biomedical researchers also appear to apply to exercise science. These mistakes cover problems from shortcomings in the design of research studies to the publication process and translation of results. Undoubtedly, it is the role of scientists to provide usable and applicable information to practitioners. However, our ability is limited if work is biased, opaque, and esoteric.

Despite the constant growth of exercise and sport sciences, there are a number of methodological problems concerning common research designs and practices that hinder the impact of research. These problems include but are not limited to inadequate validation of surrogate outcomes, too few longitudinal and replication studies, limited reporting of null or trivial results, and insufficient scientific transparency. The purpose of this review is to discuss these problems as they pertain to exercise and sports sciences and related fields such as physical therapy and sports medicine. A number of solutions are offered, some of which are practical and others more theoretical.

While discussion of problematic research practices has already taken place in exercise sciences and related fields,[1–6] this review differs in a number of ways. First, whereas different methodological problems are frequently discussed individually in separate articles, in this review we examine them as part of a bigger issue, including their potential interactions. Second, for the most part, previous articles on methodological problems in exercise and sport sciences have focused on statistical and power analyses.[1–6] The present review focuses on other, less discussed and acknowledged problems. Third, we examine these issues on a conceptual and practical level for researchers and practitioners, rather than taking a technical (and more complex) approach. By doing so, we hope to reach a broader audience, such as coaches and practitioners. Finally, in this review, we draw heavily on literature from neighboring disciplines–psychology and medicine—that have struggled with validity and reliability problems for an extended period of time and have developed effective strategies for dealing with them.[7–11] It is our belief that it is in the best interest of the exercise and sport sciences to learn from the mistakes of these other disciplines. Solving these problems will not be an easy task and will most likely take time, collaborative effort, and creative solutions. However, discussing and acknowledging them is an important step in the right direction.

Inadequate Validation of Surrogate Outcomes

Problem

A *surrogate outcome* or *end point* is a term borrowed from the medical fields, referring to a laboratory measurement used in therapeutic trials as a substitute for a clinically meaningful end point that quantifies how a subject feels, functions, or survives.[12] Note that changes induced by a therapy on a surrogate end point are expected to reflect changes in clinically meaningful end points.[12] Since surrogate outcomes are not clinically meaningful end points, they must be validated against those that are.[13,14] The validation procedure requires evidence showing that effects on the surrogate outcome can reliably predict effects on one or more clinically meaningful end points.[13,15] In medicine, surrogate validation processes are long and extensive and usually require a multilayer sequence of studies before; for example, the US Food and Drug Administration (FDA) approves a surrogate outcome as an adequate replacement for a clinical end point.[14]

In exercise and sport, hundreds of studies have been published that rely on surrogate outcomes that have not been adequately validated against meaningful, relevant outcome measures (eg, performance). For example, studies comparing the effects of various

Halperin and Pyne are with the Physiology Discipline, Australian Inst of Sport, Canberra, ACT, Australia. Halperin is with the Centre for Exercise and Sport Science Research, Edith Cowan University, Joondalup, WA, Australia. Vigotsky is with the Dept of Biomedical Engineering, Northwestern University, Evanston, IL. Foster is with the Dept of Exercise and Sport Science, University of Wisconsin–La Crosse, La Crosse, WI. Halperin (Israel.Halperin@ausport.gov.au) is corresponding author.

exercises on electromyography (EMG) amplitudes of different muscle groups are a popular study design.[16,17] It is speculated that exercises eliciting greater EMG amplitudes are superior (for an outcome of interest) than those eliciting lower amplitudes.[18] However, it is unclear whether exercises eliciting greater EMG amplitudes will necessarily lead to meaningful, superior outcomes such as muscle hypertrophy or strength.[19] Given the lack of robust longitudinal validation studies of this surrogate outcome, we do not know the answers to these important questions. Speculating that greater EMG amplitudes lead to a meaningful outcome solely based on possible physiological underpinnings is not enough. This issue has been frequently demonstrated in the medical fields, in which surrogate outcomes were deemed ineffective in predicting a clinical outcome despite a seemingly valid physiological rationale (for a powerful illustration, readers are encouraged to read the Cardiac Arrhythmia Suppression Trial[20]). While EMG is a frequently used surrogate outcome in exercise studies, other measures such as postexercise circulating hormonal levels[21,22] and muscle protein synthesis[23,24] have also been employed.

Solution

Longitudinal studies investigating the validity of commonly used surrogate outcomes in exercise and sport science are warranted. While difficult to conduct, such studies should have a substantial impact on the field, as their results could refute or confirm the conclusions of hundreds of such studies, as well as the need to continue conducting them. Until or unless common methods or approaches are validated, we urge scientists to be cautious on the degree of inference concerning surrogate outcomes. Stating, for example, that exercises that elicit greater EMG amplitudes are better than those eliciting lower EMG is premature and may lead to unwarranted conclusions. Scientists should avoid heavy use of a technology until its predictive validity has been established and subsequent implications are fully understood.

