A Bayesian approach to interpret intervention effectiveness in strength and conditioning: Part 1. A meta-analysis to derive context-specific thresholds.

Review Article

Running head: S&C context-specific effect sizes

Paul Alan Swinton[1], Katherine Burgess[1], Andy Hall[1], Leon Greig[1], John Psyllas[1], Rodrigo Aspe[1], Patrick Maughan[1,2], Andrew Murphy[1]

Doi: 10.51224/SRXIV.9

SportRχiv hosted preprint version 1

07/09/2021

PREPRINT - NOT PEER REVIEWED

Institutions

[1] School of Health Sciences, Robert Gordon University, Aberdeen, UK

[2] College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

Corresponding Author

Dr. Paul Swinton

School of Health Sciences, Robert Gordon University

Garthdee Road

Aberdeen, UK,

AB10 7QG

p.swinton@rgu.ac.uk, +44 (0) 1224 262 3361

Please cite this paper as: Swinton, PA. Burges, K. Hall, A. Greig L. Psyllas J. Aspe R. Maughan P. Murphy A. A Bayesian approach to interpret intervention effectiveness in strength and conditioning: Part 1. A meta-analysis to derive context-specific thresholds. Pre-print available from SportRχiv. https://doi.org10.51224/SRXIV.9

Abstract

Background Strength and conditioning (S&C) interventions comprising methods such as resistance, sprint and plyometrics are used to enhance athleticism and sports performance. The effectiveness of interventions can be evaluated using effect sizes calculated from physical outcomes and then compared to threshold values. The purpose of this large meta-analysis was to identify threshold values specific to S&C and assess factors that influence effect size distributions.

Methods An online database and hand search of published and unpublished S&C intervention studies from the 1950's onwards was conducted. Interventions were categorized as the following: resistance, combined, plyometric, ballistic, sprint, isokinetic, concurrent, or agility. Pre- and post-intervention data comprising means and standard deviations were extracted from outcomes categorized as: maximum strength, power, explosiveness, jump, sprint, or agility. Study and participant data including intervention length, gender and training status (untrained, recreationally trained and highly trained) were also extracted. Standardised mean difference effect sizes (SMD_{pre}) were calculated and modelled with 4-level Bayesian hierarchical meta-analysis models using 0.25-, 0.5-, and 0.75-quantiles to determine small, medium, and large threshold values, respectively.

Results Data from 679 studies comprising 8904 effect sizes were included in the analyses. Threshold values obtained across the entire data were: small - 0.12 [95%CrI: 0.11 to 0.14]; medium - 0.43 [95%CrI: 0.42 to 0.45]; and large - 0.78 [95%CrI: 0.77 to 0.80]. Effect size distributions were shown to be shifted to higher values for longer duration interventions comprising maximum strength outcomes, untrained participants, females, and higher specificity coupling between training method and outcomes. Results from analyses were synthesised to provide updated threshold values to interpret effectiveness.

Conclusions The effectiveness of S&C interventions are influenced by a range of factors creating systematic shifts in SMD_{pre} values. It is recommended that researchers and practitioners use the

S&C specific threshold values presented instead of Cohen's generic values, with scope provided for adjustment based on relevant factors.

Key Words: S&C; evaluation; effect size; Bayesian; specificity; training

1.0 Introduction

Strength and conditioning (S&C) is now a well-established discipline within sport and exercise science comprising a range of contemporary training methods and practices (1, 2). Much of the popularity of S&C originates from the perspective that muscular strength is of primary importance in athletic and sports performance and can be developed extensively with relatively limited time and cost (3, 4). Additionally, related factors including mechanical power and rate of force development are considered among the most important trainable factors for athletes (3). Research investigating training methods such as resistance, sprint and plyometrics has experienced rapid growth since the late 1990s (5, 6) where a key focus includes identifying the best regimes to improve different aspects of fitness. Intervention studies generally select surrogate parameters such as maximum repetition testing, sprinting, jumping and change of direction tasks to link training adaptations to aspects of fitness relevant to sports performance and common athletic tasks (7). An extensive research base has also emerged demonstrating strong associations between kinetic outputs and athletic tasks (8-12) including jumping, sprinting and change of direction where it is believed that the ability to perform these movements effectively may determine the outcome of important events during competition (3).

A primary aim of S&C research is to identify training methods that maximise transference of adaptations to sporting outcomes and tasks (13). It is generally accepted that the magnitude of transference and long-term performance adaptations are dependent primarily on the training principles of specificity and overload, respectively (14, 15). Specificity is often contextualised in the kinematic and kinetic similarities between the S&C training and the targeted sporting outcomes and actions (16, 17). For example, heavy resistance training comprising squat exercises may be considered moderately specific to vertical jumping but non-specific to the maximum velocity phase of sprinting. It has been identified that the principles of overload and specificity can be conflicting, where the greater the overload applied to an exercise the lower the specificity and vice versa (17). Various transference models have been developed and have been reviewed previously, with attempts to carefully balance overload and specificity (13, 14). Most models developed by practitioners have been theoretical in nature, however, researchers in S&C have primarily focused on empirical models quantifying correlations of kinematic and kinetic parameters between exercises and sporting movements and their proxies (18-21). Although, this type of observational research may provide some value in identifying possible best training practices, it does not establish causality, and correlations are sensitive to multiple methodological factors including sample size and heterogeneity, training status, exercise and load selection, test and mechanical variable selection and equipment used for data collection (19, 22, 23). In contrast, longitudinal research comparing different training interventions ranked according to the degree of specificity provides one of the best means of assessing the overall influence of specificity (24-28). A challenge in integrating the findings of longitudinal research within a general framework includes the potential for idiosyncratic findings based on experimental choices of individual studies. Findings may be influenced by a range of moderating factors including the study population, the length of intervention, the outcomes assessed, the training regimes manipulated, and the kinematic and kinetic parameters manipulated to alter the degree of specificity. Given the range of factors that could influence results, evidence synthesis approaches including meta-analysis may provide the most effective means to establish the importance of specificity and whether the phenomena is consistent across factors such as training methods (e.g. resistance, sprint, power and agility).

Evidence synthesis and the use of meta-analysis to objectively quantify general phenomena across an extensive range of studies is common in S&C, and has been influenced by the seminal research of Rhea and colleagues (29-34). Incorporating data from 140 studies Rhea et al. (29) investigated the dose-response relationship for developing maximum strength. Subsequent meta-analyses

analyses were conducted on the dose-response relationship of athletes only (32), the comparison of single versus multiple sets (33), and periodized versus non-periodized training (34). Collectively, Rhea and colleagues across their meta-analyses synthesised results from approximately 400 individual studies (30). In addition to addressing specific research questions in S&C, the large syntheses provided a means of interpreting the practical significance of intervention effects. Rhea and colleagues (29, 31) performed their meta-analyses using the pre-standardised mean difference (SMD_{pre}), dividing the mean change by the pre-intervention standard deviation. This statistic has become the most common effect size reported across sport and exercise science and comprises several conceptual advantages (35). The effect size is dimensionless and enables the synthesis of outcomes reported across different units and scales (e.g. vertical jump power in thousands of Watts, and vertical jump height in tens of centimetres). However, the major advantage of the SMD_{pre} is that it describes how future individuals performing the intervention should be expected to change in an outcome variable relative to the sampled population.

To summarise the practical significance of an intervention it is common to categorize SMD_{pre} values using qualitative labels such as "small", "medium" and "large". The standard terminology and threshold values used to demarcate these levels were proposed by Cohen (36) in the context of behavioural and social sciences. However, given Cohen himself acknowledged his thresholds were somewhat arbitrary (36), research has been conducted across various scientific disciplines to update these thresholds and enable more precise interpretations of intervention effects (37-42). Results have generally demonstrated substantive differences between Cohen's thresholds and those derived empirically, with examples of both under- (43) but predominantly over-estimation (37, 38) of thresholds. In addition, research has shown that even within a single discipline, substantive differences in the distribution of effect sizes can exist across sub-domains (44). Using almost 3000 effect sizes obtained from approximately 400 studies, Rhea (30) produced threshold values specific to strength training research and concluded that Cohen's (36) thresholds were not appropriate and effect sizes from strength training research tended to be much larger. Whilst the updated threshold values presented by Rhea (30) showed variation between untrained and recreationally trained individuals, the threshold values for highly trained individuals were similar to those originally proposed by Cohen (36). In contrast to recent attempts to create discipline specific thresholds, Rhea (30) did not clearly describe the methods used to obtain the context specific thresholds and only reported the mean SMD_{pre} value. Additionally, the statistical methods available to synthesise research findings including outlier detection, use of multi-level models to combine multiple outcomes from individual studies and use of flexible models to estimate beyond the mean and describe the underlying effect distribution (e.g. through various quantiles), have developed substantively since Rhea and colleagues (29-34) seminal work. Therefore, the purpose of this study was to update existing guidelines deriving threshold values to interpret the effectiveness of a range of contemporary outcomes and training methods in S&C. By including modern statistical methods more precise descriptions of the distribution of intervention effects can be presented and address the extent to which values are influenced by a range of factors including the training method (e.g. resistance, plyometric and sprint), the outcome type (e.g. maximum strength, power, explosiveness and jump performance), training status, gender and intervention duration. Finally, the importance of specificity was also assessed by determining whether SMD_{pre} values were greater when the training method and outcome type were identified as being specific (e.g. plyometric and vertical jump) compared with non-specific (e.g. plyometric and maximum strength).

2.0 Methods

2.1 Search strategy

A search was performed for published and unpublished studies in the English language that included S&C interventions conducted prior to January 2018. The search was performed using Embase, Medline, Web of Science, Sport Discus and Google Scholar. Hand searching of relevant journals including Medicine and Science in Sports and Exercise, the Journal of Strength and Conditioning Research, and Research Quarterly was also conducted. Database search terms were included to identify various training modes, longitudinal interventions, and a range of outcome measures. The following keywords and phrases were combined with Boolean operators; "strength" OR "resistance" OR "sprint" OR "plyometric" OR "exercise" AND "intervention" OR "training" OR "program" OR "programme" AND "1RM" OR "repetition maximum" OR "speed" OR "velocity" OR "power" OR "jump" OR "change of direction" OR "agility" OR "acceleration" OR "rate of force development". No restriction was placed on the date of the study.

