Autism intervention meta-analysis of early childhood studies (Project AIM): updated systematic review and secondary analysis

Micheal Sandbank,[1] Kristen Bottema-Beutel,[2] Shannon Crowley LaPoint,[3] Jacob I Feldman,[4,5] D Jonah Barrett,[6] Nicolette Caldwell,[7] Kacie Dunham,[4,8] Jenna Crank,[9] Suzanne Albarran,[10] Tiffany Woynaroski[4,5,8,11,12]

For numbered affiliations see end of the article

Correspondence to: M Sandbank
micheal_sandbank@med.unc.edu
(or @michealsandbank.bsky.
social on Bluesky;
ORCID 0000-0002-6562-8267)
Additional material is published online only. To view please visit the journal online.

ABSTRACT

OBJECTIVE
To summarize the breadth and quality of evidence supporting commonly recommended early childhood autism interventions and their estimated effects on developmental outcomes.

DESIGN
Updated systematic review and meta-analysis (autism intervention meta-analysis; Project AIM).

DATA SOURCES
A search was conducted in November 2021 (updating a search done in November 2017) of the following databases and registers: Academic Search Complete, CINAHL Plus with full text, Education Source, Educational Administration Abstracts, ERIC, Medline, ProQuest Dissertations and Theses, PsycINFO, Psychology and Behavioral Sciences Collection, and SocINDEX with full text, Trials, and ClinicalTrials.gov.

ELIGIBILITY CRITERIA FOR SELECTING STUDIES
Any controlled group study testing the effects of any non-pharmacological intervention on any outcome in young autistic children younger than 8 years.

REVIEW METHODS
Newly identified studies were integrated into the previous dataset and were coded for participant, intervention, and outcome characteristics. Interventions were categorized by type of approach (such as behavioral, developmental, naturalistic developmental behavioral intervention, and technology based), and outcomes were categorized by domain (such as social communication, adaptive behavior, play, and language). Risks of bias were evaluated following guidance from Cochrane. Effects were estimated for all intervention and outcome types with sufficient contributing data, stratified by risk of bias, using robust variance estimation to account for intercorrelation of effects within studies and subgroups.

RESULTS
The search yielded 289 reports of 252 studies, representing 13 304 participants and effects for 3291 outcomes. When contributing effects were restricted to those from randomized controlled trials, significant summary effects were estimated for behavioral interventions on social emotional or challenging behavior outcomes (Hedges' g=0.58, 95% confidence interval 0.11 to 1.06; P=0.02), developmental interventions on social communication (0.28, 0.12 to 0.44; P=0.003); naturalistic developmental behavioral interventions on adaptive behavior (0.23, 0.02 to 0.43; P=0.03), language (0.16, 0.01 to 0.31; P=0.04), play (0.19, 0.02 to 0.36; P=0.03), social communication (0.35, 0.23 to 0.47; P<0.001), and measures of diagnostic characteristics of autism (0.38, 0.17 to 0.59; P=0.002); and technology based interventions on social communication (0.33, 0.02 to 0.64; P=0.04) and social emotional or challenging behavior outcomes (0.57, 0.04 to 1.09; P=0.04). When effects were further restricted to exclude caregiver or teacher report outcomes, significant effects were estimated only for developmental interventions on social communication (0.31, 0.13 to 0.49; P=0.003) and naturalistic developmental behavioral interventions on social communication (0.36, 0.23 to 0.49; P<0.001) and measures of diagnostic characteristics of autism (0.44, 0.20 to 0.68; P=0.002). When effects were then restricted to exclude those at high risk of detection bias, only one significant summary effect was estimated—naturalistic developmental behavioral interventions on measures of diagnostic characteristics of autism (0.30, 0.03 to 0.57; P=0.03). Adverse events were poorly monitored, but possibly common.

CONCLUSION
The available evidence on interventions to support young autistic children has approximately doubled in four years. Some evidence from randomized controlled trials shows that behavioral interventions improve caregiver perception of challenging behavior and child social emotional functioning, and that technology based interventions support proximal improvements in specific social communication and social emotional skills. Evidence also shows that developmental interventions improve social communication in interactions with caregivers, and naturalistic

---

WHAT IS ALREADY KNOWN ON THIS TOPIC

Many different types of early childhood interventions are recommended and offered to support generalized development in young autistic children

Previous research is mixed in quality and conclusions about the effectiveness of these interventions

WHAT THIS STUDY ADDS

The evidence available has approximately doubled in only four years

Some evidence from high quality studies supports the effectiveness of specific early childhood interventions for improving certain outcomes

Researchers have inadequately monitored and reported adverse events, effects, and harms (for any intervention type), therefore physicians should guide families to watch children closely when starting an intervention

1

developmental behavioral interventions improve core challenges associated with autism, particularly difficulties with social communication. However, potential benefits of these interventions cannot be weighed against the potential for adverse effects owing to inadequate monitoring and reporting.

Introduction

Autism is a relatively common diagnosis, with current estimates suggesting approximately 1-4% of the population is affected.[1-3] Early childhood interventions are often strongly recommended for young autistic children to promote skill gain in areas that might contribute to positive long term outcomes.[4 5] Pediatricians and other physicians are often the first line of care directing families of autistic children to early childhood interventions to support their development. Therefore, physicians should be familiar with the available interventions and the landscape of evidence supporting them to make practice recommendations. However, the types of early childhood interventions recommended for this population vary widely in terms of approach and intensity, and current best practice guidelines differ across countries. For example, in the United States, the most commonly recommended treatment is early intensive behavioral intervention, an approach that incorporates operant conditioning, targets functional skills, and is characterized by a recommended intensity of 20-40 hours per week.[6] In contrast, the National Institute for Health and Care Excellence in England concluded that only two intervention approaches have sufficient evidence to support their use. These relatively low intensity interventions are pediatric autism communication therapy, and joint attention, symbolic play, engagement and regulation (JASPER), which target early social communication in the context of natural interactions.[7] Previous attempts to synthesize
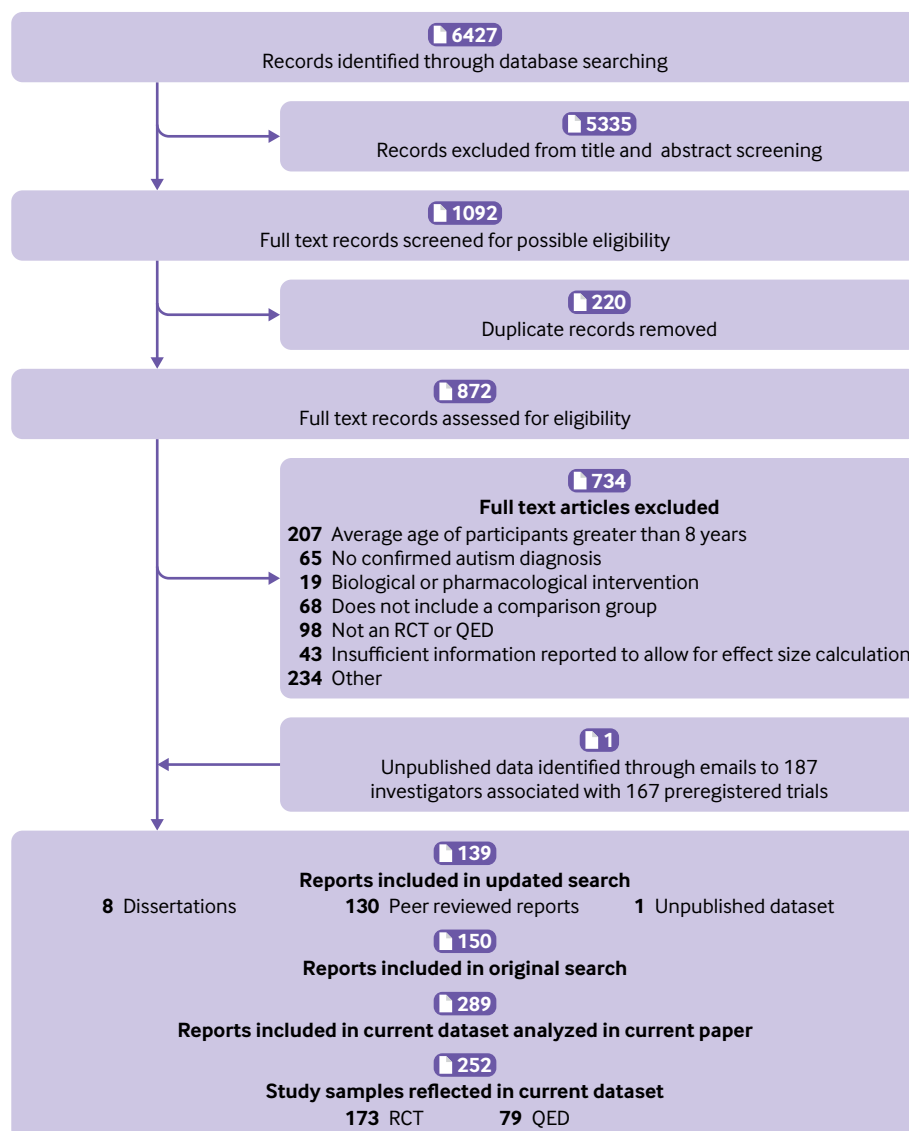

