This is a post-peer-review, pre-copyedit version of an article published in Journal of Autism and Developmental Disorders. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10803-021-05152-z

Effects of physical exercise interventions on stereotyped motor behaviours in children with ASD: A meta-analysis

Elizabeth J. TEH[1],

Ranjith VIJAYAKUMAR[2],

Timothy Xing Jun TAN[2],

Melvin J. YAP[2]

[1] Department of Otolaryngology / Division of Graduate Medical Studies, Yong Loo Lin School of Medicine, National University of Singapore

[2] Department of Psychology, National University of Singapore

Correspondence concerning this article should be addressed to:

Dr. Elizabeth J. Teh
Email: entejt@nus.edu.sg
Telephone: +65 6516 8026
ORCID ID: 0000-0001-7389-1113

Declarations

Conflicts of interest: The authors declare that they have no conflicts of interest.

Abstract

Studies have reported that physical exercise reduces maladaptive stereotyped motor behaviours (SMB) in children with ASD, but these intervention studies vary in design and outcome. The present systematic review and meta-analysis included 22 studies, involving 274 children with ASD, to quantify the effect of exercise on SMB and its potential moderators. Multi-level modelling revealed a large overall effect, Hedges' g=1.16, with significant heterogeneity across participant, treatment, and study levels. Further, a more appropriate model using between-case estimation for within-subject effects to improve comparability between single-case and group-design studies, yielded a smaller but still significant effect, g=0.51. Lastly, higher exercise intensity, but not age, exercise duration or settings, reliably predicted treatment effectiveness. Implications for clinical practice and research are discussed.

(119 words)

Keywords: autism spectrum disorder, children, stereotyped motor behaviours (SMB), exercise intervention, single-case designs, meta-analysis

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that affects an estimated one in every 54 children (Maenner et al., 2020). The key characteristics of ASD are impaired social interaction and communication skills, and restricted, repetitive patterns of behaviour, interests or activities (American Psychiatric Association, 2013). While individuals with ASD experience these symptoms to different extents, many tend to experience isolation because their symptoms reduce their ability to interact successfully with others (Kunzi, 2015). Thus, treatment that is effective and yet tailored to the individual is needed to cater to the heterogeneity of symptom profiles within the ASD population. The current review will focus on treatment of one common characteristic in ASD, namely, stereotyped motor behaviours, through the use of physical exercise interventions.

Stereotyped Motor Behaviours (SMB)

Restricted and repetitive behaviours (RRB) in ASD include stereotyped motor behaviours (SMB), insistence on sameness (IS), and circumscribed interests (CI) (Leekam, Prior, & Oljaveric, 2011). The three subtypes of RRB are significantly correlated with each other but they exhibit different relationships with other developmental characteristics, such as nonverbal IQ (Bishop et al., 2006; Richler et al., 2010). Thus, researchers have argued that it is useful to study them separately in order to better understand their etiologic mechanisms and responses to treatment (Cuccaro et al., 2003; Richler et al., 2010). Many children with ASD display forms of SMB during their development, including stereotyped movements of specific body parts such as hand-flapping, repetitive manipulation of objects, vocal stereotypy, and/or echolalia (Leekam et al., 2011; McLaughlin, 2010).

SMB may interfere with other aspects of functioning (Richler et al., 2010). Koegel and Covert (1972) reported a negative effect on learning in three children with ASD, in which their ability to focus and attend to information. Moreover, the attention allocated to specific objects or movements also leads to reduced social functioning and reduced opportunities for social interaction, potentially resulting in children with ASD not receiving critical input for normal social development (Lee, Odom, & Loftin, 2007; Richler et al., 2010). Another study by Koegel and colleagues (1974) found that while children were engaged in SMB, they were unable to exhibit appropriate play behaviours which in turn reduced their socialization. Given the negative outcomes associated with SMB, there is a need for evidence-based treatments to help reduce these maladaptive behaviours.

Theories of SMB. Researchers have proposed at least three models to explain the manifestation of SMB in people with ASD. One theory is that individuals with ASD exhibit SMB to regulate arousal and maintain homeostasis (Turner, 1999). A compensatory effect for restoring sensory homeostasis may be induced by SMB, either by increasing sensory stimulation or physiological arousal in the case of under-arousal or by providing soothing stimulation and blocking out external stimuli in the case of over-arousal (Hutt & Hutt, 1965; Zentall & Zentall, 1983). Another explanation suggests that the presence of SMB is due to reduced inhibition of inappropriate behaviours and difficulties generating novel behaviour in ASD, leading to a maintained set of restricted behaviours (Leekam et al., 2011; Turner, 1997). Lopez and colleagues (2005) show support for this theory, finding that deficits in cognitive flexibility and planning are correlated with the occurrence of stereotypic behaviours. Finally, SMB can be described as an operant behaviour which is maintained through sensory reinforcement. These behaviours provide perceptual, auditory and/or tactile stimulation, which provide pleasant internal consequences for the individual (Rapp & Vollmer, 2005) and are thus automatically reinforced (Furniss & Biswas, 2012).

All three theories above conceptualize SMB as self-maintained behaviour that either rewards the individual in terms of arousal-regulation and/or sensory reinforcement, or is sustained by a failure of inhibition. As such, it has been proposed that physical exercise can decrease an individual's need to engage in SMB by facilitating arousal modulation back to optimal levels (Boyd et al., 2011). Further, exercise and SMB that involve similar body mechanics may provide similar internal sensations. Studies have shown that interventions matching the specific topography of SMB in an individual with ASD (e.g., ball throwing exercises matching hand-flapping stereotypic behaviours) produce the greatest reduction in that form of SMB (Rapp et al., 2004; Tse et al., 2018). In this view, exercise may serve as an abolishing operation for SMB by functioning as matched-stimulation which in turn decreases the reinforcing value of SMB (Neely et al., 2015). Evidently, physical exercise shows promise as a form of intervention for SMB.

Exercise Interventions for SMB

Physical exercise interventions are an antecedent-based intervention that reportedly alleviates SMB and improves other aspects of functioning in individuals with ASD (Bremer et al., 2016; Petrus et al., 2008). Some exercises that have reportedly produced reductions in SMB include jogging (Kern et al., 1982), cycling (Olin et al., 2017), 'exergaming' or virtual reality exercise activities (Anderson-Hanley et al., 2011), ball exercises (Nazemzadegan et al., 2016), therapeutic horseback riding (Gabriels et al., 2012), yoga (Koenig et al., 2012), and martial arts (Bahrami et al., 2012). Petrus et al. (2008) conducted a systematic review of seven studies published between 1982 and 2003 which indicated that exercise provides short-term reductions of SMB in children with ASD. Another systematic review by Bremer et al. (2016), involving 13 studies from 1993 to 2013, reported that exercise interventions led to improvements in stereotypic behaviours, socio-emotional functioning, cognition and attention. Only six of the included studies measured changes in SMB, with five studies reporting significant reductions and one study by Oriel et al. (2011) reporting no significant change in SMB post-intervention. Of these six studies, only three reported effect sizes, allowing the reviewers to obtain a large average effect size of Cohen's *d* = 1.6 based on the three studies. Bremer et al. (2016) chose to conduct their systematic review without a meta-analysis due to the heterogeneous samples, interventions, and outcomes in the studies. The limitation here is that systematic reviews descriptively evaluate the quality of studies' research designs and analytical methods to gauge the robustness of their results (Mallet et al., 2012), but unlike meta-analytic reviews, do not quantitatively estimate intervention effectiveness across studies.

