The Journal of Child Psychology and Psychiatry

Journal of Child Psychology and Psychiatry **:* (2019), pp **–** doi:10.1111/jcpp.13073

Effectiveness of community-based early intervention for children with autism spectrum disorder: a meta-analysis

Allison S. Nahmias,[1] Melanie Pellecchia,[2] Aubyn C. Stahmer,[1] and David S. Mandell[2]

[1]Department of Psychiatry and Behavioral Sciences, MIND Institute, University of California, Davis, Sacramento, CA, USA; [2]Center for Mental Health, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Background: Research trials of early intervention (EI) programs for children with autism spectrum disorder (ASD) generally demonstrate medium-to-large gains, on average, compared with "treatment as usual," in different developmental domains. Almost all children with ASD receive their treatment through community-based services, however, and studies suggest that evidence-based interventions rarely make their way into community practice. Understanding the effectiveness of community-based EI and factors associated with these effects is the first step in developing strategies for wide-scale implementation of effective EI. **Methods:** Studies of community-based EI for children with ASD were identified through a systematic search. Changes in cognitive, communication, social, and adaptive functioning from pre-treatment to post-treatment were assessed using standardized mean gain scores. Effect sizes were estimated using random effects models. Moderators of interest included type of community EI program, year of publication, intervention duration, and sample selection. Moderator effects were assessed using analysis of variance of mixed-effects models and meta-regression analyses. **Results:** Forty-six groups from 33 studies met inclusion criteria (1,713 participants, mean age 37.4 months, 81.1% male). There were small but statistically significant gains in each of the four domains. Hedges's g ranged from 0.21 for adaptive behavior to 0.32 for communication outcomes, after removing outliers and correcting for publication bias. EI programs associated with universities and hospitals were superior, on average, to other community EI programs for cognitive and adaptive behavior outcomes. Intervention duration was negatively associated with effect sizes for communication and adaptive behavior outcomes. **Conclusions:** These results indicate that there remains a large gap between outcomes observed in community settings and those reported in efficacy trials. **Keywords:** Autism spectrum disorders; meta-analysis; early intervention; community programmes.

Introduction

Early intervention provided by university-based experts can result in large gains in cognition, communication, social skills, and adaptive behavior for young children with autism spectrum disorder (ASD). Early intensive behavior interventions (EIBI) and those based on applied behavior analysis (ABA) have the most evidence. Systematic reviews and meta-analyses of randomized trials find positive effects of both highly-structured, behaviorally-based early intervention and naturalistic developmental behavioral interventions on many developmental domains for young children with ASD (Murza, Schwartz, Hahs-Vaughn, & Nye, 2016; Nevill, Lecavalier, & Stratis, 2018; Reichow, 2012). For example, a meta-analysis by Ospina et al. (2008) demonstrated statistically and clinically significant positive effects on intellectual functioning, adaptive behavior, communication and language when high-intensity Lovaas-based ABA intervention was compared with either low-intensity Lovaas ABA, or special education (standardized mean difference = 0.92–0.95). A Cochrane Collaboration systematic review and meta-analysis of studies comparing EIBI to treatment as usual in the community found medium to large significant positive effects of EIBI for adaptive behavior ($g = 0.69$), language ($g = 0.50$–0.57), daily communication skills ($g = 0.74$), IQ ($g = 0.76$), socialization ($g = 0.42$), and daily living skills ($g = 0.55$; Reichow, Barton, Boyd, & Hume, 2012).

These studies are encouraging, but do not speak to probable outcomes in care received outside of research studies, otherwise known as "treatment as usual." Little research has examined outcomes for children with ASD in the community who receive routine clinical and educational care. The few studies reporting child outcomes in community-based intervention often include data from a single site or a small sample, limiting their generalizability (Ben-Itzchak, Watson, & Zachor, 2014; Magiati, Charman, & Howlin, 2007). As Kasari and Smith (2016) note, "most children with ASD have never been in a research study and most interventions applied to children with ASD have never been tested" (p. 263). Although some research demonstrates that community providers can learn to deliver evidence-based interventions for ASD effectively when research teams train them (Shire & Kasari, 2014; Shire et al.,

Conflict of interest statement: No conflicts declared

Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA 02148, USA

2017), school-based providers do not typically incorporate evidence-based practices with fidelity over the long-term (Cook, Landrum, Tankersley, & Kauffman, 2003; Mandell et al., 2013; Suhrheinrich et al., 2013). In addition to regularly altering and modifying evidence-based practices to fit child needs and resource constraints, providers also report frequently using practices that do not align with the evidence base (Hess, Morrier, Heflin, & Ivey, 2008; Stahmer, Collings, & Palinkas, 2005).

