Dear Professor Jobson,

Many thanks for considering our manuscript and placing it under review with *Journal of Sport Sciences* We are very grateful for your efforts in obtaining reviews. We are very impressed with the quality and constructiveness of the feedback provided by the reviewers which we feel has significantly improved the quality of our manuscript.

We have made changed throughout the manuscript in red to highlight where amendments were made in response to specific comments. Also, please see below for our comments in response to some of the key points raised by the reviewers.

|  |  |
| --- | --- |
| Reviewer 1 | |
| Reviewers comment | Authors response |
| The paper appears to offer a valuable insight for the field of sport science, in which practitioners work with each individual and thus individual differences are very important. In addition, many studies tend to have a small sample size due to the limited availability of participants (e.g. athletes who are willing to participate in a study). Examining multiple studies on the same topic collectively can overcome the issue of a small sample size and help better understand of individual differences. Finally, I would like to applaud the authors for making an effort to share their R analysis script. I believe this effort goes a long way in helping others utilize what is presented in the paper. Please read below comments that I hope the authors will address to improve the presentation of the information of the manuscript. | Thank you for your kind words and feedback. |
| Being in the field of sport science for over a decade, I have to speculate that this paper is likely to be overlooked by many scientists. This is not because the paper has no values. This is simply because most scientists shy away from statistics in our field. This is unfortunate. Even though some will stay away regardless of how this paper may be written, I believe the paper may be revised with great care for those who want to learn more about statistics but lack the fundamental knowledge to comprehend a paper like this manuscript. While I am not suggesting that the whole paper is re-written, please consider the following points in hopes to make this paper better received in our field. | Thank you for your suggestions which we feel have improved the accessibility of the paper. We have made efforts to include alongside mathematical formula greater exposition in plain English. We have also introduced more fundamental concepts such as fixed and random effects models, provided more footnotes explaining statistical terminology, and tried to address the order in which different models are presented such that is more logical and less of a surprise to readers. |
| Point1: Please pay great attention to the use of statistical/mathematical terms. For example, in the line 116-117 (p.4), I read "an unbiased estimator of … a population standard deviation and is sampling variance …". While this is a straightforward sentence to those who have some knowledge of statistics, terms such as unbiased, population, and sampling can be very confusing to many scientists in our field with limited statistical knowledge. Other terms that may potentially be confusing include primary research, random and fixed effects, heteroskedasticity, and proportionality. Perhaps, these simple terms that can still confuse some scientists may be defined in the footnotes? | As noted, throughout we have either included footnotes, explanations in the surrounding text, or in explanations in parentheses where there are terms that might be unfamiliar to a reader less well versed in statistical jargon. |
| Point 2: Please explain each model in a manner that is friendlier to those without much statistical knowledge. For example, the equation 17 is not likely to make sense to many scientists in our field. Is it possible to add a much simpler description/explanation of the model in without statistical terms in addition to what is already in the paper? | We have attempted to include explanation of each model both in mathematical and plain English language now, and hopefully in a manner that aids readers in understanding what the different mathematical terms refer to. |
| Please also present the results from the models in a manner that is friendlier to those without much statistical knowledge. For example, to those whose statistical knowledge is only as far as general linear model, it may be confusing where the exact values of SDir, lnV R, and lnCV R came from when a model for them typically consists of letters and coefficients and that values of a given dependent/outcome variable needs to be solved for by substituting letters with various numbers. Yet, I think it's even more confusing that an equivalent approach to the lnCV R model does not report results for an estimate of lnCV R. Instead, it reports the values of coefficients much like the same people are used to seeing (furthermore, the difference in the formats of the results makes it more difficult to understand how the equivalent approach is an equivalent approach). I will have to say that by the end of the equivalent approach to the lnCV R model, most scientists in our field will be lost. | As noted above we have tried to explain this better. Further, for this specific example we have attempted to ensure the descriptions make clear how the models aim to estimate the same thing, namely the difference between groups in variance whilst controlling for mean effects. |