More recently, Ferreira et al. (2019) and Tarr et al. (2020) conducted meta-analyses on the same topic with similar results. Ferreira et al.'s meta-analysis included eight studies and revealed an overall effect size of *d* = 1.11, suggesting that children with ASD showed a reduction in stereotyped behaviours after exercise intervention. However, the authors applied strict inclusion criteria, namely, (a) samples composed of children up to 16 years with a diagnosis of ASD; (b) physical exercise intervention without the involvement of animals as a complementary therapeutic resource or any other parallel secondary intervention (e.g., cognitive or social intervention); (c) SMB as a dependent variable; and (d) published in 1970-2017. Consequently, their review examined a relatively small sample of studies. Tarr et al. (2020) reviewed 10 studies, comprising four group-design (aggregate score) studies and six single-case studies, involving children with ASD aged 3 to 21 years old. They, too, reported positive outcomes on SMB, but with a smaller effect size of *d* = 0.456. Further, they tested 16 moderator hypotheses, including sex, age, type and function of stereotypies, intervention choice, intensity, duration, and other characteristics. Analysing the two groups of studies separately and in combination, there were no significant effects reported within any of the moderators. However, Tarr et al. (2020) applied categorical coding for each variable, which might have limited the data, and the small sample sizes in their analyses were likely under-powered to detect moderator effects with 16 variables tested. Moreover, there are methodological concerns with the conventional meta-analysis models used in both reviews, discussed next.

Current gaps and methodological problems in previous reviews

Potential moderators. Despite the range of exercise interventions reported in the literature, previous reviews have not definitively identified factors that may affect treatment outcomes on SMB. Two potential moderators that warrant further examination are exercise intensity and duration. It has been reported that vigorous exercise such as jogging might be more effective than mild exercise, such as walking or ball-play

(Kern et al., 1984; Levinson & Reid, 1993). On the other hand, Olin and colleagues (2017) found that lower- rather than higher-intensity aerobic exercise, as measured by heart rate per minute, was better for reducing SMB in adolescents with ASD. They further reported that higher-intensity exercise for 20 minutes worsened SMB symptoms compared to baselines. Commenting on exercise durations, Liu and colleagues (2016) recommended 15 minutes of moderate-to-vigorous exercise, whereas Burns and Ault (2009) proposed that just 5 to 8 minutes on the treadmill were effective for reduction of SMB. Further research is needed to clarify these mixed findings.

Another interesting question for researchers and practitioners is with regards to intervention setting. Sowa and Meulenbroek (2012) found that physical exercises delivered individually were more effective than group exercises for improving functional motor and social skills in ASD, but this moderating effect has not yet been examined on SMB outcomes. Additionally, while the majority of studies have examined physical exercise interventions conducted on children rather than adults with ASD (Bremer et al., 2016; Petrus et al., 2008), it remains unclear if children with ASD of all ages benefit from such interventions. Lastly, another variable of interest would be ASD severity in study samples. Changes in diagnostic criteria over the years makes this variable difficult to quantify consistently but it may be useful to describe samples qualitatively for review.

Previous sample limitations. The four reviews by Ferreira et al. (2019), Tarr et al., (2020), Bremer et al. (2016), and Petrus et al. (2008) – which, to the best of our knowledge, are currently the only reviews to have considered the effectiveness of exercise interventions on SMB in children with ASD – were based on small numbers of studies, partly because variability in ASD samples and interventions made it difficult to integrate findings across studies. We argue that appropriately conducted meta-analysis can be used to synthesize findings in the extant literature on the same research questions and better understand the sources of heterogeneity across studies (Beeson & Robey, 2006). We further propose that greater inclusivity in treatment reviews is important for establishing evidence-based practices for individuals with ASD. It is timely to conduct an updated review including emerging treatments such as exergaming and therapeutic horseback-riding which are increasingly used by interventionists. Additionally, besides peer-reviewed studies, unpublished dissertations often offer valuable findings from well-designed research studies (Borenstein et al., 2009), and published studies using samples drawn from autism-specific educational programmes may also contribute useful findings. A larger sample size in a well-conducted meta-analysis may increase power to detect effects of clinical interventions (Morley, 2018).

Combining single-case and group design studies appropriately. Single-case design (SCD) studies involve one or more participants undergoing one or more treatments, using a range of experimental designs to examine treatment effectiveness (Morley, 2018). With heterogeneous clinical populations such as ASD, well-designed case studies are often pivotal for investigating treatment effects, more so than randomized controlled trials (Howard et al., 2015). As such, the exercise interventions reviewed previously included a number of studies using single-case designs (SCD; e.g., Kern et al., 1982; Levinson & Reid, 1993). In SCD studies, unlike group designs, individual outcomes can be reported for each participant and treatment. However, previous meta-analyses have often inappropriately combined individual participants' outcomes in SCD studies to obtain a single average effect size from each study (e.g., Ferreira et al., 2019; Sowa & Meulenbroek, 2012). This method effectively implies that the effect sizes within a study are homogeneous, which is not a reasonable assumption in many SCD studies (Cheung, 2015).

Another problem arises when reviewers include two or more effect sizes from the same study in a meta-analysis without accounting for dependencies between them, together with single effect sizes obtained from other studies. For example, in Ferreira et al.'s (2019) meta-analysis, one study contributed separate effect sizes for male and female sub-samples, while the remaining seven studies contributed one effect size each. This method violates the assumption of independence of effect sizes and may lead to incorrect inferences about the data (Cheung, 2015). Lastly, effect sizes derived from group-designs are computed using within-group standard deviations, while effect sizes derived from SCDs use the within-phase standard deviation of each participant (Ferreira et al., 2019). This computational difference is problematic in a conventional meta-analysis which weights effect sizes by their sampling variance (Shadish et al., 2014). Tarr et al. (2020) created a between-variance based on computed Tau-*U* scores for SCDs but acknowledged that autocorrelations were not accounted for in their study. In contrast, Shadish and colleagues devised a new method for estimating the standardized difference *between* cases in SCD studies to control for autocorrelation and render the treatment effect size comparable to the group-design metric (Valentine et al., 2016). We propose using updated statistical methods involving multi-level meta-analyses and between-case effect size estimation (for SCD studies) to address all these concerns.

In summary, reviews to date have generally reported benefits of exercise interventions for reducing SMB in ASD. However, the reviews were based on small numbers of studies involving different exercise types, intensity, and duration. There is a need for an updated review of the extant literature using appropriate meta-analytic models to obtain a more accurate estimation of the true treatment effect size in the population. Concomitantly, a larger and more diverse set of studies will increase power and variability to detect and better understand moderator effects for future exercise interventions.

The present study

The present study aims to address the above empirical and methodological gaps by using updated meta-analyses methods to investigate two research aims:

1. To estimate the overall effectiveness of physical exercise interventions on SMB for children with ASD; and
2. To examine the potential influence of (i) participant age, (ii) exercise intensity, (iii) exercise duration, and (iv) intervention setting on SMB outcomes.

This review will include SCD and group intervention studies that examined exercise effectiveness on frequency of stereotyped motor behaviours in children with ASD. will include participants with a range of SMB severity, intervention methods and study designs, yielding a larger pool of studies and increasing the power of this review to detect significant effects. We will apply a multi-level meta-analysis, modelling effect sizes at participant-level, treatment-level, and/or study-level, while preserving information about effect size heterogeneity within studies (Method 1). In addition, we will present a second exploratory model using Shadish et al.'s (2014) recommendation for estimating treatment effect sizes from eligible SCD studies. This analyses will examine whether the proposed estimation method changes the pattern of results for overall effectiveness and moderator effects on SMB in children with ASD (Method 2). The findings from this review will provide deeper insights into the relationship between exercise interventions and SMB, and contribute crucial evidence-based information for designing effective exercise interventions to reduce SMB in children with ASD. Finally, it will provide insights for future reviewers interested in conducting multi-level meta-analyses involving heterogeneous clinical groups.