Fig 1 | PRISMA (preferred reporting items for systematic review and meta-analysis) diagram. QED=quasi-experimental design study; RCT=randomized controlled trial
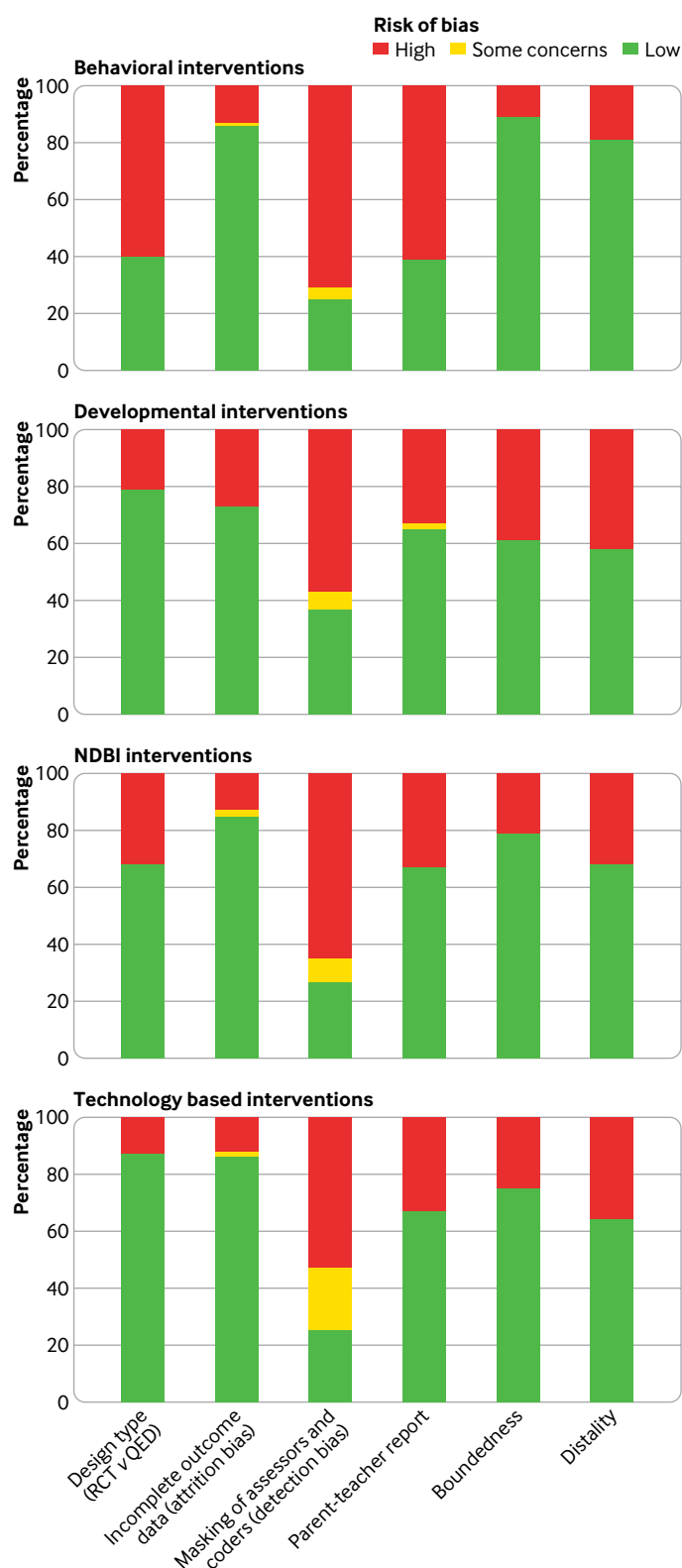
Fig 2 | Summary of risk of bias ratings for studies of behavioral, developmental, naturalistic developmental behavioral intervention (NDBI), and technology based intervention types. QED=quasi-experimental design study; RCT=randomized controlled trial

aggregation of evidence; low standards of evidence for designating practices as evidence based; limited evaluation of intervention outcomes; overreliance on vote counting over quantitative synthesis; and a rapidly transforming evidence base. Consequently, doctors, clinicians, and families have to navigate confusing and often conflicting guidance on which supports are the most likely to be efficacious for autistic children.

Project AIM: initial scope and findings

We conducted Project AIM (autism intervention meta-analysis), a scoping systematic review and meta-analysis of controlled group studies of any non-pharmacological intervention designed to support any outcome in young autistic children.[8] The initial search identified 139 studies of common intervention approaches (parsed by type) on various outcomes (categorized by domain) in young autistic children. This quantitative synthesis of evidence from group design intervention studies allowed comparison of the overall quality and findings of evidence according to intervention approach. The results of Project AIM were selected by the Interagency Autism Coordinating Committee of the US Department of Health and Human Services as an important advance in autism research.[9] Subsequently, the findings have been incorporated into clinical guidelines[5 10] and continue to shape intervention recommendations for young autistic children.

The initial Project AIM investigation and subsequent secondary analyses documented gaps in study quality, within specific intervention types and across the literature as a whole. Notable gaps included an overrepresentation of quasi-experimental studies (ie, in which participants were not randomly assigned to treatment groups), overreliance on outcomes measured by unmasked assessors and proxy reports, and inadequate monitoring of adverse events and harms.[8 11] When intervention effects were estimated from all available evidence, regardless of quality, several intervention approaches were estimated to have positive and statistically significant effects on a variety of outcomes. However, when study quality was taken into account and effects were restricted to those immune to these risks of bias (ie, selection bias, detection bias, placebo-by-proxy bias), no intervention approach was estimated to have positive and statistically significant intervention effects on any outcome. Although adverse events, effects, and harms were inadequately monitored, many studies reported information indicating they occurred (such as reasons for attrition that should have been reported as adverse events, or statistically significant negative effects on a measured outcome, which would qualify as a harm).[11]

Our initial report also found that intervention effects were larger on proximal outcomes that were specifically targeted in the intervention compared with outcomes indicating more distal developmental improvement; and on outcomes that were measured in contexts identical or similar to those of the intervention compared with those generalized to other contexts.[12]

intervention evidence to generate consistent clinical guidelines have been hindered by a number of factors: heterogenous intervention approaches that prevented

Table 1 | Estimates of variance and summary effects by intervention and outcome type with increasingly restrictive quality thresholds

	All RCTS			RCTS without caregiver report			RCTs with low detection bias		
Intervention or outcome	τ^2	g (95% CI)	P	τ^2	g (95% CI)	P	τ^2	g (95% CI)	P
Behavioral									
Social communication	0.873	0.54 (−0.24 to 1.32)	0.15	—	—	—	—	—	—
Social emotional or challenging behavior	0.334	0.58 (0.11 to 1.06)	0.02	—	—	—	—	—	—
Developmental									
Social communication	0.028	0.28 (0.12 to 0.44)	0.003	0.037	0.31 (0.13 to 0.49)	0.003	—	—	—
NDBI									
Adaptive	0.011	0.23 (0.02 to 0.43)	0.03	—	—	—.	—	—	—
Cognitive	0.028	0.18 (−0.02 to 0.38)	0.07	0.029	0.19 (−0.02 to 0.39)	0.07	0.000	0.17 (−0.02 to 0.37)	0.07
Diagnostic characteristics of autism	0.086	0.38 (0.17 to 0.59)	0.002	0.095	0.44 (0.20 to 0.68)	0.002	0.037	0.30 (0.03 to 0.57)	0.03
Language	0.065	0.16 (0.01 to 0.31)	0.04	0.075	0.13 (−0.04 to 0.30)	0.13	0.048	0.06 (−0.13 to 0.25)	0.49
Play	0.000	0.19 (0.02 to 0.36)	0.03	—	—	—.	—	—	—
Restricted and repetitive behaviors	0.029	−0.01 (−0.32 to 0.31)	0.96	—	—	—.	—	—	—
Social communication	0.047	0.35 (0.23 to 0.47)	<.001	0.046	0.36 (0.23 to 0.49)	<.001	0.009	0.11 (−0.03 to 0.26)	0.11
Technology based									
Language	0.107	0.21 (−0.13 to 0.55)	0.18	0.131	0.26 (−0.14 to 0.66)	0.16	—	—	—
Social communication	0.263	0.33 (0.02 to 0.64)	0.04	0.013	0.20 (−0.01 to 0.41)	0.06	—	—	—
Social emotional or challenging behavior	0.279	0.57 (0.04 to 1.09)	0.04	0.466	0.64 (−0.07 to 1.36)	0.07	—	—	—

95% CI=95% confidence interval; NDBI=naturalistic developmental behavioral intervention; RCT=randomized controlled trial.
Summary estimates with <5 degrees of freedom were dropped from results and reflected as —.

These findings show that conclusions drawn about intervention effectiveness, in addition to being compromised by quality concerns, are dependent on researcher measurement decisions. Interventions shown to have proximal impacts in specific contexts are often designated as effective, with little attention given to the limited scope of change quantified by these effects. These designations are then repeated in systematic reviews and meta-analyses, and eventually shape clinical guidelines until most interventions that have been designated as evidence based and recommended for clinical use are those that have been shown to effect circumscribed and specific change, rather than generalized developmental gains. Although physicians guiding families to clinical supports are often led to believe that a substantial evidence base endorses the effectiveness of interventions for enabling broad developmental improvements, our work showed that there is little evidence that this occurs. Families might then believe that their child's lack of improvement when participating in these interventions is indicative of the complexity of their child's condition, rather than the inadequacy of the interventions available.