| It seems that the manuscript takes a good detour to get to the lnCV R and its equivalent approach. After all, the authors suggest that lnCV R is superior to SDir and lnV R in dealing with the assumptions of normality and constant measurement error, and a mean-variance relationship. Consequently, the use of SDir and lnV R appears to be discouraged. It then seems redundant to spend much of the manuscript on demonstrating SDir and lnV R. Instead, the authors may spare the SDir and lnV R demonstrations to put a greater effort into explaining the lnCV R and its equivalent approach in a manner that maintains the academic rigor but is also friendlier to those without much statistical knowledge. | We agree with the need to express things in a more friendly manner to those not familiar with the statistical concepts discussed. We have attempted to address this throughout. We do feel however that it is worth presenting the other options such that readers are able to see *why* the lnCVR and multilevel model of ln\_sigma are perhaps more appropriate in certain situations. Also, the SDir has been suggested to those working in sport and exercise science previously so we feel it is appropriate to contrast it with the approaches particularly. It and the lnVR may actually be desirable for the sake of simplicity where mean-variance relationships are not present and we note this now suggesting researchers examine this and balance model complexity against assumptions. |
| Please add to the foot notes what the statistic I2 means. | We have added this now in the text and also explained what the prediction intervals provide so the reader understands the heterogeneity that the model estimates. |
| In line 251 on page 14, I believe the word seen is mis-spelled. | Thank you, this has been corrected. |
| Reviewer 2 | |
| this is a very important topic, we should be seeing papers like yours more often to improve our methods!  Which leads me to my first bit of feedback. For all equations which you did not derive yourselves, please give references, even for the most basic ones. Most readers will have to look these up and redeepen their knowledge before understanding your article. For most equations, you did provide references, but in 2. sometimes you did not e. g. SD. | We did not derive any equations ourselves but in all instances the appropriate references are provided in the accompanying text. Exceptions to this are the most basic ones as you note. We do not feel we need to provide simple descriptive statistical formula such as means and standard deviations which we assume almost all, if not all, readers of a paper like this would be familiar with or could find very quickly online. |
| Please explain why the ln RR is unaffected by the SD (equation 5). | This is because the SD doe not feature in the equation for the calculation of the lnRR, it only features in its sampling variance. This is mentioned in the text and the corresponding equations referred to. |
| 2.3: Please explain if, and if, why the ln CV R is robust against larger SD in larger groups as mentioned in 2.2. | Hopefully this is now clear and we have noted that the coefficient or variation is a standardised measure of the relative variation in a effect. |
| 3.1: line 188: So the controls had higher hypertrophy (40.15 %), or did I misread this? Please explain. | No, we think you may have misread one of the I2 statistics reported. |
| Figures do indeed extremely disrupt the text-flow. I am not sure what caused this, please check with the editor(s). | Apologies for this. It is due to the rendering of the manuscript from an Rmarkdown file to a pdf. This will hopefully be handled in proofing stage if accepted. |
| The model at the bottom of page 15 just jumps out of the dark, but is not really introduced at the beginning of the article, this is difficult in my opinion. Regarding the whole article, I find it difficult to read because of its structure. There is not really a red hering, readers are not guided through the text. I have to find out what you are planning to do while you are already doing it. For me, this is not working. Would it be possible to explain in the introduction what you will be doing at which point, how many models or variations of you model you will be introducing, or even include a kind of methods section? Please check with the editor(s). | We have now attempted to more logically present the different meta-analytic models in the paper. We had hoped it was previously clear what the connections between them were but see that this was not the case. Thus we have added to the introduction and also tried to more clearly present the links between different models as they are progressively presented. |
| P. 18, line 281: "Similar", is there any possibility showing the graphs next to each other? | Given one is a caterpillar plot and one is a meta-analytic scatterplot (i.e., bubble plot), it’s not clear that presenting them together would aid interpretation. |
| 4.: Please state more clearly where the outcomes of both your calculation models (I understood there are two...?) differ from the findings of Polito et al. (2021). | Our results for the SMD (and also the lnRR) are already in section 3.1 and are contrasted with those of Polito et al. (2021). |
| Comments from JSS Statistics and Research Design Advisory Group | |
| The paper advocates examining variation in intervention studies in sport and exercise science. The authors argue that research in sport and exercise science tends to focus on mean effects — particularly for interventions designed to improve outcomes such as. They go on to highlight that the response to interventions designed to improve fitness or performance varies considerably. They argue that interventions affect not only the mean, but also the variance in responses. The authors make a strong case for gaining insights using variances from randomised controlled trials (RCTs) or similar intervention studies, referring to interindividual heterogeneity and how this provides information on how generalisable effects might be. | Thank you, we are glad our case was strong for researcher to examine this further. |
| The authors make a coherent argument for the need for larger samples when exploring variances than is typical when only looking at mean effects. They suggest that a meta-analytic framework is appealing for sport and exercise science given the small sample sizes often involved in studies in the area. | Thank you, we hope that such efforts will also have the effect of increasing the size of primary studies, or attenuating claims made from them until sufficiently accumulated evidence for meta-analysis is available. |
