A SYSTEMATIC REVIEW AND META ANALYSIS EXAMINING THE EFFECT OF EXERCISE ON INDIVIDUALS WITH INTELLECTUAL DISABILITIES
A SYSTEMATIC REVIEW AND META ANALYSIS EXAMINING THE EFFECT OF EXERCISE ON INDIVIDUALS WITH INTELLECTUAL DISABILITIES
Abstract

This study examined the efficacy of exercise programs for individuals with ID based on experimental designs. Multiple databases were searched from inception up until March 2019. Randomised control trials were eligible for inclusion if they included a population with ID, used an exercise intervention and included performance, cardiovascular health and/or psychological measures. All data were pooled using random effects models of standardized mean differences. The review included 18 studies which represented data from 799 individuals with ID. The largest effect was seen in relation to anxiety and depression symptoms (SMD = -3.07). This study represented the first comprehensive analysis on this topic area and illustrated that exercise can play an important therapeutic role for this population.

KEYWORDS: Intellectual disability, exercise, intervention, systematic review
A systematic review and meta analyses on the effect of exercise for individuals with intellectual disabilities.

The benefits of physical activity (PA) are numerous and have been so widely accepted that even the World Health Organization presents guidelines for PA. While the evidence that PA is good for both mental and physical health and that it may help improve overall quality of life is strong, according to the WHO, few people actually engage in sufficient PA on a regular basis (World Health Organization, 2002). Policy makers work to address this issue through mass media and informational campaigns encouraging individuals to incorporate more PA into their lives. Too often however, programs are designed, and resources designated in a way that best serves the general population. Such an approach ignores the unique needs of individuals who experience impairments due to cognitive deficits, such as those with intellectual disabilities. An intellectual disability (ID) is a disability which occurs before the age of 18 and results in adaptive, cognitive, and social impairments (Schalock, Luckasson, & Shrogen, 2007).

Adults with ID already comprise a disproportionately large number of annual hospital visits and are dispensed more medications than adults without ID. One in three adults with an ID in Canada are deemed high-cost patients, a designation used to refer to patients who rank within the top 10% of annual health care spending (Lunsky et al., 2018). Given the fact that this cohort tends to experience a disproportionate number of secondary health conditions compared to typically developing adults, it would be reasonable to suggest that they, in turn, have lower quality of life (May & Kennedy, 2010). Physical inactivity is one modifiable factor which is linked to many conditions and diseases experienced by this population including motor impairments, cardiovascular disease, obesity, diabetes and types of cancer (Durstine, Gordon, Wang, & Luo, 2013). However, it is not enough to increase PA when attempting to create meaningful changes to specific outcomes. Instead it is important to increase exercise behaviour...
which is meaningful and structured and therefore, intentionally targets health and fitness related outcomes (Caspersen & Christenson, 1985).

Exercise defined as a physical activity which is planned, structured, repetitive and purposeful and focuses on improve or maintenance of one or more fitness related outcomes (Caspersen & Christenson, 1985). Exercise is one means by which individuals, including those with ID, can meet recommended guidelines for PA (Department of health and human services, 2018). Regular engagement in exercise can also have positive benefits for an individual’s mental well-being and overall quality of life which can also significantly deteriorate as this population continues to age (Ravindran et al., 2016). Approximately 4 in every 10 individuals with an ID are diagnosed with a secondary mental illness. The most common conditions are anxiety, depression, bipolar disorder and schizophrenia. Often adults with ID are often prescribed multiple medications, as opposed to more holistic alternatives such as exercise (Finlayson, Turner, & Granat, 2011), which has continually shown to regulate and improve symptoms associated with mental illnesses (Ravindran et al., 2016). Therefore, a compelling argument can be made that increasing exercise behaviour in this cohort may help to mitigate secondary conditions associated with physical and mental health.