Too Few Longitudinal Studies

Problem

Most studies in the exercise and sport sciences are of short duration, usually taking place over a few days or weeks rather than months or a season. Ideally, exercise guidelines provided by governing bodies such as the National Strength and Conditioning Association (NSCA) and American College of Sports Medicine (ACSM) should be based on a large number of studies investigating the effects of various interventions over a longer, rather than shorter, duration. The reason is that longer-duration studies have a higher degree of external validity (ie, the extent to which the results of a study can be generalized to other situations).[25,26] Longer studies mimic real-life scenarios to a greater extent than shorter ones.[25] Longer studies also have a greater degree of internal validity (ie, the degree of confidence that can be placed in the causal relationship between the intervention and the outcome). Novel interventions can affect performance in the short term yet may not lead to lasting, meaningful effects once the novel aspect vanishes and participants grow accustomed to the training intervention. Alternatively, effects can reach an early plateau. Whether the measured effects are due to novel aspects of an intervention or actual superiority can only be determined by extending the duration of the study.

Novel resistance-training programs (eg, undulating periodization) can lead to initial favorable adaptions compared with a routine program (eg, linear periodization).[27,28] The favorable initial outcomes identified with novel programs are not necessarily a result of their inherent superiority but, rather, of variations they introduce compared with more-routine programs.[27,28] Over time, the positive effects associated with such programs may diminish, leading to different conclusions about their effectiveness. This effect is illustrated in a study by Rhea et al,[29] in which resistance-trained participants were randomized into a daily undulating-periodization program (altering training variables on a daily basis) or a linear-periodization program (altering training variables on a weekly to monthly basis). It was noteworthy that all participants reported following a variation of a linear-periodization program prior to initiation of the study. Thus, they were familiar with training in a certain way. Participants following the daily undulating-periodization program improved strength to a greater extent than those following the linear-periodization program in the first 6 weeks. However, the positive effects diminished in the last 6 weeks of the study, as no statistically significant or meaningful differences were identified between the 2 groups.[29] It is likely that the initial improvements were due to the novel stimulus, high expectations, and/or effects on self-efficacy rather than an inherent superiority of the program. It is possible that a different conclusion would have been drawn if the study had lasted 6 rather than 12 weeks.

Manipulating and measuring the effect of various types of feedback on performance is another research avenue that would benefit from longitudinal studies. Specifically, more than 100 acute studies have been published on the topic of attentional focus in the past 20 years, comparing external and internal focus-of-attention instructions.[30] External focus of attention refers to instructing an individual to focus on the effects of a movement in relation to the environment, for example, instructing a person to focus on pushing a bar while completing a set of heavy squats. On the other hand, internal focus of attention refers to instructing an individual to focus on a specific body part or muscle group during the physical task, for example, instructing a person to focus on contracting the quadriceps muscles while completing a set of heavy squats. The majority of such studies report superior performance with external- compared with internal-focus instructions.[30] However, typically these studies employed short-term acute interventions.[30] Given that sport and exercise coaches tend to use internal-focus instructions more than external ones,[31,32] there is a possibility that the positive effects observed with external-focus instructions stem from their novelty. A longitudinal study investigating whether positive effects persist over time would benefit this area of research.

Solution

The simple, logical solution to the lack of longitudinal data is to conduct more longitudinal research. However, we are well aware of the difficulties in completing such studies. They are expensive, require a lot of time and resources, and, perhaps most important, they seem to receive the same weight in terms of scientific “impact” as short-term studies. Hence, exercise scientists are not often rewarded for their efforts. We believe that this is an important consideration, because without a worthwhile incentive, researchers understandably choose to conduct a short-term study rather than a long-term one. This is especially the case if, according to traditional publication metrics (publication count rather than type), short- and long-term studies carry equal weight. This is not to say that longitudinal research is inherently superior to short-term studies. However, everything else being equal, a longitudinal study is more informative and has a greater degree of internal validity, given the possibility of controlling for more confounders.[25,26] Moreover, longitudinal studies also have a greater degree of external validity, given their similarities to real-life scenarios.[25,26]

Some ways to encourage more longitudinal research include additional or targeted funding (intramural or extramural) for the addition of payment or other incentives to maintain subject compliance and involvement while limiting dropouts. Efforts to come up with creative time-tabling to ensure that longitudinal studies fit the sports' or coaches' requirements and subject availability should also increase willingness to participate and limit dropout rates. A crossover design, which reduces the number of subjects required as part of sample-size estimation, could increase the feasibility of conducting longitudinal studies. Finally, involving or embedding the researchers with the athletes or team to develop closer rapport and compliance would likely increase their willingness to participate in such studies. While sport scientists generate excellent questions concerning the effectiveness of various training inventions, the real-world questions articulated by coaches and practitioners would make them even better. The external and ecological validity of such questions would naturally be higher, and, most important, the likelihood of an effective collaboration between scientists, coaches, and athletes would increase substantially.

Reporting Nonsignificant or Trivial Results

Problem

Scientists across most fields are directly and/or indirectly encouraged to publish positive rather than negative results.[33,34] That is, they are encouraged to report that an effect is positive rather than negative or absent.[35] This practice results in a disproportionately high ratio of positive to negative outcomes published in scientific journals, and this ratio is apparently increasing with each passing year.[35] A critical problem with this practice is that it creates a false perception of "truth."[36] Whereas one of the key roles of scientists is to investigate and report how the world (in our case, exercise and/or sporting performance) works in the most objective way possible, selectively reporting positive results can lead to a distorted perception of reality.[36] This positive publication bias, which has been demonstrated in a number of disciplines, hinders the reputation of the scientific method and raises questions pertaining to the underlying rigor and credibility of science.[34,36,37] With regard to sport and exercise, we imagine that such practices influence the degree of trust that coaches and practitioners are willing to put into the research output of exercise scientists.

