Hill, A. P. (2023). Comparisons and Conversions: A Methodological Note and Caution for Meta-Analysis in Sport and Exercise Psychology. *Journal of Sport & Exercise Psychology*. Accepted 10-08-2023 [Author Accepted Version].

Comparisons and Conversions: A Methodological Note and Caution for Meta-Analysis in Sport and Exercise Psychology

Andrew P. Hill

School of Science Technology and Health, York St. John University, UK

Graduate Department of Kinesiology, University of Toronto, Canada

Author Note

Correspondence concerning this article should be addressed to Andrew Hill, School of Science Technology and Health, York St. John University, Lord Mayor's Walk, York, YO31 7EX, UK. E-mail: a.hill@yorksj.ac.uk

Abstract

Meta-analysis is a powerful tool in sport and exercise psychology. However, it has a number of pitfalls, and some lead to ill-advised comparisons and overestimation of effects. The impetus for this research note is provided by a recent systematic review of meta-analyses that examined the correlates of sport performance and has fallen foul of some of the pitfalls. Although the systematic review potentially has great value for researchers and practitioners alike, it treats effects from correlational and intervention studies as yielding equivalent information, double-counts multiple studies, and uses an effect size for correlational studies (Cohen's d) that provides an extreme contrast of unclear practical relevance. These issues impact interpretability, bias, and usefulness of the findings. This methodological note explains each pitfall and illustrates use of an appropriate equivalent effect size for correlational studies (Mathur and VanderWeele's *d*) to help researchers avoid similar issues in future work

Keywords: meta-analysis, effect sizes, reviews

Comparisons and conversions: A methodological note and caution for meta-analysis in sport and exercise psychology

Meta-analysis is an increasingly popular tool in sport and exercise psychology. The increase in popularity of the analysis is, perhaps, due to a number of reasons, not least its potential to summate large amounts of information from research, address important questions that otherwise could not be answered using single studies, and yield novel empirical and practical insights. Other reasons include the increased technical and statistical expertise of sport and exercise psychology researchers, access to user friendly software, and increasing availability of sufficient volumes of primary research on which to conduct the analysis. To the benefit of researchers and practitioners alike, for these reasons, the current upward trend in the use of the meta-analysis will likely continue, and it will become an even more prominent staple of this area of research in the future.

Collating and quantifying effects in meta-analysis can be time consuming and challenging. There are many decisions to be made when using the analysis that can have substantive implications for the findings. Some of these decisions can lead to quite different results. In wielding this tool, then, we must be mindful of these key decisions and of the major pitfalls associated with the analysis. The need for caution is, perhaps, especially so for meta-analysis because its use can give the impression of comprehensiveness and completeness and is often well cited and influential as a consequence. This methodological note pertains to three related pitfalls of meta-analysis—treating correlational and intervention studies as yielding equivalent information, double counting effects, and selecting inappropriate common effect sizes for different types of studies.

The impetus for the note is provided by a recently published systematic review of meta-analyses by Lochbaum et al. (2022). The aim of the review was to systematically collate and evaluate published meta-analyses that examined the relationship between sport psychology variables and sport performance and to report and compare effects. The thorough search spanned nearly 40 years of research and resulted in the inclusion of 30 meta-analyses of 16 distinct sport psychology variables and 61 individual effects. Along with other useful analyses, the individual effects were assessed for meaningfulness and ranked against each another. The observed effects ranged from very large beneficial effects on sport performance (mindfulness) to medium detrimental effects on sport performance (total mood disturbance). The results were discussed thoughtfully—comparing the observed effects with other existing benchmarks—and cautiously—warning against overpromising the benefits of sport psychology.

With a focus on summarizing key research, thorough and transparent reporting, and useful analyses, it is a study with great practical value and could be a major touchstone for researchers and practitioners in the future. In all, it is a piece of research for which the authors should be commended. However, despite the admirable qualities of the work and its potential value, in summating effects from different types of studies, they made three important related errors. The first error was treating correlational and intervention studies as being comparable in the information they provide. The second error was double counting of studies and effects across meta-analyses. The third was the selection of the effect size to quantify and compare effects across types of studies. All of these errors are problematic and have the potential to lead to over estimates of the effects and the precision of their estimation.

