Response to Reviewers

The effects of supervision on effort during resistance training: A Bayesian analysis of prior data and an experimental study of private strength clinic members

Dear Professor Gruet,

Many thanks for considering our manuscript and placing it under further review with *Peer Community in Health and Movement Sciences*.

We apologise sincerely for our delay in addressing these and resubmitting. Unfortunately other commitments and workloads had prevented us from returning to this. We have attempted to address or rebut the additional comments provided by the reviewers.

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

The authors

Reviewer Comments and Author Responses

Reviewer 1 - David Clarke

Reviewer Comment

Title - Suggest changing “upon” to “on.”

Author Response

We have made this change.

Reviewer Comment

Abstract - TUL, RPE, and RPD are measures of effort, not exercise performance. There are several instances in which "exercise performance" is used in place of "effort," wherein only the latter should be used. Be consistent with the constructs and operational definitions.

Author Response

In the present study exercise performance as TUL is used as an operationalisation to draw inferences about effort. As such, in places where we are drawing broader inferences about the concept of effort as informed by the exercise performance we use "effort", but in others we use more accurately to the operationalisation "exercise performance" or "TUL". We have added the following the abstract to clarify this:

"…exercise performance, measured as time under load (TUL) where a longer TUL at a given load is indicative of greater effort…"

We have also added to the introduction the text in parentheses at the end of this sentence to clarify this conceptually:

"Effort is conceptualised at the relation of task demands to the capacity to meet those demands (Steele 2020) and so in RT is determined by both the load utilised and also the proximity to momentary failure due to the fatigue (i.e., reduction in capacity) that occurs with continued performance at a given load (i.e., with more repetitions or a longer time under load)."

And clarified in the methods:

"The TUL performed (where a longer TUL was interpreted as indicating greater effort), and ratings of perceptions of effort (RPE) and discomfort (RPD) were recorded and compared between both conditions."

Reviewer Comment

Abstract - Second sentence: "We investigated supervision's role in effort…" – "in" should be "on"

Author Response

We have made this change.

Reviewer Comment

The Introduction describes the existing knowledge of the effect of supervision on resistance training outcomes, specifically strength, and summarizes the study's design and methods. While the essential content is present, it could be more clearly discussed. I suggest reorganizing the introduction as follows: • The first paragraph is good as is. • I would present the hypothesized mechanisms together up front, i.e., trainers enforce load progression and help motivate the exercisers to apply the proper amount of effort (i.e., closer proximity to momentary muscular failure). • I would then discuss the evidence for each hypothesized mechanism, which is mostly indirect. • The second-to-last paragraph should summarize the existing gaps in knowledge, i.e., the data concerning supervision effects are still limited, especially in ecological settings, and the potential mechanisms by which supervision might enhance strength are likewise unclear.
• The final paragraph should more explicitly state the study purpose and along with a hypothesis. For example, "The purpose of this study was to examine the effects of supervision on client effort in resistance training in a more ecological setting. We hypothesized that … I would then introduce the methods as the authors have done by stating that "We tested the hypothesis by …"

Author Response

We thank the reviewer for their suggested revision to the introduction. We have however opted not to make any major rewrites to it as we do feel the major rationale is presented clearly already, and we have not added any hypothesis as our analyses were not aimed at any kind of statistical hypothesis testing but instead exploratory and descriptive parameter estimation. Lastly, given the suggested restructure is largely one of preference, this is somewhat of a pragmatic choice on our part to not make changes because of available time for making revisions given other workload demands currently (and that we're already pretty late with making revisions in response to this second round of review - our apologies!).

Reviewer Comment

Minor comments on the writing: There were several errors and suboptimal wording choices present. • The word "recent" (typically followed by "meta-analysis") is repetitively used throughout. Just state the result and cite the paper. For example, the second sentence, "In a recent systematic review and meta-analysis from Fisher et al. (2022) there was a moderate standardised mean effect (0.40 [95%CI: 0.06, 0.74]) of supervised vs unsupervised RT on strength outcomes synthesised from ten studies" could be more succinctly stated as "A meta-analysis of 10 studies demonstrated a moderate benefit of supervision on strength outcomes (standardized mean effect = 0.40 [95%CI: 0.06, 0.74]; Fisher et al., 2022)." • Error, second paragraph: "Effort is conceptualized at…" – "at" should be "as" • Grammar, third paragraph:

"But a caveat is that there was limited data…" – the word "data" is plural, so "was" should be "were." There are other instances of this error in the manuscript.

Author Response

We have however made these minor changes as noted for the existing introduction.