The gap between research and practice is well-documented in psychology, medicine, and special education (Balas & Boren, 2000; Greenwood & Abbott, 2001; Williams & Beidas, 2018). To begin to bridge this gap it is important to first understand the current state of the effectiveness of routine clinical care for young children with ASD. Many factors (e.g., clinical expertise, use of evidence-based practices, geographic location, intervention dose, child characteristics) may be associated with community outcomes. If research-based evidence is making its way into routine practice for children with ASD, improvements in community outcomes over time would be expected. However, to date, no systematic review or meta-analysis has examined community outcomes for young children with ASD in public programs and routine clinical care. The present study leverages studies of community EI for ASD and control groups from trials of early intervention, combining community groups across studies to provide a rigorous assessment of outcomes from community-based interventions and explore patterns in the results.

In the present study, we: (a) determine the overall effect sizes for community-based EI for ASD in four outcome domains, and (b) assess moderators of these effects.

Methods

Search procedures and selection of studies

Studies were included that met these criteria:

1. Published, written in English;
2. Prospective, pre-test, post-test group design;
3. Presented outcomes for children with ASD separately;
4. Included more than 10 children with ASD receiving community-based intervention;
5. Child age at study intake was <73 months, which corresponds with the typical age of early intervention in the United States;
6. Provided information on outcomes of educational or behavioral services available in the community or treatment as usual (e.g., publicly funded services, routine clinical care) for a defined group or subgroup of study participants. This included school, clinic, hospital, and home-based interventions that were routinely available. Groups that received intervention provided by researchers or groups where research funds were used to pay for training to community providers were excluded as they did not reflect standard community practice and routinely available services;
7. Outcome measures included at least one of the following, reported as standard scores or developmental quotients:
   a. Cognitive: Early Learning Composite from the Mullen Scales of Early Learning (MSEL, Mullen (1995)) or Full Scale IQ from a standardized cognitive assessment that included both verbal and non-verbal IQ (e.g., Differential Abilities Scale (Elliott, 1990). Studies that only included non-verbal IQ or a cognitive measure that only assessed non-verbal IQ (e.g., Merrill-Palmer (Roid & Sampers, 2004)) were excluded to minimize measurement differences;
   b. Communication: Vineland Adaptive Behavior Scales (VABS) Communication domain (Sparrow, Balla, & Cicchetti, 1985; Sparrow, Cicchetti, & Balla, 2005);
   c. Social: VABS Socialization domain (Sparrow et al., 1985, 2005);
   d. Adaptive Behavior: VABS Composite (Sparrow et al., 1985, 2005); and
8. Reported unadjusted pre- and post- intervention means and standard deviations for outcome measures (based on recommendation from the What Works Clearinghouse (2014)).

Studies only reporting follow-up data were excluded. For studies with overlapping (or potentially overlapping) samples, the study with the largest sample for each outcome was used.

A systematic search of research databases was conducted through January 2018 to identify relevant studies. PsycINFO and Medline databases were searched for terms related to *autism* and *intervention* (see Appendix S1 for a sample search strategy). The reference list of retrieved articles, existing reviews, and meta-analyses were also examined for eligible studies.

Studies were first screened for eligibility based on title and abstract using these exclusion criteria: (a) did not include children with ASD, (b) $n \leq 10$ children with ASD, (c) not written in English, (d) participants outside the age range, and (e) did not include a behavioral intervention. The first author and two reliable coders (85% agreement between coders and first author) conducted screening and then full-text review for eligibility. The first author completed final review based on inclusion criteria (described above). See Figure 1 for the PRISMA flow diagram (Liberati et al., 2009).

Study coding

The first author extracted and coded data regarding participant, intervention, and study characteristics presented in Tables S1 and S2. Interobserver agreement between the first author and a reliable coder was calculated for 18 groups from nine studies. Percent agreement ranged from 83% to 100% for all extracted and coded moderators, except for total hours of intervention, which was 78%. EI programs were categorized as follows to reflect the range of services that children experience in the community: (a) “Model” programs (MO): intervention in the community associated with universities and/or hospitals (e.g., university laboratory schools) that offered routinely available clinical care, (b) “Specified treatment as usual” (SPC): treatment from a local school/agency or standard educational provisions that had defined procedures described in the article (including programs that reported using eclectic interventions), and (c) “Unspecified treatment as usual” (USP): a wait-list group or services as usual in the community where participants received an undefined variety of services. Study design was coded based on the following criteria: (a) “Single” (SI): reported outcomes from one community program, (b) “Comparison to other community groups” (CC): compared community programs, (c) “Comparison to experimental group” (CE): the community