Positive publication bias also wastes important resources such as time and funding committed to explore the effect of an intervention. Such effects may have already been deemed "nonsignificant" or trivial on numerous occasions, but the results were never published.[34,38] This bias encourages scientists to generate questions that are biased toward positive results to increase their chance of publication. That is, when designing a study, scientists may either consciously or unconsciously employ a design that makes it easier to find an effect, often at the expense of external validity, for example, exaggerating the dose in an intervention with the goal of finding an effect while departing from what commonly takes place in practice. This habit makes scientific output less relevant for practitioners. In more-extreme circumstances, some scientists may be tempted to manipulate their data to find a positive effect[33,39] or, alternatively, change their original hypotheses (aim or research question) according to their findings in an attempt to present the results as positive (also known as HARKing [Hypothesizing After the Results are Known]).[40] Collectively, positive publication bias hinders scientific progress and worthwhile outcomes for the general community.[33,39]

There are a number of explanations for why negative results do not get published as often. Scientists may prefer not to report or attempt to publish them, which is known as the file-drawer problem.[38,41] This action could stem from a fear that their nonsignificant or trivial results are wrong or unsuccessful and, as a result, leads to low publication potential, reluctance to upset the status quo, unwillingness to publish negative results against a theoretical model in which researchers are invested, perceived pressure from funding agencies looking for positive effects, and the desire to complete academic duties (eg, PhD completion).[38,41] Authors may also decide against attempting to publish nonsignificant results because leading journals have a high rejection rate of negative results.[35,42] Indeed, nonsignificant results are more difficult to publish and seem to suffer from an unjustified perception of inferiority compared with positive results.[35,42] Scientists may prefer to channel their limited resources to other projects that are more likely to be published. We fear that the exercise and sport sciences are no exception to this practice.

Solution

There are number of possibilities to counter the problem of publication bias. An initiative to realign and reestablish the status and importance of nonsignificant and trivial results in all of science, with exercise science being no exception, should be developed.[36,41] The issue and potential solutions need to be routinely discussed in the classroom, graduate studies, and laboratories. Role-leading academics and sport-science practitioners need to discuss the background and consequences of publication bias and emphasize the importance of transparent and even-handed reporting.[43] The second possible solution is registered reports of rationale, research design, and methods prior to experimentation. Briefly, the concept of registered reports involves submission of a proposed rationale (to establish that a study needs to be done) and research design (to document that the experimental question is appropriately addressed), which are reviewed before the study is conducted rather than after it was completed as is commonly done with the current publication model.[41,43,44] Provided the proposed research design is accepted by the reviewers and the study conducted according to the proposal, the journal essentially guarantees publication of the paper, irrespective of the results.[41,43,44]

Variations of this publication model are growing rapidly in different fields including medicine,[45] psychology,[44] and neuroscience.[41] Notably, there are early signs of registered reports in the exercise sciences.[46] This model has several clear benefits. First, scientists do not feel as pressured to report positive results, provided they follow their proposal. Second, registered reports reduce the so-called researcher's degrees of freedom, or the decision on how to analyze the data both before and after the data-collection phase, which allows scientists to implement an analysis that favors the positive rather than nonsignificant or trivial results.[33] Third, by committing to an analysis beforehand, the effects of various biased practices such as HARKing or *p*-hacking should be reduced substantially.[41] Finally, the number of nonsignificant or trivial results in clinical trials has grown substantially since registered reports have been incorporated.[47] While this solution is not perfect and does not fit all types of research questions, we believe that it is a model worth adopting in the field of exercise and sport science. Another strategy, piloted by the *BMC Psychology* journal, is "results-free" peer review, in which reviewers are asked to review a study without knowing what the results are and provisionally accept or reject the study based on the background and methods alone.[48] If accepted, the results and discussion sections are reviewed to check for proper analysis and interpretation of the data and for other minor revisions. This peer-review approach is expected to considerably reduce positive-results bias.

This review style could serve as an interim strategy until the necessary steps have been taken to switch over to the more rigorous option of registered reports.

Too Few Replication Attempts

Problem

Replication of experiments is at the heart of science.[8,34,49] Replication allows for confirmation or refutation of outcomes, exploring the boundaries of theories, and, ultimately, the progression of science.[8,34,49] One approach involves the division of replication into direct and conceptual studies.[8,50] With direct replication, researchers repeat the methods of the original study as closely as possible.[49,50] Direct replications serve to validate the results and inspect their reliability, with the goal of increasing or reducing the degree of confidence in the originally reported results.[49,50] Conceptual replication, on the other hand, investigates the boundaries of the theory assumed to be accurate.[8,51] In other words, conceptual replication seeks to validate the underlying theory rather than results.[51] With conceptual replication, one or more of the variables are intentionally modified or changed. By doing so, it is assumed implicitly that the original findings are reliable.[50] As a result, conceptual replication studies cannot refute the original results being replicated.[8,50]

