SportRXiv

Part of the Society for Transparency, Openness and Replication in Kinesiology (STORK)

Preprint
not peer reviewed

The role of supervision in resistance training; an exploratory systematic review and meta-analysis

Received: 18th Oct 2021
Supplementary materials: https://osf.io/mu8zf/
For correspondence: james.fisher@solent.ac.uk
Twitter: @jamessteeleii
@DrJpfisher

James Peter Fisher[1], James Steele[1], Milo Wolf[1], Patroklos Androulakis Korakakis[1], Dave Smith[2], Jürgen Giessing[3], Wayne L. Westcott[4]

[1]School of Sport, Health and Social Science, Solent University, East Park Terrace, Southampton, UK
[2]Manchester Metropolitan University, Crewe, UK
[3]Institute of Sport Science, University of Koblenz-Landau, Landau, Germany
[4]Quincy College, Quincy, MA, USA

ABSTRACT

Background: The body of resistance training literature appears heavily focused upon investigating *efficacy* of interventions by dint of most incorporating supervision (SUP). Authors have suggested that a lack of SUP within strength training results in inadequate workout quality and diminished results yet, since many people choose to perform resistance training unsupervised (UNSUP), it seems important to understand *effectiveness* of resistance training under such ecologically valid conditions. That is, the extent to which SUP might impact adaptation. **Objective:** To collectively explore the effects of SUP upon performance/function and body composition outcomes. **Design:** Exploratory systematic review and meta-analysis. **Search and Inclusion:** A systematic literature search using a Boolean search strategy was conducted with PubMed/MEDLINE, Scopus, and CINAHL in December 2020 and was supplemented with additional 'snowballing' searches. To be included in our analysis, studies had to be experimental trials including at least one performance/functional measure (e.g., strength, speed, power, function, endurance, and cardiorespiratory fitness) and/or body composition measure (body fat percentage, fat mass, and fat free mass). After search and screening, 12 studies were eligible for

All authors have read and approved this version of the manuscript. This article was last modified on 18th October 2021

inclusion including 301 participants in SUP groups and a further 276 participants in UNSUP groups. **Results:** The cluster robust main model for all performance/function effects (57 across 12 clusters [median = 4, range = 1-12 effects per cluster]) revealed a small, standardized point estimate favouring SUP though with relatively poor precision for the interval estimate which ranged from a trivial to a moderate effect favouring SUP (0.28 [95%CI = 0.02 to 0.55]). For sub-grouped outcome types there was very poor precision of robust estimates for speed, power, function, and endurance with all ranging from large effects supporting UNSUP to large effects supporting SUP. However, for strength there was a small, standardized point estimate favouring SUP though with moderate precision for the interval estimate which ranged from a trivial effect favouring SUP to a moderate effect favouring SUP (0.40 [95%CI = 0.06 to 0.74]). The cluster robust main model for all body composition effects (18 across 6 clusters [median = 3, range = 1-6 effects per cluster]) revealed a trivial standardized point estimate favouring SUP that was relatively precise in the interval estimate ranging only trivial effects in either direction (0.07 [95%CI = -0.01 to 0.15]). **Conclusions:** The results of the present systematic review and exploratory meta-analysis suggest that, broadly speaking, SUP resistance training might produce a small effect on increases in performance/function, most likely in strength, compared to UNSUP, and has little to no impact on body composition outcomes. However, the lack of role and purpose within supervision as well as the lack of parity in UNSUP exercise interventions make providing a conclusive and overarching recommendation difficult.

Please cite as: Fisher, J., Steele, J., Wolf, M., Androulakis-Korakakis, P., Smith, D., Giessing, J., Wescott, W.L. (2021, 18th October). The role of supervision in resistance training; an exploratory systematic review and meta-analysis. Sportrxiv. DOI: 10.51224/SRXIV.18

Introduction

The management of resistance training variables for adaptations is well established within academic literature. Empirical studies and subsequent systematic reviews and/or meta-analyses have considered manipulation of load [1], repetition duration [2], weekly volume [3], frequency [4], exercise order [5], and range of motion [6], among other variables, in an attempt to optimise exercise induced adaptations. However, in none of these reviews was training supervision (SUP) discussed as a potentially confounding variable. In fact, of the 86 empirical studies used within these reviews', SUP was mentioned (or assumed based on specific methods e.g., training on an isokinetic dynamometer) in ~80%. Further reviews have attempted to determine the intensity of effort required to optimise strength and hypertrophic adaptations, primarily by considering resistance training to failure versus not to failure [7–9]. However, people are typically poor at predicting proximity to failure based on repetitions in reserve [10], and since reaching muscular failure (MF) seems important in continued muscular adaptations [9], SUP might enhance intensity of effort [11, 12], and thus be a key stimulus for the adaptations seen in empirical research, is an important factor to consider.

The knowledge that SUP is prevalent in resistance training studies might be encouraging as to the quality of research being conducted with respect to internal validity. And in studies which have not mentioned SUP, we cannot assume that there was an absence of monitoring, coaching or encouragement, only that it was not explicitly reported in the final publication. Certainly, the inclusion of SUP supports the *efficacy* (i.e., the extent to which an intervention has the ability to bring about an intended effect under ideal circumstances) of resistance training variables for desired outcomes. That is to say that we understand the impact of the aforementioned training variables, *when exercise sessions are supervised*, and thus presumably completed with a high degree of fidelity. Furthermore, it is not surprising that SUP is commonplace in resistance training studies since it has been stated that *"the key element to effective resistance training is supervision by a qualified professional and the proper prescription of the program variables"* [13]. However, the fact that the majority of resistance training studies employ SUP might limit our knowledge of the *effectiveness* (i.e., the extent to which an intervention achieves its intended effect in its usual setting) of these variables in an everyday environment where SUP is infrequent in those participating in resistance training [14]. In fact, authors have suggested that a lack of SUP within strength training results in inadequate workout quality and diminished results [15]. Considering the inclusion of muscle strengthening activities such as resistance training in global physical activity recommendations [16], and recent arguments that sport and exercise medicine has for some time been drowning in a body of evidence regarding 'efficacy' whilst simultaneously dying of thirst from a lack of evidence regarding 'effectiveness' [17], it is important to understand the effectiveness of resistance training recommendations [18] and thus the extent to which outcomes of resistance training are impacted by the presence of SUP.

SUP within resistance training might be considered important for several reasons: (i). the accurate monitoring of adherence (attendance) and maintenance/continuation [14], (ii). the accurate monitoring and progression of strength training protocol including load progression [19], (iii). the inclusion of technical coaching which might serve to prevent injury and more effectively target specific muscles by preventing "cheating" [14], (iv). the provision of encouragement and psychological support which might enhance the positive experience of resistance exercise and (v). encouragement, motivation, or maintenance which might augment intensity of effort [14]. In contrast, in an unsupervised (UNSUP) setting, trainees might be motivated to increase the load at the detriment of technique. For example, when performing a back squat an UNSUP trainee might decrease the range of motion by not descending to required/prescribed depth and concurrently increase load. In doing so, the trainee appears to progress on paper, and can certainly manage the increased load through the now limited range of motion, but in fact might be limiting their chronic training adaptations [20]. Or they may choose to train at relatively lower efforts than those intended in resistance training recommendations. As noted, trainees may underestimate their proximity to failure and thus train at lower than intended efforts by this means [10], and also trainees typically utilise lower loads when self-selecting [11, 12].

Interestingly, many studies considering adolescent or children performing resistance training advocate a necessity of SUP by qualified and trained professionals [21, 22]. However, recommendations for adults typically lack the same emphasis for SUP, irrespective of experience. The National Strength and Conditioning Association (NSCA) provide guidance in their professional standards guidelines suggesting trainer: athlete ratios of *"…1:10 for lower junior high school, 1:15 for lower high school, 1:20 for lower college…"* and further, that *"Younger participants, novices, special populations or participants engaged in complex-movement strength and conditioning activities should be provided with greater supervision (e.g., 1:12 instead of 1:20)"* [23]. Interestingly, the authors presumably base these numbers on experience since no academic citations are provided. Whilst a professional strength coach might well be conditioned to identify where SUP is more or less important, we should be cautious in assuming that a person's maturation is paralleled by an ability to perform muscle strengthening exercise with proper technique and intensity of effort.

Numerous studies have attempted to address the impact of SUP using different methodological designs. For example, acute studies by Ratamess, et al. [11] and Dias, et al. [12] compared resistance trained females and males (respectively) self-selecting a training load they would use to complete 10-repetitions, as well as assessing maximal strength (1-repetition maximum; RM), and rating of perceived exertion. Results revealed that with the SUP of a personal trainer heavier loads were selected for the 10 repetitions, participants performed better in maximal strength testing (i.e., 1RM) and also reported a higher value for rating of perceived exertion. The authors concluded that resistance training under the SUP of a personal trainer appears to be advantageous to training efforts leading to continued progression of adaptation.

Chronic training studies have applied varied methodological approaches to assess the effects of SUP. For instance, a group of older adults underwent a period of progressive intensity of effort SUP resistance training followed by a period of training where participants could self-select to continue UNSUP or cease the intervention [24]. The data showed positive strength and functional adaptations during the period of SUP resistance training. However, strength declined to a similar extent when SUP was terminated whether participants elected to train UNSUP or to cease training altogether. Further studies have considered SUP ratio. For example, Gentil, et al. [25] reported greater strength increases for a high- (trainer: trainee; 1:5) compared to low- (1:25) SUP ratio. Participants were asked to train to 'volitional fatigue', and the authors hypothesised that the favourable strength increases for the high-SUP condition was a result of greater *"motivation or psychological reinforcement"* leading to subjects training closer to their maximal effort.

