Journal of Autism and Developmental Disorders
https://doi.org/10.1007/s10803-019-04050-9

ORIGINALPAPER

Efficacy of Group-Based Organised Physical Activity Participation for Social Outcomes in Children with Autism Spectrum Disorder: A Systematic Review and Meta-analysis

Katherine Howells[1] · Carmel Sivaratnam[1] · Tamara May[1] · Ebony Lindor[1] · Jane McGillivray[1] · Nicole Rinehart[1,2]

Abstract
The aim of this novel review and meta-analysis was to clarify the effects of group-based organised physical activity (OPA) for social and communicative outcomes in children with Autism Spectrum Disorder (ASD). Searches yielded 4347 articles. Eleven were identified for review and seven for meta-analysis. Pooled statistical results revealed a non-significant effect for communication ($k=4$; $g=0.13$, CI [−0.12, 0.38], $p=.13$) and a significant small-medium improvement in overall social functioning ($k=6$; $g=0.45$, CI [0.19, 0.72], $p=.001$). Despite acknowledged limitations, these findings are important in the context of a growing clinical and consumer-driven demand for research that determines the role of OPA as a non-medical and inclusive treatment for children with ASD.

Keywords Group-based · Organised physical activity · ASD · Autism spectrum disorder · Social functioning · Meta analysis

Autism Spectrum Disorder (ASD) is a complex condition, with a highly heterogeneous phenotypical and biological expression (Pelphrey et al. 2011; Toal et al. 2010). The disorder is characterised by marked social and communication impairments and stereotyped patterns of behaviours, interests, and activities (American Psychiatric Association 2013). In addition to the core social and behavioural features that characterise ASD, individuals also commonly exhibit motor impairments [e.g. fine and gross motor impairments Ament et al. (2015), Bhat et al. (2011) respectively and an oddness of gait Rinehart et al. (2006)].

Several papers have linked the severity of one's motor difficulties to the severity of social impairments (MacDonald et al. 2013; Zeliadt 2017). Given this established relationship, researchers have become increasingly interested in understanding whether motor interventions have developmental benefits to social functioning and communication in ASD populations (Rinehart et al. 2018). Meta-analyses in ASD child populations have revealed that a range of physical activity programs may benefit motor functioning (Sowa and Meulenbroek 2012; Yang et al. 2015). Moreover, Bremer et al.'s (2016) review and Healy et al.'s (2018) meta-analysis showed benefits to social functioning, suggesting that physical activity may facilitate peer, sibling and instructor interactions in children with ASD (Healy et al. 2018). Systematic reviews have also indicated improvements in communication, a related area of social functioning (Alhowikan 2016); however, previous meta-analyses have been unable to be conducted on this domain due to insufficient results (Sowa and Meulenbroek 2012). This indicates an updated analysis of research may be required to further clarify physical activities effects on this domain.

Despite the known physical, and potential developmental benefits of physical activity, research shows that children with neurodevelopmental disorders such as ASD engage in lower levels of physical activity than typically-developing children, which could contribute to medical sequelae such as obesity and chronic illness (McCoy et al. 2016; Rimmer and Braddock 2002). Given individuals with ASD are at significantly higher risk than the typically developing population for co-occurring psychiatric disorders (Reaven and Wainer

✉ Katherine Howells
k.howells@research.deakin.edu.au

[1] School of Psychology, Faculty of Health, Deakin Child Study Centre, Deakin University, Geelong, VIC, Australia

[2] School of Psychology, Deakin Child Study Centre, Deakin University, 221 Burwood Highway, Burwood, VIC 3125, Australia

Published online: 17 May 2019

2015) which are often complex to treat, additional physical disorder risk factors only compound the life challenges faced by this population. However, unlike complex psychiatric comorbidity, physical health risks are arguably more preventable and manageable with simple lifestyle interventions. Evidence from current lifestyle programs for adults with developmental disabilities indicate such interventions not only influenced weight loss success but improved overall lifestyles (Bazzano et al. 2009). Furthermore, the positive link between physical activities occurring in a social context (i.e. group-based physical activity), may offer an additional protective factor against psychiatric disorders. It is, therefore, unsurprising that lifestyle-based interventions running in conjunction with clinic-based interventions are receiving a growing amount of traction (Rinehart et al. 2018).

Consistent with this new focus on lifestyle-based interventions sitting alongside clinical interventions, Pellicano et al. (2014) reported that the UK autism community prioritised research focusing on programs for learning and enhancing life skills (Pellicano et al. 2014). Indeed, one of the major limitations of clinic-based interventions targeting social skill development is the generalisability of learning and results outside that setting (Leichsenring 2004). Although there are many meta-analyses examining randomised control trials (RCTs) of clinical interventions (e.g., behavioural therapy programs) for children with ASD (Reichow et al. 2018), supplementary community-based research is also necessary. Such research often involves interventions that occur within a child's natural setting, a setting which is said to provide the most effective learning opportunities (Bruder 2001). An example of such an intervention is group-based (involving two or more people in the activity) organised physical activity (OPA). OPA has been defined by Okely (1999) as physical activities involving formal and structured training sessions that are supervised by a coach/adult. This definition is consistent European and US definitions (Edwardson and Gorely 2010; Dunton et al. 2012 respectively). The definition used in the current study will also incorporate the idea that these activities must have been organised by a club (sporting body) or recreational association (Australian Bureau of Statistics 2013). A limitation of many existing reviews is their inclusion of general physical activity programs. Thus, they often include single-case designs involving, for example, an individual-based jogging activity (Lang et al. 2010), where children are not provided with opportunities to interact with peers. A focus on group-based OPA would offer a deeper insight into the effects of physical activity for social functioning.

Interestingly, results from a meta-analysis of 16 studies were more supportive of 'individual-based' exercise programs over 'group-based' for social functioning in individuals with ASD (aged 4–41.3 years) (Sowa and Meulenbroek 2012). This finding is surprising given the evidence suggesting group-based physical activity programs have the potential to increase interpersonal and social skill development (Alexander and Leather 2013; Ohrberg 2013). At the time of publication, Sowa and Meulenbroek (2012) noted that no publications examining naturalistic group-based programs were available. Given this, it is imperative more recent research on group-based programs and social functioning is explored and synthesised. To the best of our knowledge, this is the first systematic review and meta-analysis to examine the effects of group-based physical activity within an organised context specifically.

Another limitation of many existing systematic reviews and meta-analyses, as acknowledged in Bremer et al. (2016)'s review, is that they include a wide range of developmental stages (e.g., participants aged 3–25 years or 4–41.3 years) and settings (e.g., children in early education through to adults in the workforce) (Tan et al. 2016). This consequently limits the generalisability of results (Bremer et al. 2016). The current review focuses on a smaller age bracket, accepting studies with a mean age within a 5–12-year bracket (school aged children).

To address these gaps in the literature, the current review aims to investigate, systematically review, and quantitatively synthesise the literature regarding the effects of group-based OPA participation on children with ASD. A focus is placed on social functioning and communication in a narrower age band of children than has been explored in many previous reviews.

Methods

This review was registered with the International Prospective Register of Systematic Reviews (PROSPERO) (CRD42017059389) and has adhered to guidelines outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al. 2009).

Eligibility Criteria

The inclusion/exclusion of studies was based on adherence to the following criteria.

Participants

All participants of a given study were required to have a diagnosis of Pervasive Developmental Disorder (PDD) as per the ICD-10 (World Health Organisation 1993). Diagnoses of ASD as per the DSM-5 (American Psychiatric Association 2013) were also included, as well as diagnoses of Asperger's, Rett Syndrome, or PDD-Not Otherwise Specified (PDD-NOS) as per the DSM-IV (American Psychiatric Association 1994). These variations of ASD were included

to ensure for a broad search under the DSM-IV criteria, however studies consisting solely of Rett diagnoses were later excluded during screening. Studies were included if the mean age of participants was between 5 and 12.9 years, even if the full age range exceeded this bracket. Studies were also required to have a sample size of at least five.

Intervention

The review included studies that examined the effectiveness of participation in a group-based OPA intervention program. As mentioned, OPA will be defined as formal and structured training sessions that were supervised by a paid or volunteer coach/adult and were organised by a club or recreational association (inclusive of sporting bodies, community, school or educational organisations). As mentioned, this definition incorporates elements from several existing definitions (Dunton et al. 2012; Edwardson and Gorely 2010; Okely 1999; Australian Bureau of Statistics 2013). Studies were not accepted if the intervention was conducted on a 1:1 basis. The programs needed to be delivered in small groups for at least one session lasting at least 30-min a week. This time frame was chosen as a desktop review of recommended 'structured' or 'adult-let' physical activity supports 30 min as a minimum inclusion criterion (Gavin 2014), and may often be what children with disabilities are exposed to in community settings. This aligns with previous research which suggests a group-based, organised exercise program consisting of weekly 30 min sessions is feasible for children with intellectual disabilities (Bellamy 2015).