While disagreements persist on the best strategies for replication,[52] it is generally agreed that direct replication is a prerequisite to conceptual replication.[8,49,50] That is, only after confidence in the reliability of an effect is achieved should one explore its boundaries. Despite the general acceptance concerning their importance, until recently few direct replications have been pursued in most scientific disciplines.[53,54] This shortcoming may relate to journals' preference for novel results and not replications, scientists preferring to investigate topics of personal interest rather than repeating someone else's work, and fear of being perceived as hostile toward the original researcher.[34] The growing alarm pertaining to the lack of replication attempts in psychology has led to development of the Open Science Collaboration (OSC), which set a goal of conducting large-scale, multicenter, preregistered direct-replication attempts.[11] By 2015, 100 psychological studies originally published in 2008 had been directly replicated.[54] Whereas 97% of the original studies reported statistically significant results, only 36% of the replications had the same outcome.[54] In addition, the effect sizes were, on average, half as large as those reported in the original studies.[54] Comparable results are now emerging in a replication project in cancer biology.[53] Hence, the term *replication crisis* has been used to describe the current state of medical and social sciences.[7,8]

Inconsistent results could stem from a number of possibilities. For example, the original or replicated outcomes were due to chance, or, alternatively, there may be subtle differences in the investigated cohorts and/or testing environments.[8,53,54] Hence, no replication can completely confirm or refute an effect but, rather, adds or subtracts from the degree of confidence in the original finding(s).[8,53,54] Despite some worrisome results, the replication process has powerful scientific value.[50,53] Replication facilitates a deeper understanding of which effects are robust and consistent and lead to better use of limited resources, as only repeatable data will be used as a platform on which to build.[8,11,34] Fortunately, other disciplines are joining the OSC with the aim of conducting similar replication processes,[53] and journals are gradually becoming more receptive to publishing replication studies.[55,56] Given the well-deserved attention this important topic is receiving in other fields, we hope it will encourage exercise scientists to follow suit.

Solutions

First, like most other problems discussed in this review, drawing attention to and acknowledging the necessity of replication is an essential initial step. The impressive progress achieved by the OSC should influence scientists' perception of the importance and feasibility of conducting replication studies. Other disciplines are joining the OSC with similar goals, likely increasing the appreciation that novelty needs to be balanced with confirmation. Second, journal policies (and consequently editor, associate editor, and peer-reviewer attitudes) will have to change and become more receptive to replication, especially direct replication studies. This outcome can be facilitated by allocating space or special sections for a given number of replications per journal volume.[9] Early signs of this change are appearing in psychology and biomedicine journals,[55,56] but what about exercise and sport sciences? Third, replication could also be part of formal academic training. For example, replication could be discussed as part of a PhD plan or used to complete MSc theses.[9]

Insufficient Scientific Transparency

Problem

Generally, the term *open science* refers to activities designed to make the scientific process transparent and accessible.[57,58] This approach includes sharing research materials, data, exact analysis, workflow, and more.[59] Sharing research materials allows others to build on prior work, conduct robust meta-analyses, reanalyze and interpret results based on different statistical tests, control for errors, limit fraud, provide directions for replication, and investigate data in view of answering different questions.[57,58] Despite these clear benefits, data sharing is not a requirement of most exercise journals, and scientists across disciplines are not eager to share their data.[60] This disconnect can be explained by a number of factors. First, journals still employ word or page limits due to the expense of publication, which prohibits full disclosure of materials.[57] Second, the systems do not incentivize open practices. Whereas scientists are rewarded for positive and "clean" results, raw data can be messy and unclear.[33,57] Researchers may use only a subset of results that, overall, show mixed or unclear results, and sharing the full data set may bring into question their analysis and interpretation.[57] Scientists may also be hesitant to share collected data in fear that they will be used by others without proper attribution.[33,57]

Sharing data is particularly important in exercise science given the typically smaller sample sizes[5] and large interindividual responses.[61] Indeed, mean results, commonly used for statistical analysis and reporting, can be misleading in studies associated with large variability, especially when coupled with small sample sizes. For instance, despite a statistically significant group mean effect in which participants improved their $\dot{V}O_2$max in response to a similar training intervention, very large variability was recorded between participants, some of whom improved their $\dot{V}O_2$max by 100%, whereas others did not improve at all.[61] Furthermore, outliers could affect the results with relative ease in cases where small samples are investigated, such as elite athletes or participants with distinctive injuries. Hence, sharing data could help researchers examine individual responses to an intervention, in addition to the mean results, to better use the data for different questions and/or analyses.