The Canadian physical activity guidelines suggest a minimum of 150 minutes of PA per week to be considered sufficiently active (Dairo et al., 2016). Approximately 23% of typically developing Canadians are sufficiently active, comparatively research has found that as few as 2% of this individuals with intellectual disabilities are sufficiently active (Chow, Choi, & Huang, 2018), however this percentage does vary number ranges from 2% to 9% (Dairo, Collett, Dawes, & Oskrochi, 2016). (Dairo et al., 2016).
While barriers to PA for this population have been acknowledged by researchers, the most effective means to increase PA behaviour in this population remains unclear. As Temple and colleagues (2017) stated in their recent review on physical activity promotion for adults with intellectual disabilities:

It is clear from this review that experimental research focused on increasing participation in physical activity and promoting physical activity to improve the health of adults with intellectual disabilities is in its infancy…Despite the potential benefits of physical activity and low levels among adults with intellectual disabilities, this review demonstrates that research to document the process and outcomes of physical activity interventions is sadly lacking. (p.451-452)

Although this applies to PA behaviours in general, how best to increase participation in exercise is also at this point unknown.

A meta-analysis by Shin & Park (2012) assessed the effect of exercise programs on individuals with intellectual disabilities. The authors focused on outcomes related to body composition, physiological outcomes (e.g., fitness) and physical performance (i.e. balance) and found an overall positive effect of exercise programs in this population. Additionally, it was found that programs which were shorter in duration (e.g., 10 minutes) and ran 4 times per week were more effective than those that ran for longer durations, but less frequently (e.g., 3 times per week; 13). One of the major limitations of this review was related to the exclusion of individuals with Down syndrome, despite the fact that this cohort makes up a large percentage of individuals with ID. Additionally, the authors chose to focus on highlighting physical
health outcomes only, despite the fact that the psychological domain which can include anxiety, depression and self-efficacy, can all be positively impacted by exercise. Furthermore, the studies reviewed in the previous meta-analysis included a broad range of research designs. Reducing the quality of evidence to only experimental or randomized control trials can be used to examine the efficacy of exercise-based intervention; Cross-sectional and longitudinal designs cannot. Therefore, the objective of the current study is to assess the effectiveness of exercise interventions based on experimental designs on individuals with intellectual disabilities. The specific research questions are:

1. Do exercise interventions lead to positive physical and/or psychological outcomes in individuals with intellectual disabilities when compared to a control?
2. What is the magnitude (i.e., effect size) of these changes?
3. Which of these outcomes, mental or physical, are most effected by exercise?

Methods

Search strategy and selection criteria

In the Winter of 2019, a literature search of all relevant databases was conducted. Date limits were applied up to March 2019. The following databases were searched: Medical Literature Analysis and Retrieval System Online (MEDLINE), Psych Info, and SportDiscus. Search terms included physical activity and/or exercise (title and abstract), intellectual disabilities (title and abstract), adult (title and abstract), children (title and abstract). No specific publication format restrictions were set. Only studies written in English were included in the literature search.
In order to ensure the data in this analysis is of the highest methodological rigor and contains studies with the highest quality evidence, only studies which randomly assigned individuals or clusters to an experimental or control group were used.

Given the limited research on this topic, this analysis included studies on individuals from all age groups (both children and adults). In order to be inclusive of the kinds of programs created for individuals with ID, studies which focused on any intellectual disability population were included. The intervention must have been one that was specific to exercise. Any and all programs were included regardless of setting. Again, due to the relatively small body of studies in this area, any and all exercise modalities were included. However, no cross-sectional studies were included. Primary outcome variables included physical and psychological variables.

Physical outcome variables included were body composition measures (BMI, waist circumference), blood pressure, oxygen consumption and aerobic capacity. Psychological outcomes included anxiety, self-rated depression, and self-efficacy.

**Data collection and analysis**

**Selection of studies.** Two independent reviewers screened the titles and abstracts of all obtained articles. Of those identified for potential inclusion, the full texts were obtained and reviewed by two independent reviewers. The reviewers discussed and resolved any discrepancies that were found. A total of 18 studies were included in the review (see Figure 1).

**Assessment of risk of bias in included studies.** In order to grade the strength of the evidence, study quality was assessed using the Cochrane Risk of Bias 2.0 (RoB 2.0), specifically designed for cluster randomized trials. The RoB 2.0 was used to assess bias related to threats to internal validity such as flawed research design, poor study execution, and/or incomplete
reporting of results. Specifically, the RoB 2.0 assesses risks associated with randomization and allocation sequence, blinding, incomplete or missing outcome data, and selective reporting (Sterne, Egger, Moher, & Boutron, 2017). The RoB 2.0 was completed by two independent reviewers. Any disagreements were discussed until a consensus was reached.