Finally, studies have considered the impact of SUP versus UNSUP home-based exercise in clinical patients with an array of medical conditions. For example, a recent review article considering muscle hypertrophy in cancer patients devoted considerable space to the discussion of-, and reported favourable outcomes for-, SUP compared to UNSUP resistance training [26]. However, many studies have often compared SUP laboratory/fitness centrebased resistance training to an UNSUP home-based exercise condition without the same equipment or considered older adults who might be unable to access specific facilities or unwilling/unable to leave their residence [27]. We recognise the importance of evaluating the efficacy of UNSUP resistance training at home in these populations. However, the disparity in facilities and equipment confounds the issue of whether it is SUP which produces optimal adaptations, or the specifics of the prescribed protocol and environment including available equipment. We also recognise the importance of understanding the efficiency of UNSUP, home-based resistance training, especially during the recent closure of fitness centres and gyms as a result of Covid-19; indeed, many who were previously training in leisure centres continued training at home [28]. However, studies without parity in location and/or exercises performed based on access to equipment adds complexity to a research question of whether SUP itself enhances physiological adaptations to resistance training.

With this in mind, the present systematic review and meta-analysis has explored the effects of SUP versus UNSUP resistance training upon performance outcomes (i.e., strength, power, speed, function, muscular endurance and cardiorespiratory), and body composition measures (i.e., body fat %, fat mass, and fat free mass) where facilities and/or equipment/exercises did not differ between groups.

Methods

The systematic review and meta-analysis were conducted in accordance with the guidelines of the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses" (PRISMA [29]). The study was initially preregistered on the Open Science Framework (https://osf.io/ketb2) where the detailed prespecified methodological protocol can be viewed. However, in many respects we deviated from the pre-registered protocol and have detailed here where this has occurred. As a result of this, we explicitly consider this work to be exploratory in nature. Included studies were synthesised both narratively and quantitatively by meta-analysis.

Inclusion/Exclusion Criteria

Our original pre-registration implied that we would include both experimental (e.g., randomised trials), and quasi-experimental (e.g., crossover designs without randomisation) study designs. This was because we initially anticipated a low number of experimental designs comparing purely SUP versus UNSUP interventions but many more including crossover designs where UNSUP interventions followed initially SUP interventions. However, during our systematic search and screening we noticed that there were indeed several randomised trials comparing the two. Thus, we opted to limit ourselves to only including studies of this design to enhance our ability to draw inferences regarding comparative treatment effects.

We also noted that our primary outcome measures were to be broadly grouped as those pertaining to musculoskeletal function or performance (e.g., strength, power, endurance, etc.), musculoskeletal morphology (e.g., muscle size, muscle thickness, etc.), body composition (body fat mass, body fat percentage, lean mass, etc.), and other outcomes including functional or acute self-report outcomes such as affect, or rating of perceived effort. Additionally, we planned to review the outcomes captured in different studies and include any appropriate dependent upon how frequently they were captured. However, musculoskeletal morphology was not captured in any studies, nor were acute self-report outcomes. Two studies reported pain outcomes, but we limited ourselves to including these in only the narrative synthesis. As such, we ultimately opted to re-categorise outcomes into performance/function (including strength, speed, power, functional measures, endurance, and cardiorespiratory fitness), and body composition (fat mass, fat percentage, and fat free mass).

Thus, in the end we included studies that met the following criteria: a) randomized (or baseline stratified) trials that directly compared RT interventions with or without supervision (i.e., SUP vs UNSUP) reporting performance/function (including strength, speed, power, functional measures, endurance, and cardiorespiratory fitness), and body composition (fat mass, fat percentage, and fat free mass) outcomes in children[1] or adults; b) published in a peer-reviewed English language journal or on a pre-print server. Within studies with multiple SUP or UNSUP groups, training groups with different frequency and duration of training [30], or performing different exercise protocols (i.e., BodyPump [31]) were excluded to compare only groups performing the same general modality of resistance training.

Search Strategy

We carried out a comprehensive search on PubMed/MEDLINE, Scopus, and CINAHL using the following Boolean string: ("resistance training" OR "weight training" OR "weight lifting" OR "power training" OR "strength training" OR "strength exercise" OR “strength” OR "resistance exercise" OR “endurance” OR “muscle mass” OR “hypertrophy”) AND ("supervision" OR "mentoring" OR "coaching" OR "monitoring" OR "management" OR "overseeing" OR "direction"). The search was finalized on 16th December 2020; Figure 1 illustrates a flow chart of the search process.

[1] Note, one study included adolescents (sample age ~16-17 years). Given the number of studies we opted to deviate from our pre-registration and include this study also.

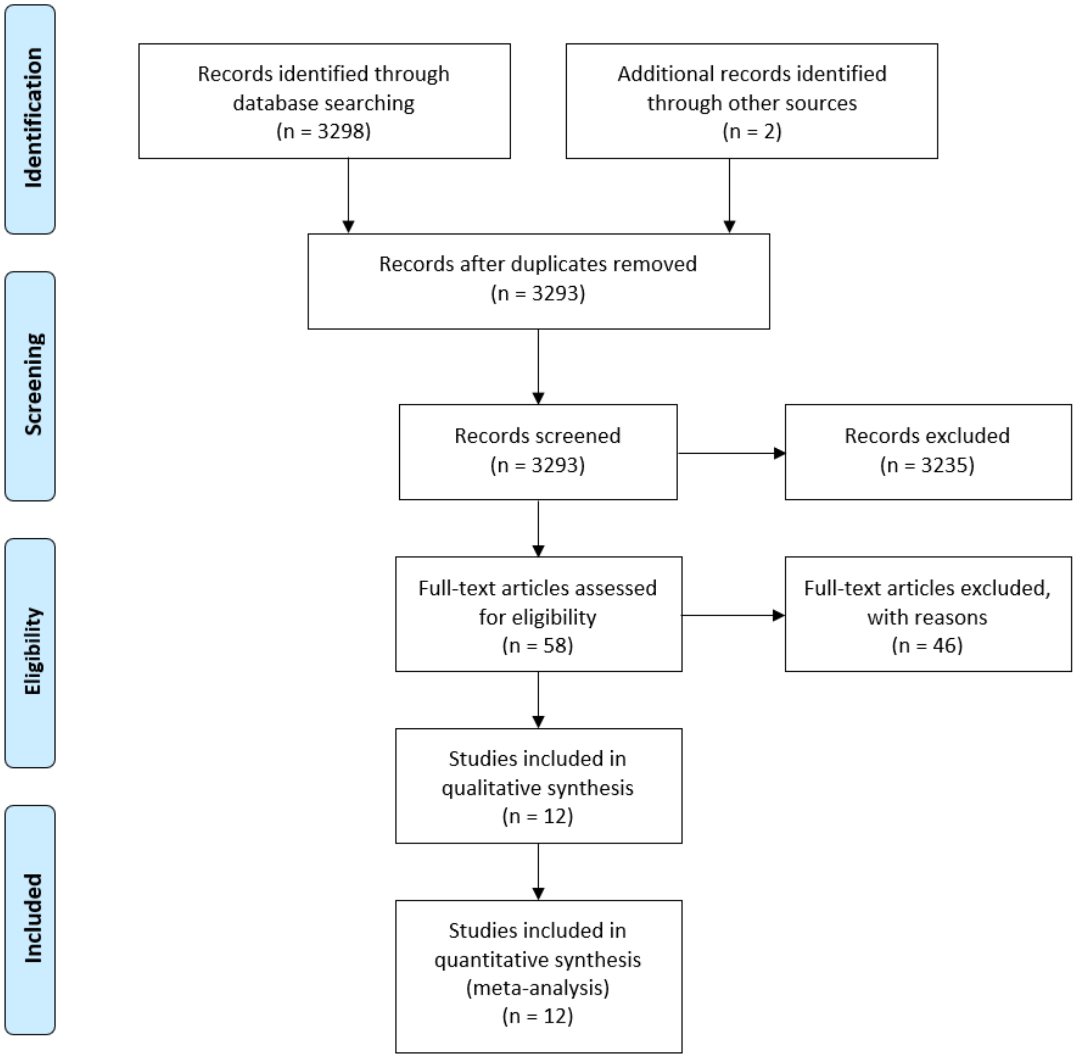

Figure 1. PRISMA flow diagram

Screening/Coding of Studies

Initial search/screening was carried out separately by three researchers (JPF, PAK, and MW). These researchers read all titles and abstracts and then reviewed full texts for papers deemed relevant based on title and abstract. Decisions then were made as to whether a study warranted inclusion based on the stated criteria. Following this, two researchers conducted a final screening of the studies to be included (JPF and JS).

After determining which studies met inclusion, one researcher (JS) separately extracted and coded the following variables for each study: authors, title and year of publication, weighted means for the sample age and body mass index, the proportion of the sample that was male, training status of the sample, what proportion of the intervention sessions were supervised in the SUP condition, what the mean supervisor: participant ratio was, whether the UNSUP condition was observed, contacted at all for check-ups during the intervention period, or required to complete a training diary, whether the location of training was the same or different for both SUP and UNSUP, description of the prescribed training intervention (duration, load, load progression rules employed, frequency, repetitions, sets per exercise, whether task failure was employed, modality), whether an adjuvant aerobic or dietary intervention was employed, adherence in both conditions, the outcome and outcome measures used, mean pre-, post-, and change scores for outcomes with the corresponding standard deviations or where these were not reported standard errors, and the number of reported dropouts and adverse events in each condition. In cases where outcome data were not reported, we either extracted the data from graphs when available via online software (WebPlotDigitizer v4.3, Ankit Rohatgi; https://apps.automeris.io/wpd/) or attempted to contact the study's authors.

Methodological quality

Two of the authors independently evaluated each study (JPF and JS) using the 11-point Physiotherapy Evidence Database (PEDro) scale, which has been validated to assess the methodologic quality of randomized trials with acceptable inter-rater reliability [32, 33]. Any discrepancies in agreement on a given scale item were settled by mutual agreement between the researchers. Given that it is infeasible to blind participants and investigators in supervised exercise interventions, we opted to remove the assessment items specific to blinding (numbers 5, 6, and 7 in the scale). After eliminating these items, this created a modified 8-point PEDro scale with a maximum value of 8 (the first item is excluded from the total score). The qualitative methodological ratings were amended similar to those used in previous exercise-related systematic reviews [34] as follows: "perfect" (8 points); "excellent" (6-7 points); "good" (5 points); "moderate" (4 points); and "poor" (0-3 points).

