A meta-analysis of resistance training in female youth: its effect on muscular strength, and shortcomings in the literature

Running head: A meta-analysis of resistance training in female youth

Jason Moran¹

Gavin Sandercock²

Rodrigo Ramirez-Campillo³

Cain C.T. Clark¹

John F.T. Fernandes¹

Benjamin Drury¹

1. Department of Sport, University Centre Hartpury (University of the West of England), Gloucestershire, United Kingdom

2. Centre for Sports and Exercise Science, School of Biological Sciences, University of Essex, Colchester, United Kingdom,

3. Department of Physical Activity Sciences, Universidad de Los Lagos, Osorno, Chile

Corresponding author contact details: jason.moran@hartpury.ac.uk, +44 1452 702482
ABSTRACT

Background: Resistance training is an effective way to enhance strength in female youth but, to date, no researcher has meta-analysed its effect on muscular strength in that population.

Objectives: This meta-analysis characterised female youths’ adaptability to resistance training (RT). A second objective was to highlight the limitations of the body of literature with a view to informing future research.

Data sources: Google Scholar, PubMed, Web of Science

Study eligibility criteria: Resistance training interventions in healthy females with a mean age between 8 and 18 years. Programmes of between 4 and 16 weeks duration that included a control group.

Study appraisal and synthesis methods: The inverse-variance random effects model for meta-analyses was used because it allocates a proportionate weight to trials based on the size of their individual standard errors and facilitates analysis whilst accounting for heterogeneity across studies. Effect sizes, calculated from a measure of muscular strength, are represented by the standardised mean difference and are presented alongside 95% confidence intervals.

Results: The magnitude of the main effect was ‘small’ (0.54, 95% confidence interval: 0.23, 0.85). Effect sizes were larger in older (> 15 yrs; ES = 0.72 [0.23, 1.21] vs. 0.38 [-0.02, 0.79]), taller (>163cm; ES = 0.67 [0.20, 1.13] vs. 0.55 [0.08, 1.02]) and heavier (<54kg; ES = 0.67 [0.30, 1.03] vs. 0.53 [-0.00, 1.06]) participants.

Conclusions and implications of key findings: Resistance training is effective in female youth. These findings can be used to inform the prescription of RT in female youth.

Key points:
• Resistance training is an effective way of increasing muscular strength in female youth.
• Older, taller or heavier female youths may be more responsive to training potentially owing to maturation-related increases in muscle mass.
• Programmes lasting 8 weeks, with 2 sessions per week and around 40 minutes per session seem most effective.
1. Introduction

Maximal strength is the maximum force or torque that can be exerted by skeletal muscles during movement [1]. The ability to exert high force is an important determinant of healthy function and athletic performance in youth [2] and resistance training (RT) is an effective way to enhance that quality [3] in this population. Indeed, previous research in youth has found that absolute strength is well correlated with certain measures of sprint and jump height ($r = 0.596$ to $0.762$) [4,5].

Intervention studies of RT in female youth are lacking in both quantity and quality. Despite this, several meta-analytical reviews have investigated the efficacy of RT in youth populations [6–9]. However, to date, no researchers have conducted meta-analyses in female youth only, with existing data often conflated with that for young males [3,10] in subgroup analyses only, thus undermining the accuracy of inferences that can be made. This is an important matter in exercise science as there is variation in how male and female youth adapt to the demands of RT [11]. For example, adaptions in strength and body composition differ based on sex [11]. Primarily, this is because of differences in circulating anabolic hormones which are higher in males than in females from the age of puberty and which denote maturation-related changes such as increases in muscle mass [11,12]. This could result in an amplification of effects in male youth: males gain around 7.2kg of muscle mass annually during the growth spurt whilst females gain just 3.5kg per year [13]. On average, the female pubertal growth spurt takes place 2 years earlier than that for males but whilst gains in muscle mass slow down in females from the age of 15 years, males can continue to gain until the age of 20 years [13]. Accordingly, as maturation-status can regulate adaptive responses to RT [14], developmental factors must be considered in the prescription of this type of exercise.

The intertwining physiological processes that underpin increases in muscular strength have not been adequately detailed in line with the training methods that induce them in female youths. This primarily relates to factors such as sex and the specificity of training stimuli and though some authors have attempted to implicitly address such issues, reviews remain
somewhat flawed. For example, Lesinski et al. [3] included studies in female youth athletes in their meta-analysis on RT but failed to separate them from male athletes for most effect estimates for chronological age, stage of maturation and training type. Indeed, dose responses were calculated independent of age and sex despite previous recommendations for age- and sex-specific approaches to RT prescription [15]. Behringer et al. [6] had previously reported that RT was effective for increasing strength in youth with an effect size of 1.1 (0.9-1.3). However, the main effect statistic was inclusive of both males and females. A later review by the same group [7] revealed beneficial effects of RT on the motor skills of running, jumping and throwing; and more recently, Harries, Lubans and Callister [9] meta-analysed the effects of RT on vertical jump in male and female youth athletes, reporting increases in vertical jump height in males during puberty, but not in females. Despite these encouraging results, the enhancement of motor skills may not necessarily be the primary effect associated with RT based on the principle of specificity [16] as such movements are heavily influenced by coordination [17] and body mass [18] whilst strength is defined by force production. In relation to this, little attention has been paid to the effects of RT on muscular strength only in female youth. Our aim was to address this gap in the literature by characterising the effects of RT on muscular strength in female youth, thus paying heed to the specificity of the response to training stimuli. A secondary aim was to inform future research and practice by describing shortcomings in the literature as it relates to RT in female youth.