2.2 Inclusion criteria

Inclusion and exclusion criteria for the current meta-analysis were kept concise in attempts to include as many relevant S&C training interventions and dependent variables as possible. Inclusion criteria comprised: 1) any intervention-based study ≥ 4 weeks; 2) healthy trained or untrained participants with a mean age between 14 and 60; 3) intervention group with a minimum of 4 participants; 4) pre and post intervention means and standard deviations; and 5) sufficient information provided to appropriately categorize the training method (identified in section 2.3). Studies comprising interventions that were predominantly aerobic-based or rehabilitation focused were excluded.

2.3 Study selection and data extraction

Following deduplication, a three-level selection process comprising title, then abstract then full-text screening was completed. Studies were screened and selected for inclusion independently by AM with discussions with PS and KB where required. A standardised extraction codebook was developed using Microsoft Excel, with data extracted and coded independently by four researchers (AM, JP, AH, LG) in duplicate with AM completing extraction for all studies to provide consistency. The following data were extracted: 1) study details (authors, year, total number of groups, and control type); 2) participant characteristics (final study n, gender, training status, and age); 3) outcome type (maximum strength, power, explosiveness, jump performance, sprinting performance, and agility performance); 4) training characteristics (intervention duration, training method: resistance, combined, plyometric, ballistic, sprint, isokinetic, concurrent, and agility); and 5) pre and post intervention means and standard deviations.

The following definitions were used to categorize outcome types: 1) maximum strength: a measure of maximum force production where time was not limited (e.g. 1-6 repetition maximum, isometric mid-thigh pull, peak torque); 2) power: a direct measurement of power output measured in Watts (absolute and normalised relative to body mass); 3) explosiveness: a measure of force or velocity where time was restricted (e.g. rate of force development, time to peak force, impulse); 4) jump performance: a measure of jump height or distance; 5) sprint performance: a measurement of the time to complete a specified linear distance or the velocity achieved; and 6) agility performance: a measurement of the time to complete a change of direction or reactive task. Categorisation of training methods were based on the dominant mode, however, where training volume in a non-dominant mode approached or exceeded approximately 20% of the total volume, the intervention was then categorized as combined (e.g. an intervention combining resistance and plyometrics). Interventions comprising aerobic activity as the non-dominant mode accounting for approximately 20% or more of total training volume were categorized as concurrent. Training status was categorized using definitions previously set by Rhea (30) based on S&C training experience: untrained (<1 year); recreationally trained (1-5 years); highly trained (>5 years). Where pre-post intervention data were not presented in text but in figures, data were extracted using PlotDigitizer 2.6.8 Windows.

2.4 Statistical analysis

A Bayesian framework was chosen over frequentist methods to provide a more flexible modelling approach and enable results to be interpreted intuitively through reporting of subjective probabilities (45). Effect sizes and their sampling variance were calculated using group mean and standard deviation values calculated pre-intervention and at any subsequent time-point. SMD_{pre} was calculated by dividing the relevant mean difference by the pre-intervention standard deviation. The sampling variance $\sigma_e^2(\text{SMD}_{\text{pre}})$ of the effect size (46) was calculated with the following formula:

$$\sigma_e^2(\text{SMD}_{\text{pre}}) = \frac{n-1}{n(n-3)}\left(2(1-r) + n\text{SMD}_{\text{pre}}^2\right) - \frac{\text{SMD}_{\text{pre}}^2}{c(n-1)^2}$$

where n is the sample size, r is the correlation between repeated measures, and $c(n-1)$ is the bias function which was approximated by $1 - \frac{3}{4n-5}$ (47). To account for the small sample sizes generally used in S&C, a bias correction was applied to the effect size and sampling variance by multiplying by the approximated bias function and its square, respectively. To account for uncertainty in σ_e^2 due to non-reporting of r, the values were allowed to vary and were estimated by including an informative Gaussian prior approximating correlation values centred on 0.7 and ranging from 0.5 to 0.9. All meta-analyses were conducted using a nested four-level mixed effects meta-analytic model. The series of nestings included the individual study (level 4), the outcome (level 3), the measurement occasion (level 2) and the sampling variance (level 1). The meta-analysis model (null model) describing the average effect and variance across each level can be expressed as:

Level1: $d_{ijk} = \beta_{0ijk} + e_{ijk}, \qquad e_{ijk} \sim N(0, \sigma_e^2)$

Level2: $\beta_{0ijk} = \eta_{0jk} + r_{ijk}, \qquad r_{ijk} \sim N(0, \sigma_r^2)$

Level3: $\eta_{0jk} = \theta_{0k} + u_{0jk}, \qquad u_{0jk} \sim N(0, \sigma_u^2)$

Level4: $\theta_{0k} = \gamma_0 + v_{0k}, \qquad v_{0k} \sim N(0, \sigma_v^2)$

To give

$$d_{ijk} = \gamma_0 + v_{0k} + u_{0jk} + r_{ijk} + e_{ijk}$$

where d_{ijk} is the observed effect size at measurement occasion i $(i = 1,2, \dots, I_{jk})$, from outcome j $(j = 1,2, \dots, J_k)$ and from study k $(k = 1,2, \dots, K)$. The indexing I_{jk} denotes that the number of measurement occasions may vary across outcomes and studies, and J_k denotes the number of outcomes may vary across studies. The random effects across the different levels $(v_{0k}, u_{0jk}, r_{ijk}, e_{ijk})$ were assumed to be independent such that the variance of an observed effect size was $\sigma^2(d_{ijk}) = \sigma_r^2 + \sigma_u^2 + \sigma_v^2 + \sigma_e^2$. The null model states that the underlying true population average effect size is γ_0, and for each study the average effect size will equal θ_{0k}, and due to the use of a normal distribution, the value for most studies will lie in the interval $\gamma_0 \pm 2\sigma_v^2$. The true effect sizes for different outcomes within studies, and across measurement occasions within outcomes can then move further from θ_{0k}, based on the magnitude of the variances σ_u^2 and σ_r^2. If we consider two observations $d_{ijk}, d_{i'j'k'}$ then the covariance of these are

$$cov(d_{ijk}, d_{i'j'k'}) = cov(v_{0k}, v_{0k'}) + cov(u_{0jk}, u_{0j'k'}) + cov(r_{ijk}, r_{i'j'k'}) + cov(e_{ijk}, e_{i'j'k'}).$$

Which is equal to $\sigma_v^2 + \sigma_u^2 + \sigma_r^2$ (for $i = i', j = j', k = k'$), $\sigma_v^2 + \sigma_u^2$ (for $i \neq i', j = j', k = k'$), and σ_v^2 (for $i \neq i', j \neq j', k = k'$). Fitting of the null model also enabled calculation of the variance partition coefficients (VPCs), such that the study level VPC is $\frac{\sigma_v^2}{\sigma_v^2+\sigma_u^2+\sigma_r^2+\sigma_e^2}$, outcome level VPC is $\frac{\sigma_u^2}{\sigma_v^2+\sigma_u^2+\sigma_r^2+\sigma_e^2}$, measurement occasion level VPC is $\frac{\sigma_r^2}{\sigma_v^2+\sigma_u^2+\sigma_r^2+\sigma_e^2}$, and observed effect VPC is $\frac{\sigma_e^2}{\sigma_v^2+\sigma_u^2+\sigma_r^2+\sigma_e^2}$. Addition of the VPCs can be used to estimate the expected (population) correlation between two randomly chosen elements within the same nesting structure (48).

To quantify small, medium, and large threshold values, parameters from the respective meta-analysis model were used to generate posterior predictions and the 0.25-, 0.5-, and 0.75-quantiles calculated, respectively. Weakly informative Student-t prior and half-t priors with 3 degrees of freedom and scale parameter equal to 2.5 were used for intercept and variance parameters (49). Inferences from all analyses were performed on posterior samples generated using the Hamiltonian Markov Chain Monte Carlo method. Interpretations were based on the median value ($SMD_{pre0.5}$: 0.5-quantile), credible intervals (CrIs) and calculation of probabilities using the posterior sample. It was determined a priori to assess the influence of training method, outcome type, training status, gender and intervention duration on effect size and threshold values. This was achieved by subset analysis for training method and outcome type, and meta-regressions for training status, gender and intervention duration. Meta-regressions were presented by selecting one level of the factor as a reference to make comparisons with the median and 95% CrI given ($\beta_{\text{Reference:Comparison}}$ = Median [95%CrI: lower bound to upper bound], such that $\beta > 0$ indicates an increased effect of the comparison relative to the reference). The initial meta-analysis model combining the entire data set was fit with a normal distribution, a skew normal distribution, and a t-distribution. No substantive improvements in model fit were obtained with the skew normal or t-distributions based on Watanabe-Akaike information criterion. Consequently, a normal distribution was used throughout all analyses which were performed using the R wrapper package brms interfaced with Stan to perform sampling (50). Outlier SMD_{pre} values were identified by adjusting the distribution by a Tukey g-and-h distribution and obtaining the 0.0035- and 0.9965-quantiles, with values beyond these points removed prior to further analysis (51). Convergence of parameter estimates were obtained for all models with Gelman-Rubin R-hat values below 1.1 (52).

3.0 Results

The search strategy returned 110,662 records which reduced to 2108 studies following deduplication and title screening. Following abstract screening this reduced to 973, and 706 following full-text screening. A total of 679 studies featured the required data to be included in the meta-analysis, generating a total of 8904 effect sizes. Most studies comprised untrained participants (n=407, 60.0%), followed by recreationally trained (n=226, 33.4%) then highly trained (n=45, 6.6%). Comparison of gender identified that most studies were comprised solely of males (n=454, 66.8%), followed by a similar number of studies comprising both males and females (n=116, 17.1%), and then females only (n=109, 16.1%).

Information describing the distribution of study characteristics are presented in table 1. Most interventions lasted between 6 and 12 weeks, with 25 studies (3.7%) including interventions longer than 25 weeks, and 14 studies including interventions longer than 52 weeks (2.1%). The median number of outcomes extracted from studies was 4 with interquartile range (IQR): 2-7. The median number of groups included in studies was 2 with IQR: 2-3. The most common training method and outcome type was standard resistance training and maximum strength, respectively (Table 1). The next most common training method was combined training, with many studies including resistance with other training modes. Following maximum strength, the most common outcome type measured was jump performance (Table 1). Interrelationships between training methods and outcome types were identified with a high degree of specificity between the variables (Figure 1). This was demonstrated for maximum strength outcomes which were measured most (~50%) in studies that comprised standard resistance or isokinetic interventions, whereas, sprint and agility outcomes were measured most in studies comprising sprint or agility-based interventions, respectively.