Method

Literature search

A systematic search was initially conducted applying Bremer et al.'s (2016) search strategy on physical exercise interventions for children with ASD. Four electronic databases were initially searched to identify published and unpublished studies for inclusion. The databases were PsycINFO, Scopus, Web of Science, and PubMed. Following Bremer et al. (2016), the definition of exercise applied was, "physical activity that is planned, structured, repetitive and purposeful" (Caspersen et al., 1985, p. 152). This definition was chosen for being broad and generalizable to many different exercise types. The keywords used in the literature search were 'Autism Spectrum Disorder', 'ASD', 'Autism', 'Asperger', 'Pervasive Developmental Disorders', 'PDD-NOS', 'children', 'physical exercise', 'exercise', 'physical activity', 'sports', and a comprehensive word string for physical exercise activities to account for the differences in terminology used in the various databases (see Appendix 1 for the full word string). Additionally, a keyword search on Google Scholar was conducted.

Inclusion criteria. Studies had to meet the following criteria: (1) published between 1900 and 2018; (2) written in English; (3) included at least one participant diagnosed with ASD; (4) independent variable was physical exercise intervention; and (5) the dependent variable was frequency of SMB. Studies were excluded if physical exercise was used as a form of punishment (contingent exercise).

{INSERT FIGURE 1 ABOUT HERE}

Figure 1 depicts the results of the systematic literature search. The initial search of four databases (PsycINFO, Scopus, Web of Science and PubMed) yielded 436 records. After removing 45 duplicates, the remaining abstracts were screened by the third author. A total of 375 records were deemed irrelevant, for reasons such as not meeting the pre-specified inclusion criteria or being commentaries and review articles. These were discussed with the first and fourth authors and agreement obtained to reject the records. The reference lists of the remaining 16 records were then scrutinized to identify additional studies and a keyword search using Google Scholar was conducted, yielding 11 further studies after screening. Thus, 27 studies were eligible for full-text appraisal. In two studies, the target population was children with developmental disabilities who exhibit stereotyped behaviours (Powers et al., 1992; Prupas & Reid, 2001). Both studies recruited children with ASD or PDD-NOS, and one child with Fragile X syndrome which is a condition that has high co-morbidity with ASD, and the authors agreed to include both studies. Four studies were excluded which did not report sufficient data (in the form of tables, graphs, exact probabilities, and/or descriptive statistics) for calculation of effect sizes, and one other study was excluded due to it having a different measure of the dependent variable, namely, whether a specific SMB improved, worsened, or stayed the same. A total of 22 studies met the inclusion criteria and were included in the meta-analysis. Summary characteristics of the included studies are presented in Table 1.

{INSERT TABLE 1 AROUND HERE}

Quality assessment

The quality of each study was evaluated using the Mixed Methods Appraisal Tool (MMAT), which is suitable for appraising methodological quality for different study designs within a single tool (Pluye et al., 2009; revised, Hong et al., 2019). The MMAT was chosen because our review includes studies using RCT (2 studies), non-randomized designs (8 group- and 10 single-case design experiments), and single-subject experiments (2 studies). The quality assessment focused on exercise interventions and data on SMB, with higher scores indicating lower risk-of-bias in studies.

Data extraction and statistical analyses

Method 1: Four-level random-effects meta-analysis.

Effect size calculation. We first computed standardized mean differences from each study following methods in previous meta-analyses (e.g., Ferreira et al., 2019; Sowa & Meulenbroek, 2012). We extracted sample sizes, treatment and control means and SDs for between-group designs; and pre-intervention and post-intervention means and standard deviations (SDs) for within-group designs and SCD designs. Where the means and SDs were not reported, the study's raw data were extracted to calculate the pre-intervention and post-intervention means and SDs. Two studies (Koenig et al., 2012; Rosenblatt et al., 2011) did not provide means and SDs or raw data, and so the effect size was calculated using the reported *t*-value and sample-size. We calculated Hedges' *g* (Hedges, 1981), which is a small-sample corrected version of standardized mean difference (Cohen's *d,* Cohen, 1988). Group studies provided one effect size per treatment, with studies having multiple treatments providing the corresponding number of effect sizes per study. SCD studies provided one effect size per participant, and multiple treatments yielded multiple effect sizes per participant. Thus, we extracted data from 22 studies (nine SCD and 13 group-design studies) which yielded 59 effect sizes. All effect size calculations were done using R statistical software (R Core Team, 2019), using the *esc_mean_sd()* function or *esc_t*() function in the *esc* package.

Four moderators of interest were extracted from the studies, as follows:

- *Participant age.* Age was treated as a continuous variable. Mean ages were coded for group-design studies.
- *Exercise intensity.* The most objective way to measure exercise intensity is by using physiological measures from participants. However, few studies used physiological measures such as participant heart rate or an accelerometer to measure intensity through participants' movements. Therefore, we instead coded exercise intensity by reference to the Compendium of Physical Activity 2011 to provide a way to standardize physical exercise intensities based on the exercises' metabolic equivalent (MET) intensities (Ainsworth et al., 2011). MET is the ratio of the work metabolic rate compared to the resting metabolic rate. The Compendium provides estimated MET values for a comprehensive list of physical activities used in the research literature, from light/low intensity (< 3.0 METs) to vigorous/high intensity (≥6 METs) (Ainsworth et al., 2011). For this review, each study's exercise description was matched to the closest activity descriptors in the Compendium to obtain the corresponding MET value, and similar activities were assigned a consistent MET value across studies. Thus, intensity was also coded as a continuous variable.

- *Exercise duration.* The duration of exercise interventions ranged from approximately 1.5 minutes to 60 minutes. Exercise duration was coded as a continuous variable.
- *Intervention setting.* Individual setting involves one participant performing the physical exercise intervention without other participants present, regardless of the number of interventionists attached to the participant. Group setting interventions have multiple participants performing the physical exercise intervention together, usually in a class. For this review, intervention setting was coded as either individual or group.

Statistical model. In a conventional meta-analysis, the sampling variance of effect sizes in a study is modelled at level-1 and between-study variation is modelled at level-2. However, this method does not account for dependencies among effect sizes from the same studies (versus from different studies) included in a meta-analysis (Cheung, 2015). The present method corrects for this by modelling the dependencies of data at separate clusters or levels, as shown by the schematic depiction in Figure 2. With sampling variance forming level-1, we modelled effect sizes (level-2) nested within treatments (level-3), which were in turn nested within studies (level-4). Thus, this four-level model accounts for dependencies among individual effect sizes in SCD and group-design studies while preserving information about effect size heterogeneity within studies, yielding a more accurate estimate of the true effect of physical exercise intervention on SMB. Additionally, we used robust variance estimation to test the significance of the effect size in consideration of complex dependencies within our multi-level structure where participants nested within treatments in some SCDs were cross-nested in other treatments within those studies (Hedges et al., 2010).

Between-study heterogeneity was indicated by the I^2 statistic, which represents the overall magnitude of true heterogeneity versus sampling error between studies. Percentages of 25%, 50%, and 75% indicate low, medium, and high heterogeneity, respectively (Hunter & Schmidt, 1990). A random-effects model was chosen over a fixed-effects model as it assumes that the average effect size within the population differs randomly across the various studies (Cheung, 2015), reflecting the heterogenous nature of the ASD population in the present investigation. Besides, a random-effects model leads to a lower Type I error compared to a fixed-effects model, generating more accurate results (Hunter & Schmidt, 1990). Lastly, using a random-effects model in the meta-analysis allows the results to be generalizable to a larger population beyond the studies in the review (Field & Gillett, 2010).

