Accuracy in predicting repetitions to task failure in resistance exercise: a scoping review and exploratory meta-analysis

Received: 16th May 2021
Supplementary materials: https://osf.io/jzuwq/
For correspondence: ihalperin@ytauex.tau.ac.ul
Twitter: @Israel_Halperin

Israel Halperin[1,2]*, Tomer Malleron[1,2], Itai Har-Nir[1,2], Patroklos Androulakis-Korakakis[3], Milo Wolf[3], James Fisher[3], James Steele[3]

[1]School of Public Health, Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel
[2]Sylvan Adams Sports Institute, Tel Aviv University, Tel-Aviv, Israel
[3]Faculty of Sport, Health, and Social Sciences, Solent University, Southampton, UK

ABSTRACT

Background: Prescribing repetitions relative to task-failure is an emerging approach to resistance training. Under this approach, participants terminate the set based on their prediction of the remaining repetitions left to task-failure. While this approach holds promise, an important step in its development is to determine how accurate participants are in their predictions. That is, what is the difference between the predicted and actual number of repetitions remaining to task-failure, which ideally should be as small as possible. **Objective:** Examine the accuracy in predicting repetitions to task-failure in resistance exercises. **Design:** Scoping review and exploratory meta-analysis. **Search and Inclusion:** A systematic literature search was conducted with PubMed, SPORTDiscus, and Google Scholar in January 2021. Inclusion criteria included studies with healthy participants who predicted the number of repetitions they can complete to task-failure in various resistance exercises, before or during an ongoing set, which was performed to task-failure. Sixteen publications were eligible for inclusion, of which 13 publications that cover 12 studies were included in our meta-analysis with a total of 414 participants. **Results:** The main multilevel meta-analysis model including all effects sizes (262 across 12 clusters) revealed that participants tended to under predict the number of repetitions to task-failure by 0.95 repetitions (95% CIs= 0.17 to 1.73), but with considerable heterogeneity ($Q_{(261)}$= 3060, p< 0.0001; I^2 = 97.9%). Meta-regressions showed that prediction accuracy slightly improved when the predictions were made closer to set failure (β= -0.025 [95% CIs= -0.05 to

All authors have read and approved this version of the manuscript. This article was last modified on 30th June 2021

0.0014]) and when the number of performed to task-failure was lower (≤12 repetitions, β= 0.06 [95% CIs= 0.04 to 0.09]; >12 repetitions, β= 0.47 [95%CIs= 0.44 to 0.49]). Set number trivially influenced prediction accuracy with slightly increased accuracy in later sets (β= -0.07 repetitions [95% CIs= -0.14 to -0.005]). In contrast, participants training status did not seem to influence prediction accuracy (β= -0.006 repetitions [95% CIs= -0.02 to 0.007]) and neither did the implementation of upper or lower body exercises (Upper body – Lower body = -0.58 repetitions [95% CIs -2.32 to 1.16]). Further, there was minimal between participant variation in predictive accuracy (standard deviation = 1.45 repetitions [95% CIs = 0.99 to 2.12]). Conclusions: Participants were imperfect in their ability to predict proximity to task-failure independent of their training background. It remains to be determined whether the observed degree of inaccuracy should be considered acceptable. Despite this, prediction accuracies can be improved if they are provided closer to task-failure, when using heavier loads, or in later sets. To reduce the heterogeneity between studies, future studies should include a clear and detailed account of how task-failure was explained to participants and how it was confirmed.

Please cite as: Halperin, I., Malleron, T., Har-Nir, I., Androulakis-Korakakis, P., Wolf, M., Fisher, J., & Steele, J., (2021). Accuracy in predicting repetitions to task failure in resistance exercise: a scoping review and exploratory meta-analysis. https://doi.org/10.31236/osf.io/x256f

Introduction

Prescribing the number of repetitions to complete per set of exercise is a key variable in resistance-training programs. Traditionally, the number of repetitions are prescribed in a predetermined and fixed manner (e.g., three sets of 10 repetitions), using a certain percentage of one repetition maximum (1RM) [1-4]. While the traditional prescription approach is effective in achieving health and performance improvements [5-7], it has some limitations. Primarily, it does not adequately account for day-to-day performance variability between and within individuals. This is evident when considering the large variability in the maximum number of repetitions that participants can complete in different exercises, even when using the same percentage of 1RM [8-10], and the effects of various nuisance variables on performance (e.g., diet, sleep, motivation) [11, 12]. To illustrate, consider two participants attempting to complete as many repetitions as they can in the squat exercise using 70% of their 1RM. The first participant completed eight repetitions while the other completed 16. If they attempt to replicate their results on different days, the number of repetitions that both will be able to complete will likely decrease or increase as a function of different nuisance variables acting upon them on that day. If both were prescribed three sets of 10 repetitions, the first participant would struggle to complete the set even at maximal effort, while the latter participant would require comparatively much less effort. Hence, if both followed the traditional prescription approach, their individual abilities, and fluctuations in these abilities, would not have been sufficiently accounted for, which could lead to sub-optimal neuromuscular adaptations.

An alternative strategy is to prescribe the number repetitions relative to task-failure [13, 14]. For example, terminating a set when one predicts to be two or three repetitions shy of task-failure[1][15]. In recent years, this approach has been frequently studied using the Repetitions in Reserve (RIR) [16-19] and the Estimated Repetitions to Failure [20, 21] approaches. While some differences exist between them, we will use the term RIR to encapsulate both approaches. The main benefit of the RIR approach is that it attempts to control for the level of effort rather than the number of repetitions completed in a set [17, 22]. Effort can be defined as the process of investing resources to complete a task, relative to the available resources or current capacity or to complete the task [23, 24]. Since acute decrement in strength capacity (i.e. muscular fatigue) occurs during an ongoing set, completing each additional repetition requires a greater investment of effort. Accordingly, the proximity to task-failure in a given set is indicative of the effort required. For example, when lifting the same load, terminating a set three repetitions away from task-failure is more effortful than terminating that set six repetitions away from task-failure, regardless of the number of repetitions completed. Since there is no strict number of repetitions that participants must complete following the RIR approach as long as they reach a certain proximity to task-failure, between- and within-subject differences in abilities can be better accounted for.

An emerging number of studies investigated the longitudinal effects of following RIR on neuromuscular adaptations, with most reporting promising results [18, 25, 26]. However, for RIR strategies to be deemed effective, trainees must demonstrate an acceptable degree of accuracy when predicting proximity to task-failure. To illustrate, consider a participant that predicts to be two repetitions from task-failure during an ongoing set, but then completes eight repetitions before reaching task-failure. This scenario demonstrates an under-prediction error of six repetitions (i.e., actual minus predicted number of repetitions). While the degree of prediction accuracy required for effective implementation of RIR remains to be determined, developing an understanding of the prediction accuracy rates is an important step in the study of the RIR approach. A growing number of studies have examined participants RIR prediction accuracy [13, 20, 22, 27-36]. However, since these studies included a wide range of populations who lifted different loads and predicted task-failure at different stages of a set (e.g., before or during a set), a clear picture of prediction accuracy remains elusive. In view of the growing popularity of RIR and the importance of developing

[1] Here we refer to task-failure as an umbrella term that includes the inability to complete another repetition despite attempting to (i.e., momentary failure, or not attempting the next repetition assuming it could not be completed (i.e., repetition maximum (RM)). Note, this is in part due to the lack of clear definitions in the literature regarding repetition prescription if resistance training, and a critical discussion of set endpoint definitions will take place in later sections.

a clear understanding of the prediction accuracy when using RIR, the goal of the current meta-analysis was to investigate the prediction accuracy estimates across studies. Additionally, we examined if the following variables influence prediction accuracy: training status, timing of prediction, repetition range (indicative of relative load), set number, and upper or lower body exercise.