Intervention and outcome type	Studies	Effect sizes	Hedges' g (95% CI)	Hedges' g (95% CI)
Behavioral				
Social communication	9	84		0.54 (-0.24 to 1.32)
Social emotional or challenging behavior	10	57		0.58 (0.11 to 1.06)
Developmental				
Social communication	14	123		0.28 (0.12 to 0.44)
NDBI				
Adaptive	11	31		0.23 (0.02 to 0.43)
Cognitive	13	48		0.18 (-0.02 to 0.38)
Diagnostic characteristics of autism	17	46		0.38 (0.17 to 0.59)
Language	26	138		0.16 (0.01 to 0.31)
Play	8	65		0.19 (0.02 to 0.36)
Restricted and repetitive behaviors	7	20		-0.01 (-0.32 to 0.31)
Social communication	32	322		0.35 (0.23 to 0.47)
Technology based				
Language	9	29		0.21 (-0.13 to 0.55)
Social communication	17	84		0.33 (0.02 to 0.64)
Social emotional or challenging behavior	8	53		0.57 (0.04 to 1.09)

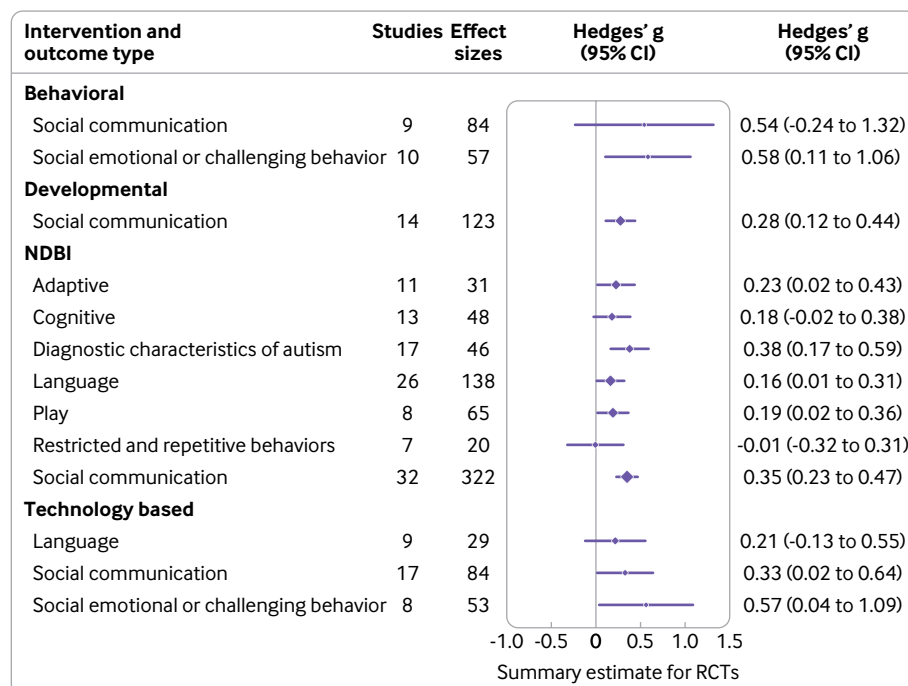-1.0 -0.5 0 0.5 1.0 1.5
Summary estimate for RCTs

Fig 3 | Forest plot of summary estimates for each outcome type by intervention type when only outcomes from randomized controlled trials (RCTs) are included. 95% CI=95% confidence interval; NDBI=naturalistic developmental behavioral intervention

Need for an updated meta-analysis

Although the initial Project AIM report was published in January 2020, the search was completed in November 2017. Therefore, studies published after the search date were not included in quality evaluations and summary effect estimation. During the decade before the search, there were considerable increases in funding and the rate of publication of autism related research.[13] Two thirds of the studies and randomized controlled trials included in Project AIM were published in the five years preceding the search date. Therefore, it was reasonable to assume that a substantial number of studies, including many randomized controlled trials, had been published since the original search was conducted. The rapidly expanding evidence base suggests that an updated review would ensure the conclusions of Project AIM reflect the most recent evidence on interventions for young autistic children and provide guidance to medical professionals, specialist clinicians, and families.

Research questions

In the current report, we sought to answer the following questions with an updated dataset that integrated studies published since 2017 (inclusive of all research conducted from 1975 to 2021).

1. What is the overall quality of evidence supporting each intervention type in terms of use of random assignment, limited reliance on proxy report, use of masked assessors, and minimal attrition? What percentage of effects reflect generalized or distal outcomes?

2. What percentage of studies monitored and reported or showed evidence of adverse events, adverse effects, or harms?

3. When all available evidence is considered from randomized controlled trials, what intervention types are estimated to have positive and statistically significant effects on targeted outcomes?

4. What intervention types are estimated to have positive and statistically significant effects on targeted outcomes when evidence from randomized controlled trials is further restricted to outcomes measured directly (ie, excluding caregiver or teacher report outcomes) and by masked assessors?

5. Are intervention effects on proximal outcomes significantly larger than intervention effects on distal outcomes? Are intervention effects on context bound outcomes significantly larger than intervention effects on generalized outcomes?

Method

Search

Search terms and databases

We completed an updated search on 17 November 2021 that replicated the initial Project AIM search in terms and databases, but was limited to studies published after 1 November 2017 (the date of our previous search). Searched databases included Academic Search Complete, CINAHL Plus with full text, Education Source, Educational Administration Abstracts, ERIC, Medline, Proquest Dissertations and Theses, PsycINFO, Psychology and Behavioral Sciences Collection, and SocINDEX with full text. Search terms are listed in the supplementary materials. This search yielded 6427 records that were then screened. In addition to searching databases that index dissertations and theses, we sought unpublished data by searching the journal *Trials* for published protocols and ClinicalTrials.gov with the search term "autism" to identify potentially relevant registered but unpublished clinical trials. Potentially relevant trials (n=168) were identified and we emailed researchers associated with those trials (n=187) with a request to share information that would allow their inclusion in the updated meta-analysis.

Screening process

All identified records were double screened at the abstract level by two of 13 independent screeners
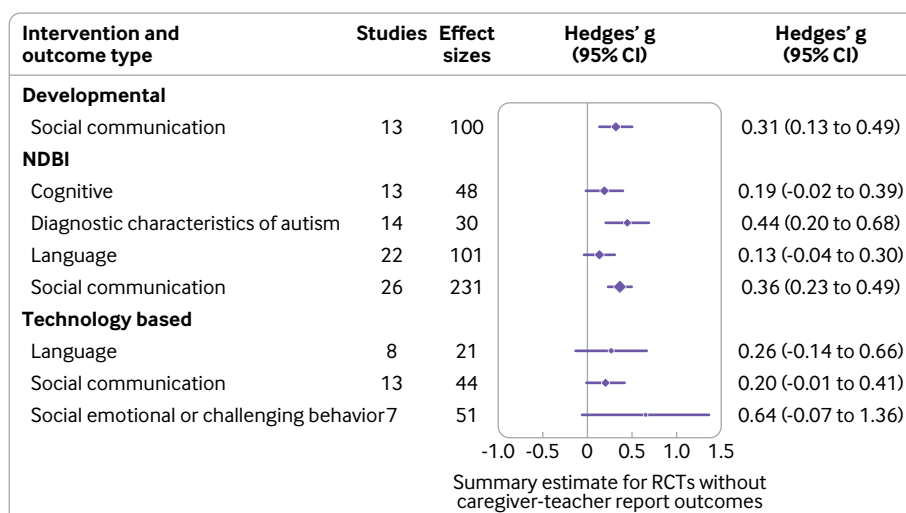
Intervention and outcome type	Studies	Effect sizes	Hedges' g (95% CI)	Hedges' g (95% CI)
Developmental				
Social communication	13	100		0.31 (0.13 to 0.49)
NDBI				
Cognitive	13	48		0.19 (-0.02 to 0.39)
Diagnostic characteristics of autism	14	30		0.44 (0.20 to 0.68)
Language	22	101		0.13 (-0.04 to 0.30)
Social communication	26	231		0.36 (0.23 to 0.49)
Technology based				
Language	8	21		0.26 (-0.14 to 0.66)
Social communication	13	44		0.20 (-0.01 to 0.41)
Social emotional or challenging behavior	7	51		0.64 (-0.07 to 1.36)

-1.0 -0.5 0 0.5 1.0 1.5
Summary estimate for RCTs without caregiver-teacher report outcomes

Fig 4 | Forest plot of summary estimates for each outcome type by intervention type when only outcomes from randomized controlled trials (RCTs) not derived from caregiver or teacher reports are included. 95%CI=95% confidence interval; NDBI=naturalistic developmental behavioral intervention

using the web application abstrackr.[14] Any record flagged as potentially eligible by at least one screener was then examined at the full text level. Studies were considered eligible if they were published in English between November 2017 and November 2021; they were experimental (ie, a randomized controlled trial) or quasi-experimental design group studies that included an intervention and a control or comparison group; they reported a simple majority of participants had autism; they included participant samples with an average age <8 years (<96 months); and they had not already been included in the previous Project AIM meta-analysis. Some studies that were eligible for inclusion reported insufficient data to allow extraction of appropriate effect sizes. For each of these studies, the first author contacted the corresponding author by email to request information that would allow effect size calculation.