A total of 103 outliers were removed from the analysis with a lower bound threshold of -1.55 (40 effect sizes below) and an upper bound threshold of 6.5 (63 effect sizes above). Direct calculation of the mean, 0.25- (small), 0.5- (medium), and 0.75-quantiles (large) from the complete empirical data returned the following SMD_{pre} values: 0.62, 0.16, 0.46 and 0.86, respectively. Based on shrinkage from application of the meta-analysis model and borrowing of information across studies and outcomes, the pooled mean estimate was reduced to $SMD_{pre0.5}$=0.52 [95%CrI: 0.49 to 0.54] and the quantile values reduced to: 0.25-quantile$_{0.5}$ = 0.12 [95%CrI: 0.11 to 0.14]; 0.5-quantile$_{0.5}$ = 0.43 [95%CrI: 0.42 to 0.45]; and 0.75-quantile$_{0.5}$ = 0.78 [95%CrI: 0.77 to 0.80]. The four-level variance parameters were $\sigma_{e_{0.5}}$=0.29 [75%CrI: 0.15 to 0.81]; $\sigma_{r_{0.5}}$=0.14 [75%CrI: 0.13 to 0.15]; $\sigma_{u_{0.5}}$=0.26 [75%CrI: 0.25 to 0.27]; and $\sigma_{v_{0.5}}$=0.33 [75%CrI: 0.31 to 0.35].

Separate meta-analysis models were conducted for each outcome type (figure 2) and each training method (Table 2). The largest threshold values were obtained for strength and agility outcomes, with substantively lower values obtained for explosiveness and sprint outcomes. A similar pattern was identified for effect size distributions across training methods, with the largest threshold values obtained for standard resistance and isokinetic interventions, and the lowest values obtained for ballistic and sprint interventions. For both sets of variables, differences in SMD_{pre} values for strength and sprint were approximately 0.2 to 0.3 (e.g. best estimates for strength vs sprint outcomes - small: 0.23 vs. 0.02; medium: 0.58 vs 0.24; and large: 0.99 vs 0.56. Best estimates for resistance vs sprint interventions - small: 0.14 vs. -0.04; medium: 0.48 vs 0.21; and large: 0.87 vs 0.48). An effect size distribution between that obtained for standard resistance and sprint interventions was obtained for a range of diverse training practices categorized as other. Example training practices in this categorization included sustained efforts with unconventional equipment, intermittent circuit style training, sensorimotor training and core stability interventions.

To assess whether effect sizes were altered by training status, intervention duration and gender, meta-regressions were performed controlling for outcome and training method based on the previous analyses establishing their influence. A rank order-effect of training status was identified with the largest pooled mean effect size obtained for untrained participants, followed by recreationally trained and then highly trained ($\beta_{\text{Untrained:Recreational}0.5}$=-0.04 [95%CrI: -0.10 to 0.02], probability less than 0 = 0.920; $\beta_{\text{Untrained:Trained}0.5}$=-0.15 [95%CrI: -0.26 to -0.05], probability less than 0 = 0.998). A similar rank order-effect was identified for intervention duration with the largest pooled mean effect size obtained for interventions longer than 10 weeks, followed by interventions between 6-10 weeks and then interventions less than 6 weeks ($\beta_{<6\text{weeks}:6-10\text{weeks}0.5}$=0.19 [95%CrI: 0.15 to 0.22], probability greater than 0 > 0.999; $\beta_{<6\text{weeks}:>10\text{weeks}0.5}$=0.29 [95%CrI: 0.25 to 0.33], probability greater than 0 > 0.999). An effect was also identified for gender, with females showing greater pooled mean SMD_{pre} values than males ($\beta_{\text{Female:Male}0.5}$=-0.09 [95%CrI: -0.10 to -0.00], probability less than 0 = 0.976).

To assess the effects of specificity on threshold values, meta-analyses were conducted comparing values obtained for matching training method and outcome type (resistance training and maximum strength; plyometric training and jump performance; agility training and agility performance; sprint training and sprint performance) with values obtained for unmatched combinations (Figure 3). Consistent and substantive differences were identified, with all analyses demonstrating greater effect sizes where specificity was included through matching of intervention and outcomes, with most effect sizes increasing by approximately 0.2 when moving from non-specific to specific combinations (mean difference: 0.23; standard deviation of difference: 0.10).

To facilitate use of meta-analytic results obtained across the present study and guide future interpretations of the effectiveness of S&C interventions, an overall summary expressed in terms of clustered outcomes is presented (Figure 4). Ranges are provided for small, medium and large thresholds, with researchers and practitioners able to select appropriate values based on moderating factors such as intervention length, training status, gender and specificity.

4.0 Discussion

The present study comprises one of the largest meta-analyses in sport and exercise science and the largest synthesis of contemporary S&C interventions to date. The analyses show that relatively short-duration S&C interventions can create improvements across a range of outcomes such that individuals can expect to progress substantively relative to their respective population from pre- to post-intervention. The results, however, show that this expected progress is influenced by a range of factors including primarily the outcome assessed, the training method and the specificity between these two factors. Effect size distributions were shifted towards higher values for resistance training and maximum strength outcomes, whereas distributions were shifted towards substantively lower values for sprint interventions and outcomes. Specificity and the matching of training methods and outcomes was found to consistently improve the success of interventions with typical increases in SMD_{pre} values of approximately 0.20 compared to low specificity coupling. Across the entire research base, Cohen's generic threshold values (0.2 = small, 0.5 = medium, 0.8 = large) were broadly in line with the values obtained from the meta-analysis presented in the current study. However, based on large potential differences in effect size distribution caused by factors such as outcome type, training method and specificity, it is suggested that context relevant threshold values are required to appropriately interpret the effectiveness of S&C interventions. The results of the present study have been synthesised to provide practical context relevant thresholds with ranges provided for researchers and practitioners to select appropriate values based on the specifics of the intervention evaluated (e.g. the training status and gender of the participants, and the specificity of the training and outcomes evaluated).

The extensive literature base included in this meta-analysis provides the most comprehensive overview of S&C interventions to date. The analysis shows that the majority (60-70%) of studies comprise intervention durations of 8 weeks or less, typically comprising two groups of 4 to 12 participants. Additionally, most interventions recruit untrained (60%) or recreationally trained (33%) individuals, with standard resistance training comprising the most frequently investigated training method. The limitation of focusing on short duration interventions with untrained participants has been identified previously (53). It has been argued that untrained individuals can make substantive improvements across a range of outcomes whilst engaging in relatively inefficient training interventions due to improvements in factors such as motor skills rather than structural or physiological adaptations (53-55). The results of the present analysis provided support for this perspective and showed a rank-order effect with the largest improvements across all outcomes obtained by untrained participants, followed by recreationally trained and then highly trained participants. It was estimated that untrained participants may be expected to experience an additional 0.15 SMD_{pre} increase compared with highly trained participants. Additionally, the results of the present meta-analysis showed a rank order effect with greater effect sizes expected for longer duration interventions, but that these differences were likely to slow. This pattern has been demonstrated in individual studies comprising intermediate and post interventions assessments over longer time periods (55). Contributing factors to the prevalence of short-term studies with lesser-trained individuals likely include the high cost (financial and time) of long duration interventions and challenges recruiting highly trained participants including lack of access and trust (56).

The results of the meta-analysis consistently demonstrated that the largest effects were obtained for resistance training methods and maximum strength outcomes. When restricted to the subset analysis of maximum strength outcomes, the threshold values were: small (0.23 [95% CrI: 0.21 to 0.26]), medium (0.58 [95% CrI: 0.56 to 0.60]), and large (0.99 [95% CrI: 0.96 to 1.02]); and when restricted to standard resistance training the threshold values were: small (0.14 [95% CrI: 0.12 to 0.17]), medium (0.48 [95% CrI: 0.46 to 0.51]), and large (0.87 [95% CrI: 0.85 to 0.90]). The adjustment in distribution compared to other training methods and outcomes is likely explained by several factors. Resistance training represented the largest proportion of interventions (35.5%) included in the analysis and due to the long history of study is the most well understood and developed training method in S&C. Greater effect sizes are likely to reflect this increased refinement and specificity linking traditional resistance training methods and maximum strength outcomes. In addition, researchers generally test maximum strength using the same exercises included in the training intervention, maximizing specificity and the improvements measured during the intervention (57). Of all training methods investigated, the largest effect sizes were obtained for isokinetic interventions with the mean SMD_{pre} value estimated to be 0.63 [95% CrI: 0.48 to 0.79], and the large threshold value estimated as 1.0 [95% CrI: 0.92 to 1.1]. Isokinetic dynamometers provide a unique mechanical stimulus with participants able to contract skeletal muscles with near-maximum or maximum effort across a range of velocities (58). Similar to standard resistance training interventions, researchers frequently use the same testing movements as featured throughout the training and also match velocities which research has identified as an important aspect of specificity (59). However, studies have also demonstrated that adaptations obtained during isokinetic interventions effectively transfer to standard maximum strength tests (58), albeit standard tests across all training methods tend to exhibit larger increases than isokinetic tests (60).

Following maximum strength, the outcome type generating the largest SMD_{pre} values was agility. Outcomes measuring agility and the related construct of change of direction speed represent a developing area within S&C (61, 62). A range of factors are thought to contribute to performance in agility and change of direction tasks including pattern recognition and reactive strength (61). Whilst reasons for the shifted effect size distribution in comparison to other outcomes requires further study, potential explanations include the complex multifaceted nature of the tasks and the scope for multiple limiting factors to be addressed. Additionally, it is recognised that many agility and change of direction tasks include substantive skill elements (62), such that failure to appropriately familiarize participants could lead to large systematic biases with regards to learning effects and subsequent overestimations of effect sizes.