Methods

The systematic search and review was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. We based our search criteria on our familiarity with common RIR terminology alongside the use of search filter containing medical subject headings (MeSH). Two reviewers (IHN and TM) performed electronic searches on Google Scholar, PubMed/MEDLINE and SPORTDiscus, harvesting any data record until January 5th, 2021. The search included the following terms: ("resistance-train*" OR “resistance-exercise*" OR “strength-train*” OR “strength-exercise*” OR "weight-train*") AND (“estim*” OR “evaluate*” OR “predict*” OR “assess*” AND “proximity-to-failure” OR "rep*-in-reserve" OR "rir” OR “rep*-to-failure” OR “failure” OR “musc*-exhaus*” OR “musc*-fatigue”). We note that Google Scholar has a 256-character search limit, which forced us to reduce the number of included terms.

Studies were included if they met the following criteria: (1) the study was published in English and was either published in a peer-reviewed journal or as a MSc or PhD thesis; (2) participants had no known medical conditions or injuries; (3) the implemented modality was resistance-exercise; (4) participants had to predict proximity to task-failure before or during a set; (5) participants had to reach task-failure in all sets that they provided a prediction for. Two reviewers (IHN and TM) assessed relevant records, which were downloaded into Endnote (version 20, Clarivate Analytics, Philadelphia, PA, USA). All duplicates were removed before screening. To enable simultaneous screening of titles and abstracts by the reviewers, all potential records were uploaded to Abstrackr [37]. When an abstract indicated inclusion and an agreement was reached by both reviewers, the full text article was assessed for eligibility. Any disagreement regarding the eligibility that arose between the reviewers was settled by IH or if consensus was reached following further discussion. This search strategy was also duplicated internally to check for consistency by two additional reviewers (PAK and MW) who did so independently.

The following data were extracted from studies found to be eligible: title, participant’s characteristics of sample size, gender, age, exercises, sets, loads, timing of task-failure prediction (before or during the set); set endpoint definition, instructions, and prediction error (actual minus predicted repetitions to task-failure). The main datum we were looking to extract was prediction error and this was extracted for all groups and conditions within each study; thus, there were multiple predictions extracted for each included study in this analysis. In the case that it was not reported, one author (JS) emailed the authors of the manuscripts requesting the raw or mean values. If the authors did not reply within one month, we resorted to calculating the prediction errors based on the figures (data was digitized using WebPlotDigitizer; v4.3, Ankit Rohatgi; https://apps.automeris.io/wpd/) and tables. Data were extracted to a csv file for meta-analysis by JS, and a Word table which was finally edited by IH.

Statistical analysis

The exploratory meta-analysis was performed using the 'metafor' package in R (v 4.0.2; R Core Team, https://www.r-project.org/) [38]. All analysis code utilised is presented in the supplementary materials (https://osf.io/grynu/).

Studies were grouped by their design and reporting (i.e., whether they reported the paired actual and predicted repetitions, or the paired difference) for appropriate calculation of raw mean change scores sizes using the 'escalc' function in 'metafor' (see analysis code). We opted to analyse using the raw, as opposed to standardized, mean changes scores given that all effects were of the same construct and measurement: number of repetitions. We examined the difference between the actual repetition number performed to task-failure and the predicted repetition number (actual minus predicted number of repetitions). Scores were calculated such that positive values indicated that underprediction had occurred. That is, the number of repetitions predicted was smaller than the number of repetitions actually performed to momentary failure. For example, if a trainee predicted to have two more repetitions prior to task-failure, but then completed six, the person under-predicted by four repetitions. Correlations between predicted and actual repetitions were reported for most studies, and when absent we were usually able to obtain access to the raw data to enable their calculation. For those studies where we were unable to obtain these, we imputed the mean correlation from across those studies where this data was available as a reasonable estimate.

Because of the nested structure of the effect sizes calculated from the studies included (i.e. studies often had multiple groups and reported effects within these for multiple conditions), multilevel mixed-effects meta-analyses with both study and intra-study groups (i.e. where there were multiple groups within a given study) were included as random effects in the model were performed. Cluster robust point estimates and precision of those estimates using 95% compatibility (confidence) intervals (CIs) were produced [39], weighted by inverse sampling variance to account for the within- and between-study variance (tau-squared). Restricted maximal likelihood estimation was used in all models. A main model was produced including all effects; that is to say, for each condition performed by each group within each study. Thus, the models included all predictions made by all groups in each of the included studies. Considering the heterogeneity of study methods used in terms of the specific resistance training protocols for which predictions were made, and the experience of participants, this main model was merely interpreted generally as to whether it seemed people tended to over- or under-estimate repetitions to task-failure or whether they were fairly accurate in their predictions. Several exploratory meta-regression and sub-group analyses of moderators (i.e., predictors of effects) were also conducted to explore study protocols and participant characteristics. Moderators examined using meta-regression included mean resistance training history of participants in months, when the prediction was made as a percentage of the total number of repetitions performed to task-failure, the mean number of repetitions performed to task-failure, the set number for which the prediction was made, and whether upper or lower body exercise were used. We observed a non-linear effect of the mean number of repetitions performed to task-failure and so to model this in an interpretable manner we employed linear splines with a knot selected at 12 repetitions using the 'lspline' package [40]. A knot position of 12 repetitions was chosen as this is historically considered the upper end of the 'hypertrophy' repetition range before moving into the 'endurance' range [41]. This in essence meant separate regression models were fit between 0-12 repetitions, and >12 repetitions. Subgroup analyses consisted of a comparison between upper and lower body exercises. Multilevel models with robust estimates were produced for each subgroup, and fixed-effects with moderator's model used to compare the models.

Note, we were not able to obtain data to permit all studies to be included in all meta-regression or sub-group models and so indicate the number included when reporting this. Further, a small number of effects (n=5) reported in studies had zero variances. In supplementary analysis we re-ran models with these included after imputing a small constant variance to them $(3\,x\,10^{-7})$. This was to check that findings were not unduly influenced by their exclusion. The results of these models were not materially different (see https://osf.io/9dhzq/) and thus the main findings reported here are those with zero variance effects excluded to not unduly overwhelm the weighting of other effects in the models and meta-analytic scatterplots.

Lastly, we included additional exploratory models examining the between participant variation in predictive accuracy. For this, we fit the same models described above to estimates of the log transformed standard deviation (ln $\hat{\sigma}$) for prediction accuracy and its variance [42]. These models were then exponentiated back to their raw scales (i.e., standard deviations) for visualization purposes.

For all models, we opted to avoid dichotomizing the existence of an effect for the main results and therefore did not employ traditional null hypothesis significance testing, which has been extensively criticized [43, 44]. Instead, we considered the implications of all results compatible with these data, from the lower limit to the upper limit of the interval estimates, with the greatest interpretive emphasis placed on the point estimate.

The risk of small study bias was examined visually through contour-enhanced funnel plots. Q and I^2 statistics were also produced and reported [45]. A significant Q statistic is typically considered indicative of effects likely not being drawn from a common population. I^2 values indicate the degree of heterogeneity in the effects: 0-40% were not important, 30-60% moderate heterogeneity, 50-90% substantial heterogeneity, and 75-100% considerable heterogeneity [46]. Further, by way of a sensitivity analysis we also replicated all models omitting any predictions made prior to the set initiating the results of which were not materially different and are included in (see https://osf.io/9yahn/ and https://osf.io/tnuwz/).

Results

Included studies

After initial searches and screening, 13 publications that cover 12 studies were identified that met the inclusion criteria. Specifically, two publications reported the same data on some of the outcomes [20, 27] but one of them [20] included additional data that was not reported in the other [27]. The same data was only used once for the analysis. Additional search approaches identified no further studies that met the inclusion criteria. Thus, the final number of studies included was 13 [13, 20, 22, 27-36]. Details of the search and inclusion process are shown in the PRISMA flow chart (Figure 1). Details of the studies can be viewed in Table 1. The Author's description of the prediction process and the set end point definitions are listed in Table 2 in the supplementary materials (https://osf.io/2fwue/). The pooled number of participants in the studies included was 414 across 25 groups within studies and with sample sizes ranging from 6 to 53 participants (median = 14) per group within each study. Full details of all included studies can be seen in the data extraction table (https://osf.io/6sc72/).