{INSERT FIGURE 2 AROUND HERE}

Method 2: Between-case random-effects meta-analysis.

Between-case effect size calculation for SCDs. Shadish and colleagues (2014) pointed out that between-groups and between-participants (or 'cases') effect sizes, as described in Method 1 above, may not be statistically comparable as they are estimated with different denominators. Group-designs use the within-group SD, while SCDs use the within-phase SD of each participant. Following Shadish et al.'s method for SCD studies, we extracted means and SDs and number of sessions in treatment and baseline conditions or calculated them using raw data across sessions available from tables or graphs. By this method, each SCD study provided one effect size per treatment, with multiple treatments resulting in multiple effect sizes per study (Figure 3). However, calculations can only be done for studies with at least three participants and available information on the specific sequence of sessions per participant. Consequently, six SCD studies were excluded from our calculations due to insufficient participants and/or session data. For Method 2, we included three SCD studies and al1 13 group studies which yielded 24 effect sizes. The calculations were done using the web-application *scdhlm* (Pustejovsky, 2016).

Statistical model. The treatment effect sizes from SCD and group studies are now modelled at the same level (level-2), nested within studies which form level-3, with sampling variance forming level-1 (see model depiction in Figure 3). Thus, a three-level random effects meta-analysis model was applied, and the overall effect size was tested using robust variance estimation.

{INSERT FIGURE 3 AROUND HERE}

For both Method 1 and Method 2, we conducted moderator analyses to examine how the effect of physical exercise intervention across levels may be moderated by participant age, exercise intensity, duration, and setting. The meta-analyses were conducted using the *metafor* package (Viechtbauer, 2010) in statistical software R (R Core Team, 2019). The R-codes for the two meta-analyses are provided in Supplement 1.

Results

Characteristics of included studies

Participants. Collectively, a total of 274 individuals with ASD took part in the physical exercise interventions. Autism was the most common diagnosis, followed by Asperger's Syndrome and then PDD-NOS. The mean age of participants across studies, where reported, was 9.60 years old, and the participant age range was from 3 to 21 years. There were 225 males (82.1%) and 35 females (12.8%). The gender of the remaining 14 individuals (5.1%) was not reported.

Outcome measures. The frequency of SMB was measured in studies either using direct or indirect methods. Researchers in 15 studies used direct methods, such as time sampling and interval recording of SMB. The remaining seven studies used indirect methods, by obtaining parent/caregiver-reported ratings of stereotypic behaviours pre- and post-intervention. The most commonly used measures were the Gilliam Autism Rating Scale (2nd ed.) which provides a 14-item subscale for stereotypic behaviours, and the Aberrant Behaviour Checklist (Aman & Singh, 1994), which provides a 7-item subscale for stereotypic behaviours. Both measures rate items on a four-point Likert scale.

Exercise interventions. A wide variety of physical exercise interventions were represented in this review, covering a range of exercise intensities, including: jogging (10 studies), walking (3 studies), ball playing (4 studies), cycling (2 studies), dance (2 studies), roller skating (1 study), trampoline jumping (1 study), yoga (2 studies), therapeutic horseback riding (1 study), *kata* martial arts (1 study) and *nei yang gong*[1] (1 study). Some studies used only one exercise intervention while others compared two or more exercise interventions.

Quality of the included studies. Following the MMAT (2018) criteria, eight papers were rated as high quality five-star studies (*****), meeting all five quality criteria; nine papers were rated as four-star (****), meeting 80% of criteria; and four papers as acceptable or three-star (***), meeting 60% of quality criteria. One study by Nazemzadegan et al. (2016) was rated lower (**) due to potential assignment bias and inconsistent reporting of the intervention sessions. All studies applied appropriate outcome measures, such as clearly defined measures for frequency of SMB, or standardised parent-rated measures for ASD-characteristic stereotypical behaviours. One common concern was rater bias, as data were often collected by parents/caregivers of participants or by researchers who were involved in conducting the intervention. Some researchers increased reliability by using two independent raters and reported high inter-rater agreements (e.g., Levinson & Reid, 1999). Another concern was confounding variables, such as individual participants' preference/dislike of exercise treatments, SMB variability, and activities immediately prior to the intervention. These factors were discussed by some authors (e.g., Lee et al., 2018), but were generally not controllable in the studies. Studies were also given lower ratings if inclusion /exclusion criteria were not specified. Lastly, potential reporting bias was identified in one study by Kern et al. (1982), where complete outcomes were reported for only two of seven participants and included in the present review.

{INSERT FIGURE 4 ABOUT HERE}

Overall exercise intervention effectiveness on SMB

Method 1: Combined four-level random-effects meta-analysis

[1] A form of Chinese mind-body exercise, based upon Zen principles of martial arts and healing, involving sets of slow, gentle, and smooth movements.

Figure 4 presents the individual effect sizes (Hedges' *g*) from the 22 included studies, and overall effect size with robust-variance estimated 95%-confidence intervals (CI_{95}) of the combined four-level random-effects model. The average effect size based on this model was $g = 1.16$ [$CI_{95} = 0.72, 1.61$], with an estimated sampling variance of 0.32. For this investigation, a positive Hedges' *g* value indicates a reduction in the frequency of SMB exhibited by children with ASD. Heterogeneity in effect sizes across studies was significant ($Q(58) = 230.20$, $p < .001$) with estimated heterogeneity variance $I^2 = 0.763$, indicating that 76.3% of the variance was accounted for by real differences in effect sizes (i.e., not attributable to sampling variance). Our multi-level meta-analysis revealed the breakdown of the variance: level-4 between-study differences contributed 23.0% of the variance; 31.4% of the variance was attributable at level-3 between-treatments within studies; and 21.9% at level-2, within treatments.

Moderator analyses. Next, we examined potential sources of variability contributing to the significant heterogeneity found, using robust variance estimation on the selected variables.

Age. The estimated regression coefficient was 0.04, *SE* = 0.08, *p* =.64, indicating that age was not significant as a moderator and did not explain any variance in intervention outcomes.

Exercise intensity. This moderator was significant, with coefficient 0.30, SE = 0.05, *p* <.001, indicating that greater exercise intensity produced more reductions in SMB.

Exercise duration. The estimated regression coefficient was -0.02, *SE* = 0.01, *p* =.02, indicating that longer duration had a negative effect on intervention outcomes.

Intervention setting. Finally, settings showed a non-significant effect on intervention outcomes, $F(1, 20) = 0.02$, *p* =.90.

Bias detection. A visual inspection of the funnel plot of effect sizes in Figure 5a showed asymmetry, which was suggestive of publication bias. We tested for bias using Egger's modified regression test for multi-level models (Egger et al., 1997) with study *SE*s as a moderator. A significant intercept suggests publication bias, possible dependencies among effect sizes that were not modelled adequately, or other inadequacy in sample selection (Sterne et al., 2011). The intercept coefficient was significant (-1.41, *p* =.001) in this analysis. A further sensitivity analysis using Cook's distance for effect sizes of individual studies was conducted to detect any cases which might exert an inordinate influence over the parameters of the model. This test was selected because it evaluates both residual statistics (influence of a case on the ability of the model to predict that case) and leverage (influence of the observed value of the outcome variable over the predicted values) in a regression analysis (Viechtbauer & Cheung, 2010). We applied a cut-off of > 4/N (where N is the number of effect sizes), which identified one study, Chan et al. (2013), as an influential outlier (Figure 5b).