**Measurement of treatment effects.** To assess the effectiveness of the intervention on the various outcome measures, standardized mean difference (SMD) and standard error (SE) were calculated for all outcome variables. The use of SMD and SE allows for the summary reporting of findings taking into account that different scales and measures were used across different studies. The degree of the standardized mean difference was assessed using Cohen’s standardized conventions (Sterne et al., 2017) for effect size; small (0.2), medium (0.5) and large (0.8). Studies typically reported pre-and post-intervention time points, however due to studies not reporting variability in the change scores (i.e., standard deviation (SD) of the change score), a comparison of the post intervention measurement score were used (Cohen, 1992). The post intervention mean score and SD for measures of performance, body composition, cardiovascular and psychological measures were all entered. A direct comparison of the exercise intervention and control group was then completed.

**Assessment of heterogeneity.** For the purpose of this review, heterogeneity is defined as follows: “Statistical heterogeneity manifests itself in the observed intervention effects being more different from each other than one would expect due to random error (chance) alone.” (Higgins & Green, 2011). A visual and statistical examination of any study estimate inconsistencies was completed by visually examining forest plots and consideration of the $X^2$ and $I^2$ values. The proposed thresholds from the Cochrane handbook chapter 9.5.2 (Higgins & Green, 2011) were used to interpret $I^2$ values.
**Data synthesis**

Random effects models were generated for each outcome using RevMan (Version 5.3, 2014) software. Forest plots of the main analyses and tables containing the results of the sensitivity analyses were also generated.

**Subgroup and sensitivity analyses.** In cases where outcomes were assessed through unique measurements (e.g., musculoskeletal fitness was measured in both upper and lower body), subgroups would be used to analyze the data. This allowed for independent analyses of effect sizes for varying measures within the same outcome. Sensitivity analyses investigated the impact of varying intraclass correlation coefficient (ICC) values.

**Results**

**Results of search**

A total of 715 records were obtained from all databases. After 129 duplicates were removed, 585 titles were screened. From this first level of screening (titles and abstracts), 489 studies were deemed irrelevant, leaving 96 full text articles to be reviewed for eligibility. Of those 96, 78 were excluded. Reason for exclusion was wrong study design (53 studies), wrong outcomes (11 studies), wrong intervention (7 studies), wrong patient population (4 studies), wrong route of administration (2 studies) and wrong setting (1 study). A total of 18 studies were included in this review (see Figure 1). The percent disagreement for the full text review was 7%, with discrepancy on seven articles. Of those seven articles, 0 were included in the final review.

**Included studies**

The 18 studies included 799 individuals with intellectual disabilities from studies conducted in Europe, the United States, Australia, and South Africa. Sixteen of the trials randomized individual participants into the intervention or control group, while two studies used cluster randomization (e.g., Day Activity centers). Intervention length ranged from 5 weeks to
52 weeks. Most interventions ran for 10-12 weeks and were typically performed 3 times per week.

**Risk of bias in included studies**
Table 1 documents data regarding the risk of bias assessments for all included studies.

Overall, two studies were low risk of bias, nine studies were deemed to have some concerns, and the remaining seven were rated as having a high risk of bias. Out of the 18 studies, 9 studies had at least some risk of bias regarding deviations due to intended intervention, usually due to lack of proper participant/trial personnel blinding.

**Results of Pooled Sample**
Participants in this review included individuals with mild to moderate ID (n=279), individuals with Down syndrome with ID (n=130), individuals with profound ID (n=37), and individuals with ID of unspecified etiology (n=353). Of those, 349 (44%) were male; however, the gender of 30 participants was unspecified (Beasley, 1982).