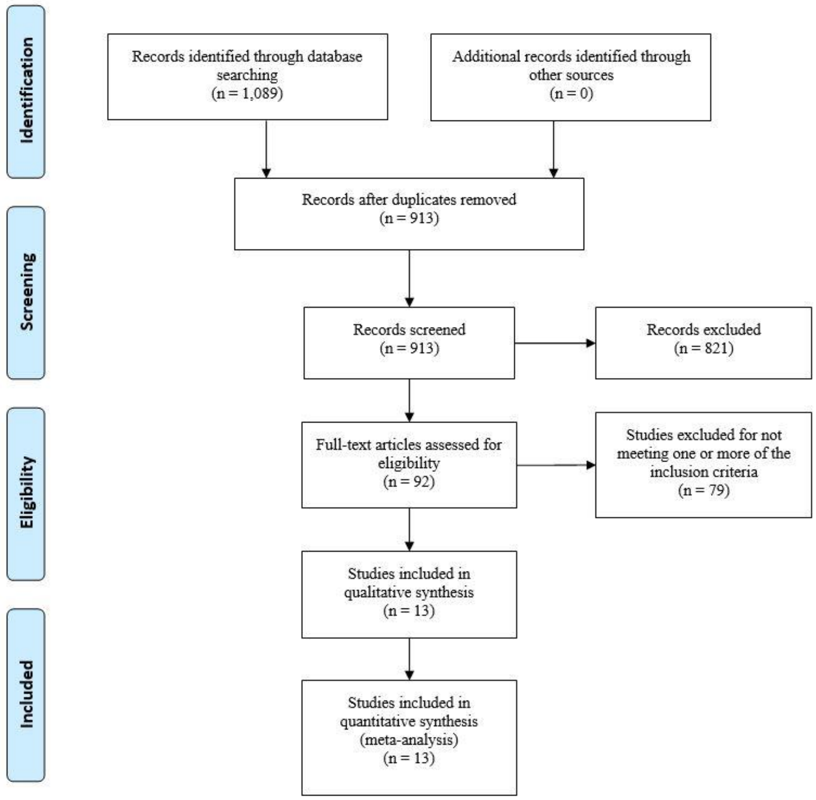

Figure 1. PRISMA flow chart illustrating different phases of the search and study selection.

Table 1. Summary of the methods and characteristics of the included studies. Note that age and training experience are presented with mean±SD.

Article	Participants (age in years)	Training Experience	Prediction timing	Sets	Load	Exercise
Hackett et al. 2012 [13]	17 M 32±5	8.2±3.2 yrs	10th rep	5	70% 1RM	Bench Press Back Squat
Lemos et al. 2017 [30]	11 F 22±1	unclear	2nd rep	1	50% 1RM 70% 1RM 90% 1RM	Chest Fly Leg Extension Pull-down Leg Curl Biceps Curl Triceps Extension Military Press
Servais, 2015 [32]	12 F 20±1 12 M 22±1	4±2 yrs	Before the set	5	65% 1RM	Bench Press
Hackett et al. 2017 [27]	28 F 28±9.5 53 M 27.3±9	3.6±4.6 yrs 5.5±6.1 yrs	10th rep	10	70% 1RM 80% 1RM	Chest Press Leg Press
Steele et al. 2017 [28]	69 F 25±8 72 M 29±10	1.5 months to 3 yrs	Before the set	1	Self-estimated 10RM	Seated row Bench press Leg press Elbow flexion Pull down Sit up
Hackett et al. 2018 [20]	21 F 30.4 27 M 26.6	4 yrs 6.7 yrs	10th rep	3	70% 1RM 80% 1RM	Chest Press Leg Press
Odgers, 2018 [33]	13 F 30±5.4 14 M	≥6 months	6 RPE 9 RPE	4	80% 1RM	Front Squat Hex Deadlift

	29±5.7					
Sousa, 2018 [29]	10 M 25±4	6±4 yrs	4 RIR 1 RIR	4	80% 1RM	Back Squat Bench Press Deadlift Back Squat Bench Press Deadlift
Ratto et al. 2019 [31]	20 M 20±2	4.7 yrs	Pre-set 4th rep 8th rep 12th rep	1	100 KG	Bench Press
Zourdos, 2021 [22]	25 M 25±3	5±3 yrs	5 RIR 3 RIR 1 RIR	1	70% 1RM	Back Squat
Emanuel et al. 2020 [35]	10 M 29.5±4	≥1 yrs	Before the set	2	70% 1RM 83% 1RM	Bench Press Back Squat Bench Press Back Squat
Mansfield et al. 2020 [34]	20 M 26±4	6±4 yrs	8th rep 3rd rep	3	60% 1RM 80% 1RM	Bench Press Prone Row Bench Press Prone Row
Hackett et al. 2021[36]	20 M 26.3±9	7±4.7 yrs	10th rep	5	70% 1RM	Bench Press Back Squat

MF: Momentary Failure; **RM**: Repetition Maximum; **RIR**: Repetition in reserve; **RPE**: Rate of Perceived Exertion (the authors used a 0-10 category ratio scale).

Main model – All effects

The main model including all effects sizes (262 across 12 clusters [median =11.5, range = 3 to 60 effects per cluster]) suggests that participants, on average, underpredicted with a point and interval estimate of 0.95 repetitions [95% CIs = 0.17 0to 1.73]. There was however considerable heterogeneity ($Q_{(261)}$ = 3060.90, p < 0.0001; I^2 = 97.89%). Figure 2 presents all effect sizes and interval estimates in an ordered caterpillar plot. Figure 3 presents the funnel plot of all effects.

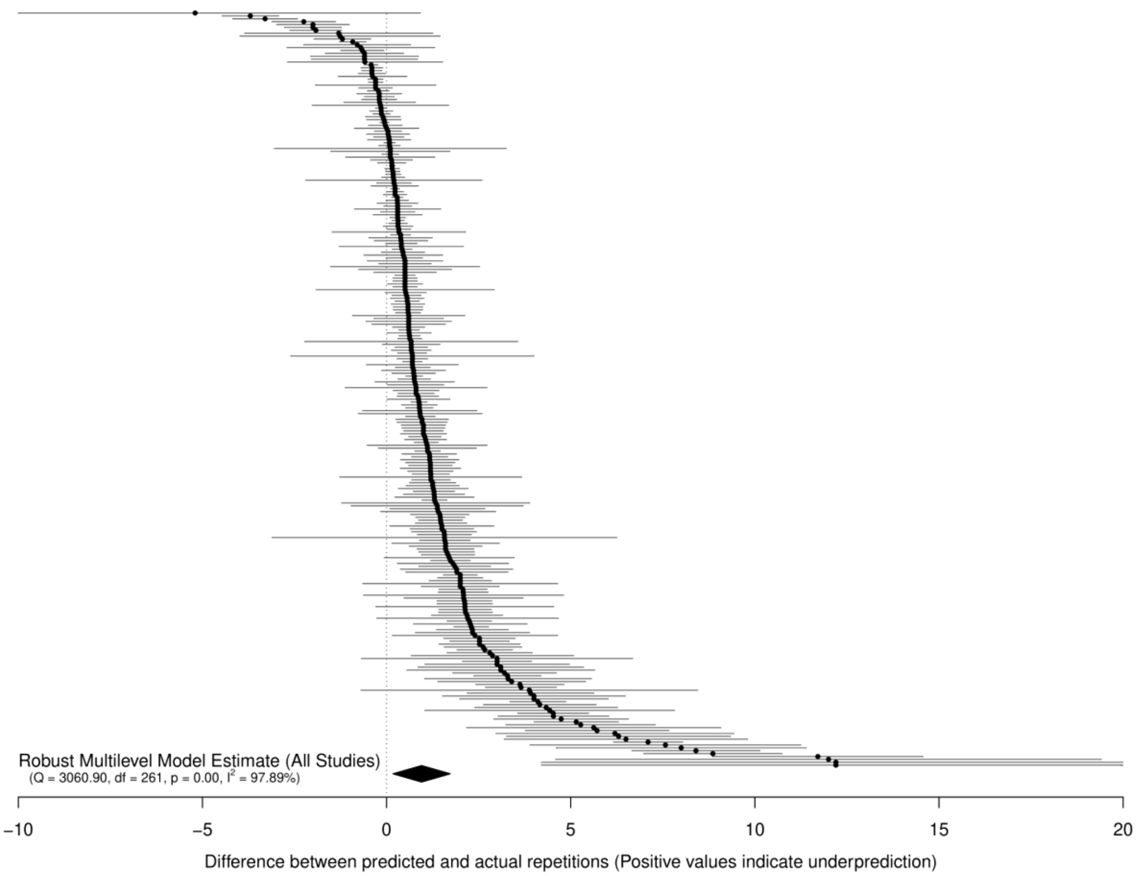

Figure 2. Ordered caterpillar plot of all effects.