We then re-ran the meta-analysis after removing this study. The overall average effect size remained significant, $g = 1.24$ [$CI_{95} = 0.82, 1.66$], with estimated sampling variance = 0.33. Effect size heterogeneity was slightly lower at 72.4% ($Q(57) = 198.98$, $p < .001$, $I^2 = 0.724$), with between-study variance at level-4 now 20.3%, variance at level-3 now 30.0%, and variance at level-2 now 22.1% . For moderators, the same two factors remained significant, i.e., exercise intensity (0.28, $SE = 0.04$, $p < .001$) and exercise duration (-0.02, $SE = 0.01$, $p = .04$). Age (0.05, $SE = 0.08$, $p = .49$) and exercise setting ($F(1,19) = 0.11$, $p = .74$) were not significant. As there was no change in the pattern of overall and moderator effects after adjusting for bias, except that overall average *g* was slightly larger in magnitude, we decided to retain all studies to avoid reducing the relatively small sample size in this meta-analysis.

{INSERT FIGURE 5 ABOUT HERE}

Method 2: Between-case three-level random-effects meta-analysis

The between-case three-level meta-analysis included 16 studies comprising effect sizes from three SCD studies calculated using Shadish et al.'s (2014) method and 13 group-design studies. Figure 6 presents a forest plot of the individual effect sizes (Hedges' *g*), overall effect size and robust-variance estimated confidence intervals. The average effect size based on this model was $g = 0.51$, [$CI_{95} = 0.19, 0.83$], with an estimated sampling variance of 0.27, indicating an overall reduction in the frequency of SMB exhibited by children with ASD obtained in these studies. Heterogeneity in effect sizes across studies was significant $Q(23) = 51.97$, $p < .001$, $I^2 = 0.409$, suggesting that 40.9% of the variance was accounted for by real differences in effect sizes not attributable to sampling variance. The level-2 between-treatment differences contributed 40.9% of the variance, while level-3 between-study variance was reduced to zero after considering the between-treatment variance.

Moderator analyses. Exercise intensity was the only significant moderator, with coefficient 0.21, $SE = 0.07$, $p = .01$. No significant effects were found for the other tested variables of participant age (-.08, $SE = 0.06$, $p = .22$) , exercise duration (-.01, $SE = 0.01$, $p = .31$), and intervention setting ($F(1, 14) = 0.80$, $p = .39$).

Bias detection. A funnel plot of the effect sizes is presented in Figure 7a. Egger's modified regression test for multi-level models, using study *SE*s as a moderator, revealed a non-significant intercept, -0.50, $p = .15$, suggesting no significant publication bias. Lastly, a sensitivity test using Cook's distance did not detect any cases above the cut-off (Figure 7b), suggesting that the overall effect size was not unduly influenced by any particular study in this analysis.[2]

{INSERT FIGURES 6 & 7 ABOUT HERE}

Post-hoc analyses. We calculated the numbers-needed-to-treat (NNT) as a clinical measure of our findings. NNT is based on the risk difference between desirable/undesirable outcomes in treatment and control groups. The ideal NNT is 1, indicating 100% improvement in the treatment group and 0% improvement in the control group (Cook & Sackett, 1995). For SMB, clinicians usually consider the impact of improvement on functioning rather than applying an absolute cut-off threshold. Thus, we simulated NNT across a range of clinical scenarios from 0.2 (more optimistic) to 0.8 (more pessimistic) probability of improvement in the control group (Thorlund et al., 2011). For model 1, with overall average $g = 1.16$, the estimated NNT ranged from 3 (0.2 probability) to 5 (0.8 probability) participants to treat. For model 2, with $g = 0.51$, the NNT range was slightly higher from 6 to 8 participants. Currently, few behavioural treatments in ASD have reported NNT values (e.g., Eldevik et al., 2010). We present this analysis for clinicians' reference and use when more NNT measures of treatments for SMB are available for comparison in future.

Discussion

This review included 22 empirical studies involving children with ASD aged 3 to 21 years old. The first aim of this review was to quantify the effectiveness of exercise interventions on SMB in children with ASD, by synthesizing a wide range of existing intervention programmes in the literature using multi-level modelling to examine variability in this heterogeneous population. The second aim was to explore the specific influence of child age, exercise intensity, duration, and intervention setting on treatment outcomes. To our knowledge, this review is the first attempt to statistically control for dependencies in small-*N* studies in this area and test moderators of exercise intervention outcomes on SMB in children with ASD. Additionally, we applied two statistical models, a four-level random-effects meta-analysis and a three-level between-case random-effects meta-analysis, to address methodological problems in the literature. The two meta-analyses indicated that exercise reduces SMB in children with ASD, but there was high variability in outcomes across treatments and studies. Exercise intensity appears to be a positive and strong predictor for effectiveness, while other variables of age, exercise duration and setting were not reliable moderators. We will discuss and compare the results from each model in turn below, together with suggested recommendations for future researchers.

---

[2] Since the study by Nazemzadegan et al. (2016) had a relatively low quality score of 2, we also conducted a sensitivity analyses by re-running all the meta-analyses without this study. The overall effect size increased slightly (Method 1: from $g = 1.16$ to $g = 1.19$; Method 2: from $g = 0.51$ to $g = 0.52$), and heterogeneity I^2 was slightly higher (Method 1: from 76.3% to 77.0%); Method 2: from 40.9% to 42.9%). The pattern of significance for overall and moderator effects remained unchanged. Thus, we decided to retain the study in the review to avoid reducing the sample sizes above.

Method 1: Four-level random-effects meta-analysis

Effectiveness of exercise on SMB

Our first model, a four-level random-effects meta-analysis with study effect sizes derived from all 22 studies using conventional estimation methods, yielded an average Hedges' *g* effect size of 1.16, which may be considered a large effect size (Cohen, 1988). Our findings are consistent with previous reviews reporting that exercise interventions can reduce SMB (e.g., Ferreira et al., 2019; Petrus et al, 2008). Further, we extend previous meta-analyses by Ferreira et al. (2019) and Tarr et al. (2020), which included fewer studies and applied conventional (but inappropriate) meta-analysis models. Our review which covers a wider range of treatments and study designs, including case studies and unpublished dissertations, supports the view that there is merit to greater inclusivity of existing treatments and studies in meta-analysis of clinical samples (Borenstein et al., 2009). Moreover, the findings are supported by the fact that most studies were of acceptable to high quality according to the MMAT tool.

Our overall findings provide indirect evidence for the arousal modulation theory of SMB in ASD, which claims that individuals with ASD exhibit SMB in order to regulate arousal to optimum levels and maintain homeostasis (Hutt & Hutt, 1964; Turner, 1999). Alternatively, exercise is known to involve similar body mechanics as SMB (Lang et al., 2010) and may generate the same pleasant sensory reinforcement induced by SMB (Rapp et al., 2005), leading to an observable reduction in SMB post-exercise. There is less support for the hypotheses of SMB maintenance due to cognitive inflexibility (Turner, 1997), which would predict a continuing need to engage in the exercise or replacement with SMB post-exercise. These theoretical questions are not directly examined in our review but are worthy of future research.

Potential moderators of exercise treatment effectiveness for SMB

Our novel use of a multi-level meta-analysis revealed that between-treatment differences accounted for the most variance (31%), while within-treatment and between-study variance were approximately 22% each. The results emphasise a need to examine differences due to variables at different levels.