Comparison

Studies with and without a comparison group were included in qualitative assessments; however, only those studies with a comparison group were assessed quantitatively. If a comparison group was used, the group must not have been exposed to the OPA program. Studies with a comparison group who received an alternative physical activity or social skills intervention were excluded except if exposed to other non-physical or everyday activities (e.g. educational programs).

Outcome

Study eligibility was dependent on two outcome measures: social functioning and communication. Social outcomes included areas such as social communication, interpersonal relationships, interactions with others, social motivation and awareness and communication (e.g. language ability). Empathy was excluded as it incorporates areas which lie outside social functioning; for example, emotion recognition and cognitive functioning (Khanjani et al. 2015). The meta-analytic component of this review only included studies with a comparison group. Two domains were separately analysed; social functioning, or social responsiveness, social motivation, socialization or social interaction skills and communication (e.g. language ability). Any measure of social functioning and communication were accepted for the meta-analysis i.e. total Social Responsiveness Scale (SRS) scores and Vineland Adaptive Behaviour Scale (VABS) communication and socialisation sub domain scores. If the total SRS score was not available, the social motivation subdomain score was used as this sub-test score most closely related to the key outcome measure of this study.

Study Design

The current review included RCTs, non-RCTs and cohort designs that used statistical procedures (not limited to significance testing) to evaluate effectiveness. Cross-sectional and single subject designs, case series and reports, qualitative studies, perspective articles, meta-analyses, and systematic reviews were excluded.

Search Strategy

The initial literature search was conducted using the following five databases (last search: 1st of March 2018): MEDLINE Complete via EBSCOhost, PsycINFO via EBSCOhost, CINAHL Complete via EBSCOhost, and SportDiscus with Full Text via EBSCOhost, and Embase via Embase (see Table 1 for search terms for MEDLINE Complete and accompanying MeSH headings).

Study Selection

The following search procedure was implemented to identify potential studies. Firstly, two independent reviewers screened studies via titles and abstracts. Once screened, any disagreements (< 1%) were discussed, and a consensus was achieved on each article. No time restrictions were placed, and the text had to be available in English. Age restrictions were placed on all databases where age limiters were available. One reviewer then completed full-text screening. Included texts were discussed with a second reviewer and, again, a consensus was achieved. A secondary search was conducted using the reference list from each included article. Two independent reviewers screened these citations via titles and abstracts. Included articles were then discussed and a consensus was achieved on articles eligible for inclusion.

Risk of Bias Assessment

The risk of bias in each included study was assessed by two independent reviewers, using an adapted version of

Table 1 Search strategy used in MEDLINE complete

1. Organised	"extra-curricular" OR Communit* Club* OR Structure* OR Intervention* "community-based"
2. Physical activity	Sport* OR Exercis* OR Train* OR Program* OR "Locomotor Activit*" OR "Physical* therap*" OR "physical* activit*" OR "motor activit*" OR Soccer OR Swim* OR Aquatic* OR Dive OR Diving OR Football OR Basketball OR Cricket OR Tennis OR Rugby OR Danc* OR Athletic* OR "Martial art*" OR Netball OR Hockey OR Gym* OR "horse rid* OR "horseback rid*" OR Equestrian OR Baseball OR Judo OR Cycling OR Surf* NOT "video game*" Mesh headings: "Youth Sports" OR "Sports for Persons with Disabilities" OR "Sports+" OR Exercise + OR Dancing
3. Autism Spectrum Disorder	ASD OR Autis* OR "Autism Spectrum Disorder" OR Asperger* OR "Asperger* syndrome" OR "Pervasive developmental disorder*" OR PDD OR PDDNOS OR "Rett Syndrome" OR "Rett* disorder" Mesh headings: "Autistic Disorder" OR "Asperger Syndrome" OR "Autism Spectrum Disorder" OR "Rett Syndrome"
4. Population	Child* OR "School-age*" OR "young child*" OR Kid* OR Boy* OR Girl* OR Male* OR Female* OR "Primary-age*" OR "5–12 year*" Mesh heading: Child+
5. 1 AND 2 AND 3 AND 4	

Limit 8 to: age group all child (0–18 years)

the "Quality in Prognostic Studies (QUIPS) Tool" (Hayden et al. 2013). This adapted tool has been used in reviews with similar baseline and follow-up measurements (Brignell et al. 2017). The QUIPS tool was modified as we were not analysing confounders or prognostic factors and these assessment categories included in the original tool did not apply. This adapted tool assessed the following areas: study participation, study attrition, outcome measurement. The timing of diagnosis (e.g., at the end, during or at baseline of study) was also included for the current review. This tool allowed reviewers to make qualitative assessments around the risk of bias in each study based on information available in the article, and hence to the public. Studies were considered to have a low risk of bias if most of the criteria in each section were met. If a component was not described, it was rated as 'unclear' and assessed as a high risk of bias across the areas. Overall risk of bias was determined by taking the most frequently occurring rating. If an equal number was achieved (e.g., two areas had a medium risk of bias, and another two had a low risk), the higher of the two ratings was used. Following independent ratings, the two reviewers discussed any discrepancies and a consensus was achieved on all items.

Data Extraction

One independent reviewer performed data extraction. Extracted data included: (1) the study design, (2) sample age range, (3) number of participants in each group, (4) diagnosis, severity, and comorbidities, (5) the type of OPA program assessed, (6) additional medication or alternative therapies used, (7) measures used to assess social outcomes, and (8) results. For quantitative assessments, both means, and standard deviations were extracted so effect sizes (as measured through Hedges *g*) could be computed. Hedges *g* provides a more accurate measure of effect size (in comparison to Cohens *d*) in sample sizes below 20 (Lakens 2013), and was used given several studies reported smaller sample sizes < 20. All conversions of means and standard deviations to effect sizes and meta-analyses were computed using Comprehensive Meta-Analysis Software Version 3.0.

Synthesis of Results

To quantify the efficacy of group-based OPA participation on social outcomes for children with ASD, effect sizes of all included studies were pooled, and Hedges *g* was calculated for the two outcomes separately: social functioning and communication. Given the nature of group-based OPA, intervention programs are often heterogeneous. In cases of heterogeneity between studies, Borenstein et al. (2011) suggest applying a random effects model, which allows the true effect size to vary across studies (Borenstein et al. 2011). Hence, each effect size (Hedges *g*) was calculated using a random-effects model. An alpha level of .05 was used to determine statistical significance. A *Q* statistic and its accompanying significance levels, tau-squared (τ^2), and an *I*-squared (I^2) statistic were computed to determine levels of heterogeneity between studies. The *Q* statistic provides a ratio of *observed* variation to the error within studies. A significant *Q* is indicative of varying true study effects (Borenstein et al. 2011). Tau-squared indicates the *variance* of the true effect sizes and the I^2 statistic indicates the *proportion* (as expressed through a percentage) of variance remaining if sampling error is removed (Borenstein et al. 2017). Values of 25, 50, and 75% could be considered low, moderate and high respectively (Higgins et al. 2003).

Ideally, subgroup analyses or a meta-regression would be performed to examine the impact of study and sample characteristics on effect sizes; however, given the small *k* in each of the meta-analyses, this was not performed.

Outliers and Publication Bias

Outliers and Sensitivity Analyses

The presence of outliers was determined via a z-score approach. Using this approach, z-scores $> \pm 1.96$ are considered 'potential' outliers, scores $> \pm 2.58$ are 'probable' outliers and scores $> \pm 3.29$ are considered to be 'extreme' outliers (Field 2013). 'Potential' outliers were present and subsequent sensitivity analyses were conducted. Sensitivity analyses were conducted using a 'one study removed' approach which allowed us to observe the impact of outlier retention or removal. Secondly, any outliers were also removed, and analyses were re-run to determine if the outlier was driving the data. Outliers were retained if the influence was negligible (e.g. no change in the interpretation of the *p*-value).

Publication Bias

Publication bias was assessed via several methods. Firstly, a visual inspection of funnel plots, which depict the relationship between study size and effect size (Borenstein et al. 2011) was performed. Secondly, Egger's test was conducted. This test provides a statistical procedure for assessing asymmetry in funnel plots (Sedgwick 2013). A 'classic fail-safe N' was also computed to determine the number of missing studies required to make the effect size statistically insignificant (Rosenthal 1979). If publication bias was present, Duval and Tweedie's trim and fill procedure was used. This procedure adjusts effect sizes and confidence intervals to account for missing studies (Duval and Tweedie 2000).

Results

Qualitative Analysis

The combined database and key article reference searches retrieved 4347 results (see Fig. 1). There were 1159 duplicates, which were identified and removed, leaving 3188 articles available for title and abstract screening. From here, 3160 articles were excluded for not meeting inclusion criteria. Twenty-eight articles were screened for eligibility at the full-text stage. Seventeen articles were then excluded for the following reasons: the physical activity program was not organised ($k = 1$), the activity was not community-based ($k = 3$), the intervention was confounded by non-physical activities ($k = 2$), the comparison group failed to meet the inclusion criteria (i.e. received some form of physical training; $k = 2$), social outcomes were not measured ($k = 7$), or the study design failed to meet the inclusion criteria (i.e. a report) ($k = 2$). Eleven articles met the inclusion criteria.