Coding procedures

All studies were independently double coded by the first author and by one of a team of five trained reliability coders. All discrepancies were identified and resolved through discussion before final codes were entered. Coding procedures were nearly identical to those used in the original Project AIM, but are briefly described here. The coding manual can be accessed through an online data repository.[15] This update was not registered but replicated procedures used in the previous meta-analysis.

Participant characteristics

When reported, we extracted the following participant sample characteristics from reports: chronological age in months; language age in months (expressive was given preference, but receptive and total language ages were also extracted in the absence of expressive); and proportion of the sample reported as male.

Intervention characteristics

Interventions were coded for type, setting, implementer, and cumulative intensity (total amount of intervention provided to participants in hours across the duration of the study). Approaches were categorized as belonging to one of nine possible types using the categorization system derived for the original Project AIM: animal

assisted; behavioral; cognitive behavioral therapy; developmental; naturalistic developmental behavioral intervention (NDBI); TEACCH (formerly treatment of autistic and related communications handicapped children); technology based; sensory based; or other. Sensory based interventions were then further coded as sensory integration therapy, other sensory based interventions, or music therapy to ensure that interventions were grouped according to a consistent theory of change. We provide a non-exhaustive list of examples for each intervention type below. In the rare event that coders were unable to agree on intervention type, the first author contacted the corresponding author of the study and asked for their input.

Animal assisted therapy—Interventions that were mediated entirely through the presence of an animal or that were characterized primarily by interaction with an animal were categorized as animal assisted therapy. Examples include equine assisted therapy and use of service dogs.

Behavioral—Interventions were categorized as traditional behavioral interventions if they relied heavily on operant principles of learning and corresponding techniques (eg, were primarily adult led, used explicit instruction and prompting, provided explicit reinforcement with tangible rewards). Examples include early intensive behavioral intervention, the picture exchange communication system, and discrete trial training.

Cognitive behavioral therapy—Interventions described as cognitive behavioral therapy, which focuses on identifying and changing thinking patterns in an effort to change behavior, were categorized as such.

Developmental—Interventions were categorized as developmental if they were primarily child led, stressed the relational or transactional and social underpinnings of development, and taught skills according to a developmental sequence with the goal of allowing developmental cascades by repairing breakdowns in relational cycles. Examples include pediatric autism communication therapy and Hanen models.

Music therapy—Interventions were categorized as music therapy if they were explicitly described as such or incorporated music and rhythm based experiences toward therapeutic ends.
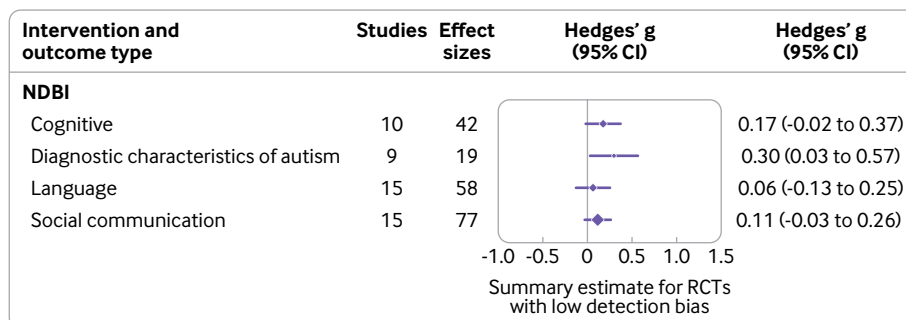
Intervention and outcome type	Studies	Effect sizes	Hedges' g (95% CI)	Hedges' g (95% CI)
NDBI				
Cognitive	10	42		0.17 (−0.02 to 0.37)
Diagnostic characteristics of autism	9	19		0.30 (0.03 to 0.57)
Language	15	58		0.06 (−0.13 to 0.25)
Social communication	15	77		0.11 (−0.03 to 0.26)

−1.0 −0.5 0 0.5 1.0 1.5
Summary estimate for RCTs with low detection bias

Fig 5 | Forest plot of summary estimates for each outcome type by intervention type when only outcomes from randomized controlled trials (RCTs) and low risk of detection bias are included. 95% CI=95% confidence interval; NDBI=naturalistic developmental behavioral intervention

Naturalistic developmental behavioral intervention—Interventions were categorized as NDBIs if they were one of several named interventions listed in the consensus paper by Schreibman and colleagues.[16] These interventions are motivated by developmental and behavioral theories of learning and characterized by child initiated interactions that give way to reciprocal social routines that are maintained through shared control by the adult and child. NDBIs are naturalistic in that they take place in environments and within routines that are already likely to occur in the child's life (eg, play in the home or community), and rely on natural antecedents and rewards. Intervention targets tend to center on early social communication skills that are thought to serve as a foundation for further developmental cascades. Examples include the early start Denver model, JASPER, and pivotal response teaching.

Sensory integration therapy—Interventions were described as sensory integration therapy if they were explicitly described as such or were characterized by structured exposure to several types of sensory opportunities (ie, tactile, vestibular, proprioceptive).[17 18] This category included Ayres sensory integration and more general sensory integration therapy.

Sensory based interventions—Interventions were categorized as sensory based interventions if they incorporated targeted exposure to sensory stimuli with the goal of enhancing processing of sensory stimuli and theoretically related outcomes, but did not include several types of sensory opportunities. Examples include auditory integration, touch therapy, and massage.

TEACCH—Interventions were categorized as TEACCH if they were explicitly named as such. TEACCH is characterized by heavy reliance on predictable environments, structured work systems, and routines (eg, visual and picture activity schedules).

Technology based interventions—Interventions were categorized as technology based if technological mediation was described as the main change agent of the intervention. Technologies included electronic devices such as computers, iPads, or robots.

Other—Interventions that could not be adequately categorized in the previous categories were coded as other for type. These studies were excluded from summary effect estimation, but included in moderator analyses.

Outcome characteristics

Outcomes were coded for domain, proximity, and boundedness.

Domain—Dependent variable names were extracted for each outcome, and outcomes were coded as representing diagnostic characteristics of autism (ie, social communication; restricted or repetitive patterns of behaviors, interests, or activities; sensory; or overall autism features—ie, total scores on diagnostic assessments) or related domains (ie, brain imaging, academic, adaptive, cognitive, language, motor, play, sleep, social emotional or challenging behavior). Further details about outcome domain coding are provided in the previous report.[8] A non-exhaustive list of example measures and associated metrics represented in each domain category for which we estimated summary effects is provided in the supplementary materials.

Proximity—Outcomes derived from measures of skills and developmental achievements directly taught or modeled in the intervention were coded as proximal, while those derived from measures of broader skill sets and developmental milestones (across the targeted domain or in a distinct untargeted domain) were coded as distal.

Boundedness—To code outcome boundedness, coders considered the context of outcome measurement and context of the intervention across the four dimensions of materials, setting, interaction partner, and interaction style to determine the degree to which they matched. Outcomes measured in contexts that were the same or highly similar (differing on only one dimension) to that of the intervention context were coded as context bound. Outcomes measured in contexts that differed across two or more dimensions from the intervention context were coded as generalized.

Risks of bias

Studies and outcomes were coded for standard risks of bias following guidance from Cochrane (ie, selection bias, detection bias, performance bias, attrition bias)[19] and reliance on caregiver or teacher report to index outcomes. We coded this additional outcome characteristic because it indicates the risk of a specific type of detection bias known as placebo-by-proxy bias, which is introduced when assessors are not only aware of participant group assignment, but also personally invested in the outcome.[20]

Adverse events, effects, and harms

Adverse events refer to any unfavorable outcomes that occur during or subsequent to participating in an intervention, but that might or might not have been caused by the intervention. Adverse effects refer to unfavorable outcomes that can be reasonably attributed to the intervention. Harms refer to sustained deterioration during or after participation in an intervention.[21] Using procedures similar to those outlined by Bottema-Beutel and colleagues,[11] two independent coders searched full text copies of each article for the following terms: "adverse events," "adverse effects," "harm," "side effect," and "complication," and coded whether adverse events were reported, whether the adverse events were described in such a way that they could reasonably be considered adverse effects, the number of adverse events reported, and whether the authors described adverse event monitoring procedures. Direct quotes describing the adverse events and monitoring procedures were copied and pasted verbatim into the coding spreadsheet.

Effect size information

For each reported outcome, postintervention means, standard deviations, and sample sizes were extracted for the intervention and counterfactual groups. These values were used to calculate and code the standardized mean difference (d) representing each intervention effect and corresponding variance. In the few cases in which reported outcomes were dichotomous, frequencies were extracted to estimate d and its variance. All calculations of d were derived using the Campbell Collaboration Practical Meta-Analysis Effect Size Calculator.[22] Standardized mean difference effect sizes were then converted to Hedges' g to correct for small sample sizes using R statistical computing software (R Core Team, 2022). Effect sizes were reflected for outcomes for which lower scores were considered adaptive, so that directionality of effect size was consistent across all effects.