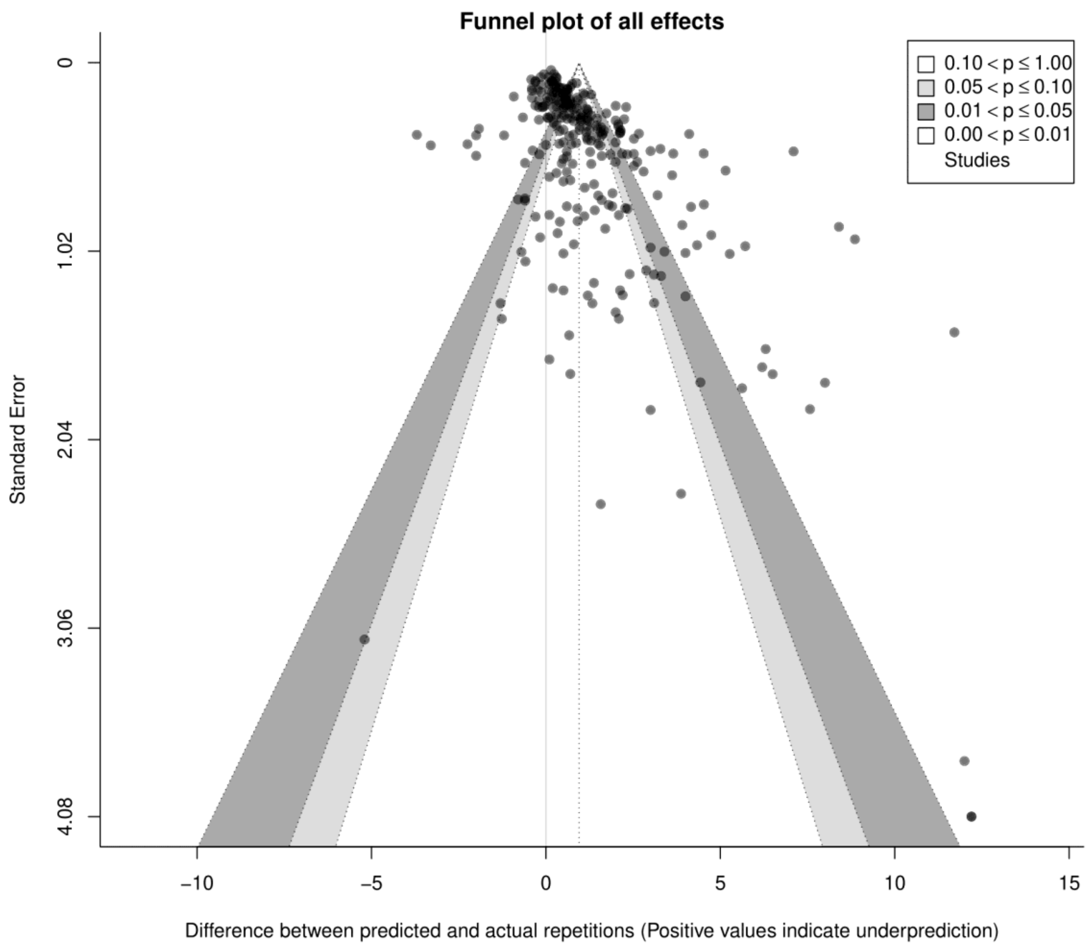

Figure 3. Contour-enhanced funnel plot for all effects.

Exploratory meta-regression analyses

Training history

Meta-regression suggested that prediction accuracy was not moderated much by the mean training history (in months) of participants in the samples (β = -0.0061 repetitions [95% CIs = -0.0195 to 0.0073]; 262 across 12 clusters [median =11.5, range = 3 to 60 effects per cluster]). There was however considerable heterogeneity (I^2 = 97.64%). Figure 4 shows the meta-analytic scatter plot for this analysis.

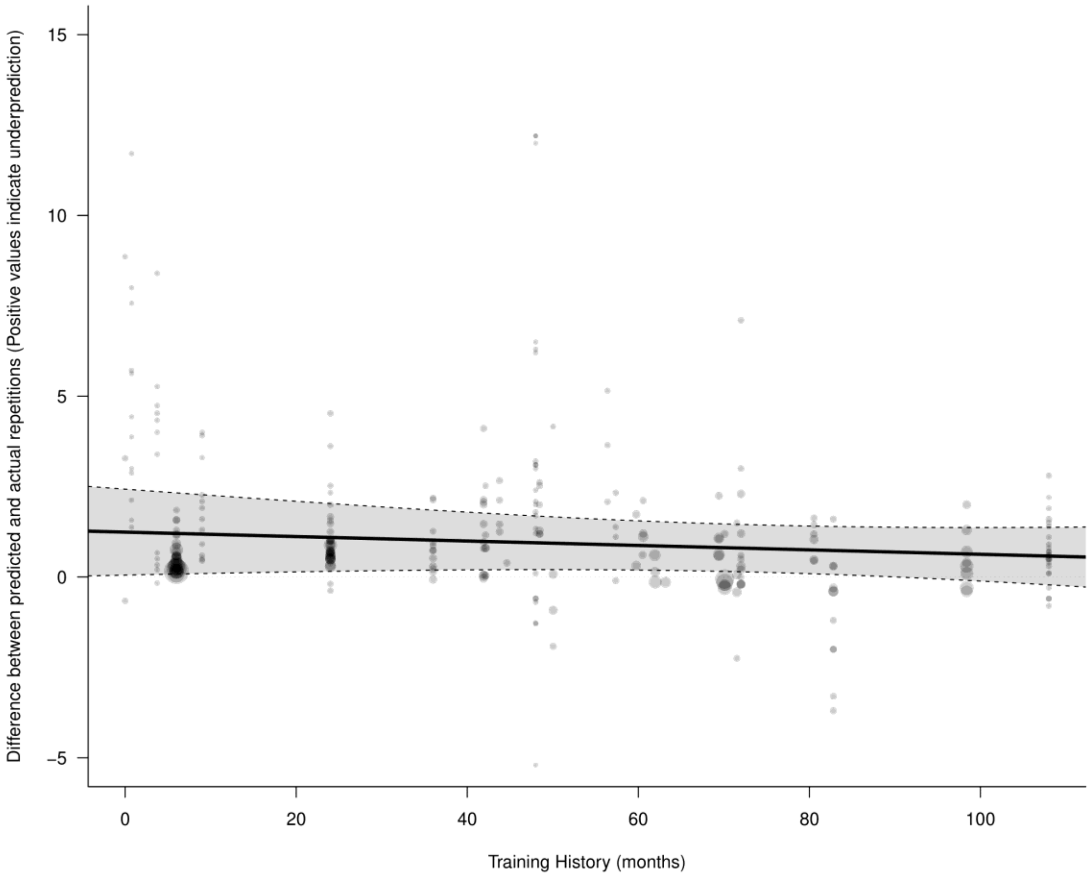

Figure 4. Meta-analytic scatter plot of training history and predictive ability.

When prediction was made

Meta-regression suggested that prediction accuracy was moderated by how close to TF participants were when they made their prediction (expressed as a percentage of total repetitions performed to TF). Accuracy increased slightly as predictions were made with closer proximity to TF (β = -0.025 repetitions [95% CIs = -0.051 to -0.001]; 238 across 11 clusters [median= 11, range= 3 to 60 effects per cluster]). There was however considerable heterogeneity (I^2 = 97.90%). Figure 5 shows the meta-analytic scatter plot for this analysis.