2. Methods

This meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [19].

2.1 Literature search

With no date restrictions, a systematic search of the Google Scholar database was initially undertaken. Searches of the PubMed and Web of Science databases were also undertaken. Only articles published in the English language were considered. These searches were
performed over three consecutive days (8th to 10th) in November, 2017. Using Boolean logic, various combinations of the following search terms were used: ‘youth’, ‘training’, ‘female’, ‘strength’, ‘resistance’, ‘weightlifting’, ‘volume’, ‘intensity’, ‘fitness’, ‘high’, ‘load’, ‘rest’, ‘sets’, ‘repetitions’, ‘plyometric’, ‘stretch-shortening cycle’, ‘jump’, ‘power’, ‘speed’, ‘velocity’, ‘agility’, ‘sprint’, ‘sprinting’, ‘alactic’, ‘acceleration’, ‘running’, ‘exercise’, ‘change of direction’, ‘paediatric’, ‘pediatric’, ‘young’, ‘children’, ‘adolescence’, ‘athletes’, ‘sport’. These combinations were searched using the following example format: ‘youth’ AND ‘training’ AND ‘female’ AND ‘strength’ OR ‘[additional search term 1]’ OR ‘[additional search term 2]’. In selecting studies for inclusion, a review of all relevant article titles was conducted before an examination of article abstracts and, then, full published articles. Only peer-reviewed articles were included in the meta-analysis. The reference lists of those studies that remained following the application of all inclusion criteria were hand-searched for further articles that could be relevant to the meta-analysis. Also, similar reviews were hand-searched for additional articles. The search process is outlined if Figure 1.

**Figure 1 Flow chart for inclusion and exclusion of studies**

### 2.2 Data extraction

The extraction of data from gathered articles was undertaken by three reviewers (JM, CC and JF) with a standardised form created in Microsoft Excel. The first reviewer collected the data before the second and third reviewers investigated its accuracy and the eligibility of studies for inclusion. Where required data were not clearly or completely reported, article authors were contacted for clarification.

The following criteria determined the eligibility of studies for inclusion in the review: cohorts of healthy females, with a mean age between 8 and 18 years. Interventions of between 4 and 16 weeks duration that included a control group. Based on a previous review [14], we defined RT as “[requiring] the musculature to contract (sic.) against an opposing force generated by some type of resistance” [6]. Effect sizes were calculated by selecting the most relevant measure of
muscular strength “based on theory or a logically defensible rationale” [20]. To account for the specificity of the training adaptation, we did not consider surrogate measures of performance such as a vertical jump [14] and tests of strength must have incorporated some form of external resistive load, and not bodyweight only. Means and standard deviations for a measure of post-intervention strength were used to calculate an effect size. The characteristics of the study participants are displayed in Table 1.

Table 1 Characteristics of study participants

2.3 Analysis and interpretation of results

Meta-analytical comparisons were carried out in RevMan version 5.3 [31]. The inverse-variance random effects model for meta-analyses was used because it allocates a proportionate weight to trials based on the size of their individual standard errors [32] and facilitates analysis whilst accounting for heterogeneity across studies [33]. Effect sizes are represented by the standardised mean difference and are presented alongside 95% confidence intervals. The calculated effect sizes were interpreted using the conventions outlined for standardised mean difference by Hopkins et al [34] (<0.2 = trivial; 0.2-0.6 = small, 0.6-1.2 = moderate, 1.2-2.0 = large, 2.0-4.0 = very large, >4.0 = extremely large). In cases in which there was more than one intervention group in a given study, the control group was proportionately divided to facilitate comparison across all participants [35].

To gauge the degree of heterogeneity amongst the included studies, the $I^2$ statistic was referred to. This represents the proportion of effects that are due to heterogeneity as opposed to chance [19]. Low, moderate and high levels of heterogeneity correspond to $I^2$ values of 25%, 50% and 75%, however, these thresholds are considered tentative [36]. The $X^2$ (chi square) statistic determines if the differences in the results of the analysis are due to chance and in such a case, a low $P$ value, or high $X^2$ statistic, relative to degrees of freedom would be apparent [32].
A risk of bias quality scale was not utilised for a number of reasons: The Cochrane Collaboration has previously discouraged the use of these scales, stating that the practice is not underpinned by empirical evidence and assessment criteria may apply inaccurate study weights [37]. Also, the subjectivity of personal opinion undermines the accuracy of such scales [37]. Blinding of study participants and trainers is undermined owing to the constraints that make such a practice difficult to implement in training intervention studies [38]. In relation to this issue, previous systematic reviews [39,40] of training amongst children and adolescents suggest that studies tend to be of low to medium quality.