On the first error, studies almost always differ in some way and, indeed, one of the benefits of meta-analysis is the way in which differences can be taken into account. However, in estimating common effects we must be sure that studies are comparable and this decision is rooted in study design. For some closely matching designs, this may not be problematic.

For example, post-intervention comparison in randomised control trials yield comparable information regardless of whether the control is a placebo-control or an active-control group. However, this is not the case in regards to intervention studies and correlational studies. Correlational studies typically have minimal levels of control so yield different (and questionable) information regarding treatment effects. For this reason, when both types of studies are included in systematic and meta-analytical reviews they are typically analysed separately (e.g., Ntoumanis et al., 2014; Oja et al., 2015; Rodriguez-Ayllon et al 2019; Wen et al., 2017).

On the second error, in including multiple meta-analyses of the same dependent variable, they have unintentionally counted stud ies and effects multiple times when aggregating the overall effects (Lochbaum et al., 2022, Table 3, p. 15). As one example, three meta-analyses of group cohesion feature in Lochbaum et al.'s (2022) review (Carron et al., 2002; Castano et al., 2013 ˜ ; Filho et al., 2014). Two are sequential reviews of the literature, which helps avoid inclusion of the same effects. However, four of 12 studies in the meta-analysis of Filho et al. are also included in the meta-analysis of Castano et al. This is obviously problematic in ˜ regard to disentangling evidence, with the influence of some factors appearing more (or less) beneficial by virtue of their repeated inclusion. It also means that the estimates of overall positive and negative effects provided in Lochbaum et al.'s (2022) review will likely be biased, with standard errors smaller and confidence intervals (CIs) narrower than otherwise would be the case. This, of course, affects whether or not effects are statistically significant and gives a greater sense of precision in estimation of effects than is actually the case.

On the third related error, Lochbaum et al. (2022) faced a common but difficult task of pooling effects from different studies. To do this job, they chose Cohen's d. Cohen's d is an effect size that expresses the difference between two groups in units of standard deviation

(SD, e.g., a control group vs. intervention group). It belongs to a family of other effect sizes that do the same with various tweaks and improvements to express standardized mean difference (e.g., Hedges' g). The comparative benefits of similar effect sizes are often noted; however, these alternatives are easily calculated using the information you would expect to see reported alongside Cohen's d. In addition, Cohen's d is, perhaps, the most widely used and reported standardized mean difference effect size. Therefore, it is an understandable choice and for the 13 intervention studies included in the meta-analysis is appropriate.

Unfortunately, though, for the correlational studies that are included in the systematic review, the use of Cohen's d is prob lematic. This is because although Cohen's d can easily be derived from Pearson's correlation coefficient (r), it cannot be interpreted easily and has some other undesirable features directly relevant to the aims of Lochbaum et al. (2022). The origins and problematic consequences of the use of Cohen's d in correlational studies have been discussed by Mathur and VanderWeele (2020). As they highlighted, the formula to convert Cohen's d from Pearson's correlation was derived for a "point-biserial" correlation; that is, the correlation between a dichotomous predictor (i.e., membership of one of two groups) and a continuous outcome, not for when both the predictor and outcome are continuous. As such, when using a continuous predictor, even though the conversion can take place from Pearson's correlation to Cohen's d, it is not clear which "groups" are being compared and to exactly what entities the standardized mean difference applies.

Interpretational difficulties aside, Mathur and VanderWeele (2020) also illustrated other notable issues with the use of Cohen's d in this context. In particular, when Cohen's d is derived from Pearson's correlation, the effect size is associated with an increase of two SDs of the predictor variable or treatment—an extreme contrast to make in most circumstances. This is the equivalent of comparing those with exposure to treatment at a mean level versus those in the 2.28 percentile. They also show that at any given contrast (−2 to 2 SDs), Cohen’s d will overestimate the contrast. Sometimes this overestimate is substantial. They showed, for instance, that the difference is largest when dichotomizing at the mean of the continuous predictor, at which point it is as much as 47% larger. The implications for the findings of Lochbaum et al. (2022) are hopefully clear—Cohen’s ds from meta-analyses of continuous predictors and sport performance (14 of 30 cases) are based on extreme contrasts and could be quite misleading.