Statistical Analyses

Quantitative synthesis of data was performed with the 'metafor' [35] package in R (v 4.0.2; R Core Team, https://www.r-project.org/). All analysis code and data are openly available in the supplementary materials (https://osf.io/mu8zf/). Where necessary studies were grouped by design (i.e., within- or between-group), and depending on reporting in individual studies either post or delta comparisons, or pre-post comparison designs [36] for the purposes of appropriate calculation of standardized effects (Hedge's *g*) using the escalc function in metafor. Standardized effect sizes were interpreted as per Cohen's [37] thresholds: trivial (<0.2), small (0.2 to <0.5), moderate (0.5 to <0.8), and large (≥0.8). Standardized effects were calculated in such a manner that a positive effect size value favours the intervention conditions (in this case, the SUP condition). Pre to post correlations for measures are often not reported in original studies; thus, where possible and for both SUP, UNSUP, and control (CON) conditions, we extracted change score standard deviations or calculated them from extracted pre-post *p* values or *t* statistics, change score standard errors, or change score confidence intervals in order to calculate pre-post correlations directly as,

$$r_{pre-post} = \frac{SD_{pre}^2 + SD_{post}^2 + SD_{change}^2}{2 * SD_{pre} * SD_{post}}$$

We then imputed the median correlation coefficient across studies as a reasonable approximation of the population parameter.

Because there was a nested structure to the effect sizes calculated from the studies included (i.e., multiple effects nested within groups and nested within studies), multilevel mixed effects meta-analyses with both study and intra-study groups/clusters included as random effects in the model were performed to explore the effect of supervised resistance training interventions upon outcome measures. Cluster (study) robust point estimates with small sample/cluster correction, and precision of those estimates using 95% compatibility (confidence) intervals (CIs), were produced weighted by the inverse sampling variance to account for the within- and between-study variance (tau-squared) [38]. Restricted maximal likelihood estimation was used in all models. Two main models were produced for both pre-registered main outcomes (performance/function, and body composition), including all standardized effect sizes to provide a general estimate of the comparative treatment effects. We then produced models sub-grouped by specific outcomes. These were presented in sub-grouped forest plots. All other models were considered secondary.

For all models, we avoided dichotomizing the existence of an effect for the main results and therefore did not employ traditional null hypothesis significance testing, which has been extensively critiqued [39, 40]. 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. Influence analyses was performed by examining Hat values and Cook's distances for the main models of performance/function and body composition and where there was evidence of influential effect sizes (Cook's *D* ~1.0, or more conservatively *D* ~ 4/K where K is the number of studies) models were rerun dropping that effect to explore sensitivity of results (only one effect was deemed influential in either main model and exclusion did not materially impact results so these are included in the supplementary materials; https://osf.io/w7kdt/ and https://osf.io/25y3r/). Q and I^2 (partitioned across levels) statistics also were produced and reported [41]. 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 and are qualitatively interpreted as: 0-40% not important, 30-60% moderate heterogeneity, 50-90% substantial heterogeneity, and 75-100% considerable heterogeneity [42].

We had planned to conduct subgroup and moderation analyses across a variety of participant, environment, and intervention characteristics. We ultimately deemed these analyses to be unnecessary for body composition outcomes as (noted in the results section), there was almost zero heterogeneity in these models. We did however explore the following meta-regression and sub-group models (including outcome type as a moderator) for the performance/function: age, proportion of sample as males, supervision ratio, difference in adherence, training status, same or different locations, whether UNSUP was observed, contacted, or completed a diary, and prescribed intervention duration, weekly frequency, number of exercises, sets per exercise, repetitions used, whether a load progression rule was employed, whether task failure was employed, and if an auxiliary aerobic intervention was prescribed. Note, due to the number of clusters being less than the number of fixed effects, these multilevel models were not produced with robust variance estimation. The results of these analyses are included in the supplementary materials including meta-analytic scatter plots and point and interval estimates across subgroups for each outcome type (see https://osf.io/5hrxa/, https://osf.io/y7mk2/, and https://osf.io/eydsq/). We also fit exploratory (not pre-registered) models to examine adherence and dropout proportions with the same multilevel structure and specifications as the main models. Further, we also explored the impact of study quality score regressed on performance/function outcomes.

As a final exploratory (not pre-registered) analysis, we examined the variation in responses between both IT and MICT conditions. We sought to identify whether there was evidence of 'true' inter-individual variation in responses to interventions by comparing the standard deviations for change scores with those of non-exercise CON conditions [43]. We have identified that there is mean-variance (on both the raw and log transformed scales) relationship across studies for change scores in RT interventions in other work (under preparation). Thus, we opted to adjust for this by employing a multilevel meta-regression of the log transformed change score standard deviations, adjusted for the log change score mean [44] calculated such that positive values showed that intervention condition variation exceeded control condition variation thus suggesting evidence of 'true' inter-individual response variation. Where studies did not report change score standard deviations, or we were unable to calculate it directly, this was estimated using the imputed median pre-post correlation coefficient noted above as,

$$SD_{change}\sqrt{SD_{pre}^2+SD_{post}^2-(2*r_{pre-post}*SD_{pre}*SD_{post})}$$

Note that, given the different measurement devices used in individual studies, we accepted pragmatically the inherent assumptions built into this comparison of a constant Gaussian measurement error (i.e., that measurement error does not scale in a non-linear fashion with measured scores).

Results

Search Results

From the initially reviewed 3298 search results, a total of 12 studies were determined to meet inclusion criteria for our analysis. Table 1 presents a summary of the interventions of the included studies. Figure 2 shows the contour enhanced funnel plot for all effects from these studies. Inspection of the funnel plot did not reveal any obvious small study bias.