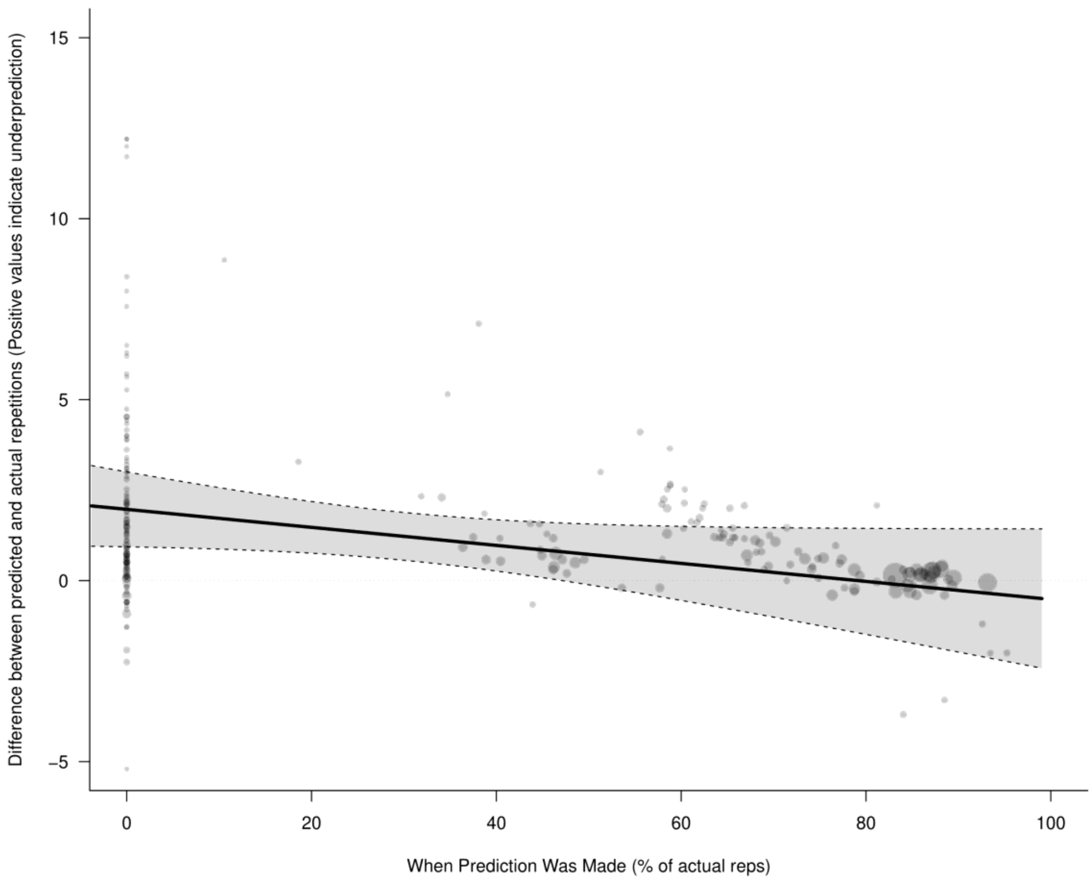

Figure 5. Meta-analytic scatter plot of when prediction was made and predictive ability.

Repetition range

Meta-regression suggested that prediction accuracy was trivially moderated by the repetition ranges performed up to 12 repetitions (first spline) but was strongly moderated by repetition ranges that included 12 or more repetitions (second spline). For the first linear spline, accuracy did not change much with performing fewer repetitions, but the second linear spline revealed that accuracy decreased as predictions in sets composed of higher repetition range (first linear spline [≤12 repetitions], β = 0.06 repetitions [95% CIs = -0.04 to 0.16]; second linear spline [>12 repetitions, β = 0.47 repetitions [95% CIs = 0.35 to 0.58]; 238 across 11 clusters [median =11, range = 3 to 60 effects per cluster]). There was however considerable heterogeneity (I^2 = 88.25%). Figure 6 shows the meta-analytic scatter plot for this analysis.

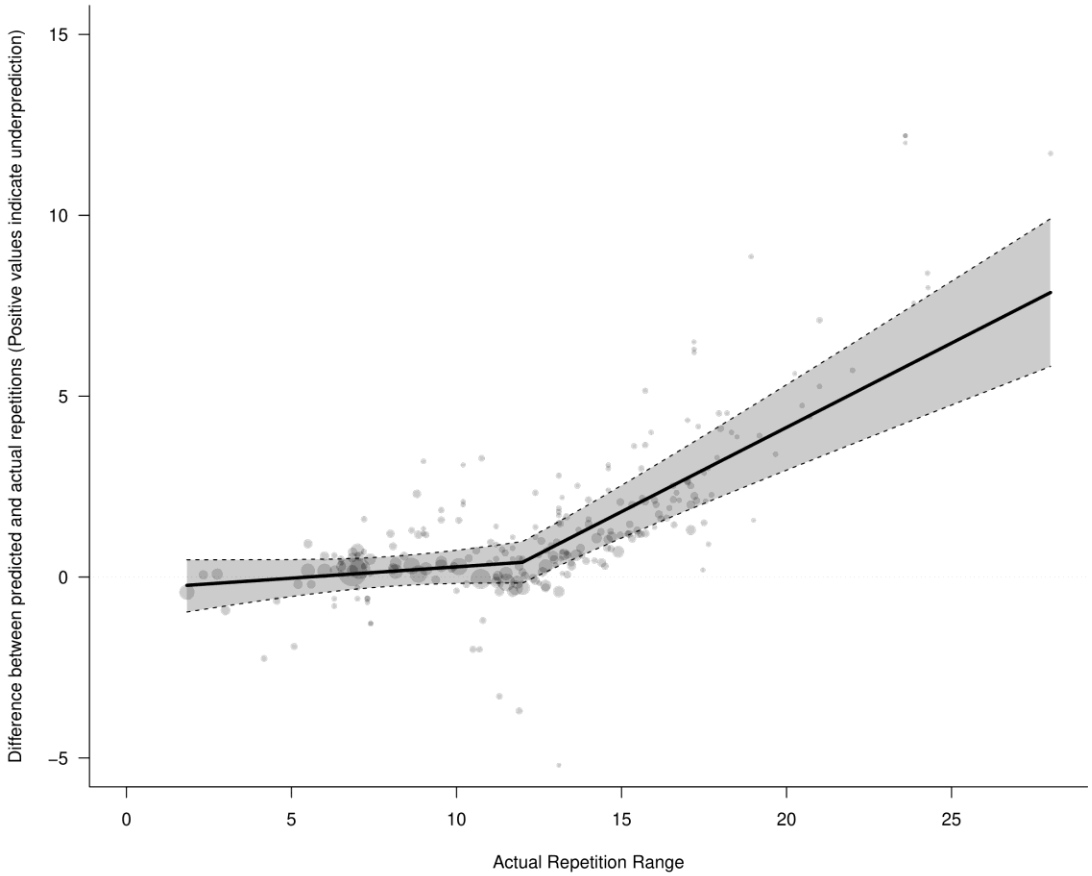

Figure 6. Meta-analytic scatter plot of the repetition ranges performed to TF and predictive ability.

Set number

Meta-regression suggested that prediction accuracy was trivially moderated by which set number the prediction was made on (β = -0.072 repetitions [95% CIs = -0.14 to -0.005]; 262 across 12 clusters [median = 11, range = 3 to 60 effects per cluster]). There was however considerable heterogeneity (I^2= 97.76%). Figure 7 shows the meta-analytic scatter plot for this analysis.