Seven articles were included in quantitative assessments. One study from the qualitative component was removed as no statistical significance testing was completed (frequencies only) and another three as no comparison group was used. Five reported sufficient data for meta-analysis. The corresponding authors of the remaining papers were contacted; two authors provided the necessary data. One paper included sufficient data for the socialisation analysis; however, there was insufficient data pertaining to communication. This author was contacted, and a follow-up email was sent 2 weeks following the initial email. After a further 2 weeks, this study was excluded from quantitative communication assessments. Six studies included data suitable for a meta-analysis of social functioning. Three studies examining socialisation from the qualitative review did not have data suitable for the meta-analysis; one showing significant improvements in overall social skills (Guest et al. 2017) and one with non-significant results (García-Gómez et al. 2014). The third did not conduct significance testing (Hayward et al. 2016). The total score of one measure was not available from one study (Gabriels et al. 2015) however sufficient data was available for subdomains. As specified in the methodology, the subdomain of social motivation was substituted into analysis. Four studies included data suitable for meta-analysis of communication. Two studies from the qualitative review did not have suitable communication data (Anderson and Meints 2016; Guest et al. 2017), both of which found non-significant results.

Study Characteristics

Table 2 presents a summary of study designs, sample characteristics, participants, types of group-based OPA interventions (inclusive of activities, duration, and frequency), comparison groups, and outcomes of reviewed studies.

Three hundred and seventy-nine participants were included across the 11 studies (*n* range 13–116). Three hundred and forty-eight of those participants were included in the quantitative analyses. The included age range was 3–16 years (mean age range 5.33–10 years). Each participant had a formal diagnosis of ASD. Nine studies reported on levels of functioning (i.e. intelligence quotient [IQ]). One sample consisted of participants with limited verbal abilities; this study failed to find significant improvements (Anderson and Meints 2016). Five studies had a mixed sample of children with average IQ's (i.e. no intellectual disability) and children who were non-verbal. These studies found either solely significant results (Bahrami et al. 2016; Movahedi et al. 2013; Gabriels et al. 2012), or a mixture of significant and non-significant results (Bass et al. 2009; García-Gómez et al. 2014). Two studies had samples with no intellectual disability (Gabriels et al. 2015; Borgi et al. 2016) (IQ's > 70 and ≥ 80 respectively); both studies also found a mix of

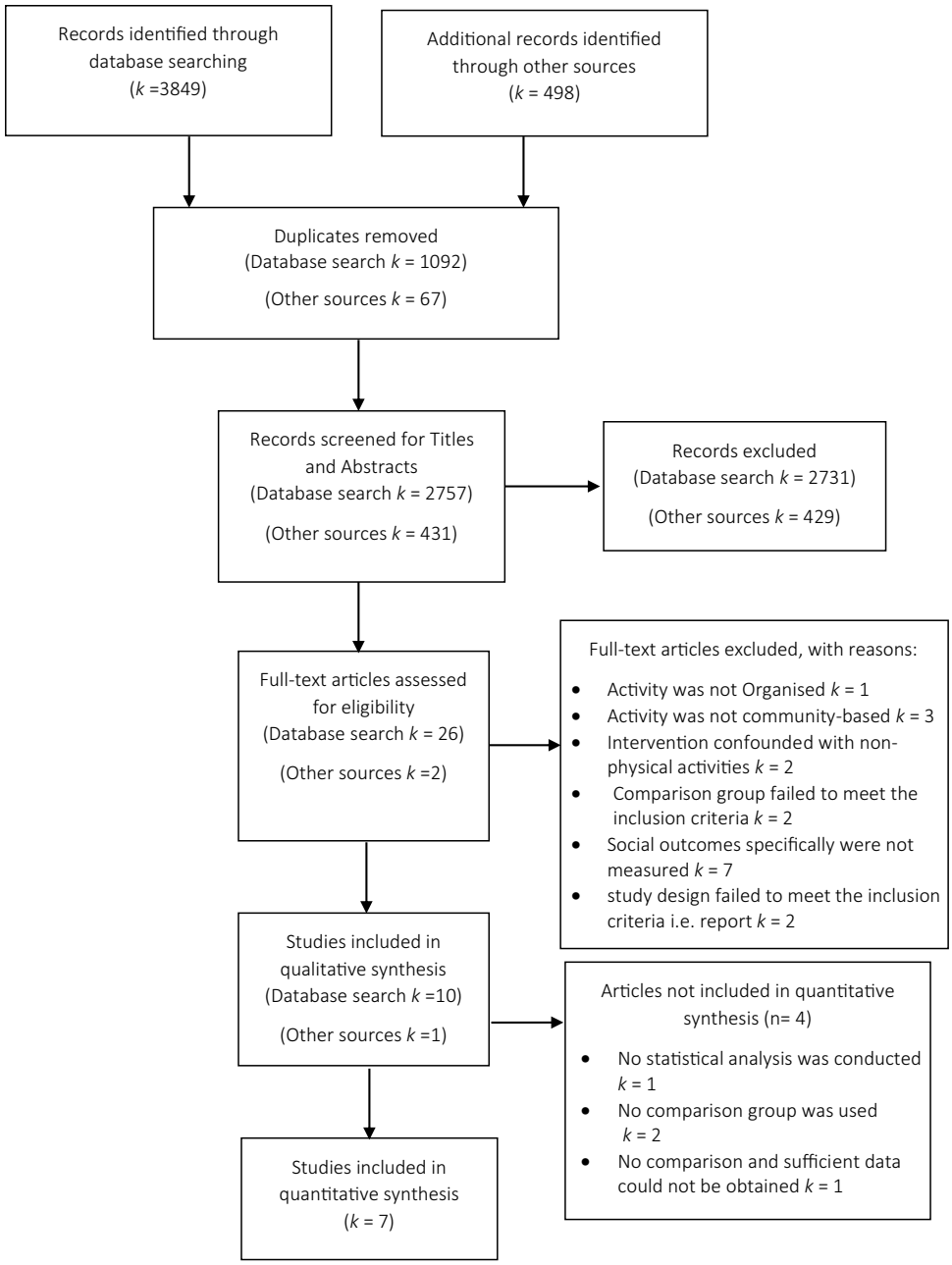


Fig. 1 PRISMA flow diagram of search results

significant and non-significant results. One study's mean cognitive scores for the intervention and control groups were 84.8 and 86.4, respectively; this study also found a mix of significant and non-significant results (Zachor et al. 2017).

The included intervention programs were as follows: horse riding ($k=6$ studies), kata techniques ($k=2$), soccer skills program ($k=1$), multi-sport camp program ($k=1$) and an outdoor adventure camp program ($k=1$). Interventions varied in both duration (from one week to 25 weeks) and frequency (from one session per week to a 5-day camp); thus, total dosage varied from 6.5 to 56 h. Two studies completed follow-up assessments (Bahrami et al. 2016; Movahedi et al. 2013). Each program evaluated some aspect of social functioning; for example, communication or socialisation. Of the 11 included studies, one (Hayward et al. 2016) did not apply a statistical procedure for assessing pre-and-post program differences.

Three of the 11 included studies did not employ a comparison group (Hayward et al. 2016; Guest et al. 2017; Anderson and Meints 2016). Of the eight studies that did, two used a waitlist comparison group (Bass et al. 2009; Borgi et al. 2016). One comparison group undertook barn activities where children learned horsemanship skills (Gabriels et al. 2015), and another two comparison groups were taught educational skills (Bahrami et al. 2016; Movahedi et al. 2013). The remaining studies utilised comparison