**Effects of intervention**
The following results come from 18 studies.

**Performance Measures**
All results for performance measures are outlined in Table 2.

**Submaximal exercise.** In this review, submaximal exercise is defined as a type of exercise that is terminated before reaching ventilatory threshold or maximum heart rate (HR). It is used to estimate VO$_2$ max or aerobic fitness (Heyward, 2009). Estimates of aerobic fitness were tested using the six-minute walk test (Boer et al., 2014; Boer, & Moss, 2016; Calder et al., 2011; Marks, Sisirak, & Chang, 2013), and the shuttle run test (Ozmen, Ryildirim, Yuktasir, & Beets, 2007; Schijndel-Speet, Evenhuis, Wijck, & Echteld, 2014). Pooled estimates of all
measures of aerobic fitness from eight studies (Beasley, 1982; Boer et al., 2014; Boer, & Moss, 2016; Calders et al., 2011; Lee, Lee, & Song, 2016; Marks, Sisirak, & Chang, 2013; Ozmen et al., 2007; Schijndel-Speet et al., 2014) with a combined sample size of 333 participants showed almost no increase in aerobic fitness when comparing exercise interventions to sedentary controls (SMD = 0.13, 95% confidence interval (CI) range from -0.11 to 0.37). The range of effects shows moderate possible harm, no effect, and small benefit. The overall $I^2 = 17\%$ indicated low heterogeneity.

**Balance.** For the purpose of this review, balance is defined as static balance or the ability to maintain the body in a fixed position (Rival, Ceyte, & Olivier, 2005). Three studies (Borji, et al., 2018; Lee et al., 2016; Schijndel-Speet et al., 2014) (pooled n=162) assessed static balance. Pooled estimates show that although the effect size is large (SMD = 1.25), CI ranged from -0.39 to 2.90, indicating no effect of the intervention. This range indicates that results show a small degree of harm, no effect, and large benefit.

**Functional fitness.** Functional fitness reflects one’s ability to perform physical activities of daily life with relative ease (Heyward, 2009). Functional fitness included five studies (Boer et al., 2014; Boer, & Moss, 2016; Calders et al., 2011; Lee et al., 2016; Marks et al., 2013) and a total of 325 participants, with pooled estimates indicating a minimal effect (SMD = -0.07). Overall, results were imprecise: showing high benefit, no effect, and a moderate degree of harm. When subgroup analysis was examined, Sit to Stand was found to favor the controls (SMD = 0.37), while the Get up and Go test did have a large, benefit observed in the intervention group (SMD = -0.77) (the decrease in the Get up and Go scores does indicate an improvement on the test, as decreased time indicates better functional fitness). The $X^2$ test of subgroup differences
was statistically significant ($p=0.02$) indicating the effects differ across between the two subgroups.

**Musculoskeletal strength.** For this review, muscular strength was defined as any activity which elicited the maximum force that a muscle or muscle group can generate at a specific velocity. Studies looked at pooled estimates from all measures from six studies (Boer et al., 2014; Calders et al., 2011, Giagazoglou et al., 2013; Schijndel-Speet et al., 2014; Shields et al., 2013; Suomi, 1998) of 351 participants showed an effect estimate indicating strong, positive increase when comparing the intervention group to the control group (SMD = 0.70, 95% CI range from 0.24 to 1.16), indicating benefit. Subgroup analyses show slightly less conclusive results for upper body strength (95% CI range from -0.17 to 1.26), while the results for low body musculoskeletal strength indicate a strong effect (SMD = 0.86, 95% CI range from 0.30 to 1.42). The overall $I^2 = 74\%$, indicating a high degree of heterogeneity.

**Flexibility.** Two studies (Giagazoglou et al., 2013; Marks et al., 2013) were included in the pooled analysis of flexibility. Results from 152 participants were inconclusive, showing possible benefit, no effect, and possible harm (SMD = -0.19, 95% CI from -1.73 to 1.34).

**Body composition.** Body composition included body mass index (Boer et al., 2014, Calders et al., 2011; Melville et al., 2015; Ozmen et al., 2007), weight in pounds (Marks et al., 2013), weight in kilograms (Schijndel-Speet et al., 2014). Pooled estimates from all measures from six studies of 343 participants showed a small effect size for weight (SMD = 0.13, 95% CI from -0.12 to 0.37). The 95% CI indicates the possibility of benefit, no effect and possible harm. Only two studies (Melville et al., 2015; Schijndel-Speet et al., 2014) measured waist circumference, and results were inconsistent: possible benefit, no effect, and possible harm (95% CI from -0.37 to 0.50). All results for body composition are outlined in Table 3.
Cardiovascular Health

All results for cardiovascular health are outlined in Table 4.