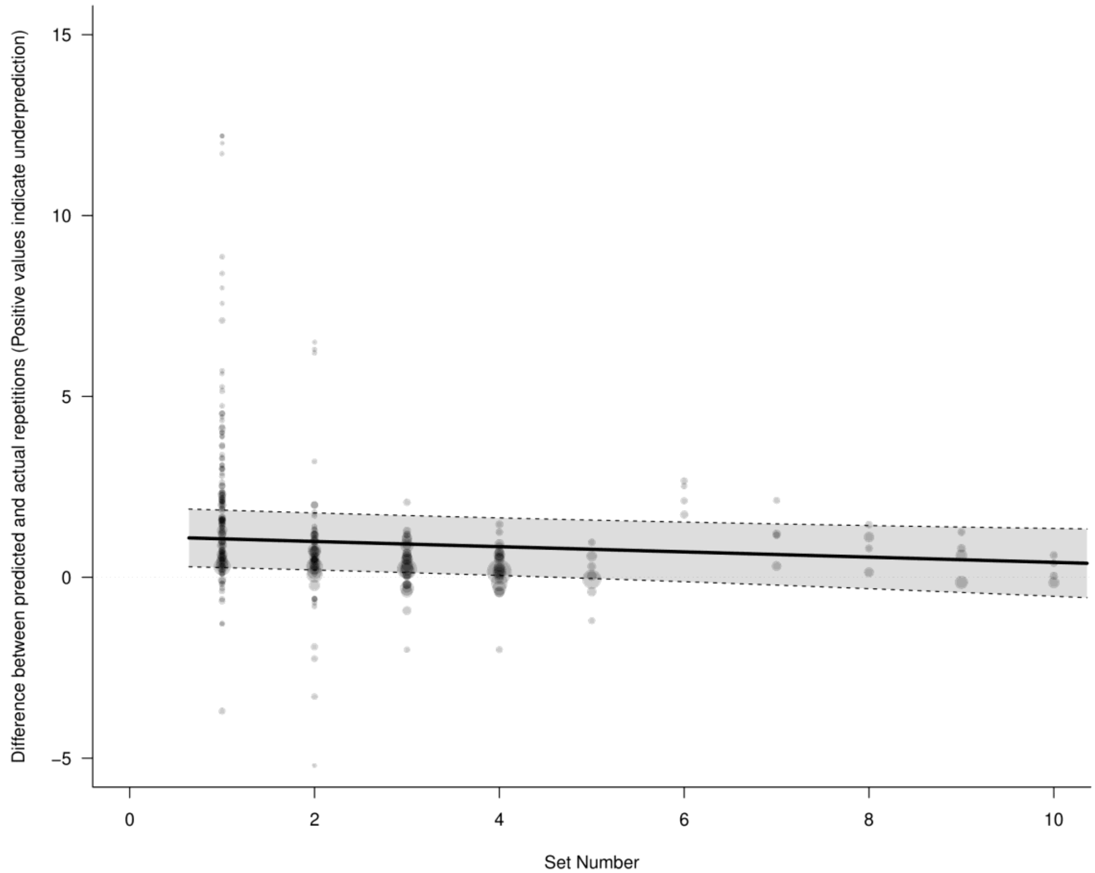

Figure 7. Meta-analytic scatter plot of set number and predictive ability.

Upper body vs lower body exercises

Subgroup models revealed prediction accuracy was slightly worse for lower body (1.51 repetitions [95% CIs = -0.38 to 3.40]; 118 effects across 8 clusters [median = 13, range = 2 to 29 effects per cluster; I^2 = 99.48%) compared to upper body effects (0.92 repetitions [95% CIs = 0.09 to 1.75]; 131 effects across 9 clusters [median = 8, range = 4 to 40 effects per cluster; I^2 = 97.29%]), but between model comparison suggested the difference was unclear and the estimate imprecise (Upper body – Lower body = -0.59 repetitions [95% CIs -2.30 to 1.13]).

Exploratory analysis of between participant variation

The main model for log transformed standard deviations in predictive accuracy revealed a relatively low between participant variation (ln $\hat{\sigma}$ = 0.37 [95% CIs = -0.01 to 0.75]) which when exponentiated was 1.45 repetitions [95% CIs = 0.99 to 2.12]. In general, the pattern of moderator effects was similar to that found in our exploratory meta-regressions of predictive accuracy. That is to say, between participant variation was lowest when predictions were made closer to task-failure, when fewer repetitions were performed per set, and in later sets. All outputs for models examining log transformed standard deviations are available in the supplementary materials (https://osf.io/7kx9e/) in addition to plots

Discussion

In this scoping review and meta-analysis we explored participant's prediction accuracy when following the RIR approach in resistance training. Overall, across studies, participants under-predicted proximity to task-failure by roughly one repetition. Prediction accuracy improved when predictions were made closer to task-failure, when fewer repetitions per set were completed, and in later sets. Conversely, and somewhat surprisingly, training status did not seem to influence prediction accuracy, nor was there much difference between upper or lower body exercises. Further, there was relatively minimal between-participant variation in predictive accuracy suggesting that the primary source of error is due to systematic underprediction.

It is not entirely clear whether the underprediction of proximity to task-failure of approximately one repetition is large enough to be considered meaningful. Mainly, the prediction error of one repetition in the context of the total number of repetitions completed per set can considerably impact the interpretation. To illustrate, an under-prediction error of one repetition can be considered small in a set composed of 20 repetitions (5% error), and large in a set composed of five repetitions (20% error). Most of the reviewed studies included more than ten repetitions per set (average of 12.6 repetitions) which can partly assist in framing and interpreting this result. While it is difficult to interpret the direction and magnitude of this prediction error, developing a deeper understanding of it can help in designing, interpreting and comparing studies. For example, future research examining the dose-response relationships of different proximities to task-failure may benefit from knowing the magnitude of prediction errors [47-50].

The finding that prediction accuracy improved when the predictions were provided towards the end of a set composed of fewer repetitions and in later sets is logical. This is because predictions early in a set coupled with a performing a greater number of repetitions allows for a wider range of errors to be made, in contrast to predictions later in a set coupled completing fewer repetitions. Further, prediction accuracy in later sets may improve due to either a practice element, or that lingering fatigue means that the sets are performed with relatively greater loads [27, 35, 36]. Moreover, it is often assumed that predictions of proximity to task-failure are made based upon either remembered or presently experienced perception of effort [16, 51]. However, there is the potential for other salient experiences, such as discomfort, to be conflated with perception of effort and influence one's prediction [52-54]. This possibility may explain the general under-prediction found across studies, and also that predictions worsened with lower loads and thus higher repetition ranges. Lower loads performed to task-failure typically elicit greater perceptions of discomfort [52, 53]. Practically, if using RIR, participants should predict proximity to task-failure as the set unfolds, rather than before it begins, for better prediction accuracies. Sets of lower repetitions, which are commonly associated with heavier loads, will also lead to better prediction accuracies. The fact that training status did not impact prediction accuracy comes as a surprise as prior studies that included participants with a range of training histories have typically shown that training background is associated with improved predictive accuracy [28, 55]; though in one study this may have also been due to more trained participants utilizing heavier loads/lower repetitions [28]. The large heterogeneity between studies may partly explain these trivial effects.

In conducting this scoping review we have identified a number methodological issues that warrant a discussion. The main one is that clear distinction between RM and momentary-failure is not always present (see Table 2 in the supplementary materials; https://osf.io/2fwue/). For example, in the Estimated Repetition to Failure scale used in studies by Hackett and colleagues [27], the following is stated in the scale's instructions *"…"0" is where the subject estimated no additional repetitions could be completed (concentric failure reached)"*. Considering the operational definitions of RM and momentary-failure provided by Steele et al [15] this explanation does not clearly differentiate between RM (the sentence before the parentheses) with momentary-failure (the part in the parentheses). If participants predict that no additional repetitions can be completed at the point of task-failure, then it can be assumed that the last repetition was successful. As such, the last repetition should be defined as RM. However, if concentric failure was achieved, then the final repetition should be defined as momentary-failure. If the definition of the task-failure they are trying to predict is not clearly explained to participants, larger prediction errors can be expected.