In revisiting Lochbaum et al. (2022), the effects impacted by this issue relate to mood (tension, depression, anger, vigor, fatigue, confusion, and total mood disturbance); mindfulness; cohesion (task and social); anxiety (cognitive, somatic, and self-confidence); achievement climate (ego and task involving); achievement goals (task and ego); perfectionism (strivings and concerns); and self-efficacy. Notably, all of the largest beneficial and largest detrimental effects reported in Lochbaum et al. (2022) appear to come from conversions of Pearson’s correlations to Cohen’s d for these relationships (e.g., mindfulness, task cohesion, self-efficacy, and total mood disturbance; see Figure 2, p. 14 in Lochbaum et al., 2022). As such, this issue may have major implications for the aims of the study and the confidence we can have in the findings as they are currently presented.

At least one sport psychology variable that is affected by this issue is proving quite controversial—perfectionistic strivings. On one hand, perfectionistic strivings have been found to be positively related to athletic performance. However, on the other hand, perfectionistic strivings have also been linked to mental health difficulties—eating disorders, depression, and suicide ideation, among others (see Limburg et al., 2016). Debates regarding the merits and “healthiness” of perfectionistic strivings are longstanding and ongoing (see Gotwals et al., 2012; Hill et al., 2018; Stoeber, 2011). Notably, the findings of Lochbaum et al. (2022) positioned perfectionistic strivings as more beneficial for performance than a large number of other psychological factors, including psychological techniques such as goal setting and self-talk. It is possible that this finding will be construed as evidential in this ongoing debate. Instead, though, it is more likely that its relative standing is a function of the approach taken in Lochbaum et al.'s study to estimating effects from correlational studies.

The inclusion of studies with different designs and the use of Cohen's d for correlational designs do not appear to be issues confined to Lochbaum et al.'s (2022) meta-analytical review. There are a number of other meta-analyses in sport and exercise psychology that appear to have been impacted by at least one of these issues and sometimes more. In the meta-analytical studies on sport performance cited by Lochbaum et al., for instance, these issues are evident in studies of cohesion (Carron et al., 2002), achievement goals (Lochbaum & Gottardy, 2015), general psychological factors (Ivarsson et al., 2020), and mood states (Lochbaum et al., 2021). It is difficult to estimate how widespread the problem is and how this may be influencing research and practice. On the basis of these examples, though, researchers and practitioners will need to be especially mindful of this error and cautious when reading and applying the findings of the research that is affected

If researchers and practitioners want to express effects from correlational studies using Cohen's *d*, Mathur and VanderWeele (2020) provide a useful alternative computation of the effect size and its standard error (I will call it MV's *d*; equation. 1.2). It is not well known and not well used. However, it is more appropriate when a continuous predictor is involved. MV's *d* is interpreted as the average increase in the standardized dependent variable associated with an increase in the predictor variable of Δ units, where delta is prespecified (and ideally a scientifically meaningful contrast of interest). In deriving MV's *d* the standard deviation (SD) of the predictor variable is required. Preferably this is the SD from all studies used to derive the estimate of the effect but, as this is quite often not reported, it can also be based on another source or subsample of studies.

To illustrate the use of MV's *d*, perfectionism and sport performance has been examined in five studies with perfectionistic strivings positively related to performance (*r* = .23, 95% CI .11 to .35) and, the other dimension of perfectionism, perfectionistic concerns unrelated to performance (*r* = .06, 95% CI -.01 to .14). In Lochbaum et al (2022), Cohen's *d* is calculated to be 0.47 and 0.12, respectively (N = 684). To recalculated the effects using MV's *d*, the SDs and sample sizes were retrieved from the four (of five) samples reporting these values in published studies (the other study examines team perfectionism) and the average SD for the studies was used (PS =0.87 and PC = 0.89, N = 156). The MV's *ds* are 0.24 (perfectionistic strivings) and 0.06 (perfectionistic concerns). As can be seen, the effects of perfectionistic strivings can be shown to be notably more modest using this estimate of effect.