Role of exercise intensity. Exercise intensity was a significant moderator of treatment effects, with higher exercise intensity producing larger effects on SMB. This finding further supports the theory of the role of SMB in arousal regulation. Kerr and van den Wollenberg (1997) reported that higher intensity exercise led to significantly greater increases in arousal compared to lower intensity exercise in a group of typical adults. For the children with ASD in this review, it is possible that higher intensity exercise caused arousal levels to return to optimum levels at a quicker rate and remain at homeostasis for longer periods of time, leading to a greater reduction in the frequency of SMB. Exercise interventions of higher intensity have also been shown to lead to greater executive functioning improvements (Tsukamoto et al., 2016). In this study, we coded intensity by reference to a standardized classification system based on metabolic equivalence (Ainsworth et al., 2011) since many studies did not report physiological measures of intensity. Future research should explore the robustness of this finding using relevant reported physiological measures such as participant heart rate.

Role of exercise duration. The meta-analysis in Method 1 found a significant negative effect of duration, but the small regression coefficient (b = -.03) suggests minimal impact on intervention outcomes that may not be clinically significant. This may partially explain the mixed recommendations in the current literature ranging from 5 to 15 minutes as the most effective exercise duration (e.g., Burns & Ault, 2009; Liu et al., 2015). It may be that a wide range of exercise durations are equally effective for reducing the frequency of SMB in children with ASD. This finding has implications for treatment dosage and further research comparing exercise durations in interventions may be recommended.

Non-significant effects of age and exercise setting. Treatment outcomes were not moderated by age, indicating that such forms of intervention will be beneficial for children and youths up to 21 years old. We also did not find a significant effect of setting, indicating that exercises were equally effective whether conducted in one-to-one or group settings. This is inconsistent with Sowa and Meulenbroek (2012), who reported that individual settings produced greater improvements in functional motor and social skills than group settings for children and adults with ASD. The authors hypothesized that outcomes were enhanced by more frequent or intensive therapist-client interactions but noted that the studies did not clearly describe the extent of interaction required in individual versus group settings. We were also unable to determine the extent to which exercise participants in group settings were required to meaningfully interact with interventionists or other participants. However, it may be hypothesized that children with ASD, who tend to experience social/communication deficits to some extent, might derive similar benefit from exercise for reducing SMB whether in an individual or group activity.

Method 2: Between-case random-effects meta-analysis

In method 2, we applied Shadish et al.'s (2014) recommendation of calculating between-case effect sizes in SCD studies with treatment data from at least three participants, which were then comparable to effect sizes from group-design studies. (We will compare and discuss the meta-analytic models in a later section.) Based on a subset of 16 eligible studies, we obtained a positive, medium average effect size of g = 0.51 [CI_{95} 0.16, 1.66]. Evidently, the finding that exercise is effective for reducing SMB in children with ASD was robust under both our methods of estimation. However, after accounting for the within-study dependencies appropriately in the between-case model, it appears that the effect size may be smaller than previously estimated using conventional methods. Further, variance at the study-level was nearly negligible (0%) after accounting for between-treatment differences (41%). In this model, exercise intensity was the only significant moderator of treatment outcome, with higher intensity associated with stronger effects on SMB. Duration of exercise was not significant (p =.30). Taken together with the small effect of duration reported in Method 1, it appears that this variable is not a reliable predictor of exercise intervention outcomes on SMB. Similarly, age and exercise setting were not significant moderators of treatment outcomes on SMB, which is consistent with the findings reported in the four-level model.

Updated meta-analysis models and effect size estimations

Previous meta-analyses have combined SCD and group studies by obtaining average effect sizes from each study, resulting in a loss of information about effect size heterogeneity within studies. Previous reviewers have also assumed independence of within-study effects (e.g., Ferreira et al., 2019; Tarr et al., 2020), which is problematic for inferential statistics. In contrast, the use of multi-level modelling in our meta-analyses allowed us to account for within-study dependencies, and preserve and examine sources of heterogeneity between cases within treatments and within studies at different levels. Additionally, in model 2, we re-estimated between-case effect sizes in SCD studies to address a methodological problem in conventional meta-analyses where effect sizes calculated from individuals (over a time-series) and groups (at pre-/post-treatment timepoints) were treated equally (Shadish et al., 2014). The between-case effect sizes were then appropriately combined with group-study effects to obtain a more accurate estimate of overall effect size and variability. The pattern of effects produced were largely similar to the previous model, but effect sizes were smaller in magnitude using the between-case model. For clinicians, the revised model indicates that efficacy of exercise interventions may be slightly lower than previously suggested.

It is possible that meta-analyses using effects derived using conventional estimation may be more vulnerable to Type I errors. However, since between-case effect sizes can only be calculated with sufficient data from study participants, there were only three SCD studies in our between-case model and inclusion of more eligible SCD studies is required before drawing firm conclusions. Hence we suggest that future researchers using SCD designs include at least three participants, and, at minimum, report outcome means, standard deviations, and sequence and number of trials per treatment. Additionally, we suggest that numbers-needed-to-treat statistics for interventions be reported to facilitate comparison of efficacy across treatments. Further, for future meta-analyses of studies involving heterogeneous ASD populations, we recommend that researchers select appropriate methods to estimate study effect sizes based on the types of study designs included, and use appropriate meta-analyses models to statistically control for dependencies in the different study designs in order to obtain more accurate estimates of true treatment effects.

Limitations and more future directions

Many studies in this review did not report whether participants were on medication or receiving other forms of intervention besides exercise during the study, which may have confounded findings in some studies. Also, since we deliberately included studies involving individuals with a range of SMB severity in order to reflect the full range of symptom severity in ASD, it is possible that our results may not equally apply to individuals who exhibit lower and higher SMB severity. In the present review, symptom severity could not be tested as a moderator due to a lack of consistency in rating scales applied by researchers across studies. Thus, sample descriptors were provided including ASD diagnoses and IQ where available for readers' review. For moderator analyses, our selected variables were mostly non-significant and the only reliable moderator, exercise intensity, was coded using a metric (Ainsworth et al., 2011). Hence further research is required on intensity and other potential moderators to better understand and design more effective exercise treatments for children with ASD.

Lastly, this review only examined immediate post-intervention effects of exercise. Only three studies measured frequency of SMB at follow-up time-points. Bahrami et al. (2012) reported that participants in their kata intervention still exhibited significantly less SMB compared to those in the control condition when measured at two days and 30 days post-intervention. Similarly, Prupas and Reid (2001) found that SMB levels at seven days post-intervention were lower than pre-intervention levels. In contrast, Levinson and Reid (1993) reported that SMB levels gradually rose back to original levels around 90 minutes after exercise. Therefore, longitudinal research is needed to better understand how well the benefits of exercise interventions are sustained over the longer-term for children with ASD.

Conclusion

This review used multi-level random-effects meta-analyses with two computation methods (conventional and between-case) for effect sizes to better understand the effects of exercise interventions on SMB in children with ASD. Since ASD is clinically heterogeneous, it is important to understand how variability in presentation might affect intervention effectiveness. Besides, there is a wide range of exercise interventions in practice. Applying multi-level meta-analyses, this study revealed that exercise interventions were effective in reducing SMB, with higher intensity exercise enhancing the effect significantly. Exercise interventions appear to be an evidence-based and sustainable approach to managing SMB in children with ASD as such interventions can be easily conducted and made affordable and accessible to clients (Chasson et al., 2007). However, more research is needed to understand moderators of exercise effectiveness on SMB symptoms in children with ASD. For future research, we recommend that researchers select appropriate methods to design single-case studies that will facilitate the use of appropriate meta-analyses in future reviews.

References

References marked with an asterisk(*) indicate studies included in the meta-analysis.

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett Jr, D. R., Tudor-Locke, C., Greer, J. L., Vezina, J., Whitt-Glover, M. C., & Leon, A. S. (2011). Compendium of physical activities: A second update of codes and MET values. *Medicine and Science in Sports and Exercise*, *43*(8), 1575-1581.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Association.