2.4 Analysis of moderator variables

To assess the potential effects of moderator variables, subgroup analyses were performed using moderating variables identified a priori. Using a random effects model we selected potential moderators likely to influence the effects of training on applied interventions. An age-based division was made between study groups whose participants were younger or older than 15 years. This division was made on the basis that this is the age at which females achieve adult height [41]. We also compared participants who were greater or less than 163 cm in stature. This represents an approximation of average female stature in the fully mature state [42]. Resistance training has previously been found to be less effective in pre-pubertal males in comparison to their pubertal and post-pubertal counterparts [14] and this was hypothesised to also be the case in females. With the high variability of body mass across various populations, body mass was divided into subgroups with a median split (greater or less than 56 kg) for the entire sample of participants. Age, height and body mass were all chosen as moderator variables because of the effect of age and biological maturation exert on performance in female youth [43]. The moderator variables of programme duration (weeks), training frequency (sessions per week), total number of training sessions and session duration (mins) were chosen based on the accepted influence of the FITT principle on adaptations to exercise [44]. These variables were divided using a median split.

3. Results
3.1 Main effect

Eleven studies were included in this meta-analysis and they comprised 15 individual experimental groups. Across all included studies, there was a small, significant improvement in strength (ES = 0.54 [0.23, 0.85], Z = 3.37 [p = 0.0008]). The overall estimate was of small magnitude and showed a significant level of between-study heterogeneity (I² = 42% [p = 0.04]). These results are displayed in Figure 2.

Figure 2 Forest plot of effect sizes with 95% confidence intervals

3.2 Effect of moderator variables

A summary of the effect of moderator variables can be viewed in Table 2. Subgroup analysis suggested highly variable levels of between-group heterogeneity. Older, taller and heavier study participants adapted to RT to a greater degree than their younger, shorter, lighter counterparts respectively with ‘moderate’ effects being achieved in the groups with larger adaptations. Interventions with a total amount of 16 sessions or fewer produced the largest effect (ES = 0.75 [0.33, 1.17], Z = 3.52 [P = 0.0004]). Those RT programmes that lasted 8 weeks or less demonstrated larger effects (ES = 0.62 [0.17, 1.07], Z = 2.71 [P = 0.007]) than those that lasted more than eight weeks (ES = 0.44 [-0.02, 0.90], Z = 1.86 [P = 0.06]). Similarly, interventions with two sessions per week were more effective (ES = 0.72 [0.34, 1.09], Z = 3.71 [P = 0.0002]) than those with more than 2 per week (ES = 0.18 [-0.26, 0.61], Z = 0.80 [P = 0.42]). Levels of heterogeneity were higher in subgroups with shorter programmes, lower training frequency, fewer training sessions and fewer minutes per session.

Table 2 Effect of moderator variables with 95% confidence intervals

4. Discussion

The purpose of this meta-analytical review was to quantify the effects of RT in female youth and to highlight limitations in the relevant body of literature with a view to improving approaches to research in the future. Based on extensive research findings in male youth
[14,45–49], the effects of RT in that population are well understood. The emergence of anabolic hormones in puberty [50] may enhance adaptations to RT due to the resultant increase in muscle mass [50]. However, because the magnitude of this adaptation is smaller in females [13], RT may be less effective in that population. This is evidenced by the results of this meta-analysis when compared to a methodologically similar meta-analysis of RT in male youth. Moran et al. [14] previously found a moderate effect (0.98 [0.70-1.27]) of RT on muscular strength in male youth aged 10 to 18 years. That contrasts with the current analysis which demonstrated effects of a far lower magnitude (0.54 [0.23, 0.85]). Given that it seems females adapt to RT to a different degree, this review is timely due to the dearth of empirical evidence in that population.

The body of research relating to RT in female youth is substantially smaller than that in male youth. A previous review of RT in male youth [14], which adopted similar study selection criteria, included 19 studies and excluded several more on the basis that it was a meta-analysis of within-group intervention effects and, thus, included no control trials or studies which included non-athletic youth. On the contrary, the current review found just 11 studies that met similar inclusion criteria, which were widened to include non-athletic youth. Given that RT can reduce injuries in females [15] and, also, that greater levels of strength are thought to prevent injury [51,52], it is curious that interventions which directly address the impact of RT on a measure of muscular strength are so scarce in female youth, a requirement for inclusion of interventions in this meta-analysis.

Based on the results of our analyses, the suggestion that RT can enhance strength in female youth is not in question but the next challenge from a research perspective is to characterise how maturation affects adaptations as the trainee develops physically. Previous meta-analyses [14,53,54] have reported variability in the degree of adaptation to various forms of training across the maturation continuum in male youth. However, the body of relevant literature is not currently large or informative enough to develop a stance on this issue in female youth. One of the primary causes of this is the relatively large amount of researchers
who have incorporated RT into their studies but have not provided measures of absolute muscular strength with which to quantify the effects of their prescribed programmes (see Figure 1). This may, justifiably, be due to study design limitations but the approach could be argued to undermine the principle of specificity of adaptation to the imposed demands of training [58]. This is not necessarily a trivial issue: the making of inferences about effects on an arguably less-relevant outcome measure, such as vertical jump, due to a training method such as traditional RT, could be considered suboptimal due to the independent nature of training adaptations to different forms of training stimuli [16], and the principle of training specificity [58]. Our rationale for this stance is founded on the basis that, despite being well correlated with absolute strength in well-trained youth [4], measures of relative strength, such as vertical jump variants, may not necessarily capture maturation-related changes in strength [53]. This is because as youths grow, relative strength can decrease as bodyweight increases, potentially resulting in reduced relative strength, despite enhanced absolute strength [53]. This could be of particular concern in heavier youth. Though we appreciate that researchers’ may have limited control over the type of test used, if possible, it is nonetheless important to consider this issue when formulating RT interventions and programmes so that the effects of training are measured with the most relevant performance tests. To measure performance following RT, researchers should consider utilising resistive apparatus which facilitates estimation of maximal strength to previously presented guidelines [59], in a variety of different exercises [60]. Such a protocol could be suitable for untrained youth as it involves the use of multiple submaximal repetitions to predict maximal strength [59]. The submaximal load used can be increased in manageable increments of between 2.5% and 5% until the participant exerts a maximal effort and whilst the study in question [59] used a submaximal resistance of 90%, lower loads could be justifiably prescribed.