Table 2 Results summary table

Authors (year)	Study design	Sample age range	Number of participants	Diagnosis and severity	OPA program
Hayward et al. (2016)	Descriptive	M = 9.66 years; SD = 2 (Male = 13; female = 5)	18 (no comparison group) 14 in analyses	ASD (diagnosis by pediatrician via DSM-5) Severity = level 1 (N = 3); Level 2 = (N = 11); Level 3 = (N = 4)	Adapted 'TOPSoccer' Program run by: Town Soccer Organization
Gabriels et al. (2012)	Controlled Trial	6–16 years (M = 8.7) (Male = 36; female = 6)	58 (E = 42 [40 in analyses]; Waitlist comparison = 16)	Autistic or Asperger's Disorder (diagnosis by community psychologist via DSM-IV, confirmed by 1st author)	Therapeutic Horseback Riding (THR) Program run at: Premier PATH International Accredited Riding Centre
Borgi et al. (2016)	RCT	6–12-year-old males (M = 8.6; SD = 1.7)	28 (E = 15; C = 13) 26 in analyses	ASD (diagnosis confirmed by DSM-IV-TR and neuropsychiatrists) Severity: all verbal with IQ's > 70 (on WISC-III)	Equine-assisted therapy (EAT) Program run at: Italian Equestrian Federation Accredited Riding Centres
Anderson and Meints (2016)	Experimental	5–16 years (M = 10; SD = 3.8) (Males = 11; females = 4)	15 (no comparison group)	ASD (clinical diagnosis by health professional) 20% had comorbid ADHD, 53% had comorbid Sensory Integration Disorder. Sample had limited verbal ability.	Equine-Assisted Activities (EAA) Program run at: 'The Stables Riding for the Disabled Activities Centre'
García-Gómez et al. (2014)	Quasi-control	7–14 years (Males = 13; females = 3)	16 (E = 8; C = 8)	ASD (no mention of diagnosis source) Severity: CARS scores indicated a mean of 31 (mild Autism). All IQ's > 50	Adapted Therapeutic Horse-riding Program run at: Monfragüe Equestrian Centre of Cáceres
Guest et al. (2017)	Quasi-experimental	8–11-year-old females (M = 9.76 ± 1.00)	13 (no comparison group)	ASD (diagnoses by physicians via DSM-5: two diagnosed with PDD-NOS under DSM-IV)	Multi-sport summer camp Program run at: local school gymnasium, with trained camp staff present.
Bass et al. (2009)	Controlled Trail	4–10 years (E: M = 6.95; SD = 1.67, C: M = 7.73; SD = 1.65) (Males = 29; females = 5)	34 (E = 19; C = 15)	ASD (met criteria under the DSM-IV-TR) Severity: Verbal (N = 15), Non-verbal (N = 19), Asperger's Disorder (N = 2), Mild ASD (N = 11), Moderate ASD (N = 16), Severe ASD (N = 5)	Therapeutic Horseback Riding Program run at: Good Hope Equestrian Training Centre
Zachor et al. (2017)	Controlled Trial	3–7 years (E:M = 5.6; SD = 0.9, C:M = 5.0; SD = 1.0) (Males = 40; females = 11)	51 (E = 30; C = 21)	ASD (diagnosis based on medical and psychological evaluations, all met DSM-IV-TR criteria) Severity: Mild-moderate based on SRS scores.	Outdoor Adventure Program Program run by: Special Education Kindergartens

Table 2 (continued)

Authors (year)	Study design	Sample age range	Number of participants	Diagnosis and severity	OPA program
Gabriels et al. (2015)	RCT	6–16 years	116 (E = 58; C = 58) 97 completed post assessments	ASD (diagnosis confirmed by meeting clinical cut-offs on the Autism Diagnostic Observation Schedule (ADOS) Severity: IQ's > 80	Therapeutic Horseback Riding (THR) Program run by: Accredited PATH international Instructor at a riding centre
Movahedi et al. (2013)	RCT	5–16 years (M = 9.13; SD = 3.27)	30 (E = 15; C = 15) 26 included in analyses	ASD (formal diagnosis under the DSM-IV-TR) Severity: This was assessed using the Gillam Autism Rating scale-second edition) (GARS-2)—average for E = 46.31; C = 44.77	Heian Shodan Kata (kata incorporates techniques from various schools of martial arts) Program run by: 20 certified Karate trainers
Bahrami et al. (2016)	RCT	5–16 years (M = 9.13; SD = 3.27) (Males = 26; females = 4)	30 (E = 15; C = 15) 8 could not complete the communication subscale	ASD (formal diagnosis under the DSM-IV-TR) Severity: This was assessed using the Gillam Autism Rating scale-second edition) (GARS-2)—average for E = 42.53; C = 47.27	Heian Shodan Kata (kata incorporates techniques from various schools of martial arts) Program run by: certified Karate trainers

Authors (year)	Program activities, duration and frequency	Medication/alternative therapies	Measures	Results
Hayward et al. (2016)	Completed soccer drills (soccer bowling, kicking into goals, 15-yard dribbling, and tunnel soccer 30-yard dash and dribbling). Duration: 90 min a week over a 6-week period. Total dosage: 9 h	Not mentioned	Parent survey designed for the program	No significance testing conducted; parent reports (N = 14) indicated 50% believed their child's relationship with siblings improved; 14.3% of parents believed it was much improved. 35.7% noted no change. Total of 64.3% reported above improved changes. Results follow up: No
Gabriels et al. (2012)	Sessions consisted of putting a helmet on, sit and wait on the bench, mount horse, THR activities, dismount horse, groom horse, put away equipment. The riding component consisted of mounting, warm up, review of skills from previous week, teaching new skill, a game or activity to practice the new skill, and cool down then dismount. Comparison group continued normal activities and did not have contact with study personnel/horses Duration: 10 weeks, 1 h a week Total dosage: 10 h	33% of sample used psychoactive medications	VABS-II Interview Edition; Survey Form (parent/caregiver report)	Significant differences were found comparing baseline and post-THR on the communication raw score ($p = 0.035$); and socialization raw score ($p = .016$). For waitlist comparison, the communication subdomain was non-significant ($p = 0.090$). No significant difference was found in any other measures for comparisons. Results follow up: No

Table 2 (continued)

Authors (year)	Program activities, duration and frequency	Medication/alternative therapies	Measures	Results
Borgi et al. (2016)	Sessions involved 20 min of grooming, 10 min of hand walking the horse, 20–30 min of horseback riding, and final phase on ground (session closure) for 10 min. Visual aids representing grooming tools and other areas were used. Duration: 25 sessions once a week over 6 months (sessions were 60–70 min each). Total dosage: 25 h	Not Mentioned	VABS (parent/caregiver report)	An ANOVA produced a significant group x time interaction on socialization subdomain (significantly increased) scores of the VABS (Vsoc t6-t0, mean ± SE, EAT: 0.72 ± 0.22. CG: 0.23 ± 0.21; $F(1,18) = 5.30$, $p = 0.034$, Tukeys test $p > 0.01$). There was no significant difference in the comparison group. No significant time × group interaction was found for the communication scale of the VABS. Results follow up: No
Anderson and Meints (2016)	The first week involved individual sessions to get a sense of the child's abilities. Following the first week 3-h weekly sessions were conducted in 3 groups of 5. The sessions were split into three sections: 1) tacking up, going to the arena, riding, dismounting, take horse back and untack. 2) Horsemanship (brushing and grooming) and 3) Stable management (feeding and cleaning). Duration: 5 weeks, 3 h per week. Total dosage: 15 h	Not mentioned	VABS (completed with the parent/caregiver by the researcher)	A t test was conducted on communication and socialization subscales of the VABS: There was no significant difference in communication scores; $t(14) = -0.792$, $p = 0.442$), and no significant difference in socialisation scores; $t(14) = 0.564$, $p = 0.582$). Results follow up: No
García-Gómez et al. (2014)	Sessions were conducted in groups of four, with three phases in sessions: (1) Preparing the equipment and horse, (2) Mounting and riding, (3) Dismounting and putting about horse and equipment away. Duration: 3-month intervention program, twice-weekly sessions (45 min each). Total dosage: 18 h	Not mentioned	BASC-T (for teachers) test battery	A Wilcoxon test found no significant difference ($p = .500$) in social skills in the experimental condition and the comparison condition ($p = .066$) using the BASC-T. Results follow up: No

Table 2 (continued)

Authors (year)	Program activities, duration and frequency	Medication/alternative therapies	Measures	Results
Guest et al. (2017)	The program involved active group games. Skills taught included: locomotor skills and object comparison. The skills progressed in difficulty throughout the week and were then implemented into translational sport settings (track and field, basketball, soccer and baseball). Duration: 5 full days. Total dosage: total hours unclear	Not mentioned	SISS and VABS (parent/caregiver reports)	Significant improvements were found among the overall social domain scores ($p = 0.005$) and interpersonal adaptive level ($p = 0.005$) between pre-and-post testing. No significant difference was found among the play and leisure time variable of the social skills domain. Significance levels for the VABS subdomains could not be obtained. Results follow up: On motor, not social functioning.
Bass et al. (2009)	Sessions included mounting/dismounting, warm up exercises (10 min), riding skills designed to help sensory skills, fine-and-gross motor skills) (15 min), mounted games (20 min), and then horsemanship skills (i.e. grooming) to finish off. Duration: one hour weekly over 12 weeks. Total dosage: 12 h	Routine therapy programs	SRS	Significant group x time interaction was present in the SRS overall scores F (1, 20) = 4.92, $p = .038$, $n^2 = .20$. Follow-up results (simple effects paired sample t test) found experimental group means significantly changed t (10) = 2.87, $p = .017$, $d = .66$, whereas means for comparison group remained unchanged t (10) = .108, $p = .916$, $d = .02$. Subscale specific results indicated a significant interaction effect in the social motivation scale F (1, 25) = 4.80, $p = .038$, $n^2 = 161$. No significant interactions were found in the social cognition and social awareness subscales. Two paired sample t-tests uncovered a significant improvement from pre-to-post testing in experimental group t (13) = 3.93, $p < .003$, but not for the comparison group t (12) = .284, $p = .782$. Results follow up: No

Table 2 (continued)