*Anderson-Hanley, C., Tureck, K., & Schneiderman, R. L. (2011). Autism and exergaming: Effects on repetitive behaviors and cognition. *Psychology Research and Behavior Management, 4*, 129-137.

*Bahrami, F., Movahedi, A., Marandi, S. M., & Abedi, A. (2012). Kata techniques training consistently decreases stereotypy in children with autism spectrum disorder. *Research in Developmental Disabilities, 33*(4), 1183-1193.

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review, 16*, 161–169.

Bishop, S. L., Richler, J., & Lord, C. (2006). Restricted and repetitive behaviors and nonverbal IQ in children with autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 48*, 1111-1121.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* West Sussex, UK: John Wiley & Sons.

Boyd, B. A., McDonough, S. G., & Bodfish, J. W. (2012). Evidence-based behavioral interventions for repetitive behaviors in autism. *Journal of Autism and Developmental Disorders*, *42*(6), 1236-1248.

Bremer, E., Crozier, M., & Lloyd, M. (2016). A systematic review of the behavioural outcomes following exercise interventions for children and youth with autism spectrum disorder. *Autism*, *20*(8), 899-915.

Burns, B. T., & Ault, R. L. (2009). Exercise and Autism Symptoms: A Case Study. *Psi Chi Journal of Psychological Research, 14*(2), 43-51.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical Activity, Exercise, and Physical Fitness: Definitions and Distinctions for Health-Related Research. *Public Health Reports (1974), 100*(2), 126-131.

*Celiberti, D. A., Bobo, H. E., Kelly, K. S., Harris, S. L., & Handleman, J. S. (1997). The differential and temporal effects of antecedent exercise on the self-stimulatory behavior of a child with autism. *Research in Developmental Disabilities, 18*(2), 139- 150.

*Chan, A. S., Sze, S. L., Siu, N. Y., Lau, E. M., & Cheung, M-C. (2013). A Chinese Mind- Body Exercise Improves Self-Control of Children with Autism: A Randomized Controlled Trial. *PloS One, 8*(7), e68184.

Chasson, G. S., Harris, G. E., & Neely, W. J. (2007). Cost Comparison of Early Intensive Behavioral Intervention and Special Education for Children with Autism. *Journal of Child and Family Studies, 16*(3), 401-413.

Cheung, M. W. (2015). metaSEM: An R package for meta-analysis using structural equation modelling. *Frontiers in Psychology*, *5*, 1521.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, R. J., & Sackett, D. L. (1995). The number needed to treat: a clinically useful measure of treatment effect. *British Medical Journal, 310*, 452–454.

Cuccaro, M. L., Shao, Y., Grubber, J., Slifer, M., Wolpert, C. M., Donnelly, S. L., Abramson, R. K., Ravan, S. A., Wright, H. H., DeLong, G. R., & Pericak-Vance, M. A. (2003). Factor analysis of restricted and repetitive behaviorsin autism using the autism diagnostic interview-R. *Child Psychiatry and Human Development*, *34*(1), 3-17.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629-634.

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2010). Using participant data to extend the evidence base for intensive behavioral intervention for children with autism. *American Journal of Intellectual and Developmental Disabilities, 115*(5), 381-405.

Ferreira, J. P., Ghiarone, T., Cabral Jr., C. R., Furtado, G. E., Carvalho, H. M., Machado-Rodrigues, A. M., & Toscano, C. V. A. (2019). Effects of physical exercise on the stereotyped behavior of children with autism spectrum disorders. *Medicina, 55*, 685. doi:10.3390/medicina55100685

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63*(3), 665-694.

Furniss, F., & Biswas, A. B. (2012). Recent research on aetiology, development and phenomenology of self injurious behaviour in people with intellectual disabilities: a systematic review and implications for treatment. *Journal of Intellectual Disability Research, 56*(5), 453-475.

*Gabriels, R. L., Agnew, J. A., Holt, K. D., Shoffner, A., Pan, Z., Ruzzano, S., Clayton, G. H., & Mesibov, G. (2012). Pilot study measuring the effects of therapeutic horseback riding on school-age children and adolescents with autism spectrum disorders. *Research in Autism Spectrum Disorders, 6*(2), 578-588.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics, 6,* 107–128.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*, 39-65.

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C. & Vedel, I. (2019). Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. *Journal of Clinical Epidemiology, 111*, 49-59.

Howard, D., Best, W., & Nickels, L. (2015). Optimising the design of intervention studies: critiques and ways forward. *Aphasiology, 29*(5), 526-562.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: correcting error and bias in research findings*. Newbury Park, California: Sage.

Hutt, C., & Hutt, S. J. (1965). Effects of environmental complexity on stereotyped behaviors of children. *Animal Behavior*, *13*(1), 1-4.

*Kern, L., Koegel, R. L., & Dunlap, G. (1984). The influence of vigorous versus mild exercise on autistic stereotyped behaviors. *Journal of Autism and Developmental Disorders, 14*(1), 57-67.

*Kern, L., Koegel, R. L., Dyer, K., Blew, P. A., & Fenton, L. R. (1982). The effects of physical exercise on self-stimulation and appropriate responding in autistic children. *Journal of Autism and Developmental Disorders, 12*(4), 399-419.

Kerr, J. H., & van den Wollenberg, A. E. (1997). High and low intensity exercise and psychological mood states. *Psychology & Health, 12*(5), 603-618.

Koegel, R. L., & Covert, A. (1972). The relationship of self-stimulation to learning in autistic children. *Journal of Applied Behavior Analysis, 5*(4), 381.

Koegel, R. L., Firestone, P. B., Kramme, K. W., & Dunlap, G. (1974). Increasing spontaneous play by suppressing self-stimulation in autistic children. *Journal of Applied Behavior Analysis, 7*(4), 521.

*Koenig, K. P., Buckley-Reen, A., & Garg, S. (2012). Efficacy of the Get Ready to Learn yoga program among children with autism spectrum disorders: a pretest-posttest control group design. *The American Journal of Occupational Therapy, 66*(5), 538-546.

Kunzi, K. (2015). Improving Social Skills of Adults With Autism Spectrum Disorder Through Physical Activity, Sports, and Games: A Review of the Literature. *Adultspan Journal, 14*(2), 100-113.

Lang, R., Koegel, L. K., Ashbaugh, K., Regester, A., Ence, W., & Smith, W. (2010). Physical exercise and individuals with autism spectrum disorders: A systematic review. *Research in Autism Spectrum Disorders*, *4*(4), 565-576.

Lee, S., Odom, S. L., & Loftin, R. (2007). Social Engagement With Peers and Stereotypic Behavior of Children With Autism. *Journal of Positive Behavior Interventions, 9*(2), 67-79.

*Lee, J., Vargo, K. K., & Porretta, D. L. (2018). An Evaluation of the Effects of Antecedent Exercise Type on Stereotypic Behaviors. *Journal of Developmental and Physical Disabilities, 30*(3), 409-426.

Leekam, S. R., Prior, M. R., & Uljarevic, M. (2011). Restricted and repetitive behaviors in autism spectrum disorders: A review of research in the last decade. *Psychological Bulletin*, *137*(4), 562-593.

*Levinson, L. J., & Reid, G. (1993). The Effects of Exercise Intensity on the Stereotypic Behaviors of Individuals with Autism. *Adapted Physical Activity Quarterly*, *10*(3), 255-268.

Liu, T., Fedak, A, T., & Hamilton, M. (2016). Effect of physical activity on stereotypic behaviors of children with autism spectrum disorder. *International Journal of School Health*, *3*(1), 17-22.