Reliability

Primary and reliability coding sheets were independently sent to a separate coding auditor who identified discrepancies using a program created in-house for this purpose. Initial codes were stored in a separate location for reliability analyses before discrepancy discussions. Reliability was calculated in R[23] with the irr package.[24] Reliability was indexed using unweighted κ for categorical variables,[25] where values over 0.6 reflect substantial agreement and values over 0.8 reflect near perfect agreement; and two way absolute intraclass correlation coefficients for continuous variables,[26] where values over 0.75 reflect strong agreement and values over 0.9 reflect excellent agreement. κ values ranged from 0.62 to 0.9, and the average κ across all included categorical variables was 0.75. Intraclass correlation coefficients ranged from 0.68 to 0.99, and the average value across all continuous variables was 0.90.

Analyses

All analyses were conducted in R.[23] Given that this meta-analysis involved a complex data structure, wherein multiple effect sizes were extracted within overlapping participant samples and these effect sizes were then categorized according to intervention and outcome type, effect sizes were analyzed using robust variance estimation meta-analysis to account for the dependence structure of the data. Specifically, subgroup correlated effects working models[27] were used to aggregate effect sizes based on type of outcome (outcome characteristics) within each type of intervention (intervention characteristics) using the clubSandwich[28] and metafor packages.[29] Summary effect sizes for each outcome and intervention type were only retained in the final models when the degrees of freedom (df) were greater than five.[30] For moderation analyses, putative categorical moderators (ie, proximal *v* distal outcome, generalized *v* context bound outcome) were added to the final model used to estimate summary effects for all studies.

To assess possible publication bias across the extant literature, we conducted Egger's regression test with cluster robust variance estimation methods (ie, the Egger MLMA with standard error approach) using the rma.mv function in the metafor package with the SCE covariance matrix obtained from the clubSandwich package.[31] Egger's regression test was conducted for each presented model, and within interventions for outcome types with df≥5 in the null models.

Patient and public involvement

Coauthors Jacob I Feldman and Tiffany Woynaroski are parents of autistic individuals and they worked on the research question, analyses, and drafting of the manuscript. Although members of the public were not directly involved in this review because of funding limitations and lack of researcher training to engage the public, the focus of this work is aligned with the research priorities of autistic people, which include rigorous evaluations of interventions designed to support development, health, and wellbeing.[32]

Results

Descriptives of included study samples, interventions, and outcomes

Figure 1 presents the PRISMA (preferred reporting items for systematic review and meta-analysis) diagram detailing the search process. From the 6427 records retrieved in the updated search, we included 139 eligible reports. Reports identified in the updated search were combined with reports from the original search to yield a dataset of 289, reporting on 252 separate study samples (173 randomized controlled trials, 79 quasi-experimental design studies) that included a total of 13 304 participants and 3291 outcomes. The mean age of participant samples was 56.11 months (range 18.9-95.2, standard deviation 19.08). The mean language age equivalent in months for samples where it was reported was 22.36 (standard deviation 12.73). The average percentage of samples reported as male was 82.57 (standard deviation 11.27). There were 10 studies (10 reports) of animal assisted interventions, 4 studies (4 reports) of cognitive behavioral therapy, 48 studies (51 reports) of behavioral interventions, 19 studies (24 reports) of developmental interventions, 6 studies (8 reports) of music therapy, 57 studies (75 reports) of NDBI, 6 studies (6 reports) of sensory integration therapy, 6 studies (7 reports) of other sensory based interventions, 9 studies (9 reports) of TEACCH, 30 studies (33 reports) of technology based interventions, and 65 studies (72 reports) of interventions categorized as other. (When summed across intervention type, the total numbers of studies and reports slightly exceed the overall totals of 252 study samples and 289 reports because a small number of studies reported effects for two separate intervention approaches tested against a control. For these studies, a single report or study is represented in two separate intervention categories.) Data and analytic code are available in an online repository.[33]

Risk of bias and outcome characteristics

Figure 2 presents the percentage of outcomes for each risk of bias rating (ie, low risk of bias, high risk of bias, unclear) for all risk of bias indicators by intervention type. Only studies included in the generation of effect sizes are summarized in terms of risk of bias. Studies of animal assisted interventions, cognitive behavioral therapy, music therapy, sensory integration therapy, and other sensory based interventions are not described here because there were too few studies of these intervention types to reliably estimate the summary effects of these approaches.[30] Overall, 173 of 252 studies were randomized controlled trials. Nearly two thirds of all outcomes (65.19%) were at risk of detection bias, and nearly half of all outcomes (45.51%) were derived from caregiver or teacher reports. Attrition bias was coded as high for 15.67% of outcomes. In most studies, the nature of the interventions prevented the possibility of fully masking participants to the intervention received. Therefore, we coded most outcomes (97.54%) as being at risk of performance bias, and we do not report risk of performance bias by intervention type because there is little variation. Most outcomes (80.56%) were coded as being generalized from the intervention context; the remainder were coded as context bound. Similarly, most outcomes (72.7%) were coded as being distal from the intervention targets; the remainder were coded as proximal.

Summary effects

We present effects stratified by risk of bias, first including all effects from randomized controlled trials, and subsequently restricting effects to those from randomized controlled trials and outcomes that were directly assessed (excluding caregiver or teacher reports), and then to effects from randomized controlled trials and outcomes subject to low risk of detection bias. Table 1 presents summary (Hedges' g), heterogeneity (τ^2), and significance estimates (P) for each of these analyses. Figures 3, 4, and 5 present forest plots of summary effect estimates associated with decreasing risks of bias and increasing levels of certainty. Summary estimates that include effects from quasi-experimental design studies, which are subject to high risk of selection bias, are provided in the supplementary materials (see figure S1).

Estimated effects from randomized controlled trials

Figure 3 reflects summary effect sizes derived from outcomes extracted only from randomized controlled trials, according to intervention and outcome type. Statistically significant effects were estimated for behavioral interventions on social emotional or challenging behavior (g=0.58, 95% confidence interval 0.11 to 1.06); for developmental interventions on social communication outcomes (0.28, 0.12 to 0.44); for NDBIs on adaptive (0.23, 0.02 to 0.43), language (0.16, 0.01 to 0.31), play (0.19, 0.02 to 0.36), and social communication outcomes (0.35, 0.23 to 0.47), and measures of diagnostic characteristics

of autism (0.38, 0.17 to 0.59); for technology based interventions on social communication outcomes (0.33, 0.02 to 0.64) and social emotional or challenging behavior (0.57, 0.04 to 1.09). There were not enough controlled studies of animal assisted interventions, cognitive behavioral therapy, music therapy, sensory integration, or sensory based interventions to generate summary estimates of their effects on any outcome for young children. Additionally, there were not enough randomized controlled trials of TEACCH to generate summary effects for any outcome.

Estimated effects from randomized controlled trials excluding outcomes from caregiver or teacher reports

Figure 4 shows summary effects estimated exclusively from outcomes that were extracted from randomized controlled trials and that were not derived from caregiver or teacher reports (ie, were not at risk of placebo-by-proxy bias). Statistically significant effects were estimated for developmental interventions on social communication outcomes (g=0.31, 95% confidence interval 0.13 to 0.49) and for NDBIs on measures of diagnostic characteristics of autism (0.44, 0.20 to 0.68) and social communication outcomes (0.36, 0.23 to 0.49).

Estimated effects from randomized controlled trials excluding all outcomes subject to high risk of detection bias

Figure 5 shows summary effects estimated exclusively from outcomes that were extracted from randomized controlled trials where assessors were naive to group assignment. Summary effect estimation was possible for NDBIs only on measures of diagnostic characteristics of autism (g=0.30, 95% confidence interval 0.03 to 0.57), and cognitive (0.17, −0.02 to 0.37), language (0.06, −0.13 to 0.025), and social communication outcomes (0.11, −0.03 to 0.26). Only the effect on measures of diagnostic characteristics of autism was statistically significant.

Moderator analyses of proximity and boundedness

Meta-regression analyses across the entire dataset suggest that summary effects were significantly smaller for distal outcomes compared with proximal outcomes (B=−0.15, P=0.002). Additionally, effect sizes coded as generalized were significantly smaller than those coded as context bound (B=−0.27, P<0.001).

Publication bias

Results from the Egger multilevel meta-analysis test[31] indicated that, across all outcomes analyzed in our models, there was evidence for funnel plot asymmetry (B=2.98, P<0.001). There was still evidence of funnel plot asymmetry when we restricted outcomes to only those from randomized controlled trials (B=5.86, P=0.01), only those from randomized controlled trials that did not include caregiver or teacher outcomes (B=8.18, P<0.001), and only those from randomized

controlled trials that were at low risk of detection bias (B=2.26, P=0.03).

Looking within interventions by outcome type (see table S1 and figures S1-S8), limited evidence was found for small study or publication bias within each outcome type with a sufficient number of effect sizes and clusters. After correcting for multiple comparisons with a Benjamini-Yekutieli false discovery rate correction,[34] no outcome type within any intervention was identified with a statistically significant result for the Egger multilevel meta-analysis test. Therefore, the evidence indicates that small sample or publication bias influenced the results, but the extent to which publication bias could have influenced the results for individual intervention types is not clear.