In none of the studies included in this meta-analysis did authors report participants’ maturation status with one of the most commonly used methods in youth sport [61]. This is a curious feature of the relevant body of literature given that altered motor coordination during puberty
may influence females’ susceptibility to damaging the anterior cruciate ligament, a risk that is deepened by deficits in strength which can predispose the knee to acute and chronic injury [62]. It is therefore logical to suggest that even an approximate estimate of the timing of peak height velocity could enable researchers to test training interventions that are specific to participants’ level of biological maturation, thus targeting the aforementioned weaknesses. Given that in the majority of studies, researchers report participants’ stature and body mass, the addition of seated stature to a typical testing battery is neither work-intensive or time-consuming. As recommended by the British Association of Sports and Exercise Sciences [63], researchers must report the biological maturity status in youth participants of both sexes.

Related to the above points, a further limitation of the current body of literature is the relatively high number of researchers who did not incorporate a control group into their study design. Several studies [64–67] were excluded on the basis that they did not provide any control group data, fulfilling other inclusion criteria. This study design feature takes on added significance in interventions in youth given that rapid changes in maturation status can result in both increases or decreases in physical capabilities [68,69]. The recruitment of individuals to studies can be difficult and the addition of a control group is not always possible. Nevertheless, with a view to progressing this area of research, study authors are encouraged to prioritise the inclusion of control groups during the conceptualisation stage of studies.

In reference to maturation-related factors, RT was more effective (moderate vs, small effects) in older, taller and heavier individuals, indicating that more biologically mature female youth may adapt to RT to a greater magnitude than their less mature counterparts. There also exists the possibility that potentially negative effects of puberty, such as impaired sensorimotor function [70], could have temporarily disrupted the progress of the less mature subgroups. This is in line with the suggestion that the existence of a maturational threshold could moderate responses to RT in youth [63]. Moreover, it is also reflective of a similar trend in male youth with the periods during and after peak height velocity seemingly an opportune time to expose well-conditioned individuals to more advanced and higher volume RT [14]. This may be partly
due to maturation-related increases in muscle mass which, in turn, can underpin gains in
strength [15]. Rising testosterone levels enhance the synthesis of muscle proteins which leads
to the greater accumulation of muscle mass [71]. Though this process is less marked in female
youth, who gain only half the amount of muscle mass that males do annually during the growth
spurt [13], it still may exert a substantial effect. In line with the age threshold for subgroup
analysis in this review, Poortmans et al. [72] found that 15 year old females possessed
substantially more muscle mass than 10 year olds (18.4 ± 0.8 kg vs. 12.5 ± 0.4 kg) indicating
a near-mature state for the older youths [73]. In females, the development of muscle mass
during puberty seems to be related to a four-fold increase in testosterone which, though lower
in magnitude than that experienced by males, still results in them achieving between 50% and
70% of the muscle mass of males [74]. Moreover, during the pubertal period, circulating
oestrogen seems to promote fat storage, also resulting in lower muscle mass [71,75]. These
factors together could go some way to explaining the size of the main effect in the current
review which was almost half of that observed in a similar meta-analysis in male youth [14].
Indeed, whilst RT seems marginally more effective in older, taller and heavier individuals,
unpublished data from our group seems to point to younger, shorter and lighter individuals
being more responsive to plyometric training. This seems logical given that relative strength
is a key factor in plyometric training whilst absolute strength is more important in RT. Taken
together, this suggests that as in male youth, females may display variable responses to
different types of training at different stages of maturation.

Contrary to what we had expected to observe from subgroup analyses, it seems that RT
interventions were more effective in studies with fewer training sessions (16 or less), shorter
study durations (8 weeks or less) and lower training frequencies (2 sessions or less per week).
The reasons for this finding are not clear but could be due to the relatively low number of
studies in this field, thus necessitating more research to clarify the time course of adaptation
to RT in female youth. We surmise that the higher effects in shorter programmes could be
indicative of a need to alter the training stimulus at the 8 week point in a training cycle: as
evidenced by our results, adaptations to resistance training can take place in a relatively short period of time necessitating an eventual change in the demands placed on body in order to sustain progress [76]. This could also be indicative of females’ lower sensitivity to the effects of RT when compared to males and could serve as a reason for coaches to plan more varied programmes of athletic development to continually drive adaptation in female youth. Despite this recommendation, the characteristics of the body of published literature could have been more influential in this finding: the lack of variation in programming characteristics such as session frequency made it somewhat difficult to divide studies with a median split for subgroup analysis. In relation to this, the dichotomisation of continuous data by median split can result in residual confounding of analyses and reduced statistical power [77].