In contrast to the upward shifted effect size distributions for strength and agility, the outcomes of explosiveness and sprint were shifted in the opposite direction with the small threshold close to a null improvement (explosiveness: -0.04 [95% CrI: -0.06 to -0.01; sprint: 0.05 [95% CrI: 0.04 to 0.08]). Previous research has demonstrated that more experienced and stronger individuals are able to express higher levels of force during time restricted tasks (3, 63). Whilst the present study demonstrated that short duration interventions can create relatively large improvements in maximum strength, it is possible that the ability to express this increased force production in tasks under restricted time is limited by factors such as the levels of neural drive (discharge rate and MU synchronisation), muscle fibre composition (type I versus type II/type IIx), connective tissue adaptions (musculotendinous stiffness) and intra and inter-muscular coordination (3, 64, 65). Additionally, one of the key results of the present study was the strong effect of specificity on SMD_{pre} values. Most studies included in the present review featured traditional resistance training where muscular actions occur over long durations. Given the low kinetic similarity between the most common training mode and explosiveness outcomes, this lack of specificity may partly explain the shift in effect size distributions. An additional explanation for the shifted distribution may include limitations in the testing process. An extensive range of tests and measurements procedures are used in S&C to obtain explosiveness outcomes with no precise definitions available. Instead, it has been noted that the explosiveness outcome categorization focusses on imprecise communication used to describe the general quality and overall effort required in the testing movements (66). In the present meta-analysis, all force and velocity measurements made under restricted time were classified as explosiveness, including: rate of force development (RFD), rate of velocity development, rate of torque development, time to peak force, peak force at varying time points, RFD at varying time points, mean and peak velocity with and without load and impulse. However, the general lack of clarity regarding the construct and the extensive range of measurements and associated data processing approaches that are used may have influenced the results obtained. It is possible that subsets of these measurements comprise effect size distributions closer to strength and power outcomes but when assessed as a whole were diluted. Despite its common use in S&C, the term explosiveness has been criticised by its imprecision (67) with recommendations that more consistent and precise constructs such as RFD be used. However, even with RFD there are concerns regarding varying levels of reliability, dependent on the specific test and metric measured (e.g. average RFD, maximal RFD, RFD at varying time points) (65, 66, 68, 69). Further research is required to delineate constructs related to explosiveness and develop appropriate tests and data processing approaches.

Many of the potential explanations for the shifted effect size distribution for explosiveness outcomes may also apply to sprint outcomes. Whilst there is generally no ambiguity in definitions and less range of sprint outcomes compared with explosiveness, there still remains variation in the underlying constructs that are measured. The most common sprint outcomes included in S&C intervention studies include the time to sprint between 5 and 50 m, with intervals comprising 10, 20 and 30 m frequently used. The majority of studies have been conducted with either team sport or untrained participants that achieve maximum velocity between 15 to 40 m, in comparison to trained sprinters that require distances of 40 to 80 m to achieve maximum velocity (70-73). Consequently, sprint data collected over 10 to 30 m may provide researchers with divergent outcomes describing both acceleration and maximum velocity. Previous studies have reported strong associations with outcomes designed to assess acceleration (e.g. 10 m) and horizontal force, power and relative strength with longer duration ground contact times (approx. 200ms) (74, 75). In contrast, maximum velocity sprinting has been shown to be dependent on the ability to maintain large horizontal and vertical forces whilst minimizing braking forces with reduced ground contact time (approx. 100ms) (72, 75, 76). Unique factors that determine acceleration or maximum velocity during sprinting may respond differently across training methods but with studies generally reporting data from multiple distances, the overall effect size may be diluted. Sprinting also comprises a substantive and complex technique element (73-77), and given the low number of studies (~10%) that included sprint specific interventions a lower effect size distribution may be expected. More broadly, the lower effect size distribution for sprint outcomes may also reflect a lack of specificity with regards to development of relevant physical outputs. Most training methods included in the review focused on bilateral production of maximum vertical forces over long durations. In contrast, sprinting activities require high forces produced over short ground contact times that are predominantly unilateral with substantive horizontal components (70, 74, 75). Previous research investigating the capacity for non-specific training moderate to improve sprint performance have shown that large increases in maximum strength (~12-18%) are required to elicit only small decreases in sprint times (~-2-8%) (78-80). Research has consistently demonstrated the effectiveness of S&C in improving morphological qualities, however, the transference of improvement from non-specific training to more complex, time restricted tasks may be limited. Therefore, use of the current proposed context specific thresholds is likely to provide better understanding and interpretation of the effectiveness of a range of interventions (figure 4).

In the present review, it was identified that specific couplings of training methods and outcomes resulted in an average increase in SMD_{pre} values of 0.23 and range of 0.11 to 0.38. This analysis provides a general quantitative description of the additional improvements that may be expected with increased specificity. In contrast, S&C research has predominantly emphasised the relationships between mechanical variables and measures of athleticism, with the application of mechanical overload as the primary means of progression. Altering coordinative factors such as intra and inter-muscular coordination and intention and direction of movement have been highlighted as key determinants when attempting to maximise transfer of training (13, 14, 17, 64). Therefore increasing specificity, while maintaining a level of kinetic overload may provide additional skill acquisition, allowing for greater application of physical capability within the context of the target outcome (17). A prototypical example of specificity in S&C can be obtained from sprint interventions. Common sprint specific interventions include resisted and assisted sprinting that allow for the overload of kinetic outputs such as vectorized force and power whilst also maintaining high levels of kinematic similarity. When performing resisted sprint training, a key consideration to balance specificity and overload has been the selection of the load used (81-85). In a recent meta-analysis, Alcaraz et al. (85) concluded resisted sprint training (sled towing) was most beneficial with loads <20% body mass. Average pooled SMD_{pre} values up to 0.61 were reported, which lay beyond the upper bound of the large threshold value for sprint outcomes identified in the present meta-analysis (0.57 [95% CrI: 0.54 to 0.60]). However, the large threshold value increased to 0.69 [95% CrI: 0.66 to 0.73] in the present study when analyses were restricted to specific sprint training and outcome couplings, better aligning with the findings of Alcaraz et al. (85) and supporting the conclusion that substantive improvements in sprint performance can be expected when using the most appropriate interventions. In addition, more recent research has suggested that optimum loading of resisted sled training based on sprint specific force-velocity and power-velocity relationships may be closer to 69-96% body mass (83, 84). Very heavy resisted sled training (80+% body mass) has been shown to improve horizontal force, power and ratio of vertical and horizontal forces (84). Therefore, in contrast to prior beliefs, maximizing sprint specific kinetics during resisted sprint training may be more beneficial than attempting to maintain high kinematic specificity including ground contact time and stride parameters (length and frequency). Whilst the results from the present meta-analysis clearly demonstrate that specificity plays an important role in increasing effectiveness of S&C interventions, further research is required to identify how best to apply specificity concepts and the relative importance of kinematic and kinetic similarities across different training methods and outcomes.

The similar effect size distributions obtained for resistance training and concurrent training is noteworthy given the diverse physiological adaptations mediated by strength and endurance activities (86, 87) and traditional perspectives of interference between the training methods (88). Fyfe and Loenneke (89) have argued that variables unrelated to training dose such as participant training status are important considerations when interpreting adaptations to concurrent training. Support for this argument was provided in a recent meta- analysis (90) that identified impaired strength development in highly trained, but not in moderately trained or untrained individuals. These findings may explain why the effect size distribution for concurrent training was not shifted downwards in the present meta-analysis, given most of the included studies comprised untrained or recreationally trained participants and their window of adaptation available for strength may nullify the potential for interference (89). Moreover, the results of the present meta-analysis reflect contemporary research (91, 92) demonstrating no or minimal interference in comparison to earlier concurrent research (93, 94). Where strength training is the dominant training modality and primary focus, the use of novel endurance methods such as interval training (91, 92) may reduce moderating factors, such as training volume that are linked to greater interference (89).

The overall findings and subsequent guidance regarding interpretation of S&C interventions is different to that of previous large meta-analyses (29, 30). Including approximately one third of the effect sizes included in the present meta-analysis, Rhea (2004) suggested the following ranges to interpret small (0.5-1.25), medium (1.25-1.9) and large (>1.9) effect sizes for interventions comprising untrained individuals. Not only are the bandings substantively wider than those presented here, but the upper bound of the small threshold proposed by Rhea (30) is beyond our threshold for a large effect. Whilst the thresholds presented by Rhea (30) for highly trained participants (small: 0.25-0.5; medium: 0.5-1.0, large: >1.0) were closer to those presented here, the intervals are still considerably wider and beyond the values we present. The recommendations presented by Rhea (30) were based on resistance training interventions and maximum strength outcomes only. When restricting analyses in the present study to this specific coupling, the small, medium and large thresholds obtained were closer (small: 0.30 [95% CrI: 0.27-0.34]; medium: 0.67 [95% CrI: 0.63 to 0.70]; large: 1.10 [95% CrI: 1.03 to 1.12]). Multiple factors specific to the analyses may explain the different recommendations made. Whilst Rhea (30) stated that the included recommendations were made "after careful and thorough examination of the effect sizes", no details were provided regarding the analysis and selection of values. Rhea (30) stated that the mean effect size calculated was approximately 1.25. In the present meta-analysis, direct calculation of the mean effect size returned values of 0.62 across the whole data and 1.09 for the sub-selection of resistance training interventions and maximum strength outcomes. However, in the present study the influence of outlying measurements was reduced by removing outliers and applying an appropriate meta-analysis model that accounted for multiple effect sizes from the same study, multiple measurements made on the same outcome within a study, and borrowed information across the data to shrink values and obtain more precise estimates of the underlying distribution. In addition, threshold values were clearly anchored selecting quantile values describing the first three quarters (i.e. majority) of the distribution. In contrast, selecting thresholds based primarily on rounded values may result in relatively uncharacteristic values being selected towards the latter portions of the distribution and potentially within the tails of a very skewed empirical distribution. Given these differences, the recommendations presented in the current meta-analysis are more likely to provide appropriate threshold values, and the updated analysis provides guidance on the interpretation of S&C outcomes beyond maximum strength, reflecting the diverse range of practices currently employed.

Whilst this is the largest meta-analysis conducted on contemporary S&C training interventions, there are multiple limitations that should be considered when interpreting the findings. Whilst an extensive literature search was implemented, the process was not exhaustive and many studies particularly those published in languages other than English are not included. In addition, the increasing volume of S&C research progressively reduces the fraction of the overall research base included. However, there is no reason to believe that the studies included are not representative and therefore descriptions of the effect size distributions are likely to be appropriate. When conducting meta-analyses of this type, there are requirements to categorize factors (e.g. training methods and outcomes) to enable statistical pooling and generalize findings. With any categorization, appropriate alternatives and ambiguities exist which may influence conclusions obtained. In the present meta-analysis ambiguities existed primarily in the categorization of outcomes as explosiveness and training methods as other. Future research is required to identify and delineate the relevant constructs underpinning force generation under restricted time, and develop valid and reliable outcomes to measure these constructs. Also, as more S&C interventions are conducted the ability to repeat similar analyses on a subset of training methods and outcomes with finer categorizations increases and these future analyses may provide more precise and generalizable information for both practitioners and researchers.