Adverse events, effects, and harms

When reports across both search periods were considered together, 10% mentioned adverse events, and of these, 66% reported that no adverse events occurred, 34% reported that adverse events occurred, and 17% reported that adverse effects occurred. Additionally, only 28% of articles that mentioned adverse events reported monitoring procedures for determining if adverse events or effects occurred. None of the reports in both search periods mentioned harms, or indicated any intention to monitor harms after the end of the intervention. The number of reported adverse events across studies ranged from 0 to 67. Interestingly, three quarters (76%) of the studies that mentioned adverse events, but did not describe any monitoring procedures, reported that no adverse events occurred. In contrast, only half (50%) of studies that described at least some adverse event monitoring procedures reported that no adverse events occurred. Therefore, it is possible that the frequency with which adverse events are reported to occur is dependent upon the robustness and transparency of procedures for monitoring them. Descriptions of adverse events extracted from studies that reported this information are provided in the supplementary materials. Examples of adverse events that could be attributed to the intervention were intense child aggression and serious adverse effects on parent mental health after participating in a parent mediated intervention (see Bottema-Beutel and colleagues[11] for additional examples of adverse events extracted from studies during the initial search period).

Discussion

The purpose of this systematic review and meta-analysis was to provide an updated summary of evidence on interventions designed for young autistic children that includes more recently published studies. In only four years after our initial search for Project AIM, the available evidence has doubled, including the number of randomized controlled trials (from 87 in our original report to 173 in the current report). Three quarters of all controlled group design tests of interventions and 80% of all randomized controlled trials were published in the past decade, which means that even intervention guidelines based on relatively recent evidence reviews[35][36] no longer reflect most available evidence from controlled group design studies.

Findings by intervention type

As in the previous meta-analysis, we estimated the effects for behavioral interventions, developmental interventions, NDBIs, and technology based interventions. However, there were still not enough studies to reliably estimate the summary effects for animal assisted interventions or cognitive behavioral therapy (although cognitive behavioral therapy is typically recommended for older populations). Because we restricted summary effect estimation to those from randomized controlled trials, we were also unable to estimate the summary effects of TEACCH on any outcome (though see supplementary figure 1 for estimates of summary effects when quasi-experimental studies were included). We further divided the previous sensory based intervention category into three groups (music therapy, sensory integration therapy, other sensory based interventions) to ensure that intervention categories had consistent theories of change. We found that these categories also had too few controlled studies to reliably estimate summary effects.

Behavioral interventions

In the US, traditional behavioral interventions are the most frequently recommended intervention approach for autistic children.[37] When evidence from randomized controlled trials is considered, it appears that behavioral interventions might have moderate positive effects on social emotional or challenging behavior outcomes. These estimates are mainly driven by effects from unmasked caregiver or teacher report measures. However, this finding differs from our previous report, in which not enough randomized controlled trials of behavioral interventions had been conducted to allow summary estimation of any effects. Several randomized controlled trials were published in the four years after the original search date; however, this increase was mostly due to randomized tests of focused behavioral interventions (eg, functional communication training, prevent teach reinforce, predictive parenting), and not by randomized tests of early intensive behavioral intervention or other comprehensive behavioral approaches. There was only one randomized controlled trial of early intensive behavioral intervention added to our current sample, which tested randomized comparisons of this treatment and the early start Denver model (an NDBI) delivered at various intensities (and found that neither intervention nor intensity was associated with superior effects on measured outcomes).[38] More studies that compare robust delivery of two different but commonly recommended comprehensive interventions are needed. Given that behavioral interventions are routinely recommended for this population, it is vital that more controlled tests with unbiased measures are

conducted, and that support recommendations are altered to reflect the current evidence base.

Developmental interventions

As in our original meta-analysis, estimates from randomized controlled trials suggest developmental interventions have positive and statistically significant effects on social communication, even when outcomes at risk of placebo-by-proxy bias are excluded. However, because developmental researchers frequently relied on observational measures of social communication derived from interactions with unmasked caregivers to indicate developmental intervention effects, our confidence in this estimate is limited by the risk of detection bias. Therefore, we recommend that intervention researchers estimate social communication effects (or effects on any domain) using assessments that use masked assessors, and not only masked coders; this can be done even when deriving variables from communication samples by using a masked clinician who is skilled at facilitating natural and responsive interactions.

Naturalistic developmental behavioral interventions

Despite only being named as a category in 2015,[16] NDBIs are now the most frequently studied intervention approach for this population. Summary effect estimates from randomized controlled trials suggest that NDBIs might improve adaptive behavior, language, play, social communication, and measures of diagnostic characteristics of autism, though our confidence in these estimates is limited by reliance on measures that are subject to high detection bias. In contrast to our previous meta-analysis, we found a statistically significant positive effect of NDBIs on measures of the diagnostic characteristics of autism, even when effects were restricted to outcomes from randomized controlled trials with low risk of detection bias. This was the only intervention effect we were able to estimate when accounting for all risks of bias considered. Given that diagnostic measures index core features of autism (ie, repetitive patterns of behaviors, interests, or activities, and social communication challenges), and that our other estimates suggest NDBIs have null effects on repetitive patterns of behaviors, interests, or activities but positive effects on social communication outcomes, it is likely that the positive and statistically significant estimates of NDBIs on overall measures of the diagnostic characteristics of autism were driven by improvements in social communication. We were unable to estimate summary effects of NDBIs on outcomes categorized as social communication alone when outcomes at high risk of detection bias were excluded because social communication outcomes were frequently measured using interactions with unmasked assessors (eg, caregivers) in studies of NDBIs. Most of these effects were excluded when detection bias was considered, and diagnostic measures of autism remained the only masked assessments indicating improvements in social communication in these studies. Therefore,

we conclude that there is relatively strong evidence that NDBIs can have positive effects on core features associated with autism, specifically social communication differences.

Using measures of diagnostic characteristics of autism (eg, Autism Diagnostic Observation Schedule, Childhood Autism Rating Scale) to index intervention outcomes could be problematic because interpreting "improved" scores on such measures as evidence of a positive intervention effect implies that the goal of an intervention is to make a child "less autistic." Although these measures indicate difficulties associated with autism diagnosis, they also capture neutral or even beneficial aspects of autism (such as special interests). If interventions are only intended to support improvements in social communication, it is important that researchers use high quality masked assessments of this specific construct to index intervention effects, such as the Communication and Symbolic Behavior Scales[39] or the Early Social Communication Scales[40] as administered and coded by people who are naive to children's group assignment.

Technology based interventions

The number of studies of technology based interventions nearly tripled (from 10 to 30 studies) in the four years between the original and updated searches. However, most of the high quality evidence supporting the potential benefits of technology based interventions reflects effects on proximal, circumscribed outcomes. Technology based interventions could have broad appeal because they tend to be new and motivating, use predictable formats, and have the potential to increase access for those who might otherwise have difficulty accessing intervention and supports. For these reasons, intervention developers might wish to integrate technological supports into more established intervention approaches to help the development of specific skills.

Other interventions

Categories of intervention approaches that were too sparsely represented in the literature to allow summary effect estimation for autistic children between birth and 8 years included animal assisted therapy, cognitive behavioral therapy, music therapy, sensory integration therapy, other sensory based interventions, and TEACCH. Given that these interventions are frequently prescribed for and used by this population, there is a need for more rigorous research evaluating the efficacy of such approaches. In the meantime, physicians guiding families toward interventions should bear in mind that the current evidence base is limited, and they should keep up to date with emerging literature.

Proximity and boundedness

As expected, we found evidence that intervention effects are greater on proximal than distal outcomes, and for context bound than generalized outcomes, replicating the findings of our previous work.[8 41] These

findings are consistent with the conclusions of a recent review of early interventions for children with or at high likelihood of receiving an autism diagnosis.[42] Therefore, researchers who measure outcomes that are not designed to show sustained developmental change are likely to observe stronger effects and to draw more positive conclusions about intervention efficacy than is warranted. Intervention is often recommended for autistic children in early childhood at high intensities and for long durations based on the assumption that this approach is necessary to support generalized developmental improvements. If the goal of evidence summaries is to determine whether early childhood interventions support such improvements, controlled group design studies that use measures well equipped to tap broad developmental change offer the most relevant evidence. Therefore, in evaluating evidence and guiding families toward early childhood interventions, it is important that physicians and other clinicians consider not only the quality of evidence, but also the scope of change reflected by the outcomes.[12]

Current trends and future research

The doubling of randomized controlled trials in only four years suggests that it is possible for autism researchers to conduct randomized controlled trials, and that randomized tests of interventions should be a minimum standard for establishing intervention efficacy. Consequently, physicians should refrain from drawing firm conclusions when available evidence is largely quasi-experimental in nature, and researchers should continue to treat random assignment as a crucial component of experimental design.

Outcomes derived from caregiver or teacher report measures comprised a substantial subset (nearly half) of all outcomes, and the percentage of these report measures even increased between the original and updated samples. Physicians evaluating evidence from randomized controlled trials of specific intervention approaches should keep this in mind and exercise caution in interpreting results that emphasize statistically significant effects on outcomes measured by proxy reports and unmasked assessments with minimal reporting of effects on more stringent outcome measures.