Lopez, B. R., Lincoln, A. J., Ozonoff, S., & Lai, Z. (2005). Examining the Relationship between Executive Functions and Restricted, Repetitive Symptoms of Autistic Disorder. *Journal of Autism and Developmental Disorders, 35*(4), 445-460.

*Losinski, M., Cook, K., Hirsch, S., & Sanders, S. (2017). The effects of deep pressure therapies and antecedent exercise on stereotypical behaviors of students withautism spectrum disorders. *Behavioral Disorders, 42*(4), 196-208.

Maenner, M. J., Shaw, K. A., Baio, J., Washington, A., Patrick, M., DiRienzo, M., Christensen, D. L., Wiggins, L. D., Pettygrove, S., Andrews, J. G., Lopez, M., Hudson, A., Baroud, T., Schwenk, Y., White, T., Rosenberg, C. R., Lee, L.-C., Harrington, R. A, Huston, M., … Dietz, P. M. (2020). Prevalence of autism spectrum disorder among children aged 8 years – Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. *MMWR Surveillance Summaries, 69*(4), 1-12. http://dx.doi.org/10.15585/mmwr.ss6904a1

*Mays, M. N. M. (2013). *Using antecedent aerobic exercise to decrease stereotypic and self- stimulatory behavior in children with autism* (Doctoral dissertaion). Retrieved from ProQuest Dissertations Publishing. (Order No. 3583653)

*McLaughlin, C. A. H. (2010). *Decreasing stereotypy in preschoolers with autism spectrum disorder: The role of increased physical activity and function.* (Doctoral dissertation). Retrieved from ProQuest Dissertations Publishing. (Order No. 3424274)

Morley, S. (2018). Single-case methods in clinical psychology: a practical guide. New York, NY: Routledge.

*Nazemzadegan, G., Babadi, A., Zeinali, Z., & Kakavandi, K. (2016). Effectiveness of ball exercises on reduction of stereotypic behavior of children with autism spectrum disorder with high performance. *Iranian Rehabilitation Journal, 14*(2), 121-126.

*Neely, L., Rispoli, M., Gerow, S., & Ninci, J. (2015). Effects of antecedent exercise on academic engagement and stereotypy during instruction. *Behavior Modification, 39*(1), 98-116.

Olin, S. S., McFadden, B. A., Golem, D. L., Pellegrino, J. K., Walker, A. J., Sanders, D. J., & Arent, S. M. (2017). The Effects of Exercise Dose on Stereotypical Behavior in Children with Autism. *Medicine and Science in Sports and Exercise, 49*(5), 983-990.

*Oriel, K. N., George, C. L., Peckus, R., & Semon, A. (2011). The effects of aerobic exercise on academic engagement in young children with autism spectrum disorder. *Paediatric Physical Therapy, 23*(2), 187-193.

Petrus, C., Adamson, S. R., Block, L., Einarson, S. J., Sharifnejad, M., & Harris, S. R. (2008). Effects of exercise interventions on stereotypic behaviors in children with autism spectrum disorder. *Physiotherapy Canada*, *60*(2), 134-145.

Pluye, P., Gagnon, M. P., Griffiths, F., & Johnson-Lafleur, J. (2009). A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *International Journal of Nursing Studies, 46*(4), 529-546.

*Powers, S., Thibadeau, S., & Rose, K. (1992). Antecedent exercise and its effects on self- stimulation. *Behavioral Interventions, 7*(1), 15-22.

*Prupas, A., & Reid, G. (2001). Effects of exercise frequency on stereotypic behaviors of children with developmental disabilities. *Education and Training in Mental Retardation and Developmental Disabilities*, *36*(2), 196-206.

Pustejovsky, J . E. (2016). scdhlm: A web-based calculator for between-case standardized mean differences (Version 0.2) [Web application]. Retrieved from: https://jepusto.shinyapps.io/scdhlm

R Development Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rapp, J. T., & Vollmer, T. R. (2005). Stereotypy I: A review of behavioral assessment and treatment. *Research in Developmental Disabilities*, *26*, 527-547.

Rapp, J. T., Vollmer, T. R., Peter, C., Dozier, C. L., & Cotnoir, N. M. (2004). Analysis of response allocation in individuals with multiple forms of stereotyped behavior. *Journal of Applied Behavior Analysis, 37*(4), 481-501.

Richler, J., Huerta, M., Bishop, S. L., & Lord, C. (2010). Developmental trajectories of restricted and repetitive behaviors and interests in children with autism spectrum disorders. *Development and Psychopathology*, *22*, 55-69.

*Richmond, L. M. (2000). *The effect of physical activity on stereotypical behavior of autistic children* (Doctoral dissertation). Retrieved from ProQuest Dissertations Publishing. (Order No. 9965850)

*Rosenblatt, L. E., Gorantla, S., Torres, J. A., Yarmush, R. S., Rao, S., Park, E. R., Denninger, J. W., Benson, H., Fricchione, G. L., Bernstein, B., & Levine, J. B. (2011). Relaxation response-based yoga improves functioning in young children with autism: a pilot study. *Journal of alternative and complementary medicine (New York, N.Y.), 17*(11), 1029-1035.

*Rosenthal-Malek, A., & Mitchell, S. (1997). Brief report: the effects of exercise on the self- stimulatory behaviors and positive responding of adolescents with autism. *Journal of Autism and Developmental Disorders, 27*(2), 193-202.

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: a primer and applications. *Journal of School Psychology, 52*(2), 123-147.

Sowa, M., & Meulenbroek, R. (2012). Effects of physical exercise on autism spectrum disorders: A meta-analysis. *Research in Autism Spectrum Disorders*, *6*(1), 46-57.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Clinical research ed.)*, *343*, d4002.

Tarr, C. W., Rineer-Hershey, A., & Larwin, K. (2020). The effects of physical exercise on stereotypic behaviours in autism: small-n meta-analyses. *Focus on Autism and Other Developmental Disabilities, 35*(1), 26-35.

Thorlund, K., Walter, S. D., Johnston, B. C., Furukawa, T. A., & Guyatt, G. H. (2011). Pooling health-related quality of life outcomes in meta-analysis—a tutorial and review of methods for enhancing interpretability. *Research Synthesis Methods, 2,* 188-203.

Tsukamoto, H., Suga, T., Takenaka, S., Tanaka, D., Takeuchi, T., Hamaoka, T., Isaka, T., & Hashimoto, T. (2015). Greater impact of acute high-intensity interval exercise on post-exercise executive function compared to moderate-intensity continuous exercise. *Physiology and Behavior, 155*, 224-230.

Turner, M. A. (1997). Towards an executive dysfunction account of repetitive behavior in autism. In J. Russell. (Ed.), *Autism as an executive disorder* (pp. 57-100). Oxford, England: Oxford University Press.

Turner, M. A. (1999). Annotation: Repetitive behavior in autism: A review of psychological research. *Journal of Child Psychology and Psychiatry*, *40*(6), 839-849.

Valentine, J.C., Tanner-Smith, E.E., Pustejovsky, J.E. and Lau, T.S. (2016). Between-case standardized mean difference effect sizes for single-case designs: a primer and tutorial using the scdhlm web application. *Campbell Systematic Reviews, 12*, 1-31.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48.

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*, 112-125.

*Watters, R. G., & Watters, W. E. (1980). Decreasing self-stimulatory behavior with physical exercise in a group of autistic boys. *Journal of Autism and Developmental Disorders, 10*(4), 379-387.

Zentall, S. S., & Zentall, T. R. (1983). Optimal stimulation: A model of disordered activity and performance in normal and deviant children. *Psychological Bulletin*, *94*, 446-471.