5.0 Practical Applications

This meta-analysis represents the most comprehensive analysis of S&C intervention studies to date. Assessment of the collective research base highlights the need for longer duration interventions conducted with highly trained participants. However, across existing studies clear patters have emerged providing insights into factors causing systematic differences in results and enabling frameworks to better interpret the effectiveness of interventions. The large-scale analysis has demonstrated that gender, training status, intervention duration, training methods, outcome type, and the degree of specificity can all influence the distribution of effect sizes expected. Given large differences in effect size distributions for outcome type especially, it is recommended where standardised mean differences are used to interpret effectiveness, Cohen's guidelines no longer be used for S&C interventions. Instead, it is recommended that researchers and practitioners use the guidelines presented herein, with scope provided for values to be adjusted based on factors specific to the intervention being evaluated, many of which are described throughout the analysis.

Acknowledgements

No funding was received for this review.

Conflicts of interest

Paul Swinton, Katherine Burgess, Andy Hall, Leon Greig, John Psyllas, Rodrigo Aspe, Patrick Maughan and Andrew Murphy declare that they have no potential conflicts of interest with the content of this article.

Author Contributions

PAS, AM and KB designed the research. AM conducted the searches and screening. AM, LG, JP and AH extracted the data. PAS performed all statistical analyses. PAS, AM, PM and RA interpreted the data analysis. PAS, AM, PM and RA wrote the manuscript with critical input from KB, LG, JP and AH. All authors read and approved the final manuscript.

References

1. Weldon A, Duncan MJ, Turner A, Sampaio J, Noon M, Wong D, Lai VW. Contemporary practices of strength and conditioning coaches in professional soccer. Biology of Sport. 2020;38(3):377-90. https://doi.org/10.5114/biolsport.2021.99328.
2. Weldon A, Duncan MJ, Turner A, Christie CJ, Pang CM. Contemporary practices of strength and conditioning coaches in professional cricket. International Journal of Sports Science & Coaching. 2021;16(3):585-600. https://doi.org/10.1177/1747954120977472.
3. Suchomel TJ, Nimphius S, Stone MH. The importance of muscular strength in athletic performance. Sports Medicine. 2016;46(10):1419-49.
4. Swinton PA, Lloyd R, Keogh JW, Agouris I, Stewart AD. Regression models of sprint, vertical jump, and change of direction performance. Journal of Strength & Conditioning Research. 2014;28(7):1839-48. https://doi.org/10.1007/s40279-016-0486-0.
5. Kraemer WJ, Ratamess NA, Flanagan SD, Shurley JP, Todd JS, Todd TC. Understanding the science of resistance training: an evolutionary perspective. Sports Medicine. 2017;47(12):2415-35. https://doi.org/10.1007/s40279-017-0779-y.
6. Ramirez-Campillo R, Álvarez C, García-Hermoso A, Ramírez-Vélez R, Gentil P, Asadi A, Chaabene H, Moran J, Meylan C, García-de-Alcaraz A, Sanchez-Sanchez J. Methodological characteristics and future directions for plyometric jump training research: a scoping review. Sports Medicine. 2018;48(5):1059-81. https://doi.org/10.1007/s40279-018-0870-z.
7. Hecksteden A, Faude O, Meyer T, Donath L. How to construct, conduct and analyze an exercise training study?. Frontiers in Physiology. 2018;9:1007. https://doi.org/10.3389/fphys.2018.01007.
8. Nimphius S, McGuigan MR, Newton RU. Relationship between strength, power, speed, and change of direction performance of female softball players. Journal of Strength & Conditioning Research. 2010;24(4):885-95. https://doi.org/10.1519/JSC.0b013e3181d4d41d.
9. Blackburn JR, Morrissey MC. The relationship between open and closed kinetic chain strength of the lower limb and jumping performance. Journal of Orthopaedic & Sports Physical Therapy. 1998;27(6):430-5. https://doi.org/10.2519/jospt.1998.27.6.430.
10. Nuzzo JL, McBride JM, Cormie P, McCaulley GO. Relationship between countermovement jump performance and multijoint isometric and dynamic tests of strength. Journal of Strength & Conditioning Research. 2008;22(3):699-707. https://doi.org/10.1519/JSC.0b013e31816d5eda.
11. Stone MH, O'Bryant HS, McCoy L, Coglianese R, Lehmkuhl MA, Schilling B. Power and maximum strength relationships during performance of dynamic and static weighted jumps. Journal of Strength & Conditioning Research. 2003;17(1):140-7. 10.1519/1533-https://doi.org/4287(2003)017<0140:pamsrd>2.0.co;2.

12. Wisløff U, Castagna C, Helgerud J, Jones R, Hoff J. Strong correlation of maximal squat strength with sprint performance and vertical jump height in elite soccer players. British Journal of Sports Medicine. 2004;38(3):285-8. https://doi.org/10.1136/bjsm.2002.002071.

13. Brearley S, Bishop C. Transfer of training: How specific should we be? Strength & Conditioning Journal. 2019;41(3):97-109. https://doi.org/10.1519/SSC.0000000000000450.

14. Suarez DG, Wagle JP, Cunanan AJ, Sausaman RW, Stone MH. Dynamic correspondence of resistance training to sport: a brief review. Strength & Conditioning Journal. 2019;41(4):80-8. https://doi.org/10.1519/SSC.0000000000000458.

15. Issurin VB. Training transfer: scientific background and insights for practical application. Sports Medicine. 2013;43(8):675-94. https://doi.org/10.1007/s40279-013-0049-6.

16. Kraemer WJ, Ratamess NA. Fundamentals of resistance training: progression and exercise prescription. Medicine & Science in Sports & Exercise. 2004;36(4):674-88. https://doi.org/10.1249/01.mss.0000121945.36635.61.

17. Bosch F, Cook K. Strength training and coordination: an integrative approach. Rotterdam: 2010 Publishers; 2015.

18. Wilson GJ, Lyttle AD, Ostrowski KJ, Murphy AJ. Assessing dynamic performance: A comparison of rate of force development tests. J Strength Cond Res. 1995;9(3):176-81.

19. Cronin JB, Hansen KT. Strength and power predictors of sports speed. Journal of Strength & Conditioning Research. 2005;19(2):349-57. https://doi.org/10.1519/14323.1.

20. Peterson MD, Alvar BA, Rhea MR. The contribution of maximal force production to explosive movement among young collegiate athletes. Journal of Strength & Conditioning Research. 2006;20(4):867-73. https://doi.org/10.1519/R-18695.1.

21. West DJ, Owen NJ, Jones MR, Bracken RM, Cook CJ, Cunningham DJ, Shearer DA, Finn CV, Newton RU, Crewther BT, Kilduff LP. Relationships between force-time characteristics of the isometric midthigh pull and dynamic performance in professional rugby league players. Journal of Strength & Conditioning Research. 2011;25(11):3070-5. https://doi.org/10.1519/JSC.0b013e318212dcd5.

22. Brughelli M, Cronin J, Levin G, Chaouachi A. Understanding change of direction ability in sport. Sports Medicine. 2008;38(12):1045-63. https://doi.org/10.2165/00007256-200838120-00007.

23. Bates BT, Zhang SO, Dufek JS, Chen FC. The effects of sample size and variability on the correlation coefficient. Medicine and Science in Sports and Exercise. 1996;28(3):386-91. https://doi.org/ 10.1097/00005768-199603000-00015.

24. Harris NK, Cronin JB, Hopkins WG, Hansen KT. Squat jump training at maximal power loads vs. heavy loads: effect on sprint ability. Journal of Strength & Conditioning Research. 2008;22(6):1742-9. https://doi.org/10.1519/JSC.0b013e318187458a.

25. Lockie RG, Murphy AJ, Schultz AB, Knight TJ, de Jonge XA. The effects of different speed training protocols on sprint acceleration kinematics and muscle strength and power

in field sport athletes. Journal of Strength & Conditioning Research. 2012;26(6):1539-50. https://doi.org/10.1519/JSC.0b013e318234e8a0.

26. de Villarreal ES, Requena B, Izquierdo M, Gonzalez-Badillo JJ. Enhancing sprint and strength performance: combined versus maximal power, traditional heavy-resistance and plyometric training. Journal of Science and Medicine in Sport. 2013;16(2):146-50. https://doi.org/10.1016/j.jsams.2012.05.007.

27. Morin JB, Petrakos G, Jiménez-Reyes P, Brown SR, Samozino P, Cross MR. Very-heavy sled training for improving horizontal-force output in soccer players. International Journal of Sports Physiology and Performance. 2017;12(6):840-4. https://doi.org/10.1123/ijspp.2016-0444.

28. Rodríguez-Rosell D, Torres-Torrelo J, Franco-Márquez F, González-Suárez JM, González-Badillo JJ. Effects of light-load maximal lifting velocity weight training vs. combined weight training and plyometrics on sprint, vertical jump and strength performance in adult soccer players. Journal of Science & Medicine in Sport. 2017;20(7):695-9. https://doi.org/10.1016/j.jsams.2016.11.010.

29. Rhea MR, Alvar BA, Burkett LN, Ball SD. A meta-analysis to determine the dose response for strength development. Medicine & Science in Sports and Exercise. 2003;35(3):456-64. https://doi.org/10.1249/01.MSS.0000053727.63505.D4.

30. Rhea MR. Determining the magnitude of treatment effects in strength training research through the use of the effect size. Journal of Strength & Conditioning Research. 2004;18:918-20. https://doi.org/10.1519/14403.1.

31. Peterson MD, Rhea MR, Alvar BA. Applications of the dose-response for muscular strength development: Are view of meta-analytic efficacy and reliability for designing training prescription. Journal of Strength & Conditioning Research. 2005;19(4):950-8. https://doi.org/10.1519/R-16874.1.

32. Peterson MD, Rhea MR, Alvar BA. Maximizing strength development in athletes: a meta-analysis to determine the dose-response relationship. The Journal of Strength & Conditioning Research. 2004;18(2):377-82. https://doi.org/10.1519/R-12842.1.