Four studies (Boer & Moss, 2016; Boer et al., 2014; Calders et al., 2011; Rosety-Rodriguez et al., 2014) examined maximal oxygen uptake (n = 109); the pooled estimates indicated a medium effect (SMD = 0.55) when comparing the exercise intervention to the sedentary control. The 95% CI ranged from 0.17 to 0.94.

Blood pressure was examined in three studies (Boer & Moss, 2016; Calders et al., 2011; Schijndel-Speet et al., 2014), and the results of pooled estimates revealed a medium effect (SMD = -0.30, 95% CI from -0.56 to -0.03). Subgroup analyses of systolic and diastolic blood pressure found a medium effect of exercise on systolic blood pressure (SMD = -0.47, 95% CI from -0.085 to -0.10), while exercise had a small effect on diastolic (SMD = -0.12, 95% CI from -0.48 to 0.23; See Figure) with no significant differences between subgroups ($ p = 0.19$).

While results from the pooled estimates of three studies (Boer & Moss, 2016; Boer et al., 2014; Calders et al., 2011) (n=89) examining heart rate did show a small effect of exercise benefiting the intervention group (SMD = 0.11) the 95% CI show indefinite results: possible harm, no effect and possible benefit.

Psychological Outcomes

All results for psychological outcomes are outlined in Table 5.

Anxiety and depression was assessed by three studies (Carraro & Gobbi, 2012; Carraro & Gobbi, 2014; Schijndel-Speet, 2014). The pooled estimates from 140 participants showed a large effect of exercise on anxiety and depression related symptoms (SMD = -3.07). However, the CI were extremely wide and provides unspecified results (95% CI from -6.81, 0.66).
Self-efficacy was examined in two studies (Marks et al., 2013; Melville et al., 2015) (n = 152) and pooled estimates indicated a large effect of exercise when comparing the intervention to the sedentary control (SMD = 0.74, 95% CI from -0.33 to 1.80).

**Discussion**

**Summary of main results and certainty of evidence**

Interest in exercise as an intervention for various outcome measures for individuals with IDs has steadily increased, with a greater number of interventions appearing in the literature. However, the inclusion of all available trials provided inconsistent results. Some evidence did indicate improvements with regards to lower body muscular endurance, blood pressure, reaction time, and self-efficacy. Results for other outcomes were inconsistent, even potentially harmful. Furthermore, the best available evidence for other outcomes is uncertain due to the quality of evidence. It is possible, however that those with ID need extra support to fully and consistently maximize the benefits of exercise.

**Overall completeness and applicability of evidence**

The studies within this review include individuals with various types of intellectual disabilities (Down syndrome, IDD, PDD-NOS), various exercise modalities, and a wide range of age groups. Very few of these studies are adequately powered, and limited number had examined multiple outcomes in the same study. The problem of low sample size resulted in confidence intervals were wide and could not provide quality, determinant results. Additionally, few studies reported ICC values, which also made it difficult to assess quality. It is important to consider, however the difficulty associated with recruiting this population, which could largely account for the small participant sample sizes. In general, studies had at least some concern or high concern with regards to risk of bias. Typically, studies did not have blinded outcome
assessors, and as a result, the risk of bias assessment with regards to measurement outcome was high.

Reporting errors were also a major cause of concern. One study, in particular, did not include any information regarding the duration, frequency, intensity or modality of the exercise intervention (Schijndel-Speet, 2014). Additionally, some studies included information for some of these categories, but not all. As a result, it was not possible to conduct post hoc analyses exploring the effect of intervention duration, frequency, and intensity on outcomes. Therefore, the body of research must become more robust in order to identify detectable differences/effects estimates resulting from various durations, frequencies, and intensities. Specifically, increases in sample size and the quality surrounding the measurement of the interventions need to be improved.