A widespread flaw in the literature relating to RT in female youth is the pooling of performance data of both females and males for analysis within the same studies [8,78–84]. In isolation, this is never an acceptable practice in research as it only determines whether a training method is effective independent of any population-specific effects. Such an approach does not consider the effects of sex and maturation level on training status given that boys and girls are biologically different and experience wholly different maturational changes at different times and tempos [85]. Given these differences, chronological age offers little basis for comparison between the sexes. Indeed, that researchers have often failed to provide separate within-study performance data for male and female youth renders many interventions’ findings somewhat flawed and, arguably, reduces their usefulness to practitioners and sports scientists alike. To continue to drive progress in this area of research, study authors must report anthropometric and performance data of males and females separately, in addition to an analysis of an overall primary effect. For an exemplary demonstration of how to present quantitative and graphical data for boys and girls within a single intervention study, we refer the reader to the work Muehlbauer, Gollhoffer and Granacher [25].

5. Conclusion
Resistance training seems to be an effective way of increasing muscular strength in female youth. However, significant limitations in the current literature prevent assured RT prescription recommendations being made. Based on our results, it seems that older, taller or heavier individuals may be more responsive to training potentially owing to maturation-related increases in muscle mass which are indicative of a maturational threshold that could moderate responses to RT [63]. Following foundational training, females of any age can be exposed to RT though because adaptations may reach an upper limit relatively quicker than in boys, workloads should be sensibly balanced. This is particularly important during the interval of maximal growth when reduced motor control can result in injury. We recommend that two sessions per week is an adequate frequency of training in female youth who should be exposed to a varied training stimuli to prevent stagnation in the longer term. In relation to obtaining more relevant data for future analyses, researchers are urged to continue to investigate the effects of RT in female youth but to devote more effort into measuring maturation status and relevant measures of performance. If researchers include both males and females in the same study, the resultant data should be presented in a way that separates one group from the other so that sex-specific inferences can be made. Data should also be presented in its raw numerical form with graphical representations being used only to add context to the reported results.

Acknowledgements, conflicts of interest and funding statements

Jason Moran, Gavin Sandercock, Rodrigo Ramirez-Campillo, Cain C.T. Clark, John F.T. Fernandes and Benjamin Drury declare that they have no conflict of interest.

No financial support was received for the conduct of this study or preparation of this manuscript.

6. Reference list


11. Yoshimoto T, Takai Y, Fukunaga Y, Fujita E, Yamamoto M, Kanehisa H. Effects of


41. Georgopoulos NA, Markou KB, Theodoropoulou A, Vagenakis GA, Benardot D, Leglise


51. Faigenbaum AD, Myer GD. Resistance training among young athletes: safety, efficacy 

52. Askling C, Karlsson J, Thorstensson A. Hamstring injury occurrence in elite soccer 

53. Moran J, Sandercock G, Rumpf MC, Parry DA. Variation in Responses to Sprint Training 

54. Moran J, Sandercock GRH, Ramirez-Campillo R, Meylan C, Collison J, Parry DA. Age-
related variation in male youth athletes’ countermovement jump following plyometric training. 

55. Siegler J, Gaskill S, Ruby B. Changes evaluated in soccer-specific power endurance 
either with or without a 10-week, in-season, intermittent, high-intensity training protocol. J. 

56. Sannicandro I, Cofano G, Rosa RA, Piccinno A. Balance training exercises decrease 

57. Pairot De Fontenay B, Lebon F, Champely S, Argaud S, Monteil1 K. ACL Injury Risk 
Factors Decrease & Jumping Performance Improvement in Female Basketball Players: A 


59. Verdijk LB, Van Loon L, Meijer K, Savelberg HHCM. One-repetition maximum strength 
test represents a valid means to assess leg strength in vivo in humans. J. Sports Sci. 