There are also inconsistencies between studies in which of the two set endpoints were used to represent task-failure. This is the case with both scale instructions and the criteria used to define task-failure. Ideally, in order to achieve higher prediction accuracies, sets that end with momentary-failure are superior to those ending with RM. This is because reaching momentary-failure is by definition the point in which no more repetitions can be completed, whereas RM is by definition an unverified prediction that the subsequent repetition cannot be completed. For example, a trainee who assumes to have reached the point of

RM, may be able to complete three more repetitions before reaching momentary-failure. Hence, there is more room for predication error unless momentary-failure is achieved. We acknowledge that requesting and ensuring that participants reach momentary-failure is not a simple task. It can be argued that it is impossible to truly verify whether momentary-failure was achieved and for ethical reasons participants cannot be forced to reach momentary-failure. The inconsistent task-failure anchors in the RIR studies could have biased the estimates in this meta-analysis as we treated task-failure to be similar across studies. Future studies should consider how task-failure is explained to participants and include a detailed account of the instructions, and how RM or momentary-failure were defined and monitored [15]. In studies that include both RM and momentary-failure as task-failure, including the ratio of sets that ended with either task-failure utilising participant self-reports and experimenter's observation, may assist explaining dissimilar results between studies.

Another methodological issue in the literature is the anchoring bias that arises when participants provide their task-failure prediction. That is, once participants report their task-failure prediction, it is possible that they set this particular number as the goal of the set. In a sense, a self-fulfilling prophecy. It is possible that if participants did not provide a prediction they would have completed more or less repetitions. Since all of the analyzed studies included in our meta-analysis suffered from this anchoring bias, the observed estimates may be smaller than what they truly are. To overcome the anchoring bias, Armes et al [56], used a deception design where participants completed sets of knee extension to both RM and momentary-failure, and were told that the purpose of the study was to inspect the reliability of their performance across trials. However, the true purpose of this study was to examine their predictive ability By bypassing the effects of the anchoring bias, the authors observed an average underprediction error of two repetitions in an internal meta-analysis of their experiments. Hence, implementing such deception designs may reveal that predictive abilities are in fact worse than the present estimate suggests.

Despite the methodological limitations of RIR literature discussed above, the RIR approach has benefits. Mainly, prescribing repetitions relative to task-failure may help to ensure that a consistent effort is reached in a given set, even if the number of repetitions are different between and within participants, sets, and exercises [16, 18, 21, 57]. In contrast, prescribing a fixed and predetermined number of repetitions using specific percentage of 1RM accounts for considerably less variability in one's abilities [58, 59]. As such, the prediction errors identified in the present study coupled with methodological concerns of RIR approaches should be viewed and weighted relative to the alternative prescription approaches. Moreover, studies comparing between sets taken to task-failure and not to task-failure on various outcomes can benefit from implementing RIR approaches in their designs. This is because such studies implement a binary task-failure and not to task-failure approach in which proximity to task-failure is not accounted for [49, 50]. By comparing groups that follow different RIR set endpoints (e.g., momentary-failure vs. 1RIR vs. 2RIR), richer and more insightful comparisons can be made. In order to strengthen RIR designs, future studies should consider the task-failure they are using and provide a clear and detailed account of how it was explained and confirmed. Additionally, including the ratio of sets that ended due to momentary-failure and RM may assist explaining different study results. Lastly, the use of deception-based designs may overcome the anchoring biases and result in better estimates of participant's prediction accuracy.

In conclusion, we found that participants typically underpredict proximity to task-failure by approximately one repetition. However, it is unclear whether this degree of underprediction represents acceptable prediction accuracy. Practitioners and trainees choosing to apply RIR based techniques can improve prediction accuracy by providing the prediction towards the end of sets composed of fewer repetitions, and in later sets. Participants background in resistance training did not seem to meaningfully impact prediction accuracy, nor were there differences between upper and lower body exercises.

Funding information
No funding was received in support of this research.

Data and Supplementary Material Accessibility
All materials, data, and code are available on the Open Science Framework project page for this study https://osf.io/jzuwq/

Author contributions
IH and JS wrote the first draft of the manuscript. TM, IHN, PAK and MW performed the literature search. JS performed the meta-analyses. All authors were involved in the interpretation of the meta-analyses, read, revised, and approved the final manuscript.

REFERENCES
1. Wilkins LW. ACSM's health-related physical fitness assessment manual (4th ed). Philadelphia ;2013.
2. Sands WA, Wurth JJ, Hewit JK. Basics of strength and conditioning manual. Colorado Springs, CO: National Strength and Conditioning Association. 2012.
3. Ratamess NA, Alvar BA, Evetoch TE, Housh TJ, Ben Kibler W, Kraemer WJ, et al. Progression models in resistance training for healthy adults. Med Sci Sports Exerc. 2009;41(3):687-708.

4. Hass CJ, Feigenbaum MS, Franklin BA. Prescription of resistance training for healthy populations. Sports Med. 2001;31(14):953-64.
5. Westcott WL. Resistance training is medicine: effects of strength training on health. Curr Sports Med Rep. 2012;11(4):209-16.
6. Kraemer WJ, Ratamess NA, French DN. Resistance training for health and performance. Curr Sports Med Rep. 2002;1(3):165-71.
7. Winett RA, Carpinelli RN. Potential health-related benefits of resistance training. Prev Med. 2001;33(5):503-13.
8. Richens B, Cleather DJ. The relationship between the number of repetitions performed at given intensities is different in endurance and strength trained athletes. Biol Sport. 2014;31(2):157-61.
9. Shimano T, Kraemer WJ, Spiering BA, Volek JS, Hatfield DL, Silvestre R, et al. Relationship between the number of repetitions and selected percentages of one repetition maximum in free weight exercises in trained and untrained men. J Strength Cond Res. 2006;20(4):819-23.
10. Hoeger WW, Barette SL, Hale DF, Hopkins DR. Relationship between repetitions and selected percentages of one repetition maximum. J Strength Cond Res. 1987;1(1):11-3.
11. Grgic J, Trexler ET, Lazinica B, Pedisic Z. Effects of caffeine intake on muscle strength and power: a systematic review and meta-analysis. J Int Soc Sports Nutr. 2018;15(1):1-10.
12. Knowles OE, Drinkwater EJ, Urwin CS, Lamon S, Aisbett B. Inadequate sleep and muscle strength: Implications for resistance training. J Sci Med Sport. 2018;21(9):959-68.
13. Hackett DA, Johnson NA, Halaki M, Chow CM. A novel scale to assess resistance-exercise effort. J Sports Sci. 2012;30(13):1405-13.
14. Zourdos MC, Klemp A, Dolan C, Quiles JM, Schau KA, Jo E, et al. Novel Resistance Training-Specific Rating of Perceived Exertion Scale Measuring Repetitions in Reserve. J Strength Cond Res. 2016;30(1):267-75.
15. Steele J, Fisher J, Giessing J, Gentil P. Clarity in reporting terminology and definitions of set endpoints in resistance training. Muscle & nerve. 2017;56(3):368-74.
16. Helms ER, Cronin J, Storey A, Zourdos MC. Application of the Repetitions in Reserve-Based Rating of Perceived Exertion Scale for Resistance Training. Strength Cond J. 2016;38(4):42-9.
17. Helms ER, Byrnes RK, Cooke DM, Haischer MH, Carzoli JP, Johnson TK, et al. RPE vs. Percentage 1RM loading in periodized programs matched for sets and repetitions. Front Physiol. 2018;9:247.
18. Graham T, Cleather DJ. Autoregulation by "Repetitions in Reserve" Leads to Greater Improvements in Strength Over a 12-Week Training Program Than Fixed Loading. J