Agreements and disagreements with other studies or reviews

Overall, the results from this meta-analysis show that while definitive, and salient changes were not observed in all domains/outcomes, some improvements were apparent. Notably, there is at least some modest evidence that exercise may lead to positive changes in musculoskeletal strength, maximal oxygen uptake, and blood pressure. Previous research supports the findings from the current meta-analysis as previous work has also found that exercise programs improved muscle force, VO2, and self-esteem in a similar population (Shin & Park, 2012). Furthermore, in the general population, sustained and ongoing exercise behaviour can help to improve total blood pressure score, and reduce symptoms associated with hypertension (Carpio-Rivera, Moncada-Jiménez, Salazar-Rojas, & Solera-Herrera, 2016) and it appears this extends to populations with intellectual disabilities as well.
The current review was the first to examine the impact that exercise can have on mental health. The evidence showed that there were large (though, somewhat imprecise) gains in mental health outcomes for adults with ID. Specifically, when investigating exercises impact on anxiety and depressive symptoms. This finding not only points out the link between mental health and exercise, but the complexity of this link in those with ID and the importance of exercise as a means of providing holistic treatment for secondary mental health conditions.

However, the results from this study show strong but variable results due to few studies actually investigating mental health in this population. This provides more reason for researchers and practitioners to continue their consideration of exercise and mental health in this group. More research needs to explore how these benefits are maintained beyond the duration of the intervention.

While research has previously shown that exercise can indeed improve body composition outcomes (Swift, Johannsen, Lavie, Earnest, & Churh, 2013), the current meta-analysis found only a small effect on body composition despite the fact that each of the six studies used exercise as a means to facilitate weight loss. Beyond the small effect size, CI indicated no effect and possibility of harm. However, when it comes to weight loss, exercise is only one small aspect related to weight loss. Individuals must also maintain a healthy diet and must reduce caloric intake while increasing energy input. Furthermore, weight loss is highly unique and differs from person to person (Swift et al., 2013). One factor which hinders weight loss and improvements in body composition for individuals with ID is medication. A large percentage of this population is prescribed antipsychotic and antidepressant medications (Doan, Lennox, Taylor-Gomez, & Ware, 2013), both of which can cause weight gain (Wharton, Raiber, Serodio, Lee, & Christensen, 2018). Medication-induced weight gain is significant and often difficult to reverse.
While none of the studies in the current review indicated whether or not participants were taking medication, it is possible that this, in combination with other individual factors (i.e., nutrition), resulted in vague findings.

Flexibility, step count, and heart rate also had unclear results, which aligns with previous work (Shin & Park, 2012). It is likely this is a result of small sample sizes as each of these outcomes had less than 100 participants included in the pooled estimates. However, outcomes with larger sample sizes (submaximal exercise, balance, functional fitness, anxiety/depression) also had imprecise results. Many of the studies included in these outcomes were 12-week exercise interventions, and it is possible that this length was not enough to produce any salient changes.

**Conclusion**

The results of this review, while inconsistent, do provide some evidence indicating the need for more research to determine the efficacy of fitness/exercise programs for this population. It is clear that within this population, exercise and exercise related behaviours can lead to some positive changes in specific outcomes. Current research is largely focused on physical outcomes while there is an extreme lack of quality evidence supporting exercise as an alternative therapy for mental health in this population. It is important to recognize that exercise may act as a primary treatment for many co morbid conditions that are prominent within this population such as anxiety and depression (Lunsky et al., 2018). While this review was not able to identify best exercise practices for this population, it does provide evidence that exercise (through any method) may be of benefit for individuals with ID.

This review is largely limited by the quality of evidence, which justifies the need for future studies to employ methodologically sound, adequately powered interventions.
Additionally, very little information could be drawn regarding optimal program frequency, timing, and length (which are fundamental to any exercise program) due to lack of evidence. This also extends to adherence, as no studies completed follow-ups to confirm if the changes they saw extended beyond the duration of the intervention. This is an important note for future researchers as evidence which supports whether or not these changes are resistant to time is lacking. Furthermore, no studies measured fidelity related to the implementation of the intervention and if there were deviations from the intended program this may have effected the final results. Therefore, it is important that future research ensure that interventions are being carried out according to the initial design.