79. Benson AC, Torode ME, Fiatarone Singh MA. The effect of high-intensity progressive


<table>
<thead>
<tr>
<th>Study</th>
<th>Study group</th>
<th>Sport</th>
<th>Age</th>
<th>Height (cm)</th>
<th>Weight</th>
<th>Participants</th>
<th>Weeks</th>
<th>Frequency (per week)</th>
<th>Total sessions</th>
<th>Session duration (mins)</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eather et al. [21]</td>
<td>Crossfit Teens</td>
<td>Physical education</td>
<td>15.5</td>
<td>165</td>
<td>58.5</td>
<td>31</td>
<td>8</td>
<td>2</td>
<td>16</td>
<td>60</td>
<td>Grip-strength (kg)</td>
</tr>
<tr>
<td></td>
<td>Control group</td>
<td>Physical education</td>
<td>15.5</td>
<td>165</td>
<td>58.5</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ignjatovic et al. [22]</td>
<td>Experimental group</td>
<td>Handball</td>
<td>16.9</td>
<td></td>
<td></td>
<td>11</td>
<td>12</td>
<td>2</td>
<td>24</td>
<td>15</td>
<td>1 repetition maximum bench press (kg)</td>
</tr>
<tr>
<td></td>
<td>Control group</td>
<td>Handball</td>
<td>16.9</td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson et al. [23]</td>
<td>High-speed treadmill (HST)</td>
<td>Soccer</td>
<td>16</td>
<td>164</td>
<td>58.1</td>
<td>12</td>
<td>6</td>
<td>2</td>
<td>12</td>
<td></td>
<td>Isometric knee flexor strength (kg) (leg extension)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard treadmill (ST)</td>
<td>Soccer</td>
<td>16</td>
<td>164</td>
<td>58.1</td>
<td>12</td>
<td>6</td>
<td>2</td>
<td>12</td>
<td></td>
<td>Isometric knee flexor strength (kg) (leg extension)</td>
</tr>
<tr>
<td></td>
<td>Control group</td>
<td>Soccer</td>
<td>16</td>
<td>164</td>
<td>58.1</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lillegard et al. [24]</td>
<td>Tanner stage 1-2 (Experimental)</td>
<td>General population</td>
<td>9.5</td>
<td>136.3</td>
<td>36.7</td>
<td>7</td>
<td>12</td>
<td>3</td>
<td>36</td>
<td>40</td>
<td>10RM leg extension (kg)</td>
</tr>
<tr>
<td></td>
<td>Tanner stage 1-2 (Control)</td>
<td>General population</td>
<td>9.6</td>
<td>137.3</td>
<td>35.3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10RM leg extension (kg)</td>
</tr>
<tr>
<td></td>
<td>Tanner stage 3-5 (Experimental)</td>
<td>General population</td>
<td>13.2</td>
<td>158.1</td>
<td>50</td>
<td>6</td>
<td>12</td>
<td>3</td>
<td>36</td>
<td>40</td>
<td>10RM leg extension (kg)</td>
</tr>
<tr>
<td></td>
<td>Tanner stage 3-5 (Control)</td>
<td>General population</td>
<td>12.5</td>
<td>153.3</td>
<td>50.3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10RM leg extension (kg)</td>
</tr>
<tr>
<td>Muehlbauer et al. [25]</td>
<td>Strength training group</td>
<td>Various</td>
<td>16.6</td>
<td>161.5</td>
<td>55.1</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>16</td>
<td></td>
<td>Maximum isometric force (N) (leg press)</td>
</tr>
<tr>
<td></td>
<td>Control group</td>
<td>Various</td>
<td>16.7</td>
<td>166</td>
<td>54.4</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Maximum isometric force (N) (leg press)</td>
</tr>
<tr>
<td>Pereira et al. [26]</td>
<td>Experimental group</td>
<td>Volleyball</td>
<td>14</td>
<td>160</td>
<td>52</td>
<td>10</td>
<td>8</td>
<td>2</td>
<td>16</td>
<td></td>
<td>1.5Kg Medicine ball throwing (m)</td>
</tr>
<tr>
<td></td>
<td>Control group</td>
<td>Volleyball</td>
<td>13.8</td>
<td>160</td>
<td>53.5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.5Kg Medicine ball throwing (m)</td>
</tr>
<tr>
<td>Santos et al. [27]</td>
<td>Combined strength and</td>
<td>Various</td>
<td>13.5</td>
<td>155.9</td>
<td>54.8</td>
<td>25</td>
<td>8</td>
<td>2</td>
<td>16</td>
<td></td>
<td>3Kg Medicine ball throwing (m)</td>
</tr>
<tr>
<td></td>
<td>endurance training (GCOM)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strength training (GR)</td>
<td>Various</td>
<td>13.5</td>
<td>159.4</td>
<td>58.9</td>
<td>21</td>
<td>8</td>
<td>2</td>
<td>16</td>
<td></td>
<td>3Kg Medicine ball throwing (m)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>Various</td>
<td>13.5</td>
<td>156.8</td>
<td>51.5</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3Kg Medicine ball throwing (m)</td>
</tr>
<tr>
<td>Siegel et al. [28]</td>
<td>Experimental group</td>
<td>General population</td>
<td>8.5</td>
<td>128.5</td>
<td>27.2</td>
<td>24</td>
<td>12</td>
<td>3</td>
<td>36</td>
<td>30</td>
<td>Cable flexion (kg)</td>
</tr>
<tr>
<td></td>
<td>Control group</td>
<td>General population</td>
<td>8.4</td>
<td>128.2</td>
<td>26.4</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cable flexion (kg)</td>
</tr>
<tr>
<td>Skattebo et al. [29]</td>
<td>Intervention group</td>
<td>Cross-country skiing</td>
<td>18</td>
<td>171</td>
<td>61</td>
<td>9</td>
<td>10</td>
<td>2</td>
<td>20</td>
<td>33</td>
<td>1RM seated pull-down (kg)</td>
</tr>
<tr>
<td></td>
<td>Control Group</td>
<td>Cross-country skiing</td>
<td>17</td>