Solutions

From a journal's perspective, requiring authors to submit research materials is an important step. Whereas word or page limits were mandatory in the past due to fees associated with paper publication, in the current digital age, uploading supplementary files with materials should not come with additional expense; in fact, such practices should be encouraged. Indeed, many journals from various fields now require authors to upload research materials with their submitted articles.[58] Another avenue encouraging open science comes from the Peer Reviewers' Openness Initiative, which is a statement researchers can sign indicating that they will refuse to conduct peer review unless data are made available.[58] Scientists should also understand that data sharing leads to greater citation rates than not sharing articles, which should increase researchers' incentive to share research materials.[59] Moreover, researchers diligently collect data, and it is fair to assume that they want to receive credit and acknowledgement when their data are used by others. Thus, developing norms for citing shared data should not just reduce the apprehension of researchers to share their work but even encourage it.[57] Another interesting strategy is to reward scientists for desirable behaviors with "badges" offered by journals, by acknowledging open practices and for following required criteria.[10] While still in its early stages, evidence demonstrating the effectiveness of this strategy is accumulating quickly.[10] For example, since 2014, when *Psychological Science* announced it would award badges for data-sharing behaviors, the average data-sharing rate increased tenfold to 38% from 2013 to 2015.[10]

Graphical presentation of numeric data is often preferred to large tables or an overload of text. However, authors should limit their use of bar graphs, as they tend to hide the shape of the distributions and presence of outliers and, accordingly, lead readers to assume a normal distribution.[62,63] This is especially the case with small sample sizes, in which outliers can substantially affect the mean.[62] Alternatively, the most transparent way to present results is with scatter plots representing the response of each individual. This option is especially appropriate for smaller samples. Box plots, violin plots, and histograms are also good options, as they allow for an appreciation of the distribution and existence of outliers.[62,63] While not as "clean" as bar graphs, the alternatives are more informative and transparent and should be encouraged by academics and journals alike (for examples, see Figure 1).

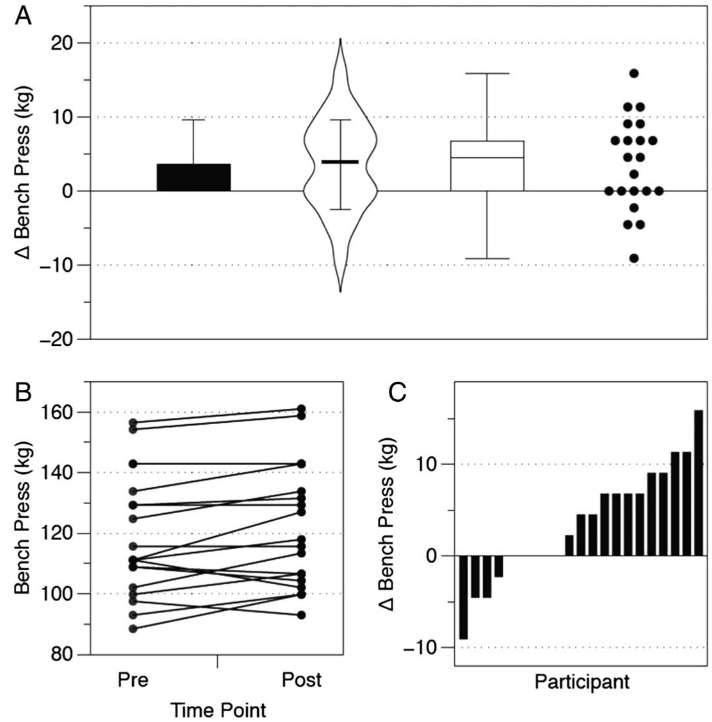

Figure 1 — Illustration of graphing options. (A) Different ways to represent group data. In this case, changes from baseline are plotted. On the left is a standard bar graph, mean ± SD, which may hide potentially important variability. Second from the left is a violin plot with mean ± SD contained within. The shape of the violin plot represents the probability density, wherein one is more likely to see a point fall in thicker parts of the plot. Second from the right is a standard box-and-whisker plot, which is useful for depicting nonparametric data, as it uses the median, range, and interquartile range rather than mean ± SD to depict variability. On the right are individual points, allowing one to observe exactly how data are distributed. (B) Individual responses to an intervention to illustrate whether there are any relationships pertaining to responses to an intervention; for example, do subjects who start with lower values exhibit larger increases? (C) Individual change scores for every participant, which allow one to appreciate the heterogeneity of responses to an intervention. Data taken from Schoenfeld et al,[64] used under CC-BY 4 (https://creativecommons.org/licenses/by/4.0/).

General Discussion

Sport performance and sport science can be enhanced by translation of study outcomes from a broad range of related scientific and medical disciplines. We have introduced and discussed a number of potential threats to the growth and impact of exercise and sport sciences and proposed relevant solutions (see Table 1 for general summary). We relied on literature from other disciplines, namely psychology and medicine, that have gone through, and are still going through, substantial changes given identification and management of these problems.[7,11,36,54] We have not investigated the extent of these problems in sport and exercise sciences, but given their prevalence in related fields with many similarities in their research designs, we think that they offer valuable insights for researchers and sport-science practitioners. Thus, it would be better to acknowledge and act on them as soon as possible to explore potential applications in research and sport-science activities.