Insufficient evidence does not allow us to draw conclusions regarding several outcomes including functional fitness, submaximal exercise performance, and heart rate. Overall this review serves a pertinent reminder that while individual studies have identified exercise as a prominent way to improve many lifestyle and health factors in those with ID this data should be subject to reproduction before it can be taken as fact.
References


https://doi.org/10.1111/jep.12145


https://doi.org/https://doi.org/10.1016/j.ridd.2013.09.022


https://doi.org/10.1016/j.ridd.2012.05.019


FOOTNOTES

Contributors: LSJ: Study conception/design, data acquisitions, data analysis and interpretation, and drafting and revision of manuscript. GB: Data acquisition, data analysis, and revision of manuscript. JC: Study conception/design, critical revision. All authors: final approval.

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests: None declared.

Patient consent: Not required.

Provenance and peer review: Not commissioned; externally peer reviewed.

Data sharing statement: Further details on studies included in this review can be retrieved by contacting the corresponding author.
Figure 1. PRISMA flow chart.
<table>
<thead>
<tr>
<th>Study (Publication Date)</th>
<th>Randomization process bias over all judgement</th>
<th>Bias due to deviations from intended interventions</th>
<th>Bias due to missing outcome data</th>
<th>Bias in measurement of the outcome</th>
<th>Bias in selection of the reported result</th>
<th>Overall Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beasley (1982)</td>
<td>Some concerns</td>
<td>Some concerns</td>
<td>Low</td>
<td>Some concerns</td>
<td>Low</td>
<td>Some concerns</td>
</tr>
<tr>
<td>Boer (2014)</td>
<td>Low</td>
<td>Low</td>
<td>Some concerns</td>
<td>Low</td>
<td>Low</td>
<td>Some concerns</td>
</tr>
<tr>
<td>Boer (2016)</td>
<td>Low</td>
<td>Some concerns</td>
<td>Some concerns</td>
<td>Low</td>
<td>Low</td>
<td>Some concerns</td>
</tr>
<tr>
<td>Borji (2018)</td>
<td>High risk</td>
<td>Low</td>
<td>Low</td>
<td>Some concerns</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Bossink (2017)</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Calders (2011)</td>
<td>High risk</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Carraro (2012)</td>
<td>Low</td>
<td>Some concerns</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Carraro (2014)</td>
<td>Low</td>
<td>Some concerns</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Giagazoglou (2013)</td>
<td>Low</td>
<td>Low</td>
<td>Some concerns</td>
<td>Low</td>
<td>Some concerns</td>
<td>Low</td>
</tr>
<tr>
<td>Khalili (2009)</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Lee (2016)</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Marks (2013)</td>
<td>Some concern</td>
<td>Low</td>
<td>Some concerns</td>
<td>Low</td>
<td>Low</td>
<td>Some concerns</td>
</tr>
<tr>
<td>Melville (2015)</td>
<td>Low</td>
<td>Some concerns</td>
<td>Some Concerns</td>
<td>Low</td>
<td>Low</td>
<td>Some concerns</td>
</tr>
<tr>
<td>Ozmen (2007)</td>
<td>Some concern</td>
<td>Some concerns</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Some concerns</td>
</tr>
<tr>
<td>Rosety-Rodriguez (2014)</td>
<td>Low</td>
<td>Some concerns</td>
<td>Low</td>
<td>Some concerns</td>
<td>Low</td>
<td>Some concerns</td>
</tr>
<tr>
<td>Shield (2013)</td>
<td>Some concern</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Some concerns</td>
</tr>
<tr>
<td>Suomi (1998)</td>
<td>Low</td>
<td>Some concerns</td>
<td>Low</td>
<td>Some concerns</td>
<td>Low</td>
<td>Some concerns</td>
</tr>
<tr>
<td>Van Schijndel Speet (2017)</td>
<td>Low</td>
<td>Some concerns</td>
<td>Some concerns</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>
Table 1
Risk of bias assessment for included studies
<table>
<thead>
<tr>
<th>Outcome or Subgroup Title</th>
<th>No. of studies</th>
<th>No. of participants</th>
<th>Statistical Method</th>
<th>Effect Size (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Aerobic Fitness</td>
<td>8</td>
<td>333</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>0.13 [-0.11, 0.37]</td>
</tr>
<tr>
<td>1.2 Balance</td>
<td>3</td>
<td>162</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>1.25 [-0.39, 2.90]</td>
</tr>
<tr>
<td>1.3 Functional Fitness</td>
<td>5</td>
<td>325</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>-0.07 [-0.60, 0.46]</td>
</tr>
<tr>
<td>1.3.1 Sit-to Stand</td>
<td>5</td>
<td>198</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>0.37 [-0.20, 0.94]</td>
</tr>
<tr>
<td>1.3.2 Get up and Go</td>
<td>3</td>
<td>127</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>-0.77 [-1.34, -0.19]</td>
</tr>
<tr>
<td>1.4 Musculoskeletal Strength</td>
<td>6</td>
<td>351</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>0.70 [0.24, 1.16]</td>
</tr>
<tr>
<td>1.4.1 Upper body</td>
<td>4</td>
<td>223</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>0.55 [-0.17, 1.26]</td>
</tr>
<tr>
<td>Section</td>
<td>Trials</td>
<td>N</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>--------</td>
<td>----</td>
<td>----------------------------------------</td>
<td></td>
</tr>
<tr>
<td>1.4.2 Lower body</td>
<td>4</td>
<td>128</td>
<td>0.86 [0.30, 1.42]</td>
<td></td>
</tr>
<tr>
<td>1.5 Flexibility</td>
<td>2</td>
<td>152</td>
<td>-0.19 [-1.73, 1.34]</td>
<td></td>
</tr>
<tr>
<td>1.5.1 Upper Body</td>
<td>1</td>
<td>67</td>
<td>0.08 [-0.40, 0.56]</td>
<td></td>
</tr>
<tr>
<td>1.5.2 Lower Body</td>
<td>2</td>
<td>85</td>
<td>-0.30 [-3.16, 2.56]</td>
<td></td>
</tr>
<tr>
<td>1.7 Step Count</td>
<td>3</td>
<td>188</td>
<td>0.30 [-0.15, 0.75]</td>
<td></td>
</tr>
</tbody>
</table>
Table 3