Reviewer Comment

As with the Introduction, the Methods are thoroughly described but could be better organized to highlight critical features. In particular, the operational definitions of effort in this study were TUL, RPE, and RPD. These operational definitions should be more explicitly defined and justified as measures of effort. Currently, their description is an afterthought buried in a paragraph: "The TUL performed, and ratings of perceptions of effort (RPE) and discomfort (RPD) were recorded and compared between both conditions." Furthermore, whether session RPE or RPE post each exercise is not made clear until the end of the first paragraph in the Protocols section ("In addition, they recorded their RPE and RPD in that order immediately upon completing the exercise using previously validated scales for differentiating these perceptions.") Such critical details about the measures should be mentioned upfront in their own paragraph.

Author Response

As above with the introduction revisions, we have opted not to make further structural changes as we believe that the required information is contained in the methods section. As noted above though we have more clearly noted that, given the conceptualisation of effort, the exercise performance as TUL is indicative as an operationalisation of effort in the present study.

Reviewer Comment

I question the following sentence: "This study was not pre-registered and as described below the sample size was justified based on logistical concerns and the analysis is considered exploratory." To me, it seems a bit disingenuous to state the analysis is exploratory when there are meta-analyses of supervision effects on RT outcomes and a clear testable hypothesis that the study addresses using a very detailed Bayesian analysis. Why was the experimental intervention not pre-registered? Was this a deliberate choice? Could the study be retrospectively registered? (I have seen others do so.)

Author Response

We have tried to be as transparent as possible with this study. The original experimental study was conducted several years previously, and before any of the authors were fully cognisant of the value of pre-registration. Further, we did not have a sufficiently precise statistical hypothesis to test for which we could for example have performed a power analysis (Frequentist or Bayesian) to help justify sample size for the experimental hypothesis. We would prefer not to retrospectively register given this would be merely performative at this stage and does not serve the primary purpose of pre-registration (i.e., the transparent reporting of research to enable the third-party evaluation of the severity of tests for claims made based on hypothesis tests). In retrospect we could have looked to convert the standardised effect size metrics to the relevant scales for the parameter estimates presented contrasting between conditions, and then conduct either an analysis akin to examining the overlap with the region of practical equivalence (ROPE), or calculated Bayes Factors for this. But we feel this would be disingenuous to add (and also take more time than we have currently) as we never planned to do this a priori. As such, having the full dataset open and being transparent about the nature of the study is appropriate in our opinion (notably, these responses will also be publicly available so people can see why we made this choice).

Reviewer Comment

Experimental approach section: This sentence is run-on: "This consists of a single set of the resistance machine exercises prescribed on their current training programme card using a load that should permit them to reach momentary failure within a time-under-load (TUL) of 90-120 seconds (though an upper limit of 180 seconds TUL is enforced to avoid machines being occupied for too long on the clinic floor preventing other members from using them) using a ~ 12 seconds repetition duration (i.e., ~4:4 seconds concentric:eccentric actions with a 2 isometric second hold whilst still under load with tension, not "locked-out", on the involved musculature at the end of each concentric and eccentric muscle action)."

Author Response

We have split this sentence now.

"This consists of a single set of the resistance machine exercises prescribed on their current training programme card. A load is used that should permit them to reach momentary failure within a time-under-load (TUL) of 90-120 seconds (though an upper limit of 180 seconds TUL is enforced to avoid machines being occupied for too long on the clinic floor preventing other members from using them) using a ~ 12 seconds repetition duration (i.e., ~4:4 seconds concentric:eccentric actions with a 2 isometric second hold whilst still under load with tension, not"locked-out", on the involved musculature at the end of each concentric and eccentric muscle action)."

Reviewer Comment

Previous observational sample subsection: The following sentence is run on, contains several awkward passages, and contains punctuation errors: "For TUL we limited data to Core and Assisted members training sessions that either were not led by an exercise scientist or were respectively, took the first training session after at least 6 months of previous training at Kieser Australia had been completed by each member, randomly sampled 1000 Core and 1000 Assisted members and then filtered to the resistance machines used in by members in the experimental study so that we had a selection of members across varied clinic locations and completing sessions with a selection of resistance machine exercises and had TUL data for each exercise." Suggest rewriting as a numbered list of inclusion criteria for the TUL data.

Author Response

We have rewritten this as:

"For TUL we limited data to Core and Assisted members training sessions that either were not, or were, led by an exercise scientist respectively[^3]. We took the first training session after at least 6 months of previous training at Kieser Australia had been completed by each member. Then we randomly sampled 1000 Core and 1000 Assisted members and filtered to the resistance machines used by members in the experimental study."