Ultimately, implementing the proposed recommendations depends on challenging and changing the culture and contemporary practices of sport and exercise sciences. From the publishing perspective, policies will have to evolve and be modified.[9,55] Journals need to become more accepting of replication studies. If this is done, scientists will feel more confident conducting replication studies knowing that they are not "inferior" and could be published.[9,11] A balance is needed between novelty or original research and confirmation research for progression in a scientific field. Otherwise, bricks will continue to be laid over a potentially unreliable foundation.[8,50]

Table 1 Summary of Contemporary Problems and Possible Solutions to Enhance Sport and Exercise Science

	Problems	Solutions
Inadequate validation of surrogate outcomes	• Not clear if associated with a meaningful outcome • Misleading	• Validate against meaningful outcome(s) • Explicitly state their status and justify reason for using them
Too few longitudinal studies	• Acute studies suffer from lower external and internal validity	• Reward long-term studies • Collaborate • Award badges of excellence • Provide dedicated space in journals
Reporting nonsignificant or trivial results	• False perception of truth • Distorts knowledge	• Registered reports • Blind-results peer review • Change negative perception of null results • Award badges of excellence
Too few replication attempts	• Difficult to conclude if original results are due to chance or bias. • Building on shaky ground	• Reward replications, especially direct ones • Include in academic training • Create space for replication studies in journals
Insufficient scientific transparency	• Prohibits proper replications, meta-analysis, and deeper investigation of data	• Change journal polices • Provide raw data and detailed analytical procedures • Change citation practices • Reward data sharing

Journals also need to become more accepting of trivial or nonsignificant outcomes.[35,37,42] Good science should not be defined by studies' results but, rather, on the underlying questions, quality of the methods and analyses, and the likely impact of the outcomes.[34,36] Given that most journals prefer publishing positive rather than trivial results,[42] scientists have been encouraged to search for novel/positive results at the expense of relevant and important (real-world) questions.[36] Moreover, chasing statistically significant results could encourage scientists to engage in inappropriate scientific behaviors such as *p*-hacking, needlessly excluding outliers, and even fraud.[33] Similar to replication, this problem can be solved by accepting trivial and nonsignificant results more frequently and working toward changing the negative perception of null results in scientific culture.[42] Avoiding or limiting trivial results will distort the effective real-world solutions scientists are seeking to identify.[35,37] In addition to a cultural shift, implementing registered reports and/or blinded results to peer reviews will be helpful in reducing the frequency of these negative occurrences.[41,43] Rewarding scientists for desirable behaviors with "badges" is also a novel and effective strategy.[10] Journal editors, associate editors, and especially peer reviewers (and thesis examiners) will need to be educated and up-skilled in these issues.

Journals will also need to develop clearer guidelines concerning the analysis sections of studies. Supplementary material, methods, raw data, and detailed analytical procedures of studies can be published online. First, online publication will lead to greater transparency, allowing others to reanalyze the results and conduct meta-analysis. Second, this approach provides the blueprint for robust direct replications.[9,11] Effect sizes tied to a meaningful real-life reference or threshold values together with confidence internals will provide more useful outcomes.[3] Grant-funding agencies will need to revise submission procedures that incorporate these elements. This means, for example, that replication, long-term, and surrogate validation studies should be properly incentivized and encouraged.

Finally, scientists themselves should work collaboratively to expose, acknowledge, and address these problems and develop ways to resolve them. Scientists should offer lectures and courses dedicated to these issues, expand the length and number of courses pertaining to methodology and statistics, include replication studies in academic training as part of MSc and PhD programs, address the issue of validating surrogate outcomes, and encourage journals and professional societies to modify and evolve publication and professional practices and culture. Despite the great progress that sport and exercise sciences has made as a discipline or group of disciplines, there is room for improvement. Acknowledging and developing awareness of challenges to publication and science is an important first step. Learning from neighboring disciplines that have already identified and confronted these issues could save precious time and resources and provide better service to coaches, athletes, and the sporting community.

References

1. Hopkins W, Marshall S, Batterham A, Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc*. 2009;41:3–13. PubMed doi:10.1249/MSS.0b013e31818cb278
2. Halperin I, Pyne DB, Martin DT. Threats to internal validity in exercise science: a review of overlooked confounding variables. *Int J Sports Physiol Perform*. 2015;10(7):823–829. PubMed doi:10.1123/ijspp.2014-0566
3. Andersen MB, McCullagh P, Wilson GJ. But what do the numbers really tell us?: arbitrary metrics and effect size reporting in sport psychology research. *J Sport Exerc Psychol*. 2007;29(5):664–672. PubMed doi:10.1123/jsep.29.5.664
4. Ivarsson A, Andersen MB, Stenling A, Johnson U, Lindwall M. Things we still haven't learned (so far). *J Sport Exerc Psychol*. 2015;37(4):449–461. PubMed doi:10.1123/jsep.2015-0015
5. 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
6. Ivarsson A, Andersen MB. What counts as "evidence" in evidence-based practice?: searching for some fire behind all the smoke. *J Sport Psychol Action*. 2016;7(1):11–22. doi:10.1080/21520704.2015.1123206
7. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124. doi:10.1371/journal.pmed.0020124