<td>166</td>
<td>60</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1RM seated pull-down (kg)</td>
</tr>
<tr>
<td>Study</td>
<td>Group/Activity</td>
<td>Sport</td>
<td>Peak Torque (ft-lbs)</td>
<td>Isokinetic Dynamometer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------</td>
<td>----------------</td>
<td>----------------------</td>
<td>------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sugimoto et al. [30]</td>
<td>Moderate-compliance</td>
<td>Volleyball</td>
<td>15.5</td>
<td>170.3</td>
<td>59.4</td>
<td>7</td>
<td>10</td>
<td>3</td>
<td>30</td>
<td>Hip abductor peak torque (ft-lbs)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(isokinetic dynamometer)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High-compliance</td>
<td>Volleyball</td>
<td>15.5</td>
<td>171.6</td>
<td>68.8</td>
<td>7</td>
<td>10</td>
<td>3</td>
<td>30</td>
<td>Hip abductor peak torque (ft-lbs)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(isokinetic dynamometer)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noncompliance group</td>
<td>Volleyball</td>
<td>16</td>
<td>173.4</td>
<td>63.9</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>Hip abductor peak torque (ft-lbs)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(isokinetic dynamometer)</td>
<td></td>
</tr>
<tr>
<td>Yoshimoto et al. [11]</td>
<td>Training group</td>
<td>Various</td>
<td>13.8</td>
<td>154.9</td>
<td>48.2</td>
<td>27</td>
<td>8</td>
<td>5.5</td>
<td>44</td>
<td>Maximal knee extension (kg)</td>
<td></td>
</tr>
<tr>
<td>Control group</td>
<td></td>
<td>Various</td>
<td>13.8</td>
<td>154.4</td>
<td>46.4</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td>Maximal knee extension (kg)</td>
<td></td>
</tr>
<tr>
<td>Subgroup</td>
<td>Effect size and confidence interval</td>
<td>Effect descriptor</td>
<td>P</td>
<td>Groups</td>
<td>N</td>
<td>Between group P</td>
<td>Between group P</td>
<td>Within group P</td>
<td>Within group P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------</td>
<td>-------------------</td>
<td>----</td>
<td>--------</td>
<td>----</td>
<td>-----------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 15yrs</td>
<td>0.38 [-0.02, 0.79]</td>
<td>Small</td>
<td>0.06</td>
<td>7</td>
<td>196</td>
<td>7.6%</td>
<td>0.30</td>
<td>41%</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 15yrs</td>
<td>0.72 [0.23, 1.21]</td>
<td>Moderate</td>
<td>0.004</td>
<td>8</td>
<td>155</td>
<td></td>
<td></td>
<td>42%</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 163cm</td>
<td>0.55 [0.08, 1.02]</td>
<td>Small</td>
<td>0.02</td>
<td>8</td>
<td>211</td>
<td>0%</td>
<td>0.73</td>
<td>57%</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 163cm</td>
<td>0.67 [0.20, 1.13]</td>
<td>Moderate</td>
<td>0.005</td>
<td>6</td>
<td>119</td>
<td></td>
<td></td>
<td>18%</td>
<td>0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 56kg</td>
<td>0.53 [-0.00, 1.06]</td>
<td>Small</td>
<td>0.05</td>
<td>7</td>
<td>180</td>
<td>0%</td>
<td>0.68</td>
<td>60%</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 56kg</td>
<td>0.67 [0.30, 1.03]</td>
<td>Moderate</td>
<td>0.0004</td>
<td>7</td>
<td>150</td>
<td></td>
<td></td>
<td>3%</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 8wks</td>
<td>0.62 [0.17, 1.07]</td>
<td>Moderate</td>
<td>0.007</td>
<td>8</td>
<td>231</td>
<td>0%</td>
<td>0.58</td>
<td>56%</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 8wks</td>
<td>0.44 [-0.02, 0.90]</td>
<td>Small</td>
<td>0.06</td>
<td>7</td>
<td>120</td>
<td></td>
<td></td>
<td>25%</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 2 sess p/week</td>
<td>0.72 [0.34, 1.09]</td>
<td>Moderate</td>
<td>0.0002</td>
<td>9</td>
<td>221</td>
<td>70.1%</td>
<td>0.07</td>
<td>36%</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; than 2 sess p/week</td>
<td>0.18 [-0.26, 0.61]</td>
<td>Trivial</td>
<td>0.42</td>
<td>6</td>
<td>130</td>
<td></td>
<td></td>
<td>21%</td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 16 sessions</td>
<td>0.75 [0.33, 1.17]</td>
<td>Moderate</td>
<td>0.0004</td>
<td>7</td>
<td>184</td>
<td>54.6%</td>
<td>0.14</td>
<td>35%</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 16 sessions</td>
<td>0.30 [-0.11, 0.72]</td>
<td>Small</td>
<td>0.16</td>
<td>8</td>
<td>167</td>
<td></td>
<td></td>
<td>33%</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 40 mins per session</td>
<td>0.34 [-0.38, 1.06]</td>
<td>Small</td>
<td>0.03</td>
<td>3</td>
<td>77</td>
<td>0%</td>
<td>0.52</td>
<td>53%</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 40 mins per session</td>
<td>0.64 [0.11, 1.16]</td>
<td>Moderate</td>
<td>0.02</td>
<td>4</td>
<td>72</td>
<td></td>
<td></td>
<td>4%</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Effect of moderator variables with 95% confidence intervals
Records identified through database searching: 397