Note, in recalculating the effect, Δ was set to one SD of the predictor variable (PS or PC). This value was selected to make the interpretation similar to a beta coefficient in regression analyses— the use of which is more common in this area and which is, arguably, a more reasonable contrast than two SDs. Researchers and practi tioners may or may not agree on this point, but it hopefully serves to illustrate the ambiguity in determining these types of effects. Ideally, Δ should have practical relevance, but determining whether it does is also challenging. Based on typical SDs in published research, a Δ of one SD would be similar to comparing an athlete who responded "agree" with an athlete who responded "strongly agree" on the response formats used to measure perfectionism (e.g., 1–5 or 1–7 Likert scales). As a reference point, research has found similar increases in standardized PS and PC to produce statistically significant changes in a range of motivation, performance, and well-being outcomes in athletes (see Hill et al., 2018).

What, then, should be done with the effects reported in Lochbaum et al. (2022)? The estimates of effects using Cohen's d for studies wherein there is a dichotomous predictor variable remain useful estimates of effects. However, effects from studies using correlation designs should be separated from effects from intervention studies. An accurate comparison cannot be made between them, even when using MV’s d. Second, careful consideration needs to be given to what studies and effects have been included multiple times to avoid underestimating standard errors and overestimating precision. This can be avoided by including estimates once and changing how effects are weighted (e.g., Rebar et al., 2015). Third, the effects (Cohen’s d) for the correlation designs will need to be re-estimated using MV’s d or another appropriate effect size. Pearson’s correlation is common and could be used. In deciding between the two, MV’s d provides estimates of the effect in the same metric as Cohen’s d (SDs of the dependent variable), whereas Pearson’s correlation is “unit free.” As such, MV’s d may be of more practical value for researchers and practitioners depending on the intended use of the information.

To support researchers and practitioners in regard to calculating MV’s d, Mathur and VanderWeele (2020) provided a function in the R package MetaUtility (“r_to_d”) for this conversion. I have also provided simple R code that calculates MV’s d, its standard error, and 95% CIs and shows the calculations for the perfectionistic strivings example (Hill, 2022). Researchers and practitioners are encouraged to consider using these tools and to check effects of particular interest to their work. This includes revisiting effects in Lochbaum et al. (2022) and other meta-analyses that have fallen foul of this error. In doing so, MV’s d also has the advantage of specifying a fixed contrast in the predictor variable. Being able to decide on this contrast is likely to be practically useful in sport and exercise psychology and when determining to what degree sport psychology variables influence sport performance.

Concluding remarks

No research study is perfect—mistakes are common and easily made. Meta-analysis includes a number of decisions, technical and otherwise, that make it especially prone to errors. Some of these errors will have minor consequences. As shown in the current research note, though, sometimes errors contribute to misleading findings and can have important implications for contentious issues and future research and practice. This does not diminish the efforts of researchers as we seek to scrutinize evidence in sport and exercise psychology. However, we should be watchful for these types of errors and provide solutions when possible. When under   taking meta-analysis in the future, researchers can hopefully use the information in this methodological note to help guide their work and guard against problematic comparisons and conversions.

References

Carron, A. V., Colman, M. M., Wheeler, J., & Stevens, D. (2002). Cohesion and performance in sport: A meta analysis. *Journal of Sport and Exercise Psychology*, *24*(2), 168–188. https://doi.org/10.1123/jsep.24.2.168

Castaño, N., Watts, T., & Tekleab, A. G. (2013). A reexamination of the cohesion–performance relationship meta-analyses: A comprehensive approach. *Group Dynamics: Theory, Research, and Practice*, *17*(4), 207. http://doi.org/10.1037/a0034142

Filho, E., Dobersek, U., Gershgoren, L., Becker, B., & Tenenbaum, G. (2014). The cohesion–performance relationship in sport: A 10-year retrospective meta-analysis. *Sport Sciences for Health*, *10*, 165-177. https://doi.org/10.1007/s11332-014-0188-7