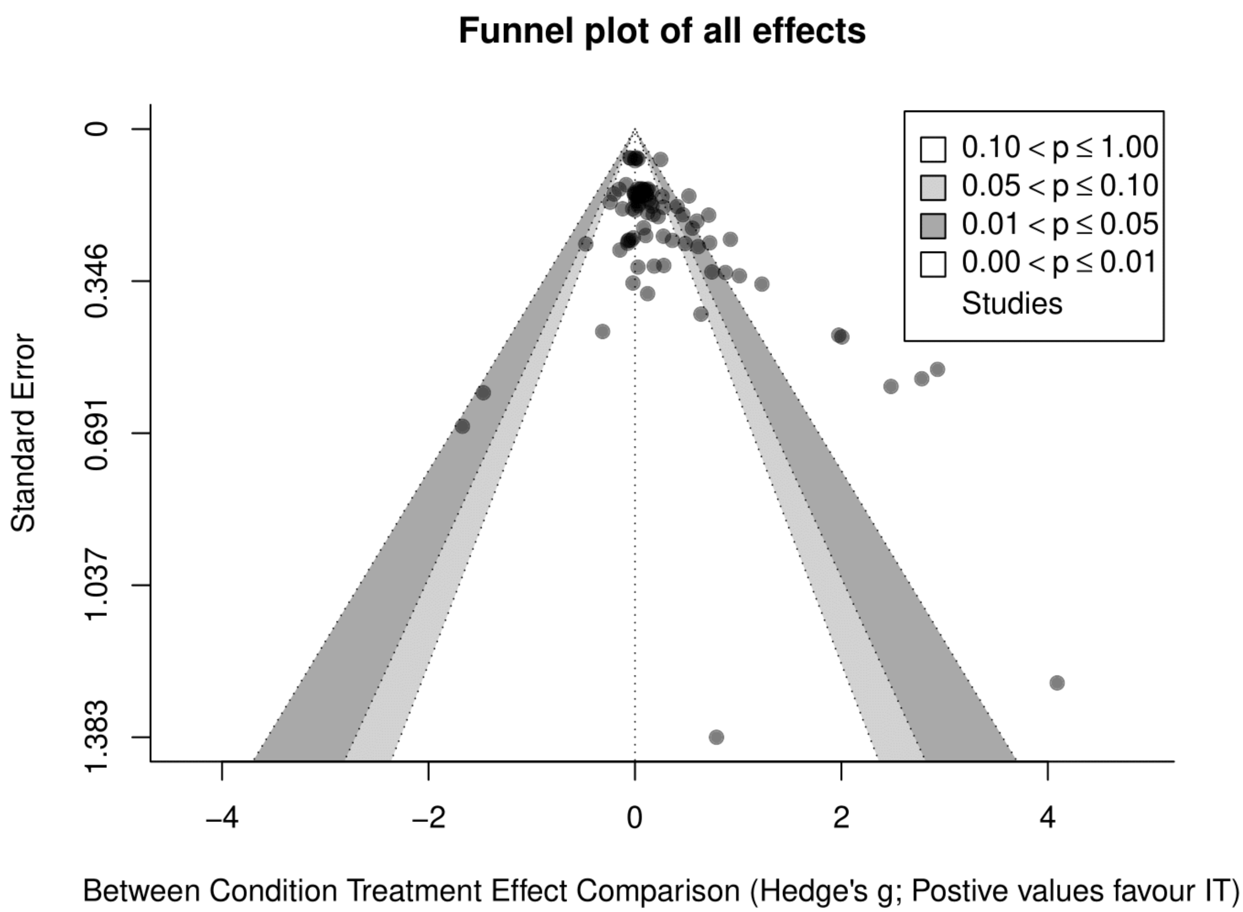

Figure 2. Contour enhances funnel plot of all effects

Table 1. Studies meeting inclusion criteria

Study	Participant characteristics *n,* (age; mean ± SD years)	Training experience	Frequency and Duration	Adherence / Attendance	Protocol (inc. differences between SUP and UNSUP) and effort	Resistance Training Modality
Mazzetti, et al. (2000)	Trained males SUP=10 (25.2 ±1.5 years) UNSUP=8 (23.8 ±1.3 years)	1-2 years	Week 1; 2 x/week Weeks 2,7-12; 3 x/week Weeks 3-6; 4x/week 12 weeks	SUP=100% UNSUP=100%	Both groups performed a prescribed protocol: Weeks 1-2; 3 sets of 8-12RM Weeks 3-6; 3 sets of 8-10RM Weeks 7-10; 3-4 sets of 6-8RM Weeks 11-12; 2-3 sets of 3-6RM Both groups train to RM and choose their own training load	Free weights Bodyweight Resistance Machines
Coutts, et al. (2004)	Trained male rugby league players SUP=21 (16.6 ±1.2 years) UNSUP=21 (16.8 ±1.0 years)	SUP = 3.1 ±4.5 months UNSUP =3.4 ±5.6months	3x/week 12 weeks	SUP=94.5% UNSUP= 84.7%	Both groups performed a prescribed protocol. Repetitions and load adapted based on intended RM Both groups train to RM and choose their own training load	Free Weights Bodyweight Plyometric
Enoksen, et al. (2013)	Junior elite soccer players SUP=9 UNSUP=8 Combined = 19.1 ±3.5 years	Not stated	2 x/week 10 weeks	Not stated	Both groups performed a prescribed protocol: 2-4 sets of 4, 6, 8, 10 or 12RM Both groups train to RM and choose their own training load	Free weights Resistance Machines Bodyweight
Stefanov, et al. (2013)	Sedentary men (*n*=27) and women (*n*=58) BMI $\geq$25 SUP=29 (males=10, females=19) (47.8 ±1 years) UNSUP=22 (males=8, females=14) (47.8 ±1.3 years)	None	Weeks 1-10; 2x/week Weeks 11-20; 3x/week Weeks 21-24; 4x/week 6 months	SUP=73.4% UNSUP=54.8%	Both groups performed a prescribed protocol: Weeks 1-10; 2 sets of 8-14 RM and aerobic exercise at 50-60% MHR Weeks 11-24; 3 sets of 8-14RM and aerobic exercise at 60-70% MHR Both groups train to RM and choose their own training load	Free weights Bodyweight Resistance bands
Storer, et al. (2014)	Trained males SUP=17 (36.3 ±4.3 years) UNSUP=17(36.3 ±4.3 years)	3 months	3x/week 12 weeks	SUP=>100% UNSUP=>100%	SUP: *"…a 3-cycle, non-linear program in which acute program variables including exercise selection, volume and intensity*	Not stated

				(Regardless of instruction not to, participants in both groups performed additional unsupervised training (approx. 2x/week)	*were varied over both the 4-week mesocycles and within the weekly microcycles."* UNSUP: *"... subjects were permitted to train using methods of their own choosing but with the understanding that increased lean mass was the primary objective."*	
Dalager, et al. (2015)	Office workers SUP=81 (males=25, females=56) (46.4 ±10.3 years) UNSUP=65 (males=22, females=43) (44.7 ±10.8 years)	Not stated	3x/week 20 weeks	SUP=39% UNSUP=33%	Both groups performed a prescribed protocol: Week1; 20RM, progressing to 8RM in Week 20 **Both groups train to RM and choose their own training load**	Free weights
Hunter, et al. (2017)	University employees SUP=25 (males=5, females=20) (42.2 ±4.3 years) UNSUP=25 (males=5, females=20) (42.8 ±4.9 years)	Not stated	>1 and <u><</u>5 sessions/week Participants were able to select their frequency of participation 8 weeks	SUP and UNSUP both averaged 1.6 sessions/week	Both groups performed a prescribed protocol: 3 sets of 8-12RM **15-18 on 6-20 RPE scale**	Free weights
Rustaden, et al. (2017)	Overweight and obese women BMI>25 SUP=35 (39 ±10 years) UNSUP=35 (42 ±11 years)	Untrained	3x/week 12 weeks	SUP=89% UNSUP=74%	Both groups performed a prescribed protocol: Session 1; 2sets of 8-10RM Session 2; 2-4sets of 13-15 RM Session 3; 2-4 sets of 3-6 RM Weeks 1-4 = 2 sets Weeks 5-8 = 3 sets Weeks 9-12 = 4 sets **Both groups train to RM and choose their own training load**	Free weights
Cergel, et al. (2019)	Postmenopausal osteoporotic women with vertebral fractures SUP =20 (58.9 ±4.7 years) UNSUP=20 (60.2 ±7.6 years)	Not stated	3x/week 6 weeks	SUP=100% UNSUP=>85% *"Although all participants were fully compliant in the supervised exercise group,*	Both groups performed a prescribed protocol: Weeks 1&2; 3 sets of 8 repetitions Weeks 3&4; 3 sets of 10 repetitions Weeks 5&6; 3 sets of 12 repetitions	Floor based spinal stability exercises

				compliance to exercise was not clearly defined in the home-based exercise group…"	Both groups performed a specific number of repetitions	
Orange, et al. (2019)	Healthy aging adults SUP=17 (males=4, females=13) (53.6 ±3.6 years) UNSUP=19 (males=7, females=12) (54.2 ±3.3 years)	Untrained	3x/week 4 weeks	SUP=94.6% UNSUP= 98.7%	Both groups performed a prescribed protocol: Week1; 1set of 8 repetitions Week 2; 2sets of 8 repetitions Week 3; 2 sets of 10 repetitions Week 4; 3 sets of 10 repetitions 4-6 on 10-point RPE scale	Resistance bands Bodyweight
Hunter, et al. (2020)	University employees SUP=28 (males=8, females=20) (41.6 ±9.5 years) UNSUP=28 (males=7, females=21) (46.1 ±9.1 years)	Untrained	2x/week 16-week	SUP=94% UNSUP=68%	Both groups performed a prescribed protocol: 3 sets of 8-12RM 15-18 on 6-20 RPE scale	Free weights Resistance Machines
Kullman, et al. (2020)	Healthy subjects (8 males and 9 females) SUP=9 (sex undefined) (23.0 ±4.2 years) UNSUP=8 (sex undefined) (20.5 ±1.6 years)	Untrained	2x/week 8 weeks	SUP=94% UNSUP=98%	Both groups performed a prescribed protocol: 3 sets of 10 exercises No details for repetitions were provided, however the authors stated: *"All subjects were encouraged to adjust their effort level as they became stronger by either increasing the number of repetitions or adjusting body positioning to increase resistance."*	Suspension training exercises: Low row, chest press, Y-fly, triceps press, biceps curl, squat, lunge, hamstring curl, calf press, side plank

SUP=supervised, UNSUP=unsupervised, RM=repetition maximum, RPE= rating of perceived exertion

Participant characteristics and intervention length

The current review included 12 randomised controlled trials consisting of a total of 301 participants in SUP groups and a further 276 participants in UNSUP groups. Our pre-registration of this review originally intended to include studies with symptomatic/clinical patients as participants as well as compare between SUP and home-based exercise interventions. This was based on a low expectation of studies comparing SUP and UNSUP resistance training. However, following the searches, we elected to refine the search criteria to better evaluate the impact of supervision alone, with data being confounded by exercise modality, location, and the inclusion of clinical patients. A range of training statuses were present within the included studies which might hinder the degree to which we can effectively conclude as to whether SUP is more or less important in trained or untrained persons.

Most of the studies considered untrained participants [30, 31, 45–50], whilst two studies considered trained males, the longest with 1-2years [19], and the shortest with ~3months training experience [51]. The remaining two studies considered athletic populations [14, 52]. At extremes of a spectrum, one study considered adolescent rugby league players (mean=16.7 ±1.1years old) [14], while another considered postmenopausal osteoporotic women with vertebral fractures (mean=60.3. ±9.3years old) [47]. In between, two studies included overweight and obese participants (BMI≥25) [31, 45], while 6 of 12 studies included male and female participants [30, 45, 46, 48–50], 2 studies included only females [31, 47], and 4 studies included only males [14, 19, 51, 52]. Of the studies, 12 studies identified intervention duration varied from 4 weeks [48], 6-weeks [47], 8-weeks [46, 50], 10 weeks [52], 12 weeks [14, 19, 31, 51], 16 weeks [49], and 20 weeks [30], up to 6 months [45]. Training frequency was 2-4x/week varying within, and between studies. See table 1 for participant characteristics, training frequency, and intervention duration.

Resistance Training Modality and Effort

Resistance type varied between studies and was often a combination of free weight, bodyweight, and resistance machine training. Free weights were the most frequently used resistance modality, appearing in 8 of 12 studies [14, 19, 30, 31, 45, 46, 49, 52]. Resistance machines were used in 3 of 12 studies [19, 49, 52], bodyweight resistance was used in 5 of 12 studies [14, 19, 45, 48, 52], and resistance bands in 2 studies [45, 48]. Finally, plyometric exercise was programmed in 1 study [14], suspension training in 1 study [50], and floor-based spinal stability exercise in 1 study [47]. Storer, et al. [51] did not state the modality of resistance since participants in the UNSUP were not programmed specific methods of training, but rather permitted to choose their own methods in context of the primary objective of increasing lean mass.

In the present review 4 studies prescribed exercise based on repetition maximum [14, 19, 31, 45], 5 studies prescribed a specific number of repetitions to be completed [30, 47, 48, 50, 52], 2 studies prescribed a repetition range equating to a rating of perceived exertion of

All authors have read and approved this version of the manuscript. This article was last modified on 18th October 2021

15-18 on the 6-20 Borg scale [46, 49], whilst a final study did not prescribe repetition ranges [51]. See table 1 for resistance type and protocol, including differences in effort.

Quality Assessment

Study quality was assessed through the use of the Physiotherapy Evidence Database (PEDro) scale with blinding omitted (though of note, 3 studies blinded assessors [31, 47, 51], while none blinded subjects or those delivering interventions for obvious reasons). The included studies had a median PEDro score of 5 indicating "good" quality but ranging from 3 to 8 indicating a range from "poor" to "perfect" according to the adapted PEDro criteria. Individual scores are available in the online materials (https://osf.io/tdje3/). Meta-regression did not suggest that study quality materially impacted effect estimates for performance/function outcomes (see https://osf.io/sajkn/ and https://osf.io/buhzy/).