Robust intervention evaluations require that researchers assess, interpret, and report adverse events, adverse effects, and harms.[43] However, these remain infrequently monitored or reported in the autism intervention literature. Despite this lack of monitoring, however, there is some evidence to suggest that adverse events and adverse effects might be relatively common.[11] One promising finding from our updated review is the increased frequency of describing adverse event monitoring procedures: 44% of studies that mentioned adverse events provided at least some description of their approach to monitoring procedures. Moving forward, researchers should develop definitions of adverse events and procedures for monitoring them that are tailored by intervention type, and shared across research groups. These measures could include procedures for active monitoring for adverse events that might occur within the intervention sessions (eg, child distress, injury from intervention equipment, aggression), and additional procedures for active monitoring of events that could occur outside of the intervention sessions (eg, sleep disturbance, changes in eating habits, anxiety, parental distress). Because anecdotal reports and qualitative evidence suggest autistic adults might have experienced long term harms from participation in specific interventions,[44] and because they have expressed that research about potential intervention harms is among their top research priorities,[43 45] researchers should make concerted efforts to follow participants over longer periods of time to document the potential for sustained negative impacts of interventions. Until such information is explored by researchers, families and practitioners will continue to have little basis on which to weigh the potentially positive effects of interventions against the potential for negative impacts.

Strengths and limitations

This updated meta-analysis has several strengths. Our search of published and unpublished literature ensured that we identified and retrieved as many potentially eligible studies as possible. Our double screening and double coding process and high reliability suggest our data reliably reflect attributes of included studies and outcomes. Additionally, our statistical methods ensured that we estimated summary effects as precisely as possible, while accounting for the intercorrelated structure of the data and retaining power in subgroup analyses.[27] Our coding process also accounted for study and outcome level risks of bias, and outcome characteristics that are often ignored (ie, boundedness and proximity), ensuring appropriate caution in interpreting results.

Our investigation also has a number of limitations. Although our procedures closely followed those used in the previous meta-analysis, this update was not registered. Even though we reliably categorized most interventions and our intervention categories have proved useful to other reviews,[10] approximately a quarter of studies featured interventions that were categorized as other because they did not cluster together in terms of their theory of change and were not reflected in the current summary estimates. However, effect sizes from these studies were reflected in moderator analyses (ie, for proximity and boundedness) and can be included in future meta-regressions that investigate whether specific participant or intervention characteristics are associated with stronger effects. There were a range of studies in this catch-all category which describe interventions that might deserve further attention given that they support improvements in domains that are often regarded as impairments by autistic people[46] and implicated in important developmental processes, but rarely targeted by intervention. These range from specialized intervention approaches that support sleep[47] and management of eating

aversions[48] to academic programs that support reading acquisition.[49]

Clinical recommendations

Authors of other prominent autism evidence reviews based their clinical recommendations on rubrics specifying different thresholds of study quality.[35 50] For example, the National Clearinghouse on Autism Evidence and Practice designated practices as evidence based if they were supported by two peer reviewed quasi-experimental or randomized controlled group studies conducted by at least two research groups; five single case design studies conducted by at least three research groups and featuring a minimum of 20 participants in total; or a combination of one controlled group study and three single case design studies conducted by at least two research groups.[50] In contrast, the National Autism Center designated interventions as established if they were supported by two quasi-experimental or randomized controlled group studies or four single case design studies with at least 12 participants and consistent effects.[35] Because we recognize that the designation of a single threshold as sufficient for drawing conclusions is somewhat arbitrary and varies by reviewer, we elected to transparently present summary effects associated with increasingly lower risks of bias (and higher levels of certainty). Consistent with our previous findings which were derived using a similar analytic approach, we observed that because contributing effects were increasingly excluded based on risk of bias, fewer summary effects could be estimated, and associated confidence intervals were wider. Consequently, as quality thresholds increased, fewer summary effects could be synthesized with a high degree of confidence (df≥5), and crossed the threshold for statistical significance, even though the magnitude of summary effect estimates was often similar. To some extent, this observation is a function of our analytic approach. Raising quality standards will often reduce the sample of studies and effects that meet such standards, and so reduce power to detect a true effect that might be present. By presenting estimates associated with increasing levels of confidence, we hoped to show which clinical recommendations might be supported by some evidence (ie, evidence from randomized controlled trials but including effects at risk of detection and placebo-by-proxy bias), and which might be supported by the best available evidence (ie, evidence from randomized controlled trials excluding effects at high risk of detection bias).

Given that there are few intervention approaches at present with the best available evidence supporting their efficacy for improving developmental outcomes, what intervention recommendations should medical professionals make? The Australian government recently updated their national guideline for supporting autistic children and their families, and integrated a framework for making ethical support recommendations.[10 51] The framework suggests that supports should be plausible (have a clear mechanism of effectiveness and be supported by the best available evidence), practical (feasible to deliver in local conditions), desirable (consistent with child wants and needs, and family priorities), and defensible (benefits outweigh effort and opportunity costs, and will be viewed positively by the child later in life). Drawing on this framework, we recommend that physicians guide families toward interventions with the most robust evidence supporting their efficacy for improving the intended outcomes, provided that the supports can be offered in a way that integrates and strengthens child wellbeing and family routines rather than disrupting them. For example, interventions provided within the home, embedded in daily routines, and focused on strengthening caregiver capacities to support development are less likely to disrupt child and family wellbeing than interventions provided at high intensities in clinics directly to the child by clinicians. Clinicians should ensure that they have adequate systems for monitoring whether the selected intervention promotes progress in terms of acquiring specific skills in specific contexts, and also in the broader, generalized development of these children. Finally, it cannot be assumed that interventions and supports are harmless, so physicians should advise families to monitor for indicators of negative effects and child or family distress.

Conclusion

Studies investigating interventions for young autistic children have proliferated at an astonishing rate, but corresponding improvements in study quality have not kept pace. Some high quality evidence exists, which suggests that NDBIs can improve core features associated with autism. However, it is not clear if such outcomes are desirable for autistic people given that measures of core features of autism are not restricted to impairments that need to be addressed to positively influence autistic development. Interventions tend to have larger effects on small and specific changes in specific contexts, and smaller effects on distal and generalized developmental improvement. We are unable to weigh the potential benefits of any intervention against the potential for unintended negative consequences because most researchers are not adequately monitoring and reporting adverse events.

AUTHOR AFFILIATIONS

[1]Division of Occupational Science and Occupational Therapy, Department of Health Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[2]Lynch School of Education and Human Development, Boston College, Chestnut Hill, MA, USA

[3]TEACCH Autism Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[4]Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

[5]Frist Center for Autism and Innovation, Vanderbilt University, Nashville, TN, USA

[6]University of Alabama at Birmingham, Birmingham, AL, USA

[7]Department of Curriculum and Instruction, University of Arkansas, Fayetteville, AR, USA

[8]Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN, USA

[9]Austin, TX, USA

[10]Department of Special Education, University of Texas at Austin, Austin, TX, USA

[11]Vanderbilt Kennedy Center, Nashville, TN, USA

[12]Department of Communication Sciences and Disorders, John A Burns School of Medicine, University of Hawaii at Manoa, Honolulu, HI, USA

1 Bradshaw J, Eberth JM, Zgodic A, Federico A, Flory K, McLain AC. County-level prevalence estimates of autism spectrum disorder in children in the United States. *J Autism Dev Disord* 2023 4 May. https://link.springer.com/10.1007/s10803-023-05920-z

2 O'Nions E, Petersen I, Buckman JE, et al. Autism in England: assessing underdiagnosis in a population-based cohort study of prospectively collected primary care data. *Lancet Reg Heal* 2023. https://www.thelancet.com/journals/lanepe/article/PIIS2666-7762(23)00045-5/fulltext

3 Wallis KE, Adebajo T, Bennett AE. Prevalence of autism spectrum disorder in a large pediatric primary care network[Internet]. *Autism* 2023;27:1840-6. https://journals.sagepub.com/doi/10.1177/13623613221147396. doi:10.1177/13623613221147396

4 Hyman SL, Levy SE, Myers SM, et al. Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics* 2020;145: e2019344710.1542/peds.2019-3447.

5 Lord C, Charman T, Havdahl A, et al. The Lancet Commission on the future of care and clinical research in autism. *Lancet* 2022;399:271-334. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)01541-5/fulltext. doi:10.1016/S0140-6736(21)01541-5

6 Reichow B, Hume K, Barton EE, Boyd BA. Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). Cochrane Database Syst CD009260. 2018.

7 National Institute for Health and Care Excellence. Autism spectrum disorder in under 19s: support and management. 2021. www.nice.org.uk/guidance/cg170

8   Sandbank M, Bottema-Beutel K, Crowley S, et al. Project AIM: Autism intervention meta-analysis for studies of young children. *Psychol Bull* 2020;146:1-29. doi:10.1037/bul0000215

9   Interagency Autism Coordinating Committee. 2020 IACC Summary of Advances in Autism Research. 2022 May; Available from: https://iacc.hhs.gov/publications/summary-of-advances/2020/

10 Trembath D, Varcin K, Waddington H, et al. *National guideline for supporting the learning, participation, and wellbeing of autistic children and their families in Australia.* Autism CRC, 2022.