Authors (year)	Program activities, duration and frequency	Medication/alternative therapies	Measures	Results
Zachor et al. (2017)	Sessions took place in urban parks near the children's kindergartens. Children completed four physical activities: two way climbing rope ladder, rope elevator (Children pull one child [connected to a harness with a rope pulley] up a tree, a rope bridge activity (involving children passing across three sections of rope) and hammock rope swing. The comparison group completed normal kindergarten routines. Duration: 13 weekly sessions (30 min each). Total dosage: 6.5 h	No mention	VABS and SRS	SRS: A significant group x time interaction was found for the comparison group only (scores increased [$F1, 20 = 5.3$ $p = 0.028$, $n^2 = 0.210$]). A significant time effect for the experimental condition was found in the social communication sub domain (F [1, 29] = 4.2, p = 0.05, $n^2 = 0.126$). Non-significant results were found for social motivation, communication and awareness. VABS: A significant time increase was found in the VABS communication subscale (F [4, 46] = 5.7, p = 0.021). No significant difference was found in socialization. Results were not separated by group, separate scores were unable to be obtained. Results follow up: No
Gabriels et al. (2015)	The horse-riding program consisted of horse riding skills (i.e. trotting, steering, turning, mounting and halting) and horsemanship skills (caring for horse). The riding consisted of a warm-up, skill review, learning a new skill, lesson review and a cool-down activity. The comparison group did barn activities (BA) and completed a brief art project. Duration: 10 weeks, once a week (45 min each session). Total dosage: 7.5 h	Not mentioned	SRS (caregiver reports), VABS-II (completed by researchers with a caregiver) and SALT (measured by a speech therapist)	The social cognition and communication subscales of SRS significantly improved in the THR group compared to the BA comparison group ($ES = .41$ [$p = .05$]; $ES = .63$ [$p = .003$] respectively). No significant difference was found on the social awareness and social motivation subscales or subscales of the VABS-II. Communication (SALT): The THR group significantly improved post treatment in terms of: an increase in the number of different words used, ES = .54 (p = .01) and more words used ES = .54 (p = .01). After controlling for age, IQ and gender-significantly greater improvements were found on social communication subscale (ES = .72 [p = .002]) number of words spoken, and number of new words (ES = .52 [p = .02]). Results follow up: No

Table 2 (continued)

Authors (year)	Program activities, duration and frequency	Medication/alternative therapies	Measures	Results
Movahedi et al. (2013)	The program consisted of 15 min of warm-up, 65 min for the activity and 10 min to cool down. Techniques learnt included a range of blocking, punching, sticking, and kicking techniques. Post assessments were conducted two days after the program finished, with follow up assessments conducted 30 days later. Comparison group = learnt educational skills (cognitive and language). Duration: 4 sessions per week over 14 weeks. Total dosage: 56 h	Routine medical care	GARS-2—social interaction subscale (in person interviews with parents and caregivers)	A repeated measures ANOVA found significant within participant differences in the experimental group: F (2, 48) = 17.02, $p < .001$ and showed significant between participant differences F (1, 24) = 4.24, $p < .01$. There was a significant group by time interaction F (2, 48) = 14.91, $p < .001$. Paired samples t test indicated significant decrease in social dysfunction in the exercise group post intervention t (12) = 6.17, $p < .001$ but no significant difference was found in the comparison group. Results follow up: Yes—results remained significantly unchanged.
Bahrami et al. (2016)	The program consisted of 15 min of warm-up, 65 min for the activity and 10 min to cool down. Techniques learnt included a range of blocking, punching, sticking, and kicking techniques. Post assessments were conducted two days after the program finished, with follow up assessments conducted 30 days later. Comparison = learnt educational skills (cognitive and language). Duration: 4 sessions per week over 14 weeks Total dosage: 56 h	Routine medical care	GARS-2—communication subscale (in person interviews with parents and caregivers)	A two-factor mixed-model ANOVA indicated a significant group x time interaction F (1, 20) = 22.35, $p < .001$. More specifically, communication deficits significantly decreased in the exercise group t (10) = 6.70, $p < .001$. No significant difference was found in the comparison group. Results follow up: Yes—results remained significantly unchanged.

**GARS-2* Gillam Autism Rating Scale, *SRS* Social Responsiveness Scale, *VABS* Vineland Adaptive Behaviour Scale, *SALT* Systematic Analysis of Language Transcripts, *BASC-T* Behaviour Assessment System for Children–Teacher, *SISS* Social Skills Improvement System, *ADOS* Autism Diagnostic Observation Schedule, *E* experimental condition, *C* comparison

Table 3 Risk of bias assessment ratings for included studies

Study	Study participation	Study attrition	Outcome measurement	Timing of diagnosis	Overall risk of bias
Hayward et al. (2016)	Medium	Medium	Medium	Low	Medium
Gabriels et al. (2012)	Medium	Low	Low	Low	Low
Borgi et al. (2016)	Medium	Low	Low	Low	Low
Anderson and Meints (2016)	Medium	Low	Medium	Low	Medium
García-Gómez et al. (2014)	Medium	High	High	Low	High
Bass et al. (2009)	Medium	Low	Medium	Low	Medium
Zachor et al. (2017)	Medium	Low	Medium	Low	Medium
Guest et al. (2017)	High	High	Medium	Low	High
Gabriels et al. (2015)	Medium	Low	Low	Low	Low
Movahedi et al. (2013)	Medium	Medium	Medium	Low	Medium
Bahrami et al. (2016)	Medium	Low	Medium	Low	Medium

groups who completed usual routine activities (Zachor et al. 2017; Gabriels et al. 2012) or did not report on the activities of the comparison group (García-Gómez et al. 2014).

Outcomes of Reviewed Studies

Social Functioning

Eight of the 10 studies that conducted significance testing found significant improvements in some aspect of social functioning. In the three studies that utilised the SRS and reported a total score, one found significant improvements (Bass et al. 2009), one did not (Zachor et al. 2017) and the other did not report the overall score (Gabriels et al. 2015). Regarding SRS sub-domains, two found improvements in social communication (Gabriels et al. 2015; Zachor et al. 2017), and one found significant improvements in social motivation (Bass et al. 2009). No studies found significant improvements in social awareness. For studies that examined the VABS ($k=6$), two found significant improvements on the socialisation subscale (Guest et al. 2017). Significance levels were unable to be obtained in one study (Guest et al. 2017) and separate group data was unable to be obtained in another (Zachor et al. 2017). One study found improvements in a general social skills domain following a multi-sport camp program (Guest et al. 2017), whereas García-Gómez et al. (2014) did not find significant improvements in general social skills following a horse-riding program. Lastly, Movahedi et al. (2013) found significant improvements (using the Gilliam Autism Rating Scale-Second Edition) in social interactions following a karate techniques program.

Communication

Six studies examined communication using the Vineland Adaptive Behaviour Scale. Two of these found significant improvements (Gabriels et al. 2012; Zachor et al. 2017); however, the significant increase noted in Zachor et al. (2017) were for participants as a whole. Separate data for the intervention and control groups could not be obtained. Significance levels for Guest et al. (2017) were not able to be obtained. A further study found a significant reduction (using the Gilliam Autism Rating Scale-Second Edition) in communication deficits following a kata techniques program (Bahrami et al. 2016).

Risk of Bias Assessments

Risk of bias assessments indicated four of the seven studies included in the meta-analysis carried a medium risk of bias (Bahrami et al. 2016; Bass et al. 2009; Zachor et al. 2017; Movahedi et al. 2013). A further two studies in the qualitative component carried a high risk of bias (Guest et al. 2017; García-Gómez et al. 2014), and two were a medium risk (Hayward et al. 2016; Anderson and Meints 2016) (see Table 3). Three included studies were considered a low risk of bias (Gabriels et al. 2012, 2015; Borgi et al. 2016).

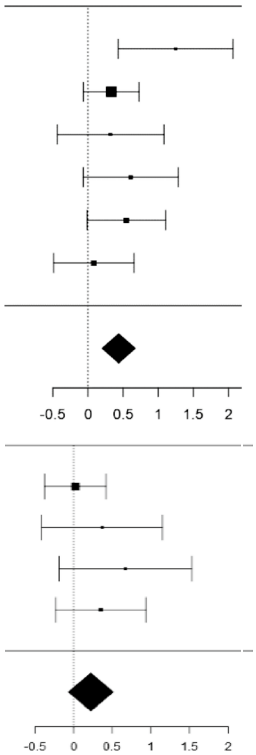
Meta-analysis

Publication Bias and Outliers

A visual inspection of funnel plots pertaining to overall social functioning indicated publication bias was unlikely. However, the funnel plot for communication was asymmetrical. These observations were further supported through Egger's test. Results were non-significant for the socialisation funnel plot ($p=.26$) indicating no significant asymmetry. For communication, however, Egger's test produced a significant result ($p=.05$). Fail-safe N indicated 18 studies were deemed necessary to produce non-significant results for social functioning. A fail-safe N of 0 was obtained for

Table 4 Depicts effect sizes and significance levels for meta-analyses

Domain and study name	Hedges' *g*	*S.E.*'s	95% C.I.		z-score	*p*-value	Forest Plots
			Lower	Upper			Hedges' *g* and 95% C.I. s
Social Functioning (n = 234)							
Movahedi et al. (2013)	1.25	.42	2.07	2.99	2.99	.003*	
Gabriels et al. (2012)	.33	.20	-.07	.73	1.63	.10	
Borgi et al. (2016)	.32	.39	-.44	1.08	.83	.40	
Bass et al. (2009)	.61	.35	-.07	1.27	1.77	.08	
Zachor et al. (2017)	.55	.28	-.01	1.11	1.92	.06	
Gabriels et al. (2015)	.08	.29	-.49	.66	.29	.78	
Summary Effect Size	.45	.14	.19	.72	3.33	.001*	
Communication (n = 201)							
Gabriels et al. (2012)	.02	.20	-.38	.42	.11	.91	
Borgi et al. (2016)	.36	.39	-.40	1.16	.92	.36	
Bahrami et al. (2016)	.65	.42	-.18	1.48	1.54	.13	
Gabriels et al. (2012)	.35	.29	-.23	.92	1.19	.24	
Adjusted Summary Effect Size	.13	.14	-.12	.38	1.54	.13	



S.E. standard error, *CI* Confidence Interval, *n* number of participants
*$p < .05$

communication, indicating publication bias may be present. Duval and Tweedie's trim and fill procedure was used to adjust the effect size and confidence intervals for the communication domain.