8. Earp BD, Trafimow D. Replication, falsification, and the crisis of confidence in social psychology. *Front Psychol*. 2015;6(621):1–11. PubMed doi:10.3389/fpsyg.2015.00621
9. Koole SL, Lakens D. Rewarding replications: a sure and simple way to improve psychological science. *Perspect Psychol Sci*. 2012; 7(6):608–614. PubMed doi:10.1177/1745691612462586
10. Kidwell MC, Lazarević LB, Baranski E, et al. Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. *PLoS Biol*. 2016;14(5):e1002456. PubMed. doi:10.1371/journal.pbio.1002456
11. Collaboration OS. An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspect Psychol Sci*. 2012;7(6):657–660. PubMed doi:10.1177/1745691612462588
12. Nimmo WS, Tucker GT. *Clinical Measurement in Drug Evaluation*. New York, NY: Wiley; 1995.
13. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Stat Med*. 2012;31(25):2973–2984. PubMed doi:10.1002/sim.5403
14. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. *Health Aff*. 2005;24(1):67–78. PubMed doi:10.1377/hlthaff.24.1.67
15. Aronson J. Biomarkers and surrogate endpoints. *Br J Clin Pharmacol*. 2005;59(5):491–494. PubMed doi:10.1111/j.1365-2125.2005.02435.x
16. Stastny P, Gołaś A, Blazek D, et al. A systematic review of surface electromyography analyses of the bench press movement task. *PLoS ONE*. 2017;12(2):e0171632. PubMed doi:10.1371/journal.pone.0171632
17. Clark DR, Lambert MI, Hunter AM. Muscle activation in the loaded free barbell squat: a brief review. *J Strength Cond Res*. 2012;26(4): 1169–1178. PubMed doi:10.1519/JSC.0b013e31822d533d
18. Journal of Orthopaedic & Sports Physical Therapy. Strengthening your hip muscles: some exercises may be better than others. *J Orthop Sports Phys Ther*. 2013;43(2):65. doi:10.2519/jospt.2013.0501
19. Halperin I, Vigotsky AD. The mind–muscle connection in resistance training: friend or foe? *Eur J Appl Physiol*. 2016;116(4):863–864. PubMed doi:10.1007/s00421-016-3341-y
20. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. *N Engl J Med*. 1991;324(12):781–788. PubMed doi:10.1056/NEJM199103213241201
21. Phillips SM. Strength and hypertrophy with resistance training: chasing a hormonal ghost. *Eur J Appl Physiol*. 2012;112(5):1981–1983. PubMed doi:10.1007/s00421-011-2148-0
22. Schoenfeld BJ. Postexercise hypertrophic adaptations: a reexamination of the hormone hypothesis and its applicability to resistance training program design. *J Strength Cond Res*. 2013;27(6): 1720–1730. PubMed doi:10.1519/JSC.0b013e31828ddd53
23. Damas F, Phillips S, Vechin FC, Ugrinowitsch C. A review of resistance training-induced changes in skeletal muscle protein synthesis and their contribution to hypertrophy. *Sports Med*. 2015;45(6): 801–807. PubMed doi:10.1007/s40279-015-0320-0
24. Atherton PJ, Miller BF, Burd NA, et al. Commentaries on viewpoint: what is the relationship between acute measure of muscle protein synthesis and changes in muscle mass? *J Appl Physiol*. 2015;118(4): 498–503. doi:10.1152/japplphysiol.01069.2014
25. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin Trials*. 2006;1(1):e9. doi:10.1371/journal.pctr.0010009
26. Pincus T, Stein C. Why randomized controlled clinical trials do not depict accurately long-term outcomes in rheumatoid arthritis: some explanations and suggestions for future studies. *Clin Exp Rheumatol*. 1996;15:S27–S28. doi:10.1007/BF03342657
27. Kiely J. Periodization paradigms in the 21st century: evidence-led or tradition-driven? *Int J Sports Physiol Perform*. 2012;7(3):242–250. PubMed doi:10.1123/ijspp.7.3.242
28. Harries SK, Lubans DR, Callister R. Systematic review and meta-analysis of linear and undulating periodized resistance training programs on muscular strength. *J Strength Cond Res*. 2015;29(4): 1113–1125. PubMed doi:10.1519/JSC.0000000000000712
29. Rhea MR, Ball SD, Phillips WT, Burkett LN. A comparison of linear and daily undulating periodized programs with equated volume and intensity for strength. *J Strength Cond Res*. 2002;16(2):250–255. PubMed
30. Wulf G. Attentional focus and motor learning: a review of 15 years. *Int J Sports Psychol*. 2013;6(1):77–104. doi:10.1080/1750984x.2012.723728
31. Porter J, Wu W, Partridge J. Focus of attention and verbal instructions: strategies of elite track and field coaches and athletes. *Sport Sci Rev*. 2010;19(4):77–89.
32. Halperin I, Chapman DW, Martin DT, Abbiss C, Wulf G. Coaching cues in amateur boxing: an analysis of ringside feedback provided between rounds of competition. *Psychol Sport Exerc*. 2016;25:44–50. doi:10.1016/j.psychsport.2016.04.003
33. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci*. 2011;22(11):1359–1366. PubMed doi:10.1177/0956797611417632
34. Forstmeier W, Wagenmakers EJ, Parker TH. Detecting and avoiding likely false-positive findings—a practical guide. *Biol Rev Camb Philos Soc*. 2016;92(4):1941–1968. PubMed doi:10.1111/brv.12315
35. Fanelli D. Negative results are disappearing from most disciplines and countries. *Scientometrics*. 2011;90(3):891–904. doi:10.1007/s11192-011-0494-7
36. Nosek BA, Spies JR, Motyl M. Scientific utopia II: restructuring incentives and practices to promote truth over publishability. *Perspect Psychol Sci*. 2012;7(6):615–631. doi:10.1177/1745691612459058
37. Fanelli D. "Positive" results increase down the hierarchy of the sciences. *PLoS ONE*. 2010;5(4):e10068. PubMed doi:10.1371/journal.pone.0010068
38. Franco A, Malhotra N, Simonovits G. Publication bias in the social sciences: unlocking the file drawer. *Science*. 2014;345(6203): 1502–1505. PubMed doi:10.1126/science.1255484
39. Fanelli D. How many scientists fabricate and falsify research?: a systematic review and meta-analysis of survey data. *PLoS ONE*. 2009;4(5):e5738. PubMed doi:10.1371/journal.pone.0005738
40. Kerr NL. HARKing: hypothesizing after the results are known. *Pers Soc Psychol Rev*. 1998;2(3):196–217. PubMed doi:10.1207/s15327957pspr0203_4
41. Chambers CD, Feredoes E, Muthukumaraswamy SD, Etchells P. Instead of "playing the game" it is time to change the rules: registered reports at AIMS Neuroscience and beyond. *AIMS Neurosci*. 2014; 1(1):4–17. doi:10.3934/Neuroscience.2014.1.4
42. Matosin N, Frank E, Engel M, Lum JS, Newell KA. Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture. *Dis Model Mech*. 2014;7(2): 171–173. PubMed doi:10.1242/dmm.015123
43. Chambers C. *Registered reports: a step change in scientific publishing*. Elsevier website. http://www.elsevier.com/reviewers-update/story/innovation-in-publishing/registered-reports-a-step-change-in-scientific-publishing; 2014.
44. Nosek BA, Lakens D. Registered reports. *Soc Psychol*. 2014;45(3): 137–141. doi:10.1027/1864-9335/a000192