Strength Cond Res. 2019;doi: 10.1519/JSC.0000000000003164. Epub ahead of print. PMID: 31009432.
19. Shattock K, Tee JC. Autoregulation in Resistance Training: A Comparison of Subjective Versus Objective Methods. J Strength Cond Res. 2020;doi: 10.1519/JSC.0000000000003530 Publish Ahead of Print.
20. Hackett DA, Cobley SP, Halaki M. Estimation of repetitions to failure for monitoring resistance exercise intensity: Building a case for application. J Strength Cond Res. 2018;32(5):1352-9.
21. Kemmler W, Kohl M, Fröhlich M, Jakob F, Engelke K, von Stengel S, et al. Effects of High-Intensity Resistance Training on Osteopenia and Sarcopenia Parameters in Older Men with Osteosarcopenia—One-Year Results of the Randomized Controlled Franconian Osteopenia and Sarcopenia Trial (FrOST). J Bone Miner Res. 2020;35(9):1634-44.
22. Zourdos MC, Goldsmith JA, Helms ER, Trepeck C, Halle JL, Mendez KM, et al. Proximity to Failure and Total Repetitions Performed in a Set Influences Accuracy of Intraset Repetitions in Reserve-Based Rating of Perceived Exertion. J Strength Cond Res. 2021;35:S158-S65.
23. Halperin I, Emanuel A. Rating of perceived effort: methodological concerns and future directions. Sports Med. 2020;50(4):679-87.
24. Steele J. What is (perception of) effort? Objective and subjective effort during task performance. PsyArXiv. 2021;doi: 10.31234/osf.io/kbyhm Epub ahead of print.
25. Arede J, Vaz R, Gonzalo-Skok O, Balsalobre-Fernandéz C, Varela-Olalla D, Madruga-Parera M, et al. Repetitions in reserve vs. maximum effort resistance training programs in youth female athletes. J Sports Med Phys Fitness. 2020;60(9):1231-9.
26. Buskard ANL JK, Eltoukhy MM, Strand KL, Villanueva L, Desai PP, Signorile JF. Optimal Approach to Load Progressions during Strength Training in Older Adults. Med Sci Sports Exerc. 2019;51(11):2224-33.
27. Hackett DA, Cobley SP, Davies TB, Michael SW, Halaki M. Accuracy in estimating repetitions to failure during resistance exercise. J Strength Cond Res. 2017;31(8):2162-8.
28. Steele J, Endres A, Fisher J, Gentil P, Giessing J. Ability to predict repetitions to momentary failure is not perfectly accurate, though improves with resistance training experience. PeerJ. 2017;5:e4105.
29. Sousa CA. Assessment of Accuracy of Intra-set Rating of Perceived Exertion in the Squat, Bench Press, and Deadlift [Master's theses]: Florida Atlantic University Boca Raton, FL; 2018.
30. Lemos EA, Caldas LC, Leopoldo APL, Leopoldo AS, Ferreira LG, Lunz W. The perception of effort is not a valid tool for establishing the strength-training zone. J Hum Sport Exerc. 2017;12(3):593-606.

31. Ratto AG. Application of the predicted repetitions-to-failure perceived exertion scale for the NFL-225lb bench press test [Master's theses];. Humboldt State University. 2019.
32. Servais B. Regulating resistance exercise intensity using perceptual response and the “anticipatory feedback” model [Master's theses]: Humboldt State University; 2015.
33. Odgers JB. Exertion/Velocity Profiling and Assessment of Accuracy of Intra-set Rating of Perceived Exertion in the Front Squat and Hexagonal Bar Deadlift [Master's theses]: Faculty of Graduate Studies and Research, University of Regina; 2020.
34. Mansfield SK, Peiffer JJ, Hughes LJ, Scott BR. Estimating Repetitions in Reserve for Resistance Exercise: An Analysis of Factors Which Impact on Prediction Accuracy. J Strength Cond Res. 2020;doi: 10.1016/j.jshs.2021.01.007. Epub ahead of print. PMID: 33497853.
35. Emanuel A, Rozen S, II, Halperin I. The Effects of Lifting Lighter and Heavier Loads on Subjective Measures. Int J Sports Physiol Perform. 2020;16(2):176-83.
36. Hackett DA. Influence of Movement Velocity on Accuracy of Estimated Repetitions to Failure in Resistance-Trained Men. J Strength Cond Res. 2021;doi: 10.1519/JSC.0000000000003978. Epub ahead of print. PMID: 33938720.
37. Wallace BC, Small K, Brodley CE, Lau J, Trikalinos TA. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. Proceedings of the 2nd ACM SIGHIT international health informatics symposium 2012. p. 819-24.
38. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36(3):1-48.
39. Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. Res Synth Methods. 2010;1(1):39-65.
40. Bojanowski M. lspline: Linear Splines with Convenient Parametrisations. R package version. 10-0 https://CRANR-projectorg/package=lspline. 2017.
41. Schoenfeld BJ, Grgic J, Van Every DW, Plotkin DL. Loading Recommendations for Muscle Strength, Hypertrophy, and Local Endurance: A Re-Examination of the Repetition Continuum. Sports (Basel). 2021 Feb 22;9(2).
42. Nakagawa S, Poulin R, Mengersen K, Reinhold K, Engqvist L, Lagisz M, et al. Meta-analysis of variation: ecological and evolutionary applications and beyond. Methods Ecol Evol. 2015;6(2):143-52.
43. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon Statistical Significance. Am Stat. 2019;73(sup1):235-45.
44. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature. 2019;567(7748):305-7.
45. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Bmj. 2003;327(7414):557-60.

46. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions Version 5.1. 0 (updated March 2011). London, UK: The Cochrane Collaboration. 2011.
47. Giebetasing J, Fisher J, Steele J, Rothe F, Raubold K, Eichmann B. The effects of low-volume resistance training with and without advanced techniques in trained subjects. J Sports Med Phys Fitness. 2016;56(3):249-58.
48. Giessing J, Eichmann B, Steele J, Fisher J. A comparison of low volume 'high-intensity-training' and high volume traditional resistance training methods on muscular performance, body composition, and subjective assessments of training. Biol Sport. 2016;33(3):241-9.
49. Grgic J, Schoenfeld BJ, Orazem J, Sabol F. Effects of resistance training performed to repetition failure or non-failure on muscular strength and hypertrophy: A systematic review and meta-analysis. J Sport Health Sci. 2021;S2095-2546(21)00007-7. doi: 10.1016/j.jshs.2021.01.007. Epub ahead of print. PMID: 33497853.:2095-546.
50. Vieira AF, Umpierre D, Teodoro JL, Lisboa SC, Baroni BM, Izquierdo M, et al. Effects of Resistance Training Performed to Failure or Not to Failure on Muscle Strength, Hypertrophy, and Power Output: A Systematic Review With Meta-Analysis. J Strength Cond Res. 2021;35(4):1165-75.
51. Coquart JB, Eston RG, Noakes TD, Tourny-Chollet C, L'Hermette M, Lemaître F, et al. Estimated time limit: a brief review of a perceptually based scale. Sports Med. 2012;42(10):845-55.
52. Stuart C, Steele J, Gentil P, Giessing J, Fisher JP. Fatigue and perceptual responses of heavier-and lighter-load isolated lumbar extension resistance exercise in males and females. PeerJ. 2018;6:e4523.
53. Fisher JP, Farrow J, Steele J. Acute fatigue, and perceptual responses to resistance exercise. Muscle & Nerve. 2017;56(6):E141-E6.
54. Emanuel A, Smukas IIR, Halperin I. An analysis of the perceived causes leading to task-failure in resistance-exercises. PeerJ. 2020;8:e9611.
55. Hackett DA, Cobley SP, Davies TB, Michael SW, Halaki M. Accuracy in estimating repetitions to failure during resistance exercise. J Strength Cond Res. 2017;31(8):2162-8.
56. Armes C, Standish-Hunt H, Androulakis-Korakakis P, Michalopoulos N, Georgieva T, Hammond A, et al. “Just one more rep!”–Ability to predict proximity to task failure in resistance trained persons. Front Psychol. 2020;11:3760.
57. Fairman CM, Zourdos MC, Helms ER, Focht BC. A scientific rationale to improve resistance training prescription in exercise oncology. Sports Medicine. 2017;47(8):1457-65.
58. Phillips SM, Winett RA. Uncomplicated resistance training and health-related outcomes: evidence for a public health mandate. Curr Sports Med Rep. 2010;9(4):208-13.

59. Steele J, Fisher J, Skivington M, Dunn C, Arnold J, Tew G, et al. A higher effort-based paradigm in physical activity and exercise for public health: making the case for a greater emphasis on resistance training. BMC Public Health. 2017;17(1):1-8.