Performance/Function Outcomes

The main model for all performance/function effects (57 across 12 clusters [median = 4, range = 1-12 effects per cluster]) revealed a small standardized point estimate favouring SUP though with relatively poor precision for the interval estimate which ranged from a trivial to a moderate effect favouring SUP (0.28 [95%CI = 0.02 to 0.55]), with substantial/considerable heterogeneity the majority of which fell between-studies ($Q_{(56)}$ = 184.31, $p < 0.0001$, $I^2_{between}$ = 68.52%, I^2_{within} = 11.45%). For sub-grouped outcome types there was very poor precision of robust estimates (likely due to the correction for small cluster numbers – see output comparing multilevel model estimates prior to, and after cluster robust estimation in supplementary materials (https://osf.io/jert7/) for speed, power, function, and endurance with all ranging from large effects supporting UNSUP to large effects supporting SUP. However, for strength there was a small standardized point estimate favouring SUP though with moderate precision for the interval estimate which ranged from a trivial effect favouring SUP to a moderate effect favouring SUP (0.40 [95%CI = 0.06 to 0.74]), with similarly substantial/considerable heterogeneity the majority of which fell between-studies ($Q_{(23)}$ = 76.02, $p < 0.0001$, $I^2_{between}$ = 77.08%, I^2_{within} = 3.92%). Cardiorespiratory fitness also revealed a more precise estimate compared with other outcome types, though only trivially favoured SUP in its point estimate and still ranged from a small effect favouring UNSUP to a large effect favouring SUP (0.18 [95%CI = -0.31 to 0.67]), with similarly substantial/considerable heterogeneity all of which fell between-studies ($Q_{(6)}$ = 11.24, p = 0.0814, $I^2_{between}$ = 73.93%). Figure 3 shows the sub-grouped forest plot for performance/function outcomes.

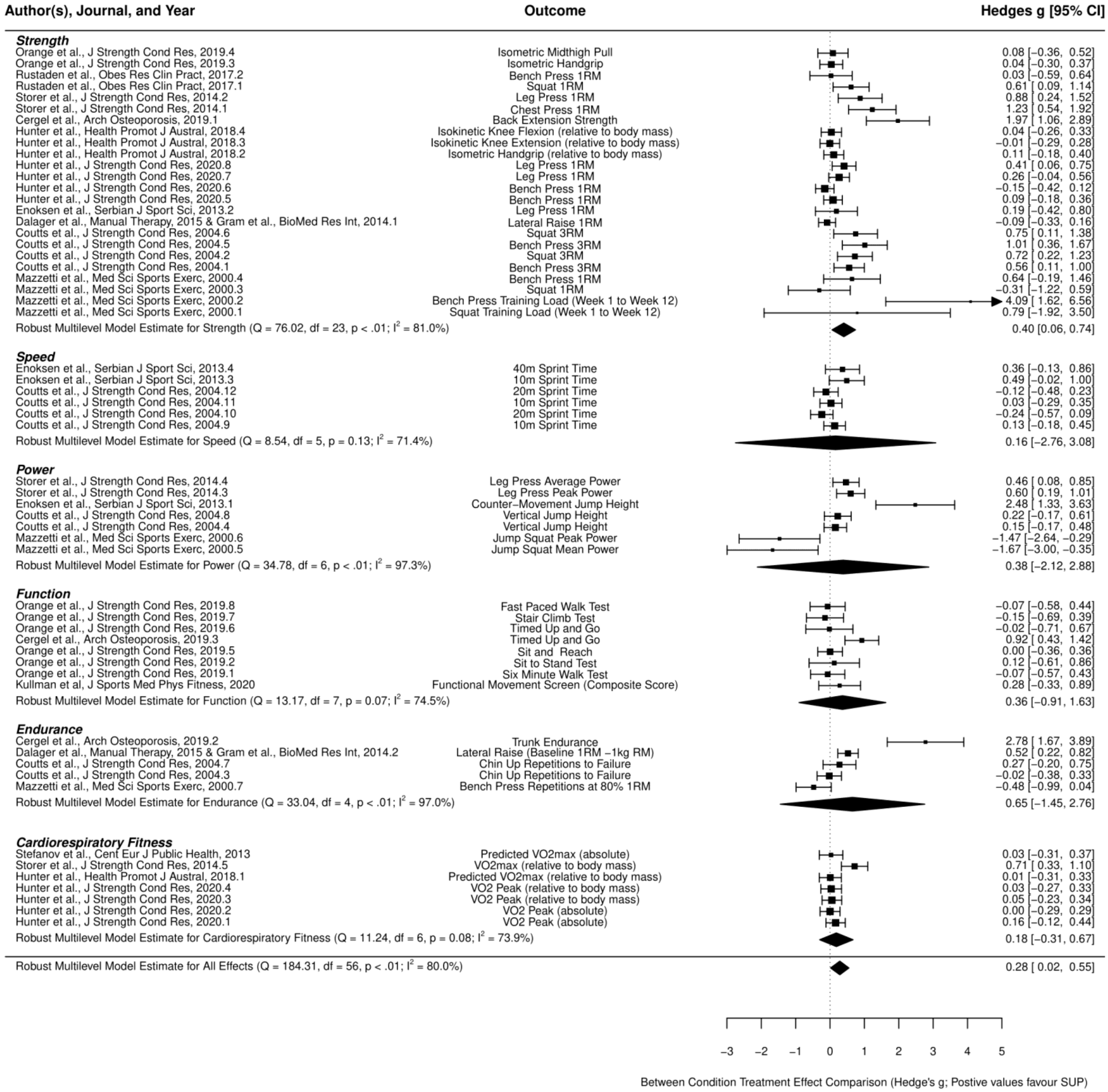

Figure 3. Sub-grouped forest plot for performance/function outcomes

All authors have read and approved this version of the manuscript. This article was last modified on 18th October 2021

Body Composition Outcomes

The main model for all body composition effects (18 across 6 clusters [median = 3, range = 1-6 effects per cluster]) revealed a trivial standardized point estimate favouring SUP that was relatively precise in the interval estimate ranging only trivial effects in either direction (0.07 [95%CI = -0.01 to 0.15]), with essentially no heterogeneity ($Q_{(17)}$ = 6.76, p = 0.9865, $I^2 \approx 0\%$). This similarly held across all sub-grouped outcome types. Figure 4 shows the sub-grouped forest plot for body composition outcomes.

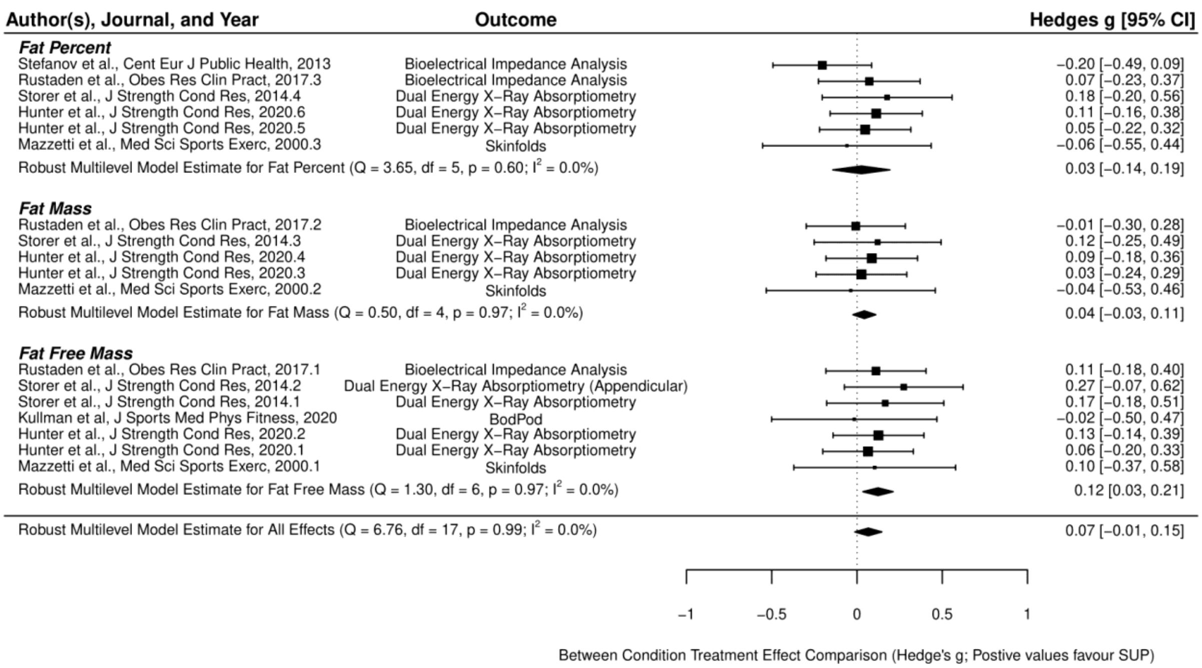

Figure 4. Sub-grouped forest plot for body composition outcomes

Adherence and Dropouts

There was minimal difference in adherence or dropout proportions between conditions which were relatively high and low, respectively. Adherence for SUP was 91.5% [95%CI = 82.7% to 96.0%] and for UNSUP was 87.1% [95%CI = 71.2% to 94.9%], and dropouts for SUP were 14.6% [95%CI = 7.2% to 27.2%] and for MICT were 17.9% [95%CI = 8.1% to 34.9%].

Inter-Individual Response Variation

There was no clear evidence of 'true' inter-individual variation in responses from examination of the log change score standard deviations adjusted for log change score means for either SUP or UNSUP conditions. The difference in intercepts when compared with

CON conditions were -0.19 [95%CI = -7.57 to 7.18] and -0.13 [95%CI = -5.64 to 5.37] for SUP and UNSUP respectively (see figure in supplementary materials: https://osf.io/rxbhs/).

Discussion

The aim of this review was to collectively explore studies which have compared resistance training interventions with or without SUP. To our knowledge this is the first review to consider this area of research and present both exploratory meta-analytic and narrative discussion.