11 Bottema-Beutel K, Crowley S, Sandbank M, Woynaroski TG. Adverse event reporting in intervention research for young autistic children. *Autism* 2021;25:322-35. doi:10.1177/1362361320965331

12 Sandbank M, Chow J, Bottema-Beutel K, Woynaroski T. Evaluating evidence-based practice in light of the boundedness and proximity of outcomes: capturing the scope of change. *Autism Res* 2021;14:1536-42. https://onlinelibrary.wiley.com/doi/10.1002/aur.2527. doi:10.1002/aur.2527

13 Office of Autism Research Coordination, National Institutes of Health on behalf of the Office of the Secretary, Department of Health and Human Services. Report to Congress on Activities Related to Autism Spectrum Disorder and Other Developmental Disabilities Under the Combating Autism Act of 2006 (FY 2006–FY 2009). 2010. https://iacc.hhs.gov/publications/report-to-congress/2009/report_to_congress_2009.pdf

14 Wallace BC, Small K, Brodley CE, Lau J, Trikalinos TA. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In: *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium.* ACM. 2012: 819-24. https://dl.acm.org/doi/10.1145/2110363.2110464. doi:10.1145/2110363.2110464.

15 Sandbank M. Project AIM. 2022. osf.io/kr2cd.

16 Schreibman L, Dawson G, Stahmer AC, et al. Naturalistic developmental behavioral interventions: empirically validated treatments for autism spectrum disorder. *J Autism Dev Disord* 2015;45:2411-28. http://link.springer.com/10.1007/s10803-015-2407-8. doi:10.1007/s10803-015-2407-8

17 Case-Smith J, Weaver LL, Fristad MA. A systematic review of sensory processing interventions for children with autism spectrum disorders. *Autism* 2015;19:133-48. https://journals.sagepub.com/doi/10.1177/1362361313517762. doi:10.1177/1362361313517762

18 Parham LD, Roley SS, May-Benson TA, et al. Development of a fidelity measure for research on the effectiveness of the Ayres Sensory Integration intervention. *Am J Occup Ther* 2011;65:133-42. https://research.aota.org/ajot/article-abstract/65/2/133/5467. doi:10.5014/ajot.2011.000745

19 Higgins JPT, Sterne JAC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials. In: Chandler J, McKenzie J, Boutron I, Welch V, et al, eds. *Cochrane Methods. Cochrane Database Syst Rev* CD201601. 2016. doi:10.1002/14651858.CD201601

20 Grelotti DJ, Kaptchuk TJ. Placebo by proxy. *BMJ* 2011;343:d4345. 2011. https://www.bmj.com/content/343/bmj.d4345.full

21 Peryer G, Golder S, Junqueira DR, Vohra S, Loke YK, et al, Cochrane Adverse Effects Methods Group. Adverse effects. In: Higgins JPT, Thomas J, Chandler J, eds. *Cochrane Handbook for Systematic Reviews of Interventions.* 1st ed. Wiley, 2019: 493-505, Available from https://onlinelibrary.wiley.com/doi/10.1002/9781119536604.ch19. doi:10.1002/9781119536604.ch19.

22 Wilson DB. Practical Meta-Analysis Effect Size Calculator [online calculator]. https://campbellcollaboration.org/research-resources/effect-size-calculator.html

23 R Core Team R. R. A language and environment for statistical computing. 2013. https://apps.dtic.mil/sti/citations/AD1039033

24 Gamer M. irr: Various coefficients of interrater reliability and agreement. Httpcran R-Proj Orgwebpackagesirrirr Pdf. 2010. https://cir.nii.ac.jp/crid/1572824499179920768

25 Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37-46. https://journals.sagepub.com/doi/10.1177/001316446002000104. doi:10.1177/001316446002000104.

26 McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1:30. https://psycnet.apa.org/journals/met/1/1/30/. doi:L10.1037/1082-989X.1.1.30.

27 Pustejovsky JE, Tipton E. Meta-analysis with robust variance estimation: Expanding the range of working models. *Prev Sci* 2022;23:425-38. doi:10.1007/s11121-021-01246-3

28 Pustejovsky J. clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections (R package Version 0.5. 5; computer software). 2022.

29 Viechtbauer W. metafor: meta-analysis package for R. Version 2.4-0 (computer software). 2022.

30 Fisher Z, Tipton E. robumeta: An R-package for robust variance estimation in meta-analysis. arXiv; 2015. https://arxiv.org/abs/1503.02220

31 Rodgers MA, Pustejovsky JE. Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychol Methods* 2020;26:141. https://psycnet.apa.org/record/2020-52356-001. doi:10.1037/met0000300

32 Interagency Autism Coordinating Committee (IACC). 2021-2023 IACC strategic plan for autism research, services, and policy. September 2023. Retrieved from US Department of Health and Human Services Interagency Autism Coordinating Committee website. https://iacc.hhs.gov/publications/strategic-plan/2023

33 Sandbank M. Project AIM. 2023. doi:10.17605/OSF.IO/KR2CD

34 Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 2001;1165-88. https://www.jstor.org/stable/267407510.1214/aos/1013699998.

35 National Autism Center. *Findings and conclusions: national standards project, phase 2.* Randolph MA Natl Autism Cent, 2015.

36 Roberts JM, Williams K, Smith K, Campbell L. *Autism spectrum disorder: Evidence-based/evidence-informed good practice for supports provided to preschool children, their families and carers.* Rep Prep Natl Disabil Insur Agency NDIA, 2016.

37 Reichow B. Overview of meta-analyses on early intensive behavioral intervention for young children with autism spectrum disorders. *J Autism Dev Disord* 2012;42:512-20. doi:10.1007/s10803-011-1218-9

38 Rogers SJ, Yoder P, Estes A, et al. A multisite randomized controlled trial comparing the effects of intervention intensity and intervention style on outcomes for young children with autism. *J Am Acad Child Adolesc Psychiatry* 2021;60:710-22. doi:10.1016/j.jaac.2020.06.013

39 Wetherby AM, Prizant BM, Barry M. Communication and symbolic behavior scales. 1990. https://psycnet.apa.org/doiLanding?doi=10.1037/t11527-000

40 Mundy P, Delgado C, Block J, Venezia M, Hogan A, Seibert J. Early social communication scales (ESCS). Coral Gables, FL. 2003. http://education.ucdavis.edu/sites/main/files/file-attachments/escs_manual_2003_2013.pdf

41 Yoder PJ, Bottema-Beutel K, Woynaroski T, Chandrasekhar R, Sandbank M. Social communication intervention effects vary by dependent variable type in preschoolers with autism spectrum disorders. *Evid Based Commun Assess Interv* 2013;7:150-74. https://www.tandfonline.com/doi/abs/10.1080/17489539.2014.917780. doi:10.1080/17489539.2014.917780

42 Franz L, Goodwin CD, Rieder A, Matheis M, Damiano DL. Early intervention for very young children with or at high likelihood for autism spectrum disorder: an overview of reviews. *Dev Med Child Neurol* 2022;64:1063-76. https://onlinelibrary.wiley.com/doi/10.1111/dmcn.15258. doi:10.1111/dmcn.15258

43 Dawson M, Fletcher-Watson S. When autism researchers disregard harms: a commentary. *Autism* 2022;26:564-6. https://journals.sagepub.com/doi/10.1177/13623613211031403. doi:10.1177/13623613211031403

44 Anderson LK. Autistic experiences of applied behavior analysis. *Autism* 2023;27:737-50. https://journals.sagepub.com/doi/10.1177/13623613221118216. doi:10.1177/13623613221118216

45 Benevides TW, Shore SM, Palmer K, et al. Listening to the autistic voice: mental health priorities to guide research and practice in autism from a stakeholder-driven project. *Autism* 2020;24:822-33. https://journals.sagepub.com/doi/10.1177/1362361320908410. doi:10.1177/1362361320908410

46 Pellicano E, Dinsmore A, Charman T. What should autism research focus upon? Community views and priorities from the United Kingdom. *Autism* 2014;18:756-70. https://journals.sagepub.com/doi/10.1177/1362361314529627. doi:10.1177/1362361314529627

47 Malow BA, Adkins KW, Reynolds A, et al. Parent-based sleep education for children with autism spectrum disorders. *J Autism Dev Disord* 2014;44:216-28. http://link.springer.com/10.1007/s10803-013-1866-z. doi:10.1007/s10803-013-1866-z

48 Sharp WG, Burrell TL, Berry RC, et al. The autism managing eating aversions and limited variety plan vs parent education: a randomized clinical trial. *J Pediatr* 2019;211:185-192.e1. https://www.sciencedirect.com/science/article/pii/S0022347619304032. doi:10.1016/j.jpeds.2019.03.046

49 Hudson RF, Sanders EA, Greenway R, et al. Effects of emergent literacy interventions for preschoolers with autism spectrum disorder. *Except Child* 2017;84:55-75. https://journals.sagepub.com/doi/10.1177/0014402917705855. doi:10.1177/0014402917705855

50 Steinbrenner JR, Hume K, Odom SL, et al. Evidence-based practices for children, youth, and young adults with autism. FPG Child Dev Inst. 2020. https://eric.ed.gov/?id=ED609029

51 Trembath D, Waddington H, Sulek R, et al. An evidence-based framework for determining the optimal amount of intervention for autistic children. *Lancet Child Adolesc Health* 2021;5:896-904. https://www.thelancet.com/journals/lanchi/article/PIIS2352-4642(21)00285-6/fulltext. doi:10.1016/S2352-4642(21)00285-6

Web appendix: Supplementary materials