Regarding outliers, one sample (Movahedi et al. 2013) was found to be a 'probable' outlier using the z-score approach ($z = 2.99$), however, did not exceed the 'extreme' outlier value of 3.29 (in accordance with Field 2009). A 'one study removed' analysis did not change the direction or the significance of the finding for social functioning. The outlier was also removed from the analysis, which was re-run. As expected, the overall effect size reduced (− .09), however, results remained significant ($p = .004$). Consequently, the study was retained in analysis. No outliers were found for the communication domain.

Social Functioning

The forest plot below summarizes the effect sizes pertaining to social functioning across all included studies (Table 4). The meta-analysis found significant improvement in overall social functioning following participation in a form of group-based OPA program, in comparison to a group who did not participate ($k = 6$; $g = 0.45$, CI [0.19, 0.72], $p = .001$). This effect size demonstrated low levels of heterogeneity ($Q = 6.00$, $p = 3.1$; $\tau^2 = .02$;;$I^2 = 16.69\%$).

Communication

The forest plot below summarizes the effect sizes pertaining to communication across all included studies (Table 4). Results indicated no significant improvement in children's levels of communication following participation in a form of group-based OPA program, in comparison to a group who did not participate ($k = 4$; $g = 0.13$, CI [− 0.12, 0.38], $p = .13$). This effect size demonstrated very little observed inconsistencies or heterogeneity ($Q = 2.29$, p = .51; $\tau^2 = 0.00$; $I^2 = 0.00$). However, in accordance with Higgins et al. (2003), this observed homogeneity may be a result of the small *k* in this analysis.

Discussion

There is growing acknowledgment of the positive influence OPA programs may have on developmental outcomes in children with ASD, including motor and social benefits (Rinehart et al. 2018). Interventions such as group-based OPA, which take place within a child's natural environment, were of specific interest given they provide increased opportunities for social interactions and communication. This is the first review and meta-analysis that examines the social and communication effects of group-based OPA programs for children with ASD. The qualitative component of the review indicated several types of group-based OPA, given at varying dosages, may benefit social skills in children with ASD. While the number of studies eligible for inclusion in the quantitative component was limited, the results indicated group-based OPA programs may have a small to medium effect for children with ASD on social outcomes, with non-significant findings in communication.

Pooled statistical analyses found a significant small-medium effect in overall social functioning with 234 participants. This suggests participation in several types of group-based OPA programs can make a small-medium improvement in social functioning for children with ASD. As previously mentioned, movement may promote the development of social understanding and interaction skills (Trevarthen and Delafield-Butt 2013). When children participate in group-based OPA programs, they have the potential to engage in social interactions, which may increase overall social functioning and allow for a more fully integrated community experience (Özer et al. 2012). This is an important finding given the aforementioned push for more naturalistic, community-based research which can have benefits for children, families and broader communities alike (Rinehart et al. 2018), with youth community sport participation positively predicting adult community engagement (Perks 2007).

In addition to overall social functioning, the meta-analysis examined communication. Pooled statistical analyses did not reach significance. There may be several explanations for this. Firstly, the small sample of studies ($k = 4$) in this analysis may have contributed to the non-significant results. Some communication data from included studies (Zachor et al. 2017) could not be obtained and was therefore not included in analyses; potentially reducing the effect size. Secondly, it may be that OPA does not benefit communication skills. The International Classification of Functioning, Disability, and Health classifies communication as a distinct, but a related area to social functioning, encompassing areas such as speaking, listening, reading and writing (World Health Organisation 2001). While Gabriels et al. (2015) did find significant increases in aspects of speaking [amount of words used, and different words spoken (see Table 3)] following a horse-riding program, it is also possible the broader areas involved in communication which many measures such as the VABS assess (e.g. reading and writing) are less influenced by physical activity programs than other areas of communication that are more social in nature (such as the number of words spoken by a child). More research is needed to clarify group-based OPAs effects on the numerous subareas that comprise communication.

The qualitative component of this review noted there were mixed results for social functioning and communication across studies. For example, one study found significant improvements in social motivation, while two did not. Furthermore, two of three found significant improvements in the SRS social communication subscale. There are several possible explanations for the discrepancies between study results; these include (1) heterogeneity among individuals within and across samples, (2) considerable variability in terms of dosage (3) intervention type, and (4) the quality of study designs. Each of these explanations have been elaborated below.

Heterogeneity

ASD is heterogeneous in nature (Toal et al. 2010). The extent and area of impairments in cognitive functioning and symptom severity (Masi et al. 2017) differ across individuals. The heterogeneity rationale has been used to explain pattern differences across studies examining executive functioning (Geurts et al. 2014) and can be applied to many areas; for example, emotional disturbances (Matson and Nebel-Schwalm 2007) and degrees of motor impairment (Ming et al. 2007). While not enough data existed to include cognitive functioning and ASD symptom severity as moderators, the qualitative component indicated several studies had samples consisting of verbal and non-verbal children, either finding solely significant results or a mixture of significant and non-significant results. Two examined samples of children with no intellectual impairments (IQ's > 70) (Borgi et al. 2016; Gabriels et al. 2015) and found a mixture of significant and non-significant results. It is currently unclear whether it is only individuals with particular presentations or levels of ASD severity who benefit from OPA based on the existing literature.

Program Dosage

Despite the variability across studies in program dosage (6.5–56 h), eight of the 10 included studies that performed significance testing revealed significant improvements in at least one aspect of social functioning following a group-based OPA program. It is important to note significant results were still obtained in the study with the minimum dosage (6.5 h) (Zachor et al. 2017). Given only two studies (Bahrami et al. 2016; Movahedi et al. 2013) completed follow up assessments, it is unclear whether the benefits associated with shorter programs (e.g. of 6.5 h) are maintained long-term, and what the optimal or ideal program dosage should be to achieve maximum benefits for social functioning. In saying this, studies with the highest dosage (Bahrami et al. 2016; Movahedi et al. 2013) had the largest effect size on social functioning and communication. Given the small *k* in the meta-analyses, examining the effects of dosage as a moderator was not completed. More research examining varying program lengths would be required to ensure an adequate number of studies are available for future moderator analyses.

Types of Programs Examined

The current review examined a range of OPA activities, indicating several OPA programs may benefit social functioning. The qualitative component indicated significant improvements following horse riding, a 'kata techniques' training program, a multi-sport camp, and an outdoor adventure camp program.

It is important to note six of the 11 studies involved an equine therapy program. This is unsurprising given its steady growth in popularity for children with ASD (Kersten and Thomas 2005). There have been several recent reviews focusing on equine-assisted interventions alone for children with ASD. These reviews concluded equine-assisted therapy may benefit motor functioning and self-care (Peters and Wood 2017), as well as behavioural skills (Srinivasan et al. 2018). There is also preliminary evidence that suggests equine therapy improves socio-communication skills (Srinivasan et al. 2018). The qualitative aspect of this review complements this existing body of reviews and suggests this equine-assisted therapy may be beneficial for social functioning and communication for children with ASD. Only one program examining a 'team sport' such as soccer met the inclusion criteria (Hayward et al. 2016) i.e. involved statistical procedures. While the soccer program examined in the current review did not conduct significance testing on social domains and therefore could not be included in the meta analyses, parent reports suggest there may be social benefits (Hayward et al. 2016). To date, research examining the benefits of team sports for children with ASD is minimal. It would be beneficial for future studies to elaborate on this area, and examine the impact of team sports such as football, basketball and the like for children with ASD. This is important as some researchers have suggested participation in team sports has the potential to enhance life skills such as goal setting, problem solving, and constructive thinking (Papacharisis et al. 2005), skills which aid with employability in adulthood. Research into skill development in these areas are, as mentioned, prioritised by Autism communities (Pellicano et al. 2014). Moreover, team sports, which allow groups of people to work together towards a common goal (Wolf-Wendel et al. 2001), are said to 'facilitate prosocial behaviour and interpersonal communication' (Cei et al. 2017), and hence have the potential to benefit social functioning in children with ASD.