33. Rhea MR, Alvar BA, Burkett LN. Single versus multiple sets for strength: a meta-analysis to address the controversy. Research quarterly for exercise and sport. 2002;73(4):485-8. https://doi.org/10.1519/R-12842.1.

34. Rhea MR, Alderman BL. A meta-analysis of periodized versus nonperiodized strength and power training programs. Research Quarterly for Exercise and Sport. 2004;75(4):413-22. https://doi.org/0.1080/02701367.2004.10609174.

35. Caldwell A, Vigotsky AD. 2020. A case against default effect sizes in sport and exercise science. PeerJ 8:e10314 https://doi.org/10.7717/peerj.10314.

36. Cohen, J. Statistical power analysis for the behavioral sciences. Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associate; 1988.

37. Brydges CR. Effect size guidelines, sample size calculations, and statistical power in gerontology. Innovation in Aging. 2019;3(4):igz036. https://doi.org/10.1093/geroni/igz036.

38. Paterson TA, Harms PD, Steel P, Credé M. An assessment of the magnitude of effect sizes: Evidence from 30 years of meta-analysis in management. Journal of Leadership and Organizational Studies. 2016;23(1):66-81. https://doi.org/10.1177/1548051815614321.

39. Gignac GE, Szodorai ET. Effect size guidelines for individual differences researchers. Personality and Individual Differences. 2016;102:74-8. https://doi.org/10.1016/j.paid.2016.06.069.

40. Lovakov A, Agadullina, ER. Empirically derived guidelines for effect size interpretation in social psychology. European Journal of Social Psychology. 2021;00:1-20. https://doi.org/10.1002/ejsp.2752.

41. Quintana DS. Statistical considerations for reporting and planning heart rate variability case-control studies. Psychophysiology. 2017;54(3):344-9. https://doi.org/10.1111/psyp.12798.

42. Rubio-Aparicio M, Marin-Martinez F, Sanchez-Meca J, Lopez-Lopez JA. A methodological review of meta-analyses of the effectiveness of clinical psychology treatments. Behavior Research Methods. 2018;50(5):2057-73. https://doi.org/10.3758/s13428-017-0973-8.

43. Kraft MA. Interpreting effect sizes of education interventions. Educational Researcher. 2020;49(4):241-53. ttps://doi.org/10.3102/0013189X20912798.

44. Schäfer T, Schwarz MA. The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. Frontiers in Psychology. 2019;10:813. https://doi.org/10.3389/fpsyg.2019.00813.

45. Kruschke JK, Liddell TM. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychonomic Bulletin and Review. 2018;25(1):178-206. https://doi.org/10.3758/s13423-016-1221-4.

46. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychological Methods. 2002;7(1):105-25. https://doi.org/10.1037/1082-989x.7.1.105.

47. Hedges LV. Estimation of effect size from a series of independent experiments. Psychological Bulletin. 1982;92(2):490-99. https://doi.org/10.1037/0033-2909.92.2.490.

48. Hox, JJ, Moerbeek M, Van de Schoot R. Multilevel Analysis. Techniques and applications. 3rd edition. 2018. Routledge.

49. Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis. 2006;1(3):515-34.

50. Bürkner PC. brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software. 2017;80(1):1-28. https://doi.org/ 10.18637/jss.v080.i01.

51. Verardi V, Vermandele C. Univariate and multivariate outlier identification for skewed or heavy-tailed distributions. The Stata Journal. 2018;18(3):517-32. https://doi.org/10.1177/1536867X1801800303.

52. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis: Taylor & Francis; 2014.

53. Stone MH, Collins D, Plisk S, Haff G, Stone ME. Training principles: Evaluation of modes and methods of resistance training. Strength & Conditioning Journal. 2000;22(3):65-76. https://doi.org/10.1080/14763140208522788.

54. Appleby B, Newton RU, Cormie P. Changes in strength over a 2-year period in professional rugby union players. Journal of Strength & Conditioning Research. 2012;26(9):2538-46. https://doi.org/10.1519/JSC.0b013e31823f8b86.

55. Ahtiainen JP, Pakarinen A, Alen M, Kraemer WJ, Häkkinen K. Muscle hypertrophy, hormonal adaptations and strength development during strength training in strength-trained and untrained men. European Journal of Applied Physiology. 2003;89(6):555-63. https://doi.org/10.1007/s00421-003-0833-3.

56. Mikecz R. Interviewing elites: Addressing methodological issues. Qualitative Inquiry. 2012;18(6):482-93. https://doi.org/10.1177/1077800412442818.

57. 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. 2021;9(2): https://doi.org/10.3390/sports9020032.

58. Ratamess NA, Beller NA, Gonzalez AM, Spatz GE, Hoffman JR, Ross RE, Faigenbaum AD, Kang J. The effects of multiple-joint isokinetic resistance training on maximal isokinetic and dynamic strength and local muscular endurance. Journal of Sports Science and Medicine. 2016;15(1):34-40.

59. Behm DG, Sale DG. Velocity specificity of resistance training. Sports Medicine. 1993;15(6):374-388. https://doi.org/10.2165/00007256-199315060-00003.

60. Gentil P, Del Vecchio FB, Paoli A, Schoenfeld BJ, Bottaro M. Isokinetic dynamometry and 1RM tests produce conflicting results for assessing alterations in muscle strength. Journal of Human Kinetics. 2017;56(1):19-27. https://doi.org/10.1515/hukin-2017-0019.

61. Sheppard JM, Young WB. Agility literature review: Classifications, training and testing. Journal of Sports Sciences. 2006;24(9):919-32. https://doi.org/10.1080/02640410500457109.

62. Nimphius S, Callaghan SJ, Bezodis NE, Lockie RG. Change of direction and agility tests: Challenging our current measures of performance. Strength & Conditioning Journal. 2018;40(1):26-38. https://doi.org/10.1519/SSC.0000000000000309.

63. Turner AN, Comfort P, McMahon J, Bishop C, Chavda S, Read P, Mundy P, Lake J. Developing powerful athletes Part 2: Practical applications. Strength & Conditioning Journal. 2021;43(1):23-31. https://doi.org/10.1519/SSC.0000000000000544.

64. Young WB. Transfer of strength and power training to sports performance. International Journal of Sports Physiology and Performance. 2006;1(2):74-83. https://doi.org/10.1123/ijspp.1.2.74.

65. Maffiuletti NA, Aagaard P, Blazevich AJ, Folland J, Tillin N, Duchateau J. Rate of force development: physiological and methodological considerations. European Journal of Applied Physiology. 2016;116(6):1091-116. https://doi.org/10.1007/s00421-016-3346-6.

66. Turner AN, Comfort P, McMahon J, Bishop C, Chavda S, Read P, Mundy P, Lake J. Developing powerful athletes, Part 1: mechanical underpinnings. Strength & Conditioning Journal. 2020;42(3):30-9. https://doi.org/10.1519/SSC.0000000000000543.

67. Winter EM, Abt G, Brookes FC, Challis JH, Fowler NE, Knudson DV, Knuttgen HG, Kraemer WJ, Lane AM, Van Mechelen W, Morton RH. Misuse of "power" and other mechanical terms in sport and exercise science research. Journal of Strength & Conditioning Research. 2016;30(1):292-300. https://doi.org/10.1519/JSC.0000000000001101.

68. Hernández-Davó JL, Sabido R. Rate of force development: reliability, improvements and influence on performance. A review. European Journal of Human Movement. 2014;33:46-69.

69. Haff GG, Ruben RP, Lider J, Twine C, Cormie P. A comparison of methods for determining the rate of force development during isometric midthigh clean pulls. The Journal of Strength & Conditioning Research. 2015;29(2):386-95. https://doi.org/10.1519/JSC.0000000000000705.

70. Morin JB, Slawinski J, Dorel S, Couturier A, Samozino P, Brughelli M, Rabita G. Acceleration capability in elite sprinters and ground impulse: push more, brake less? Journal of Biomechanics. 2015;48(12):3149-54. https://doi.org/10.1016/j.jbiomech.2015.07.009.

71. Rumpf MC, Lockie RG, Cronin JB, Jalilvand F. Effect of different sprint training methods on sprint performance over various distances: a brief review. Journal of Strength & Conditioning research. 2016;30(6):1767-85. https://doi.org/10.1519/JSC.0000000000001245.

72. Haugen T, McGhie D, Ettema G. Sprint running: From fundamental mechanics to practice—A review. European Journal of Applied Physiology. 2019;119(6):1273-87. https://doi.org/10.1007/s00421-019-04139-0.

73. Nicholson B, Dinsdale A, Jones B, Till K. The Training of Short Distance Sprint Performance in Football Code Athletes: A Systematic Review and Meta-Analysis. Sports Medicine. 2020;51:1179-207. https://doi.org/10.1007/s40279-020-01372-y.

74. Morin JB, Edouard P, Samozino P. Technical ability of force application as a determinant factor of sprint performance. Medicine and Science in Sports and Exercise. 2011;43(9):1680-8. https://doi.org/10.1249/MSS.0b013e318216ea37.

75. Morin JB, Bourdin M, Edouard P, Peyrot N, Samozino P, Lacour JR. Mechanical determinants of 100-m sprint running performance. European Journal of Applied Physiology. 2012;112(11):3921-30. https://doi.org/0.1249/MSS.0b013e318216ea37.

76. Colyer SL, Nagahara R, Takai Y, Salo AI. How sprinters accelerate beyond the velocity plateau of soccer players: Waveform analysis of ground reaction forces. Scandinavian

Journal of Medicine & Science in Sports. 2018;28(12):2527-35. https://doi.org/10.1111/sms.13302.

77. Mann R, Murphy A. The mechanics of sprinting and hurdling. Long Beach: Ralph V. Mann; 2018.

78. Cronin J, Ogden T, Lawton T, Brughelli M. Does increasing maximal strength improve sprint running performance? Strength & Conditioning Journal. 2007;29(3):86-95. https://doi.org/10.1519/1533-4295(2007)29[86:DIMSIS]2.0.CO;2.

79. Harris NK, Cronin JB, Hopkins WG, Hansen KT. Squat jump training at maximal power loads vs. heavy loads: effect on sprint ability. Journal of Strength & Conditioning Research. 2008;22(6):1742-9. https://doi.org/10.1519/JSC.0b013e318187458a.