| As well as introducing and evaluating some effect size measures that are used in other disciplines, but rarely if ever used in sport and exercise science, the authors discuss how variation can be explored alongside mean effects in meta-analysis research. They illustrate one approach for meta-analysis where response variation is modelled using examples from resistance training Randomised Controlled Trials. | Thank you. It was a pleasure to stumble on Polito et al. (2021) which prompted the idea to explore this using RCTs. |
| The r code for models, data summaries, and supplementary materials presented on the Open Science Framework (OSF) was really useful and welcome. Currently, the “Polito et al. RT Extracted Data.csv” is not available on the OSF, although it can be extracted from a non-anonymised file on GitHub. | We are not sure why this is and in fact we cannot seem to access it ourselves from OSF now. So we have added the github repository link to the manuscript aswell. |
| The analysis methods are appropriate for ratio level data, which are used in the examples provided. | Thank you. |
| Abstract  Some suggested very minor changes to the abstract, these are merely suggestions to improve clarity but the authors can choose to ignore them.  Meta-analysis has become commonplace within sport and exercise science for synthesising and summarising empirical studies. However, most research in the field focuses upon mean effects; particularly the effects of interventions aimed at improving outcomes such as fitness or performance. CUTCUT It is well known that individual responses to interventions vary considerably. Hence, interest has increased in exploring precision or personalised exercise approaches. CUTCUT Not only are the mean effects affected by interventions, but variances may also be impacted. Exploration of variances in studies such as randomised controlled trials (RCTs) can yield insight into CUTCUT interindividual heterogeneity in response to interventions and help determine the CUTCUT generalisability of effects. Yet, larger sample sizes than those used for typical mean effects are required when probing variances. Thus, in a field with small samples such as sport and exercise science, exploration of variance through a meta-analytic framework is appealing. CUTCUT Despite the value of embracing and exploring variation alongside mean effects in sport and exercise science, CUTCUT it is rarely applied to research synthesis through meta-analysis. CUT< we introduce and discuss effect size approaches and models for meta-analysis of variation using relatable examples from resistance training RCTs.>CUT We introduce and evaluate different effect size calculations along with models for meta-analysis of variation using relatable examples from resistance training RCTs. |  |
| Thank you, we have made these minor changes to enhance clarity. |
| Page 1 lines 27-30: suggested reword: “…likely the most common aim in meta-analysis in sport and exercise science, and indeed primary empirical research too, is to compare the means of measurements taken across different categorical grouping variables...” | We have made this change. |
| In discussing versions of Cohen’s d standardised effect, there might be merit in critically evaluating Glass’ Delta for RCTs. I realise this quite deliberately provides no information whatsoever on variation in the intervention group, but it is a measure frequently recommended in text books and is built into software such as JASP. Note that this is merely a suggestion and not a requirement. | We have added to the accompanying footnote mention of Glass’ Delta. |
| The data used was from a recent systematic review and meta-analysis of randomised trials by Polito et al., (2021) that included a randomised intervention group(s) and a non-training control comparison which was published in the Journal of Sport Sciences. The authors argue that given the review only included studies with a non-training control group, this study selection was ideal to examine variation of interindividual responses specifically by means of comparing the variances in change scores between the intervention group(s) and control group(s). I think the authors make a good argument for using this data. | Thank you again. |
| Besides modelling interindividual response variation, the authors used an appropriate multilevel mixed-effects meta-analysis model with random intercepts for study, arm, and effect. | Thank you. We hope that by being explicit about such models and their assumptions and utility researchers might recognise that they have wide applications within our field. Particularly as it is rare for studies to only include single outcome effects. |
| The r package used (metafor) is appropriate for the methods applied in the study. | Thank you. Wolfgang’s package is always a lifesaver. |
| The authors conclude by highlighting the advantages of acknowledging variability in meta-analyses and focusing on more than the mean differences between groups or conditions as intervention and control comparisons. This is a very reasonable conclusion. As such, the authors finish by recommending that future meta-analyses researchers should analyse both means and variances, but also when re-analysing past meta-analyses. | Thank you. |
| Something I would recommend the authors consider is findings of Ilyas Bakbergenuly et al (2020) simulation study 'Estimation in meta-analyses of response ratios’ and how they may impact on results and recommendations. | Thank you for sharing this paper. We considered whether to mention it in the manuscript, but given our focus was not on the lnRR and it was merely presented as an alternative effect size alongside the SMD for mean differences we opted not to. The estimator described by Bakbergenuly et al. also is not yet extended to the multilevel models used here where multiple variance parameters must be estimated. |

Again we would like to thank all involved in the process of reviewing this manuscript and look forward to the next round of reviews in considering our revisions.

Many thanks