Risk of Bias

This review was limited to studies with an English translation available. Given this, there is potential that relevant studies were missed. Similarly, while grey literature and unpublished work were not excluded, no efforts were made to search in these areas specifically. While it may be that no such relevant work exists, there was evidence to suggest analyses carried out on communication were influenced by publication bias; that is, studies producing statistically significant results have a higher likelihood of being published (Onishi and Furukawa 2014).

A lack of high-quality studies limited the number of statistical analyses that could be completed. Only three of the included studies were considered to have a low risk of bias (Gabriels et al. 2012, 2015; Borgi et al. 2016). A lack of blinding and randomisation, or a lack of clarity in reporting these and other procedures, for example; adequate descriptions of outcome measures, reasons for attrition and its subsequent effects on data (as seen in many included studies) affects the robustness of the study design and subsequent results. Furthermore, study designs ranged from uncontrolled quasi-experimental designs to RCTs (see Table 3). Given (Eccles et al. 2003) suggests the results of uncontrolled designs, without a comparison group, may overestimate effects in pre–post studies, we excluded studies without a comparison group from the meta-analytic component. Future studies should endeavour to include comparison groups.

While there are clear biases and design limitations in many of the existing community-based studies in this area, they nevertheless offer an insight into the impact these programs may have for children with ASD in a more naturalistic setting. Clinical intervention studies typically employ strict procedures which limit the generalizability of results outside of the clinical context (Leichsenring 2004). Naturalistic

designs instead provide evidence of a program's effectiveness under the conditions of regular practice (Leichsenring 2004). Despite the risk of bias associated with many papers in the current review, results based on the existing literature provide evidence for the effectiveness of group-based OPA for children with ASD. To ensure more robust research, it is important that future community-based research studies attempt to employ randomisation techniques where possible and ensure all potential biases are clearly reported.

Summary of Limitations

There is a paucity of research concerning group-based OPA programs and only a small number of studies could be included for qualitative and quantitative analyses. It is important to note the dosage needed for long-term effects cannot be ascertained from included studies due to the lack of longitudinal designs or designs which provide follow up beyond the immediate program end. Furthermore, no moderator analyses were undertaken despite studies varying considerably in terms of study characteristics and participant characteristics. Further research in this area is recommended so the impact of study and sample characteristics on effect sizes can be evaluated in future reviews and meta-analyses. We acknowledge these limitations influence the strength of our conclusions; however, there is a tendency for people to make conclusions based on summaries provided in qualitative reviews and these 'ad hoc summaries' are often misleading (Borenstein et al. 2011). Given this, we believed adding a quantitative component to be advantageous over simply a qualitative review.

Conclusion

In sum, the current review and meta-analysis provides evidence in support of group-based OPA participation. These activities have the potential to allow for a more integrated community experience for children with ASD and may subsequently increase community engagement later in life. This is an important advancement in the area given the increased consumer demand for more naturalistic, life skill enhancing treatment options (Pellicano et al. 2014). The combined qualitative and quantitative results suggest group-based movement programs provide opportunities for social skill development. This is the first review to examine group-based physical activity specifically within an organised context. The naturalistic context of included studies provides evidence for the effectiveness of group-based OPA in practice. Overall, larger scale studies and further RCTs are needed in this field to explore they overall efficacy for children with ASD. Moreover, only one 'team sport' program was able to be found. Future research efforts should focus on examining the impact of team sports, given the potential to enhance social behaviours, and the avenues provided to build teamwork related skills necessary for adulthood. Overall, further research in this space which employs more robust techniques may provide further justification for the prescription of these programs, sitting alongside current clinical interventions (Rinehart et al. 2018).

Acknowledgments KH is supported by a scholarship funded by Jonathan and Simone Wenig. Funding related to this paper has also been received from the National Disability Insurance Agency (NDIA) and the Moose Foundation.

Author Contributions KH, CS, TM, JM and NR were involved in the conception and study design of this paper. KH and EL completed risk of bias assessments. KH and TM were involved in data analysis. All authors contributed to analysis interpretation and manuscript drafting. All authors approved the final manuscript.

Compliance with Ethical Standards

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Alexander, M., & Leather, R. (2013). Parents' perspectives on appropriate sports programs for children with autism spectrum disorders. *Palestra, 27*(4), 20–24.

Alhowikan, A. (2016). Benefits of physical activity for autism spectrum disorders: A systematic review. *Saudi Journal of Sports Medicine, 16*(3), 163.

Ament, K., Mejia, A., Buhlman, R., Erklin, S., Caffo, B., Mostofsky, S., et al. (2015). Evidence for specificity of motor impairments in catching and balance in children with autism. *Journal of Autism and Developmental Disorders, 45*(3), 742.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders—DSM IV*. Washington DC: American Psychiatric Publishing.

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

Anderson, S., & Meints, K. (2016). Brief report: The effects of equine-assisted activities on the social functioning in Children and Adolescents with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 46*(10), 3344–3352.

Australian Bureau of Statistics (2013). Stats & facts: Sport and physical recreaction—Differentials in participation. www.abs.gov.au/ausstats/abs@.nsf/Lookup/4156.0.55.001Main+Features5Nov%202013.

Bahrami, F., Movahedi, A., Marandi, S., & Sorensen, C. (2016). The effect of Karate techniques training on communication deficit of children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 46*(3), 978–986.

Bass, M. M., Duchowny, C. A., & Llabre, M. M. (2009). The effect of therapeutic horseback riding on social functioning in children with autism. *Journal of Autism and Developmental Disorders, 39*(9), 1261–1267.

Bazzano, A. T., Zeldin, A. S., Diab, I. R. S., Garro, N. M., Allevato, N. A., Lehrer, D., et al. (2009). The Healthy Lifestyle Change Program: A pilot of a community-based health promotion intervention for adults with developmental disabilities. *American Journal of Preventive Medicine, 37*(6), S201–S208.

Bellamy, J. (2015). Feasibility of a group-based school exercise intervention for children with intellectual disabilities. *Journal of Science and Medicine in Sport, 19,* e12.

Bhat, A. N., Landa, R. J., & Galloway, J. C. (2011). Current perspectives on motor functioning in infants, children, and adults with Autism Spectrum Disorders. *Physical Therapy, 91*(7), 1116–1129. https://doi.org/10.2522/ptj.20100294.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. Hoboken: Wiley.

Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods, 8*(1), 5–18. https://doi.org/10.1002/jrsm.1230.

Borgi, M., Loliva, D., Cerino, S., Chiarotti, F., Venerosi, A., Bramini, M., et al. (2016). Effectiveness of a standardized equine-assisted therapy program for children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 46*(1), 1–9.

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. https://doi.org/10.1177/1362361315616002.

Brignell, A., Albein-Urios, N., Woolfenden, S., Hayen, A., Iorio, A., & Williams, K. (2017). Overall prognosis of preschool autism spectrum disorder diagnoses. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD012749.

Bruder, M. (2001). *Infants and toddlers: Outcomes and ecology. Early childhood inclusion: Focus on change* (pp. 203–228). Baltimore: Sage.

Cei, A., Franceschi, P., Rosci, M., Sepio, D., & Ruscello, B. (2017). Motor and psychosocial development in children with autism spectrum disorder through soccer. *International Journal of Sport Psychology, 48*(5), 485–507.

Dunton, G., McConnell, R., Jerrett, M., Wolch, J., Lam, C., Gilliland, F., et al. (2012). Organized physical activity in young school children and subsequent 4-year change in body mass index. *Archives of Pediatrics and Adolescent Medicine, 166*(8), 713–718.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.

Eccles, M., Grimshaw, J., Campbell, M., & Ramsay, C. (2003). Research designs for studies evaluating the effectiveness of change and improvement strategies. *BMJ Quality & Safety, 12*(1), 47–52.

Edwardson, C. L., & Gorely, T. (2010). Parental influences on different types and intensities of physical activity in youth: A systematic review. *Psychology of Sport and Exercise, 11*(6), 522–535.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

Gabriels, A. J. A., Holt, K. D., Shoffner, A., Zhaoxing, P., Ruzzano, S., et al. (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. https://doi.org/10.1016/j.rasd.2011.09.007.

Gabriels, R., Pan, Z., Dechant, B., Agnew, J. A., Brim, N., & Mesibov, G. (2015). Randomized controlled trial of therapeutic horseback riding in children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 54*(7), 541–549. https://doi.org/10.1016/j.jaac.2015.04.007.

García-Gómez, A., Risco, M. L., Rubi, J. C., Guerrero, E., & García-Peña, I. M. (2014). Effects of a program of adapted therapeutic horse-riding in a group of autism spectrum disorder children. *Electronic Journal of Research in Educational Psychology, 12*(1), 107–128.