Reviewer Comment

Protocols section:
• The first paragraph is overlong – a new paragraph could begin at the "Participants were instructed to ensure…"
• Typo: "These TUL recordings where then input…" – "where" should be "were"

Author Response

We have made both of these changes.

Reviewer Comment

Statistical Analysis:
• "All posterior estimates and their precision,…" Precision should be pluralized as "precisions"
• Paragraph starting with "For both the previous observational sample…" The authors define "session type" (core or assisted) but then refer to it as "condition" in the model description. Suggest helping the reader by using consistent terminology.

Author Response

We have pluralised "precisions" and also added *"…session type (i.e., condition)…"* to clarify the terminology.

Reviewer Comment

Figure 2 title: "Prior Sample Data Distributions" should be changed to "Previous observational sample histograms"

Author Response

We have changed the title.

Reviewer Comment

An issue I sensed with the first version of the paper but failed to articulate in my original comments was the lack of validation of the modeling approach. The study features many assumptions and modeling decisions, and the reader is left to wonder about the extent to which the conclusions are robust to them. A comment on the sensitivity of the conclusions to the modeling decisions would be appropriate in the Discussion.

Author Response

Whilst we agree to some extent that all elements of model selection and the assumptions underlying them should be well justified and explained the standards are often double in this regard - researchers in our field using more traditional t-test or ANOVA type analyses (and which probably don't even understand the assumptions underlying them) don't get asked to justify these. We've explained quite clearly, and also provided the lay summary for our modelling choices. Adding a comment to the discussion about the possible sensitivity of our conclusions to our choice of models is in one sense tautological as our results, as are any using a single modelling approach, conditional on that model (the exception being in the case of Bayesian Model Averaging) but also that making that point leads to the question why not take "multiverse" type approach and try out lots of different models to see what happens. In principle this would be great, but as we've noted we don't have time to continue working on this in that fashion. If readers are curious then the data is openly available and they are welcome to examine whether a different, perhaps more justifiable, modelling choice would lead to substantively different estimates.

Reviewer Comment

In addition, diagnostic analyses were performed and included as a supplementary file, which I commend, but they were not mentioned in the Results or Discussion. A quick sentence or two summarizing the diagnostics would help assure the reader about the model validities.

Author Response

We have added to the analysis section where we link to the diagnostic plots:

"Trace plots were produced along with $\hat{R}$ values to examine whether chains had converged (all were $\hat{R} < 1.05$ indicating convergence)…"

Reviewer Comment

As I alluded to above, the study purpose could be more explicitly articulated in the first paragraph of the Discussion. I suggest modifying the first sentence as follows: "Our study is a first to offer insights into the role of supervision on exerciser effort during RT in an ecologically valid real-world setting. These insights were derived from a unique and strong study design involving observational and experimental components."

I suggest modifying the second sentence as follows to focus on the outcomes measures as operational definitions of effort: "…we assessed effort as the TUL, RPE, and RPD."

Author Response

We have changed these sentence to the ones suggested.

Reviewer Comment

Some paragraphs are overlong. The limitations paragraph, for example, takes up a whole page. It could be broken up into single paragraphs for each of the ~4 limitations discussed by the authors.

Author Response

We have broken up the limitations paragraph.

Reviewer 2 - Anonymous

Reviewer Comment

I encourage the authors to revise the conclusion so that it: 1. Maintains the statement that under real-world settings there was little difference in exercise performance. 2. Clearly states that the observed benefit in the experimental setup was based on an acute effect of a single session. 3. Avoids speculation and focuses on direct interpretations of the data.

Author Response

We have made changes to the conclusions to try to more accurately reflect the findings as suggested (and also minor change to the abstract). The conclusion now reads:

"In real world settings there was little difference in exercise performance suggesting similar degrees of effort during both supervised and unsupervised conditions. Yet, in our experimental study there was a clear benefit to performance when under supervision suggesting trainees performed with a higher degree of effort. There was a small difference in RPE reported with and without supervision under both real world and experimental conditions suggesting that under supervision trainees train with greater proximity to failure, which was also supported by greater RPD under supervision; though in real world conditions the RPE reported suggested that effort may be sub-optimal irrespective of condition. These results in general support previous work highlighting the importance of supervision during RT and that under experimental conditions trainees likely train with greater effort when supervised."

Steele, James. 2020. "What Is (Perception of) Effort? Objective and Subjective Effort During Attempted Task Performance." PsyArXiv. https://doi.org/10.31234/osf.io/kbyhm.