The main results from the meta-analysis were that the estimate for SUP upon performance was, at best, compatible with only small effects (0.28). When considered based on the different performance/function outcome types, estimates were very imprecise for speed, power, function, endurance, and cardiorespiratory fitness prohibiting any confident inferences to be made. However, there was a small point estimate of effect for strength ranging from trivial to moderate favouring SUP (0.40 [95%CI = 0.06 to 0.74]). For body composition outcomes, though point estimates tended towards favouring SUP, all interval estimates were precise and mostly ranged only trivial effects suggesting little impact of SUP on these outcomes.

The results of the present systematic review and exploratory meta-analysis suggest that SUP resistance training might produce small increases in performance/function, most likely to occur for strength outcomes, compared to UNSUP. No individual effects or studies appeared particularly influential in our model for strength outcome. However, some studies did show quite large (though imprecise) point estimates favouring SUP which are worth considering. For example, Çergel, et al. [47], considered postmenopausal osteoporotic women with vertebral fractures performing spinal stability exercises. In context of the present review this represents an atypical population group which might be more subject to psychological factors impacting exercise adherence and fear avoidance. There is a large body of research considering fear avoidance in persons with low-back pain [53, 54], and as such even where adherence might be similar, effort as a result of confidence is likely to be different between groups prescribed an exercise program and supervised through performance of those exercises.

Other studies showing large strength increases, and particularly so for trained participants, were Coutts, et al. [14] and Mazzetti, et al. [19]. Both studies reported greater adherence and greater load increases, respectively. These factors might plausibly play a role in the learning of the skill of the tested exercises for strength. For example, motor control research has shown that a motor schema is highly task- [55] and load-/force-specific [56]. In this sense, the more frequent practice (e.g., greater adherence [14]) and the practice of a test with a heavier load (e.g., greater load progression [19]) would likely impact post-intervention performance favourably [57]. As such the benefits of SUP for strength might be a product of greater adherence and load progression. Exploratory analysis did suggest that the application of a prescribed load progression rule (which would presumably aid UNSUP in knowing when to increase training loads) had a precise trivial point estimate with essentially zero heterogeneity (see https://osf.io/9mfcx/ and https://osf.io/hqyjm/). In the absence of a load progression rule it seems reasonable that a supervisor might pay closer attention to load progression and thus SUP may impact strength gains via this means. Adherence however did not clearly impact upon strength effects in our exploratory analyses (https://osf.io/qkp96/).

Although SUP may have a moderate effect on strength potentially moderated by studies where UNSUP conditions did not receive specific instruction on how to progress loads, it is not wholly clear from the current body of evidence what other aspects of either participants or interventions might influence its impact. Thus, it is worth considering our other exploratory analyses as well as narratively exploring the included studies to identify potential factors that might explain this.

Participant characteristics and intervention length

Participant characteristics, including training status, and intervention length might be of importance. The majority of studies included untrained persons, which, training over a short duration, would be more likely to experience early adaptations and might be less impacted by SUP (and factors which SUP is might enhance, such as load progression [19] and adherence [45]). In contrast, trained persons might require a greater stimulus (e.g., heavier load, or greater intensity of effort) to continue making positive adaptations [9]. Indeed, some of the larger effect sizes in favour of SUP were seen in trained participants [19, 51]. Furthermore, as a person's training status evolves, so the adaptations might evidence divergence between SUP and UNSUP groups. That being said, though plausible, we did not identify in our exploratory analyses of performance/function outcomes any clear difference between trained or untrained participants or impact of intervention duration in weeks.

Location and Resistance Training Modality

Studies included in this review differed as to whether the location of training was the same, or different, for the UNSUP conditions. In all cases where they differed, UNSUP participants trained at home. Indeed, where the location of training was the same SUP seemed less likely to have an impact upon outcomes. However, though we did not quantitatively explore it due to the fact most used mixed approaches, the modality used for resistance training (often tied to location) is worth considering since a key role of SUP is the technical instruction of complex movements [58]. Furthermore, whilst safety bars can be used to prevent weights from falling/dropping and injuring a trainee (e.g., for a bench press or back squat exercise), should the trainee not be able to complete a repetition, confidence might be increased by performing free weight exercises under SUP where spotting is possible [31, 48, 52]. In contrast, performing exercises with a heavier load, or at a higher intensity of effort (i.e., close to or at MF) might be more attainable when using resistance machines which pose a lower risk of serious injury [59]. Alternately, confidence might increase with training status, as would technical proficiency.

Intensity of Effort

A key variable which might be impacted by SUP is that of effort [11, 12], specifically proximity to MF. Research has suggested that similar adaptations occur irrespective of training to- or not to- MF in previously untrained persons, whilst training to MF appears important in trained persons wishing to continue making muscular adaptations [9]. A difficulty in the discussion of effort has been the lack of clarity over terminology. In a recent narrative [60] self-determined repetition maximum (*"set endpoint when trainee determines they could not compete the next repetition if it were attempted"*) and repetition maximum (RM; the *"set endpoint when trainees complete the final repetition possible whereby if the next repetition was attempted, they would achieve MF"*) were identified and discussed. The disparity between them is particularly noteworthy in the present context since evidence suggests that a trainee typically under predicts the number of repetitions possible and thus their proximity to MF [10].

The aforementioned study by Gentil et al. [25] (which did not meet inclusion criteria) reported greater adaptations for more favourable SUP ratios (trainer: trainee of 1:5 vs. 1:25). In this study participants were encouraged to train to volitional failure (identified and discussed as comparable to self-determined RM [60]), and the authors suggested the differing adaptations were a result of higher intensity of effort (and thus proximity to MF) for the favourable SUP ratio. As such, in the studies included herein, effort might have been greater where participants were encouraged to train to RM in a SUP condition. This may have led to a greater progression in load as participants exceeded the desired repetition range [19, 31, 49]. Furthermore, previous evidence has shown that load selection is higher in SUP compared to UNSUP conditions [11, 12]. In contrast, in an UNSUP condition, where persons are poor at predicting the number of repetitions possible, they might not have progressed load to the same degree. Despite this, our exploratory analyses did not indicate any clear impact of whether participants trained to task failure (either an self-determined RM or MF) as indicated in the intervention descriptions. Although, it is worth noting that typically reporting of set endpoints is vague and unclear in most studies [60].

Adherence

Though overall there was little difference between SUP and UNSUP with respect to the adherence to the prescribed frequency of training, across studies it appeared higher in trained persons (i.e., SUP and UNSUP≥100% [51], SUP and UNSUP=100% [19], and SUP=94.5%, UNSUP=84.7% [14]) compared to untrained persons (mean over 7 studies: SUP=83.4%, UNSUP=73.1% [30, 45, 47–50]. This is as expected; if a person already has the motivation to engage in training UNSUP (indeed the majority of trained persons tend to train alone [28]) then it seems reasonable that they would likely continue to do so, and the degree of SUP would be unlikely to be a determining factor. In contrast, adherence was lower in previously untrained participants who may not to have had the same motivation to participate in resistance training (by dint of the fact they previously were not). This was particularly low in the study by Dalager, et al. [30] considering male and female office workers (SUP=39%, UNSUP=33%). Multiple authors have attributed favourable adaptations for SUP compared to UNSUP to significantly greater adherence [14, 31, 45, 49]. Further, Çergel, et al. [47], reported: *"Although all participants were fully compliant in the supervised exercise group, compliance to exercise was not clearly defined in the home-based exercise group…"*. In other studies, there was a similarity in attendance and training volume [46, 48, 50]. However, limitations exist. For example, the study by Orange, et al. [48] was only 4 weeks in duration and recruited untrained older adults; thus, even a small dose of moderate exercise is likely to produce strength increases. Furthermore, whilst adherence was similar over a short period, we cannot be certain that differences in adherence would not occur over a longer duration. Should disparity in attendance occur over a longer intervention period, it might result in differing adaptations catalysed by the significantly greater intensity of effort in the SUP compared to the UNSUP group. The authors state that the average heart rate for the SUP group was ~14b·p·m^{-1} higher, and equivalent to 70% of age-predicted maximum heart rate; meeting the American College of Sports Medicine physical activity guidelines for moderate intensity aerobic exercise [61]. Taking this as an indication of the intensity of effort achieved, prolonged, and potentially more frequent exercise at higher intensity of effort may produce positive health and fitness adaptations.

Between study differences in UNSUP conditions

Interestingly, the nature of UNSUP conditions was not consistent between studies. We did not identify clear differences in our exploratory analyses based upon supervision ratios, whether training was alone or group, or whether UNSUP participants were observed, contacted regularly, or completed training diaries. However, given the diversity of UNSUP conditions across studies, it is worth looking more closely at their methods.

For example, across a number of the studies, the UNSUP group received instructions/technical guidance on intensity, technique, and progression prior to beginning the intervention [31, 45, 47, 48]. However, this varied in detail; in the study by Stefanov, et al. [45] this consisted of a 1-week exercise course including 2 lectures and 3 practical sessions to acquaint participants with basic principles and execution of different exercises. In contrast, in the studies by Çergal, et al. [47] and Orange, et al. [48] – both of which used bodyweight and resistance band exercise – participants were given instructions and pictures in a booklet, and attended a single session where exercises were demonstrated and performed under SUP, and technique adjusted as necessary. The degree of exercise coaching prior to beginning the intervention would be expected to play a role, both in the adherence and progression of the exercise program and, as a result, in the adaptations. A greater amount or quality of initial coaching might improve self-efficacy and confidence in a person's ability to complete and/or progress an exercise program. A person who has greater confidence and enjoyment would probably be more likely to adhere to a resistance training program.