Gavin, M. (2014). Fitness and your 2-3 year old. https://kidshealth.org/en/parents/fitness-2-3.html.

Geurts, H., Sinzig, J., Booth, R., & HappÉ, F. (2014). Neuropsychological heterogeneity in executive functioning in autism spectrum disorders. *International Journal of Developmental Disabilities, 60*(3), 155–162.

Guest, L., Balogh, R., Dogra, S., & Lloyd, M. (2017). Examining the impact of a multi-sport camp for girls ages 8-11 with Autism Spectrum Disorder. *Therapeutic Recreation Journal, 51*(2), 109–126.

Hayden, J. A., van der Windt, D. A., Cartwright, J. L., Côté, P., & Bombardier, C. (2013). Assessing bias in studies of prognostic factors. *Annals of Internal Medicine, 158*(4), 280–286.

Hayward, L. M., Fragala-Pinkham, M., Johnson, K., & Torres, A. (2016). A community based, adaptive soccer program for children with Autism: Design, implementation and evaluation. *Palaestra, 30*(4), 44–50.

Healy, S., Nacario, A., Braithwaite, R. E., & Hopper, C. (2018). The effect of physical activity interventions on youth with autism spectrum disorder: A meta-analysis. *Autism Research, 11*(6), 818–833.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327,* 557.

Kersten, G., & Thomas, L. (2005). *Equine assisted mental health resource handbook*. Utah: Equine Assisted Growth Learning Association (EAGLA).

Khanjani, Z., Mosanezhad Jeddi, E., Hekmati, I., Khalilzade, S., Etemadi Nia, M., Andalib, M., et al. (2015). Comparison of cognitive empathy, emotional empathy, and social functioning in different age groups. *Australian Psychologist, 50*(1), 80–85.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4,* 863.

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. https://doi.org/10.1016/j.rasd.2010.01.006.

Leichsenring, F. (2004). Randomized controlled versus naturalistic studies: A new research agenda. *Bulletin of the Menninger Clinic, 68*(2), 137–151.

MacDonald, M., Lord, C., & Ulrich, D. A. (2013). The relationship of motor skills and social communicative Skills in school-aged children with autism spectrum disorder. *Adapted Physical Activity Quarterly, 30*(3), 271–282.

Masi, A., DeMayo, M. M., Glozier, N., & Guastella, A. J. (2017). An overview of autism spectrum disorder, heterogeneity and treatment options. *Neuroscience bulletin, 33*(2), 183–193.

Matson, J. L., & Nebel-Schwalm, M. S. (2007). Comorbid psychopathology with autism spectrum disorder in children: An overview. *Research in Developmental Disabilities, 28*(4), 341–352. https://doi.org/10.1016/j.ridd.2005.12.004.

McCoy, S. M., Jakicic, J. M., & Gibbs, B. B. (2016). Comparison of obesity, physical activity, and sedentary behaviors between adolescents with autism spectrum disorders and without. *Journal of Autism and Developmental Disorders, 46*(7), 2317–2326. https://doi.org/10.1007/s10803-016-2762-0.

Ming, X., Brimacombe, M., & Wagner, G. C. (2007). Prevalence of motor impairment in autism spectrum disorders. *Brain and Development, 29*(9), 565–570. https://doi.org/10.1016/j.braindev.2007.03.002.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The, P. G. (2009). Preferred reporting items for systematic reviews and

meta-analyses: The PRISMA statement. *PLOS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097.

Movahedi, A., Bahrami, F., Marandi, S., & Abedi, A. (2013). Improvement in social dysfunction of children with autism spectrum disorder following long term kata techniques training. *Research in Autism Spectrum Disorders, 7*(9), 1054–1061.

Ohrberg, N. J. (2013). Autism Spectrum Disorder and youth sports: The role of the sports manager and coach. *Journal of Physical Education, Recreation & Dance, 84*(9), 52–56.

Okely, A. D. (1999). The relationship of participation in organised sports and games, participation in nonorganised physical activity, and cardiorespiratory endurance to fundamental motor skill ability among adolescents.

Onishi, A., & Furukawa, T. A. (2014). Publication bias is underreported in systematic reviews published in high-impact-factor journals: Metaepidemiologic study. *Journal of Clinical Epidemiology, 67*(12), 1320–1326.

Özer, D., Baran, F., Aktop, A., Nalbant, S., Ağlamış, E., & Hutzler, Y. S. W. A. I. (2012). Effects of a Special Olympics Unified Sports soccer program on psycho-social attributes of youth with and without intellectual disability. *Research in Developmental Disabilities, 33*(1), 229–239. https://doi.org/10.1016/j.ridd.2011.09.011.

Papacharisis, V., Goudas, M., Danish, S. J., & Theodorakis, Y. (2005). The effectiveness of teaching a life skills program in a sport context. *Journal of Applied Sport Psychology, 17*(3), 247–254.

Pellicano, E., Dinsmore, A., & Charman, T. (2014). What should autism research focus upon Community views and priorities from the United Kingdom. *Autism: The International Journal of Research & Practice, 18*(7), 756–770.

Pelphrey, K. A., Shultz, S., Hudac, C. M., & Vander Wyk, B. C. (2011). Research review: Constraining heterogeneity: The social brain and its development in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 52*(6), 631–644.

Perks, T. (2007). Does Sport Foster Social Capital? The Contribution of Sport to a Lifestyle of Community Participation. *Sociology of Sport Journal, 24*(4), 378–401.

Peters, B. C. M., & Wood, W. (2017). Autism and equine-assisted interventions: A systematic mapping review. *Journal of Autism and Developmental Disorders, 47*(10), 3220–3242.

Reaven, J., & Wainer, A. L. (2015). Children and adolescents with ASD and co-occurring psychiatric conditions: Current trends in intervention. *International Review of Research in Developmental Disabilities, 49,* 45–90.

Reichow, B., Hume, K., Barton, E. E., & Boyd, B. A. (2018). Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD009260.pub2.

Rimmer, J. H., & Braddock, D. (2002). Health promotion for people with physical, cognitive, and sensory disabilities: An emerging national priority. *American Journal of Health Promotion, 16*(4), 220–224.

Rinehart, N. J., Tonge, B. J., Iansek, R., McGinley, J., Brereton, A. V., Enticott, P. G., et al. (2006). Gait function in newly diagnosed children with autism: Cerebellar and basal ganglia related motor disorder. *Developmental Medicine and Child Neurology, 48*(10), 819–824. https://doi.org/10.1017/S0012162206001769.

Rinehart, N. J., Jeste, S., & Wilson, R. B. (2018). Organized physical activity programs: Improving motor and non-motor symptoms in neurodevelopmental disorders. *Developmental Medicine and Child Neurology, 60,* 856–857. https://doi.org/10.1111/dmcn.13962.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638.

Sedgwick, P. (2013). Meta-analyses: How to read a funnel plot. *BMJ, 346,* f1342.

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.

Srinivasan, S. M., Cavagnino, D. T., & Bhat, A. N. (2018). Effects of equine therapy on individuals with autism spectrum disorder: A systematic review. *Review Journal of Autism and Developmental Disorders, 5,* 1–20.

Tan, B., Pooley, J., & Speelman, C. (2016). A meta-analytic review of the efficacy of physical exercise interventions on Cognition in individuals with autism spectrum disorder and ADHD. *Journal of Autism and Developmental Disorders, 46*(9), 3126–3143. https://doi.org/10.1007/s10803-016-2854-x.

Toal, F., Daly, E., Page, L., Deeley, Q., Hallahan, B., Bloemen, O., et al. (2010). Clinical and anatomical heterogeneity in autistic spectrum disorder: A structural MRI study. *Psychological Medicine, 40*(7), 1171–1181.

Trevarthen, C., & Delafield-Butt, J. T. (2013). Autism as a developmental disorder in intentional movement and affective engagement. *Frontiers in Integrative Neuroscience, 7,* 49.

Wolf-Wendel, L., Toma, J. D., & Morphew, C. C. (2001). There's no" I" in" team": Lessons from athletics on community building. *The Review of Higher Education, 24*(4), 369–396.

World Health Organisation (1993). The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research.

World Health Organisation (2001). *International Classification of Functioning, Disability and Health: ICF*: World Health Organization.

Yang, Y. J., Siao, M. R., Tsai, F. T., & Luo, H. J. (2015). Effect of physical activity interventions on children and adolescents with autism spectrum disorder: A systematic review and meta-analysis. *Physiotherapy, 101*(Supplement 1), e1685–e1686. https://doi.org/10.1016/j.physio.2015.03.091.

Zachor, D. A., Vardi, S., Baron-Eitan, S., Brodai-Meir, I., Ginossar, N., & Ben-Itzchak, E. (2017). The effectiveness of an outdoor adventure programme for young children with autism spectrum disorder: A controlled study. *Developmental Medicine and Child Neurology, 59*(5), 550–556.

Zeliadt, N. (2017). Autism in motion: Could motor problems trigger social ones? https://www.scientificamerican.com/article/autism-in-motion-could-motor-problems-trigger-social-ones/