45. Association WM. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191. doi:10.1001/jama.2013.281053
46. Dessing JC, Beek PJ. *Human Movement Science* adopts registered reports for hypothesis-driven research. *Hum Mov Sci*. 2015;44: A1–A2. doi:10.1016/j.humov.2015.09.011
47. Kaplan RM, Irvin VL. Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLoS ONE*. 2015;10(8): e0132382. PubMed doi:10.1371/journal.pone.0132382
48. Button KS, Bal L, Clark A, Shipley T. Preventing the ends from justifying the means: withholding results to address publication bias in peer-review. *BMC Psychol*. 2016;4(1):59. PubMed doi:10.1186/s40359-016-0167-7
49. Simons DJ. The value of direct replication. *Perspect Psychol Sci*. 2014;9(1):76–80. PubMed doi:10.1177/1745691613514755
50. Schmidt S. Shall we really do it again?: the powerful concept of replication is neglected in the social sciences. *Rev Gen Psychol*. 2009;13(2):90–100. doi:10.1037/a0015108
51. Pashler H, Harris CR. Is the replicability crisis overblown?: three arguments examined. *Perspect Psychol Sci*. 2012;7(6):531–536. PubMed doi:10.1177/1745691612463401
52. Stroebe W, Strack F. The alleged crisis and the illusion of exact replication. *Perspect Psychol Sci*. 2014;9(1):59–71. PubMed doi:10.1177/1745691613514450
53. Nosek BA, Errington TM. Making sense of replications. *Elife*. 2017; 6:e23383. PubMed doi:10.7554/eLife.23383
54. Collaboration OS. Estimating the reproducibility of psychological science. *Science*. 2015;349(6251):aac4716. doi:10.1126/science.aac4716
55. Editorial. Receptive to replication. *Nat Biotechnol*. 2013;31(11):943. PubMed doi:10.1038/nbt.2748
56. Pashler H, Wagenmakers EJ. Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? *Perspect Psychol Sci*. 2012;7(6):528–530. PubMed doi:10.1177/1745691612465253
57. Nosek BA, Bar-Anan Y. Scientific utopia: I: opening scientific communication. *Psychol Inq*. 2012;23(3):217–243. doi:10.1080/1047840X.2012.692215
58. Morey RD, Chambers CD, Etchells PJ, et al. The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review. *R Soc Open Sci*. 2016;3(1):150547. PubMed doi:10.1098/rsos.150547
59. McKiernan EC, Bourne PE, Brown CT, et al. How open science helps researchers succeed. *Elife*. 2016;5:e16800. PubMed doi:10.7554/eLife.16800
60. Vanpaemel W, Vermorgen M, Deriemaecker L, Storms G. Are we wasting a good crisis?: the availability of psychological research data after the storm. *Collabra Psychol*. 2015;1(1):1–5. doi:10.1525/collabra.13
61. Bouchard C, Rankinen T. Individual differences in response to regular physical activity. *Med Sci Sports Exerc*. 2001;33(6):S446–S451. doi:10.1097/00005768-200106001-00013
62. Weissgerber TL, Milic NM, Winham SJ, Garovic VD. Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol*. 2015;13(4):e1002128. doi:10.1371/journal.pbio.1002128
63. Saxon E. Beyond bar charts. *BMC Biol*. 2015;13(1):60. doi:10.1186/s12915-015-0169-6
64. Schoenfeld BJ, Aragon A, Wilborn C, Urbina SL, Hayward SE, Krieger J. Pre- versus post-exercise protein intake has similar effects on muscular adaptations. *Peer J*. 2017;5:e2825. doi:10.7717/peerj.2825