*Body Composition Comparison*

<table>
<thead>
<tr>
<th>Outcome or Subgroup Title</th>
<th>No. of studies</th>
<th>No. of participants</th>
<th>Statistical Method</th>
<th>Effect Size (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Weight</td>
<td>6</td>
<td>343</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>0.13 [-0.12, 0.37]</td>
</tr>
<tr>
<td>2.2 Waist circumference</td>
<td>2</td>
<td>189</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>0.06 [-0.37, 0.50]</td>
</tr>
</tbody>
</table>
Table 4

*Cardiovascular Fitness Comparisons*

<table>
<thead>
<tr>
<th>Outcome or Subgroup Title</th>
<th>No. of studies</th>
<th>No. of participants</th>
<th>Statistical Method</th>
<th>Effect Size (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Maximal Oxygen Uptake</td>
<td>4</td>
<td>109</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>0.55 [0.17, 0.94]</td>
</tr>
<tr>
<td>3.2 Blood Pressure</td>
<td>3</td>
<td>242</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>-0.30 [-0.56, -0.03]</td>
</tr>
<tr>
<td>3.2.1 Systolic</td>
<td>3</td>
<td>121</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>-0.47 [-0.85, -0.10]</td>
</tr>
<tr>
<td>3.2.2 Diastolic</td>
<td>3</td>
<td>121</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>-0.12 [-0.48, 0.23]</td>
</tr>
<tr>
<td>3.3 Heart Rate</td>
<td>3</td>
<td>89</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>0.11 [-0.44, 0.65]</td>
</tr>
</tbody>
</table>
### Table 5

**Psychological Comparisons**

<table>
<thead>
<tr>
<th>Outcome or Subgroup Title</th>
<th>No. of studies</th>
<th>No. of participants</th>
<th>Statistical Method</th>
<th>Effect Size (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Anxiety and Depression</td>
<td>3</td>
<td>140</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>-3.07 [-6.81, 0.66]</td>
</tr>
<tr>
<td>4.2.1 Self Efficacy</td>
<td>2</td>
<td>152</td>
<td>Std. Mean Difference (IV, Random, 95% CI)</td>
<td>0.74 [-0.33, 1.80]</td>
</tr>
</tbody>
</table>