The UNSUP conditions also varied in the degree of supervision they experienced throughout the intervention. For example, perhaps due to the age of the participants, the UNSUP group of young rugby players (mean age of 16.7 ± 1.1 years) studied by Coutts et al. [14] were observed by a team manager who was not trained in strength and conditioning coaching but monitored attendance and program administration. Whilst we might expect the adherence to be similar when there is a degree of supervision by a team manager, the authors reported significantly lower attendance in the UNSUP compared to the SUP group (84.7% and 94.5%, respectively). However, the observation of a team manager might have encouraged a greater intensity of effort and motivation during the resistance exercise. Other differences include; (i) provision of tutelage at an intermediary follow-up (after 6 weeks of a 12-week intervention [14, 31], at weeks 5, 9 and 13 of a 16-week intervention [49], and once every 3 weeks throughout a 6-month intervention [45]), (ii) participants being telephoned weekly by an instructor to answer questions about their training and intensity of effort [48], and (iii) participants in the UNSUP groups being observed and/or able to seek guidance and assistance from gym instructors throughout the intervention [14, 19, 46, 49]. Once again, this variety might impact the degree of adaptation experienced by UNSUP participants. Access to a personal trainer at each session might be a provision that instils or enhances confidence in participants, and regular check-ups to provide encouragement and query intensity of effort would be likely to improve adherence and effort beyond that of someone without the same management. Ultimately, we might start to consider whether the UNSUP groups in many studies were truly UNSUP, or – based on a recent commentary on accurate definitions – were *facilitated* (i.e., *"Exercise or physical activity undertaken without the presence of a healthcare professional or qualified fitness instructor but with scheduled meetings or check-ins between sessions to monitor progress and provide support (virtually or in person)"*) [62].

In many of the studies the UNSUP group received a training program to follow, which might have resulted in similarity in adaptations between SUP and UNSUP training groups. Certainly, we might assume that the more detailed or better understood a training program the more accurately it can be followed. If parity exists in following a training program and the supervised experience, then we would expect similarity in adaptation. However, in the study by Storer, et al. [51] , the SUP group followed a prescribed a *"…3-cycle, nonlinear program in which program variables including exercise selection, volume, and intensity were varied…"* whilst the UNSUP group were not provided a program and rather were instructed to *"…train using methods of their own choosing."* This might better reflect a real-world condition where prescription is not often provided to those choosing to exercise UNSUP.

A further study is that of Kullman, et al. [50], who compared SUP and UNSUP whole-body suspension training. Notably, the outcomes of the intervention were improvement in functional movement screen score and lean body mass with no significant between group differences. Little detail is provided as to the level of SUP and both groups were encouraged to *"…adjust their effort level as they became stronger by either increasing the number of repetitions or adjusting body positioning to increase resistance."* Adherence was similar between groups (94% and 98% for SUP and UNSUP, respectively), and by the nature of the exercise modality it might be likely that training 2x/week for 8 weeks served to increase competency in the exercises programmed in both groups. The authors clarify that the FMS is used to assess movement quality and predict likelihood of injury, however, they also state that, whilst statistically significant, the small increases seen in this subjectively scored test failed to meet the minimal detectable change (MDC) identified in previous research (i.e., a composite score increase of 1.1, compared to an MDC of 2.07) [63].

Identified role of the coaches/personal trainers

Previous research has raised some interesting dialogue as to the purpose of SUP within strength training. For example, Hillmann and Pearson [15] suggested *"Each athlete needs to be supervised and pushed through workouts in order to achieve optimal strength development".* The authors surveyed NCAA Div 1-A university strength coaches about details and practices around strength training SUP, reporting on the themes of coach-to-athlete ratio, scheduling, size of facility, training protocol, and equipment [15]. However, none of those themes identify the *role* or *purpose* of SUP. Baker [58] talked more about the role of a strength and conditioning coach and identified a process of instruction, performance, feedback. However, while the article, provides focus upon coaching technical elements of strength training including verbal reinforcement of technique, there is no discussion of the role of intensity of effort, or encouragement. Interestingly, motivation is mentioned in context of adherence and rate of occurrence of exercise training, and is similarly discussed by Mazzetti, et al. [19], but is not considered in view of the motivation to apply effort. In a later study, Massey, et al. [64] observed and analysed strength and conditioning coaches' behaviour. In observing 6 coaches over 120 minutes each and identifying 8,640 individual behaviours, the most frequently observed were silent monitoring (22%), management (15%), instruction (17%), and feedback (17%) of which hustle – later described as *"verbal efforts to intensify athletic effort"* (11%).

This area is particularly noteworthy since of the 12 studies identified and included herein, 9 of them failed to mention any role or purpose of SUP [14, 19, 30, 45–47, 49–51]. Of the three studies which did discuss the intended role of SUP; Enoksen, et al. stated: "*The duties of the expert coach were to follow up every strength training session throughout the 10 weeks providing technical instructions, training methodological advice, motivation and optimal social and mental support."* [52], Rustaden, et al. stated *"The personal trainers could spot/secure and verbally motivate the participants during the weightlifting exercises, while forced repetitions were prohibited."* [31], and Orange, et al. stated: *"Participants received real-time encouragement and feedback on exercise technique with form being adjusted by the CSCS if necessary."* [48]. In fact, in 4 of the studies the UNSUP group were in an environment where a gym instructor observed and was available to seek guidance from, which, in some research, constitutes themes within SUP [14, 19, 46, 49].

Limitations

Ultimately many of the inconsistencies between studies represent an important limitation in this area of research and to the extent to which we can draw firm conclusions from the meta-analytic findings. Of course, we employed appropriate meta-analytic techniques including clustered random effects and robust estimation to enhance our ability to draw inferences. Yet these are still undoubtedly limited to conclusions regarding the role of SUP 'in general' and not in any specific context.

Notably a lack of clarity as to the role of SUP within respective studies makes it difficult to appreciate whether and how SUP might have impacted adaptations to resistance training. We might consider that, in previously untrained persons, a focus upon technical guidance and proficiency might be of greater importance and dominate SUP – which might, in turn, be less likely to impact physiologic response. In contrast, in people with increasing resistance training experience, and thus existent technical expertise, encouragement to exercise at greater effort levels might represent a more important and valued coaching input. However, that is not to say that untrained persons do not need encouragement to work hard and might also attain greater results with correct SUP. Future research should consider the discrepancy in these coaching approaches during SUP resistance training as well as client preferences across the spectrum of training experiences.

The lack of parity in UNSUP resistance training groups also limits the extent to which we can consider the efficacy of UNSUP resistance exercise. For example, training UNSUP might be best thought of as training alone, with a self-written program, without the intermittent monitoring of an expert/practitioner/researcher. This was best identified in the study by Storer, et al. [51] who identified a goal to the UNSUP participants and then allowed them to train however they deemed appropriate in view of this goal. In contrast, monitoring by a team manager, access to a personal trainer on the gym floor, and having a training program prescribed along with remote but consistent contact by a trainer might be more akin to *degrees* of SUP, rather than UNSUP. In a real-world environment these represent some of the services that are paid for by gym memberships or online/remote personal training services, rather than reflecting the habits and responses to training completely UNSUP. Once again, future research might consider preferences to and perceptual responses to degrees of supervision and perhaps as well as adaptations. Finally, this area of SUP has become more contemporary over the recent years with the growing popularity in virtual personal training as a result of gym closures and subsequent Covid-19 lockdown protocols [28], representing another element of SUP. Whilst some studies have considered virtual personal training in older adults [65], future research might consider the efficacy of this type of SUP by comparison to face to face personal strength training.

Conclusions

The results of the present systematic review and exploratory meta-analysis suggest that, broadly speaking, SUP resistance training might produce a small effect on increases in performance/function, most likely in strength, compared to UNSUP, and has little to no impact on body composition outcomes. However, the lack of role and purpose within supervision as well as the lack of parity in UNSUP exercise interventions make providing a conclusive and overarching recommendation difficult. Future research should consider the limitations with the present literature discussed here and, in line with recent definitional taxonomies, look to investigate the role of SUP in a more systematic fashion to support future confirmatory meta-analysis.

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/mu8zf/

Author contributions

JF and JS conceived of the study; JF, JS, MW, and PAK collected the data; JS carried out statistical analyses; all authors were meaningfully involved in interpreting data, and drafting and critically revising the manuscript for intellectually important content.

REFERENCES

1. Schoenfeld BJ, Grgic J, Ogborn D, Krieger JW. Strength and hypertrophy adaptations between low-vs. high-load resistance training: a systematic review and meta-analysis. J Strength Cond Res. 2017;31:3508–3523.
2. Schoenfeld BJ, Ogborn DI, Krieger JW. Effect of repetition duration during resistance training on muscle hypertrophy: a systematic review and meta-analysis. Sports Med. 2015;45:577–585.
3. Schoenfeld BJ, Ogborn D, Krieger JW. Dose-response relationship between weekly resistance training volume and increases in muscle mass: A systematic review and meta-analysis. J Sports Sci. 2017;35:1073–1082.
4. Schoenfeld BJ, Grgic J, Krieger J. How many times per week should a muscle be trained to maximize muscle hypertrophy? A systematic review and meta-analysis of studies examining the effects of resistance training frequency. J Sports Sci. 2019;37:1286–1295.
5. Nunes JP, Grgic J, Cunha PM, Ribeiro AS, Schoenfeld BS, de Salles BF, Cyrino ES. What influence does resistance exercise order have on muscular strength gains and muscle hypertrophy? A systematic review and meta-analysis. Eur J Sport Sci. 2021;21(2):149-157.