Gotwals, J. K., Stoeber, J., Dunn, J. G. H., & Stoll, O. (2012). Are perfectionistic strivings in sport adaptive? A systematic review of confirmatory, contradictory, and mixed evidence. *Canadian Psychology/Psychologie Canadienne*, *53*(4), 263–279. https://doi.org/10.1037/a0030288

Hill, A. (2022). Mathur and VanderWeele's d (R code). York St. John University. Software. https://doi.org/10.25421/yorksj.22332172.v1

Hill, A. P., Mallinson-Howard, S. H., & Jowett, G. E. (2018). Multidimensional perfectionism in sport: A meta-analytical review. *Sport, Exercise, and Performance Psychology*, *7*(3), 235–270. https://doi.org/10.1037/spy0000125

Ivarsson, A., Kilhage-Persson, A., Martindale, R., Priestley, D., Huijgen, B., Ardern, C., & McCall, A. (2020). Psychological factors and future performance of football players: A systematic review with meta-analysis. *Journal of Science and Medicine in Sport*, *23*(4), 415–420. https://doi.org/10.1016/j.jsams.2019.10.021

Limburg, K., Watson, H. J., Hagger, M. S., & Egan, S. J. (2016). The relationship between perfectionism and psychopathology: A meta-analysis. *Journal of Clinical Psychology*, *73*(10), 1301–1326. https://doi.org/10.1002/jclp.22435

Lochbaum, M., Stoner, E., Hefner, T., Cooper, S., Lane, A. M., & Terry, P. C. (2022). Sport psychology and performance meta-analyses: A systematic review of the literature. *PLOS ONE*, *17*(2), e0263408. https://doi.org/10.1371/journal.pone.0263408

Lochbaum, M., Zanatta, T., Kirschling, D., & May, E. (2021). The profile of moods states and athletic performance: A meta-analysis of published studies. *European Journal of Investigation in Health, Psychology and Education*, *11*(1), 50–70. https://doi.org/10.3390/ejihpe11010005

Mathur, M. B., & VanderWeele, T. J. (2020). A simple, interpretable conversion from Pearson's correlation to Cohen's for *d* continuous exposures. *Epidemiology*, *31*(2), e16–e18. https://doi.org/10.1097/ede.0000000000001105

Ntoumanis, N., Ng, J. Y., Barkoukis, V., & Backhouse, S. (2014). Personal and psychosocial predictors of doping use in physical activity settings: a meta-analysis. *Sports Medicine*, *44*, 1603-1624. http://doi.org/10.1007/s40279-014-0240-4

Oja, P., Titze, S., Kokko, S., Kujala, U. M., Heinonen, A., Kelly, P., ... & Foster, C. (2015). Health benefits of different sport disciplines for adults: systematic review of observational and intervention studies with meta-analysis. *British Journal of Sports Medicine*, *49*(7), 434-440. http://doi.org/10.1136/bjsports-2014-093885

Rebar, A. L., Stanton, R., Geard, D., Short, C., Duncan, M. J., & Vandelanotte, C. (2015). A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. *Health psychology review*, *9*(3), 366-378. http://doi.org/10.1080/17437199.2015.1022901

Rodriguez-Ayllon, María, Cristina Cadenas-Sánchez, Fernando Estévez-López, Nicolas E. Muñoz, Jose Mora-Gonzalez, Jairo H. Migueles, Pablo Molina-García et al. (2019) Role of physical activity and sedentary behavior in the mental health of preschoolers, children and adolescents: a systematic review and meta-analysis. *Sports Medicine, 49* (9), 1383-1410. http://doi.org/10.1007/s40279-019-01099-5

Stoeber, J. (2011). The dual nature of perfectionism in sports: Relationships with emotion, motivation, and performance. *International Review of Sport and Exercise Psychology*, *4*(2), 128–145. https://doi.org/10.1080/1750984x.2011.604789

Acknowledgments

The author would like to thank Dr Malak Olamaie, York St. John University, for checking the calculations of MV’s d.