80. Comfort P, Haigh A, Matthews MJ. Are changes in maximal squat strength during preseason training reflected in changes in sprint performance in rugby league players?. Journal of Strength & Conditioning Research. 2012;26(3):772-6. https://doi.org/10.1519/JSC.0b013e31822a5cbf.

81. Alcaraz PE, Palao JM, Elvira JL. Determining the optimal load for resisted sprint training with sled towing. Journal of Strength & Conditioning Research. 2009;23(2):480-5. https://doi.org/10.1519/JSC.0b013e318198f92c.

82. Petrakos G, Morin JB, Egan B. Resisted sled sprint training to improve sprint performance: a systematic review. Sports Medicine. 2016;46(3):381-400. https://doi.org/10.1007/s40279-015-0422-8.

83. Cross MR, Brughelli M, Samozino P, Brown SR, Morin JB. Optimal loading for maximizing power during sled-resisted sprinting. International Journal of Sports Physiology and Performance. 2017;12(8):1069-77. https://doi.org/10.1123/ijspp.2016-0362

84. Morin JB, Petrakos G, Jiménez-Reyes P, Brown SR, Samozino P, Cross MR. Very-heavy sled training for improving horizontal-force output in soccer players. International Journal of Sports Physiology and Performance. 2017;12(6):840-4. https://doi.org/10.1123/ijspp.2016-0444.

85. Alcaraz PE, Carlos-Vivas J, Oponjuru BO, Martinez-Rodriguez A. The effectiveness of resisted sled training (RST) for sprint performance: a systematic review and meta-analysis. Sports Medicine. 2018;48(9):2143-65. https://doi.org/10.1007/s40279-018-0947-8.

86. Baar K. Using molecular biology to maximize concurrent training. Sports Medicine. 2014;44(2):117-25. https://doi.org/10.1007/s40279-014-0252-0.

87. Fyfe JJ, Bishop DJ, Stepto NK. Interference between concurrent resistance and endurance exercise: molecular bases and the role of individual training variables. Sports medicine. 2014;44(6):743-62. https://doi.org/10.1007/s40279-014-0162-1.

88. Nader GA. Concurrent strength and endurance training: From molecules to man. Medicine and Science in Sports and Exercise. 2006;38(11):1965-1970. https://doi.org/10.1249/01.mss.0000233795.39282.33.

89. Fyfe JJ, Loenneke JP. Interpreting adaptation to concurrent compared with single-mode exercise training: some methodological considerations. Sports Medicine. 2018;48(2):289-97. https://doi.org/10.1007/s40279-017-0812-1.

90. Petré H, Hemmingsson E, Rosdahl H, Psilander N. Development of Maximal Dynamic Strength During Concurrent Resistance and Endurance Training in Untrained, Moderately Trained, and Trained Individuals: A Systematic Review and Meta-analysis. Sports Medicine. 2021;51:991-1010. https://doi.org/10.1007/s40279-021-01426-9.

91. Laird RH, Elmer DJ, Barberio MD, Salom LP, Lee KA, Pascoe DD. Evaluation of Performance Improvements After Either Resistance Training or Sprint Interval–Based Concurrent Training. Journal of Strength & Conditioning Research. 2016;30(11):3057-65. https://doi.org/10.1519/JSC.0000000000001412.

92. Robineau J, Lacome M, Piscione J, Bigard X, Babault N. Concurrent training in rugby sevens: effects of high-intensity interval exercises. International Journal of Sports Physiology and Performance. 2017;12(3):336-44. https://doi.org/10.1123/ijspp.2015-0370.

93. Hickson RC. Interference of strength development by simultaneously training for strength and endurance. European Journal of Applied Physiology and Occupational Physiology. 1980;45(2):255-63. https://doi.org/10.1007/BF00421333.

94. Kraemer WJ, Patton JF, Gordon SE, Harman EA, Deschenes MR, Reynolds KA, Newton RU, Triplett NT, Dziados JE. Compatibility of high-intensity strength and endurance training on hormonal and skeletal muscle adaptations. Journal of Applied Physiology. 1995;78(3):976-89. https://doi.org/10.1152/jappl.1995.78.3.976.

Table 1: Distribution (percentiles) of study characteristics, training method and outcome type.

Study Characteristic	**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Participants per study	4	14	17	19	21	24	28	32	37	52	177
Participants per group	4	7	8	9	10	11	12	14	16	21	94
Mean Age	14	14.6	16.9	18.5	20.0	20.6	21.6	22.5	23.5	24.9	60
Publication Year	1962	1996	2005	2008	2014	2012	2014	2015	2016	2017	2018
Intervention duration (weeks)	4	5	6	6	8	8	8	10	12	14	520

Training method	**Number of studies (%)**	**Number of effects (%)**	**Outcome type**	**Number of studies (%)**	**Number of effects (%)**
Resistance	241 (35.5)	2764 (31.0)	Maximum strength	421 (62.0)	2608 (29.3)
Combined	227 (33.4)	2479 (27.8)	Jump	381 (56.1)	1564 (17.6)
Plyometric	127 (18.7)	1227 (13.8)	Explosiveness	260 (38.3)	1913 (21.5)
Other	84 (12.4)	615 (6.9)	Sprint	256 (37.7)	1260 (14.2)
Sprint	74 (10.9)	605 (6.8)	Power	204 (30.0)	1205 (13.5)
Ballistic	34 (5.0)	441 (5.0)	Agility	116 (17.1)	354 (4.0)
Isokinetic	32 (4.7)	366 (4.1)			
Concurrent	29 (4.3)	277 (3.1)			
Agility	14 (2.1)	130 (1.5)			

Table 2: Meta-analysis results and effect size thresholds categorised by training method.

Training method	Mean [95%CrI]	Small: 0.25q [95%CrI]	Medium: 0.5q [95%CrI]	Large: 0.75q [95%CrI]	VPC observed effect [75%CrI]	VPC measure Occasion [75%CrI]	VPC outcome level [75%CrI]	VPC study level [75%CrI]
Resistance	0.57 [0.51 to 0.62]	0.14 [0.12 to 0.17]	0.48 [0.46 to 0.51]	0.87 [0.85 to 0.90]	0.26 [0.24 to 0.27]	0.08 [0.07 to 0.10]	0.26 [0.24 to 0.29]	0.40 [0.36 to 0.43]
Combined	0.51 [0.46 to 0.55]	0.13 [0.11 to 0.15]	0.41 [0.39 to 0.43]	0.74 [0.71 to 0.77]	0.33 [0.31 to 0.36]	0.00 [0.00 to 0.01]	0.24 [0.22 to 0.27]	0.42 [0.38 to 0.46]
Plyometric	0.56 [0.50 to 0.63]	0.14 [0.11 to 0.18]	0.47 [0.44 to 0.50]	0.82 [0.78 to 0.86]	0.29 [0.27 to 0.32]	0.01 [0.00 to 0.12]	0.25 [0.21 to 0.30]	0.36 [0.30 to 0.41]
Other	0.43 [0.35 to 0.50]	0.10 [0.05 to 0.14]	0.39 [0.35 to 0.43]	0.72 [0.67 to 0.77]	0.30 [0.27 to 0.33]	0.0 [0.00 to 0.01]	0.32 [0.28 to 0.37]	0.37 [0.31 to 0.43]
Ballistic	0.43 [0.34 to 0.53]	0.10 [0.05 to 0.15]	0.36 [0.31 to 0.41]	0.65 [0.59 to 0.71]	0.45 [0.39 to 0.51]	0.09 [0.03 to 0.16]	0.12 [0.05 to 0.19]	0.33 [0.25 to 0.43]
Sprint	0.29 [0.20 to 0.37]	-0.04 [-0.09 to 0.01]	0.21 [0.17 to 0.26]	0.48 [0.43 to 0.54]	0.53 [0.46 to 0.60]	0.10 [0.03 to 0.16]	0.03 [0.00 to 0.09]	0.32 [0.25 to 0.41]
Isokinetic	0.63 [0.48 to 0.79]	0.25 [0.18 to 0.32]	0.60 [0.53 to 0.67]	1.0 [0.92 to 1.1]	0.32 [0.26 to 0.38]	0.06 [0.02 to 0.12]	0.06 [0.02 to 0.11]	0.55 [0.46 to 0.64]
Concurrent	0.52 [0.37 to 0.69]	0.11 [0.03 to 0.18]	0.47 [0.40 to 0.54]	0.87 [0.78 to 0.97]	0.24 [0.19 to 0.29]	0.02 [0.00 to 0.06]	0.43 [0.34 to 0.52]	0.31 [0.21 to 0.42]
Agility	0.43 [0.26 to 0.61]	0.16 [0.08 to 0.25]	0.40 [0.32 to 0.48]	0.67 [0.57 to 0.79]	0.39 [0.28 to 0.50]	0.02 [0.00 to 0.08]	0.11 [0.04 to 0.20]	0.46 [0.30 to 0.62]

CrI: Bayesian credible intervals. q: quantile. VPC: variance partition coefficients

Figure 1: Percentage distribution of outcome types (legend) analysed for each training method (x-axis)