Records remaining after removal of duplicates: 378

Studies removed after abstract inspection
- Inaccessible source (n = 4)
- Inappropriate intervention (n = 2)
- Inappropriate source (n = 5)
- Inappropriate study design (n = 82)
- Male participants (n = 51)
- No measure of strength (n = 1)
- Not peer-reviewed (n = 33)
- Participants wrong age (n = 35)
- Prospective study notice (n = 1)
- Study duration too long (n = 11)

Records remaining after inspection of abstracts: 153

Studies removed:
- Inaccessible source (n = 4)
- Inappropriate intervention (n = 27)
- Inappropriate source (n = 4)
- Inappropriate study design (n = 9)
- Insufficient data provided (n = 4)
- Male and female data grouped (n = 15)
- Male participants (n = 16)
- No control group (n = 5)
- No measure of strength (n = 1)
- Non-English language source (n = 1)
- Participants wrong age (n = 22)
- Study duration too long (n = 4)
- Unable to establish sex (n = 4)

Records remaining after inspection of full published papers: 10

Studies included in meta-analysis: 11

Articles added through supplementary searches (n = 1)
Figure

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Experimental Mean</th>
<th>Experimental SD</th>
<th>Experimental Total</th>
<th>Control Mean</th>
<th>Control SD</th>
<th>Control Total</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ether et al. [21]</td>
<td>30.5</td>
<td>8.9</td>
<td>31</td>
<td>30.8</td>
<td>8.9</td>
<td>19</td>
<td>0.41 [0.17, 0.99]</td>
<td></td>
</tr>
<tr>
<td>Ignjatovic et al. [22]</td>
<td>40.1</td>
<td>6.6</td>
<td>11</td>
<td>39.3</td>
<td>4.7</td>
<td>10</td>
<td>0.05 [0.01, 0.91]</td>
<td></td>
</tr>
<tr>
<td>Johnson et al. [23] (HST)</td>
<td>31.55</td>
<td>8.9</td>
<td>12</td>
<td>24.75</td>
<td>7.4</td>
<td>4</td>
<td>0.75 [0.42, 1.12]</td>
<td></td>
</tr>
<tr>
<td>Johnson et al. [23] (ST)</td>
<td>38.01</td>
<td>7.1</td>
<td>12</td>
<td>24.76</td>
<td>7.4</td>
<td>4</td>
<td>0.16 [0.04, 0.38]</td>
<td></td>
</tr>
<tr>
<td>Lillegard et al. [24] (Tanner 1-2)</td>
<td>19.3</td>
<td>9.6</td>
<td>7</td>
<td>10.4</td>
<td>4.7</td>
<td>5</td>
<td>1.03 [0.23, 2.10]</td>
<td></td>
</tr>
<tr>
<td>Lillegard et al. [24] (Tanner 3-5)</td>
<td>37.4</td>
<td>7.3</td>
<td>6</td>
<td>23.8</td>
<td>10.1</td>
<td>4</td>
<td>1.45 [0.05, 2.96]</td>
<td></td>
</tr>
<tr>
<td>Muichlauer et al. [25]</td>
<td>1,328</td>
<td>324.5</td>
<td>6</td>
<td>1,061.3</td>
<td>190.3</td>
<td>7</td>
<td>4.35 [0.62, 8.09]</td>
<td></td>
</tr>
<tr>
<td>Pereira et al. [26]</td>
<td>7.8</td>
<td>14.3</td>
<td>10</td>
<td>0.43</td>
<td>10</td>
<td>7.4</td>
<td>0.06 [-0.62, 0.75]</td>
<td></td>
</tr>
<tr>
<td>Santos et al. [27] (GCOM)</td>
<td>4.25</td>
<td>0.73</td>
<td>26</td>
<td>3.76</td>
<td>0.43</td>
<td>11</td>
<td>0.73 [0.00, 1.46]</td>
<td></td>
</tr>
<tr>
<td>Santos et al. [27] (OR)</td>
<td>4.28</td>
<td>0.74</td>
<td>21</td>
<td>3.70</td>
<td>0.43</td>
<td>10</td>
<td>0.78 [0.00, 1.56]</td>
<td></td>
</tr>
<tr>
<td>Siegel et al. [28]</td>
<td>11.8</td>
<td>1.9</td>
<td>24</td>
<td>11.8</td>
<td>2</td>
<td>16</td>
<td>10.1 [0.00, 0.63]</td>
<td></td>
</tr>
<tr>
<td>Skatzoop et al. [29]</td>
<td>34.67</td>
<td>3.12</td>
<td>9</td>
<td>29.23</td>
<td>4.66</td>
<td>7</td>
<td>5.46 [0.20, 2.43]</td>
<td></td>
</tr>
<tr>
<td>Sugimoto et al. [30] (High compliance)</td>
<td>55.4</td>
<td>13.1</td>
<td>7</td>
<td>51.6</td>
<td>6.1</td>
<td>3</td>
<td>4.1 [1.07, 1.66]</td>
<td></td>
</tr>
<tr>
<td>Sugimoto et al. [30] (Moderate compliance)</td>
<td>51.4</td>
<td>7.8</td>
<td>7</td>
<td>51.6</td>
<td>6.1</td>
<td>4</td>
<td>4.7 [1.26, 1.20]</td>
<td></td>
</tr>
<tr>
<td>Yoshimoto et al. [11]</td>
<td>39.8</td>
<td>9.8</td>
<td>27</td>
<td>41.5</td>
<td>7.4</td>
<td>20</td>
<td>10.96 [-0.79, 0.39]</td>
<td></td>
</tr>
</tbody>
</table>

Total (95% CI) 217 134 100.0% 0.54 [0.23, 0.85]

Heterogeneity: Tau² = 0.15, Chi² = 24.13, df = 14 (P = 0.04), % = 42%
Test for overall effect: Z = 3.37 (P = 0.0009)