6. Schoenfeld BJ, Grgic J. Effects of range of motion on muscle development during resistance training interventions: A systematic review. SAGE Open Med. 2020;8: 2050312120901559.
7. Davies T, Orr R, Halaki M, Hackett D. Effect of training leading to repetition failure on muscular strength: a systematic review and meta-analysis. Sports Med. 2016. 1;46(4):487-502.
8. Schoenfeld BJ, Grgic J. Does training to failure maximize muscle hypertrophy? Strength Cond J. 2019;41:108–113.
9. 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. 23;S2095-2546(21)00007-7.
10. Halperin I, Malleron T, Har-Nir I, Androulakis-Korakakis P, Wolf M, Fisher J, Steele J. Accuracy in predicting repetitions to task failure in resistance exercise: a scoping review and exploratory meta-analysis. Sports Med. 2021; Online ahead of print
11. Ratamess NA, Faigenbaum AD, Hoffman JR, Kang J. Self-selected resistance training intensity in healthy women: the influence of a personal trainer. J Strength Cond Res. 2008;22(1):103–111.
12. Dias MRC, Simão RF, Saavedra FJ, Ratamess NA. Influence of a personal trainer on self-selected loading during resistance exercise. J Strength Cond Res. 2017;31(7):1925–1930.
13. Kraemer WJ, Ratamess NA, French DN. Resistance training for health and performance. Curr Sports Med Rep. 2002;1(3):165–171.
14. Coutts AJ, Murphy AJ, Dascombe BJ. Effect of direct supervision of a strength coach on measures of muscular strength and power in young rugby league players. J Strength Cond Res. 2004;18(2):316-323.
15. Hillmann A, Pearson DR. Supervision: The key to strength training success. Strength Cond J. 1995;17(5):67–71.
16. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, Carty C, Chaput JP, Chastin S, Chou R, Dempsey PC. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med. 2020;54(24):1451-62.
17. Beedie C, Mann S, Jimenez A, Kennedy L, Lane AM, Domone S, Wilson S, Whyte G. Death by effectiveness: exercise as medicine caught in the efficacy trap! British Journal of Sports Med. 2016;50(6):323.
18. Steele J, Fisher J, Skivington M, Dunn C, Arnold J, Tew G, Batterham AM, Nunan D, O'Driscoll JM, Mann S, Beedie C. 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.
19. Mazzetti SA, Kraemer WJ, Volek JS, Duncan ND, Ratamess NA, Gomez AL, Newton RU, Hakkinen KE, Fleck SJ. The influence of direct supervision of resistance training on strength performance. Med Sci Sports Exerc. 2000;32(6):1175-84.

20. Bloomquist K, Langberg H, Karlsen S, Madsgaard S, Boesen M, Raastad T. Effect of range of motion in heavy load squatting on muscle and tendon adaptations. Eur J Appl Physiol. 2013;113(8):2133-42.
21. Faigenbaum AD, Myer GD. Resistance training among young athletes: safety, efficacy and injury prevention effects. Br J Sports Med. 2010;44(1):56–63.
22. Stricker PR, Faigenbaum AD, McCambridge TM. Resistance training for children and adolescents. Pediatrics. 2020;145(6).
23. Travis TN, Brown V, Caulfield S, Doscher M, McHenry P, Statler T. NSCA Strength and conditioning professional standards and guidelines. Strength Cond J. 2017;39:1-24.
24. Steele J, Raubold K, Kemmler W, Fisher J, Gentil P, Giessing J. The effects of 6 months of progressive high effort resistance training methods upon strength, body composition, function, and wellbeing of elderly adults. BioMed Research International. 2017.
25. Gentil P, Bottaro M. Influence of supervision ratio on muscle adaptations to resistance training in nontrained subjects. J Strength Cond Res. 2010;24(3):639–643.
26. Köppel M, Mathis K, Schmitz KH, Wiskemann J. Muscle Hypertrophy in Cancer Patients and Survivors via Strength Training. A Meta-Analysis and Meta-Regression. Crit Rev Oncol Hematol. 2021;103371.
27. Lacroix A, Hortobagyi T, Beurskens R, Granacher U. Effects of supervised vs. unsupervised training programs on balance and muscle strength in older adults: a systematic review and meta-analysis. Sports Med. 2017;47(11):2341-61.
28. Steele J, Androulakis-Korakakis P, Carlson L, Williams D, Phillips S, Smith D, Schoenfeld BJ, Loenneke JP, Winett R, Abe T, Dufour S, et al. The Impact of Coronavirus (COVID-19) Related Public-Health Measures on Training Behaviours of Individuals Previously Participating in Resistance Training: A Cross-Sectional Survey Study. Sports Med. 2021;51(7):1561-80.
29. Moher D, Liberati A, Tetzlaff J, Altman DG, Prisma Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine. 2009;6(7):e1000097.
30. Dalager T, Bredahl TG, Pedersen MT, Boyle E, Andersen LL, Sjøgaard G. Does training frequency and supervision affect compliance, performance and muscular health? A cluster randomized controlled trial. Man Ther. 2015;20(5):657-65.
31. Rustaden AM, Haakstad LA, Paulsen G, Bø K. Effects of BodyPump and resistance training with and without a personal trainer on muscle strength and body composition in overweight and obese women—A randomised controlled trial. Obes Res Clin Pract. 2017;11(6):728-39.
32. Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). Aust J Physiother. 2002;48(1):43-9.

33. Elkins MR, Herbert RD, Moseley AM, Sherrington C, Maher C. Rating the quality of trials in systematic reviews of physical therapy interventions. Cardiopulm Phys Ther J. 2010;21(3):20.
34. Kümmel J, Kramer A, Giboin LS, Gruber M. Specificity of balance training in healthy individuals: a systematic review and meta-analysis. Sports Med. 2016;46(9):1261-71.
35. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Soft. 2010;36(3):1-48.
36. Morris SB. Estimating effect sizes from pretest-posttest-control group designs. Organ Res Methods. 2008;11(2):364-86.
37. Cohen J. Statistical power analysis for the behavioral sciences. Academic press; 2013.
38. 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.
39. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature. 2019(7748):305-7.
40. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon statistical significance. The Am Stat. 2019;73(sup1):235-45.
41. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-60.
42. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savović J, Schulz KF, Weeks L, Sterne JA. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011 29;343(7829).
43. Atkinson G, Williamson P, Batterham AM. Issues in the determination of 'responders' and 'non-responders' in physiological research. Exp Physiol. 2019;104(8):1215-25.
44. Nakagawa S, Poulin R, Mengersen K, Reinhold K, Engqvist L, Lagisz M, Senior AM. Meta-analysis of variation: ecological and evolutionary applications and beyond. Methods Ecol Evol. 2015;6(2):143-52.
45. Stefanov T, Vekova A, Bonova I, Tzvetkov S, Kurktschiev D, Blüher M, Temelkova-Kurktschiev T. Effects of supervised vs non-supervised combined aerobic and resistance exercise programme on cardiometabolic risk factors. Cent Eur J Public Health. 2013;21(1):8-16.
46. Hunter JR, Gordon BA, Lythgo N, Bird SR, Benson AC. Exercise at an onsite facility with or without direct exercise supervision improves health-related physical fitness and exercise participation: An 8-week randomised controlled trial with 15-month follow-up. Health Prom J Aust. 2018;29(1):84-92.
47. Çergel Y, Topuz O, Alkan H, Sarsan A, Akkoyunlu NS. The effects of short-term back extensor strength training in postmenopausal osteoporotic women with vertebral fractures: comparison of supervised and home exercise program. Arch Osteoporos. 2019;14(1): 1–8.

48. Orange ST, Marshall P, Madden LA, Vince RV. Short-term training and detraining effects of supervised vs. unsupervised resistance exercise in aging adults. J Strength Cond Res. 2019;33(10):2733–2742.
49. Hunter JR, Gordon BA, Bird SR, Benson AC. Exercise supervision is important for cardiometabolic health improvements: a 16-week randomized controlled trial. J Strength Cond Res. 2020;34(3): 866–877.
50. Kullman EL, Saylor SM, Little KD. Efficacy of whole-body suspension training on enhancing functional movement abilities following a supervised or home-based training program. J Sports Med Phys Fit. 2019;60(2):244–250.
51. Storer TW, Dolezal BA, Berenc MN, Timmins JE, Cooper CB. Effect of supervised, periodized exercise training vs. self-directed training on lean body mass and other fitness variables in health club members. J Strength Cond Res. 2014;28(7):1995-2006.
52. Enoksen E, Staxrud M, Tønnessen E, Shalfawi SA. The effect of supervised strength training on young elite male soccer players' physical performance. Serb J Sports Sci. 2013;7(4).
53. Grotle M, Vøllestad NK, Veierød MB, Brox JI. Fear-avoidance beliefs and distress in relation to disability in acute and chronic low back pain. Pain. 2004;112(3):343–352.
54. Khodadad B, Letafatkar A, Hadadnezhad M, Shojaedin S. Comparing the effectiveness of cognitive functional treatment and lumbar stabilization treatment on pain and movement control in patients with low back pain. Sports Health. 2020;12(3):289–295.
55. Schmidt RA. Motor schema theory after 27 years: Reflections and implications for a new theory. Res Quart Exerc Sport. 2003;74(4):366–375.
56. Brown LE, Weir JP. ASEP procedures recommendation I: accurate assessment of muscular strength and power. JEP Online. 2001;4(3).
57. Fisher J, Steele J, Smith D. High-and low-load resistance training: interpretation and practical application of current research findings. Sports Med. 2017;47(3):393–400.
58. Baker D. Science and practice of coaching a strength training program for novice and intermediate-level athletes. Strength Cond J. 2001;23(2):61.
59. Kerr ZY, Collins CL, Dawn Comstock R. Epidemiology of weight training-related injuries presenting to United States emergency departments, 1990 to 2007. Am J Sports Med. 2010;38(4):765–771.
60. 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):368–374.
61. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, Nieman DC, Swain DP. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc. 2011;43(7):1334-59.

62. Denton F, Power S, Waddell A, Birkett S, Duncan M, Harwood A, McGregor G, Rowley N, Broom D. Is It Really Home-Based? A Commentary on the Necessity for Accurate Definitions across Exercise and Physical Activity Programmes. Int J Environ Res Public Health. 2021;18(17):9244.
63. Teyhen DS, Shaffer SW, Lorenson CL, Halfpap JP, Donofry DF, Walker MJ, Dugan JL, Childs JD. The functional movement screen: a reliability study. J Orthop Sports Phys Ther. 2012;42(6):530-40.
64. Massey CD, Maneval MW, Phillips J, Vincent J, White G, Zoeller B. An analysis of teaching and coaching behaviors of elite strength and conditioning coaches. J Strength Cond Res. 2002;16(3):456-60.
65. Hong J, Kim J, Kim SW, Hong HJ. Effects of home-based tele-exercise on sarcopenia among community-dwelling elderly adults: Body composition and functional fitness. Exp Gerontol. 2017;87:33–39.