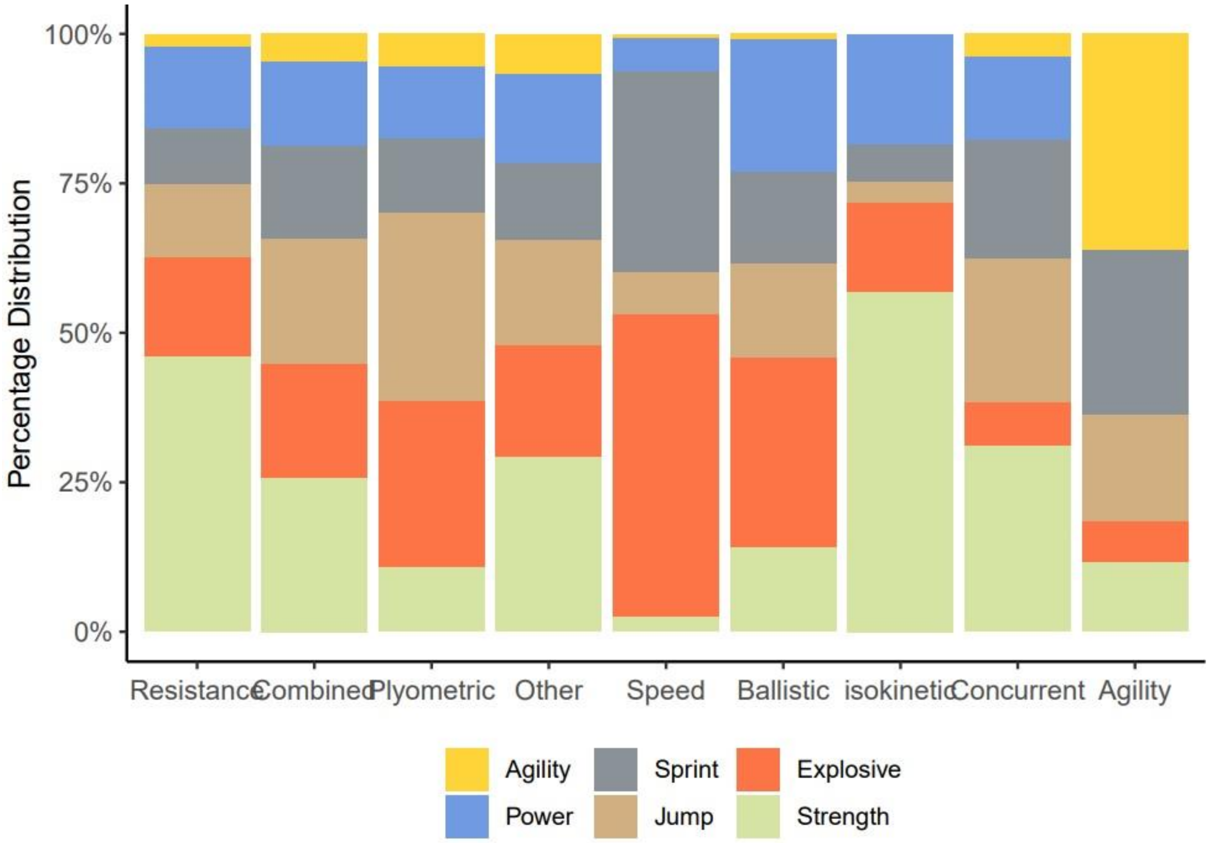
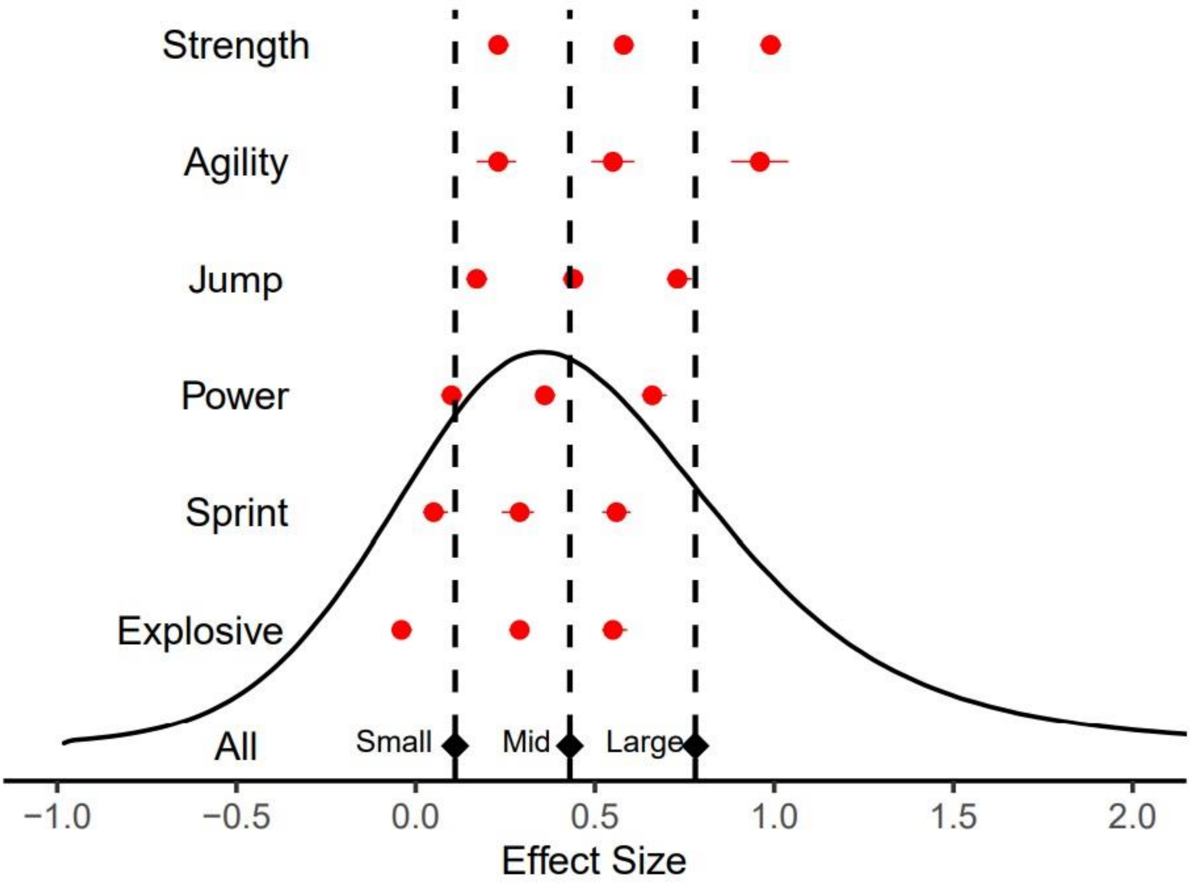
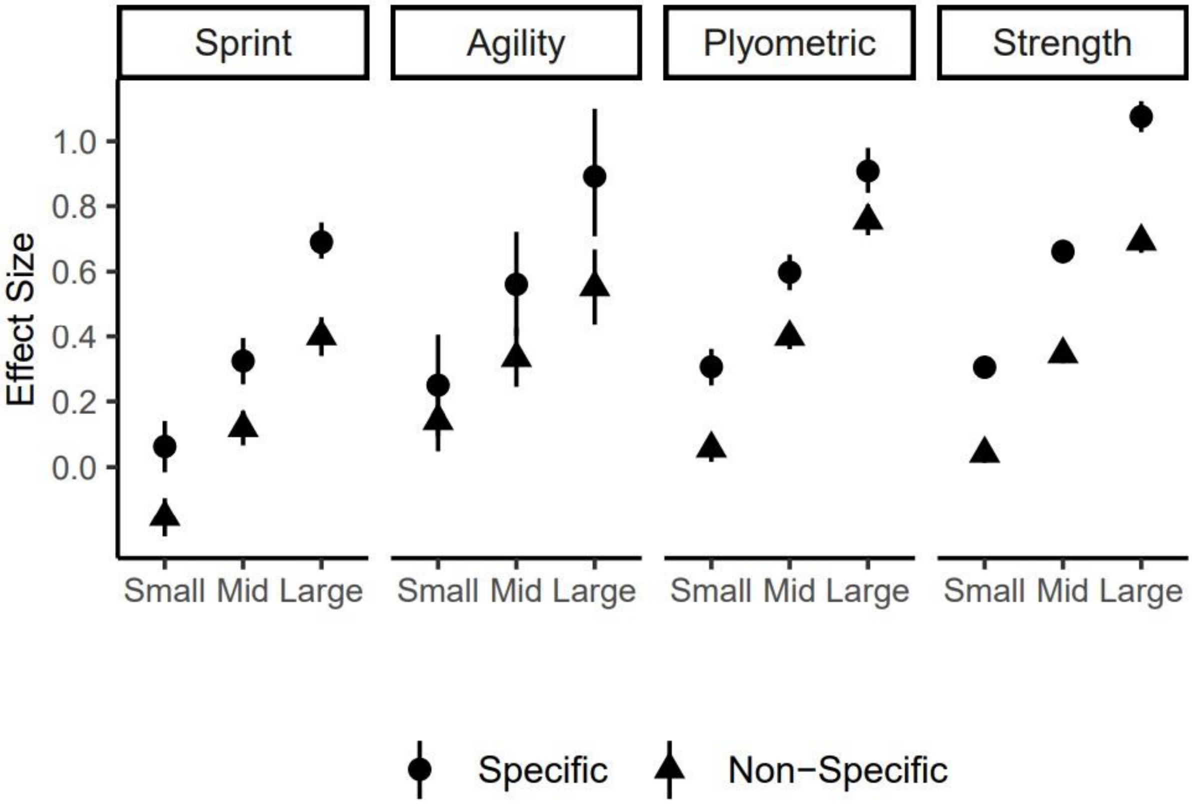


Figure 2: Empirical distribution and modelled outcome-specific effect size thresholds



Black curve is a density plot of the directly calculated empirical effect size values across all outcomes. Small, mid, and large thresholds represent the 0.25, 0.5 and 0.75-quantiles of predicted draws. Black diamonds are based on direct calculation and red intervals illustrate uncertainty in estimates through the median value (circle) and 95% credible interval.

Figure 3: Comparison of effect sizes across specific and non-specific combinations of training method and outcome types.



Small, mid and large thresholds represent the 0.25, 0.5 and 0.75-quantiles of predicted draws. Variation in values is illustrated by the median value of the quantile (circle or triangle) and 95% credible interval.

Figure 4: Summary of recommended threshold values to interpret effectiveness of S&C interventions focusing on outcome types.

Outcome Type:	Strength			Athletic			Sprint / Explosiveness		
	(e.g. 1RM, 3RM, Isokinetic, Isometric)			(e.g. agility, jump, power)			(e.g. 10 m, 20 m 30 m sprint time, RFD)		
Small Threshold	**[0.10**	**0.25**	**0.40]**	**[0.00**	**0.15**	**0.30]**	**[0.0**	**0.05**	**0.15]**
Medium Threshold	**[0.45**	**0.60**	**0.75]**	**[0.30**	**0.45**	**0.60]**	**[0.15**	**0.25**	**0.35]**
Large Threshold	**[0.85**	**1.0**	**1.15]**	**[0.60**	**0.75**	**0.90]**	**[0.40**	**0.55**	**0.70]**

Lower values in threshold range

Higher values in threshold range

Highly trained
Males
Low specificity
Short duration

Untrained
Females
High specificity
Long duration

An a priori value should be selected from each small/medium/large interval based on the specifics of the context (e.g. intervention duration, training status, gender and degree of specificity between training method and outcome).