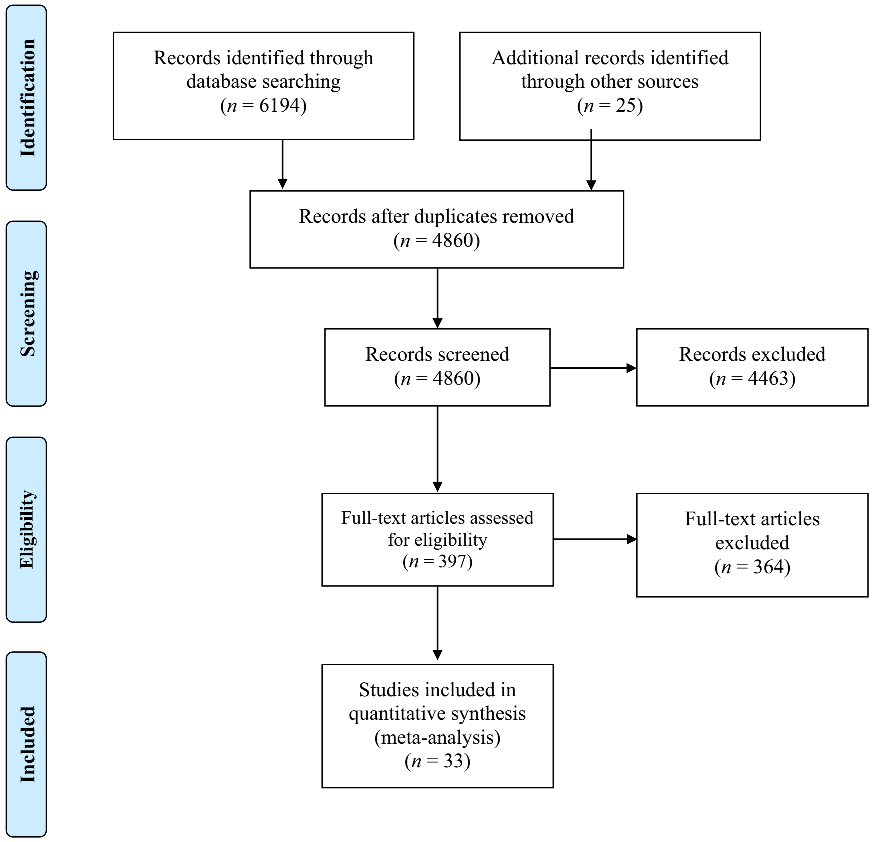

Figure 1 PRISMA flow diagram

group was the control group for a researcher-implemented treatment, and (d) "Other:" did not fit the other categories. Treatment group allocation was coded based on the following criteria: (a) "Random assignment" (RA): participants assigned to groups utilizing random assignment, (b) "Standard assignment" (SA): participants assigned to groups based on standard community practice, (c) "Not applicable" (NA): included a single treatment group, and (d) "Other:" did not fit the other categories.

Analyses

All outcome data were continuous. Changes between baseline and posttreatment assessments were evaluated using standardized mean gain scores. As shown in the equation below, effect sizes were calculated by dividing the mean change from baseline to post-treatment by the pooled standard deviation of the difference score and transformed to Hedges's g estimates (Hedges, 1981) to correct potential bias from small sample sizes:

$$g = \left(1 - \frac{3}{4(n-1)}\right) \left[\frac{M_{\text{Post}} - M_{\text{Pre}}}{\sqrt{\left(SD_{\text{Pre}}^2 + SD_{\text{Post}}^2 - 2rSD_{\text{Pre}}SD_{\text{post}}\right)}/\sqrt{(2(1-r))}}\right]$$

No study reported the data needed to calculate pretest-posttest correlations among outcome measures (r in the equation above), therefore test-retest reliabilities from test manuals and published papers were used as proxies (Lipsey and Wilson (2001). Because test-retest reliability may overestimate pretest-posttest correlations, sensitivity analyses with r values of 0.3 (low), 0.5 (medium), and 0.8 (high) were conducted. The calculated effect sizes were similar, so the test-retest reliabilities were used.

Potential outliers were detected using the sample-adjusted meta-analytic deviancy (SAMD) statistic; including extreme studies may overestimate true variability (Huffcutt & Arthur, 1995). A conservative cutoff of $|2.58|$ was used to exclude groups from analyses (Beal, Corey, & Dunlap, 2002). SAMDs were rank-ordered and scree plots were examined to confirm outlier status.

Weighted mean effect sizes, heterogeneity, moderators, and publication bias statistics were calculated using Comprehensive Meta-Analysis Version 3.3.070 (Borenstein, Hedges, Higgins, & Rothstein, 2014). Separate random effects models were conducted for cognitive, communication, social, and adaptive behavior outcomes, as recommended by Borenstein, Hedges, Higgins, and Rothstein (2009) and Lipsey and Wilson (2001). Effect sizes were weighted to account for its relative precision based on the standard error of the effect size and tau-squared using the reciprocal of the squared standard error plus tau-squared. Study quality was not used to weight effect sizes due to inconsistent reporting.

Heterogeneity of effect sizes was examined using the Q and I^2 statistics. Exploratory moderator analyses were conducted for models with a significant Q statistic or an $I^2 \geq 50$, indicating at least moderate heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003). Categorical moderators were examined using an analysis-of-variance (ANOVA) of mixed-effect models for each variable hypothesized to moderate the overall effect size. Meta-regression analyses were used to examine continuous moderators: age of the sample at intake, intervention duration, approximate hours of intervention, and year of publication. Year of publication was used as a proxy for recency of intervention, as <50% of studies reported when the intervention occurred. Adjusted meta-regression was used to examine differences by EI category controlling for other covariates. Due to the small number of studies, only statistically significant moderators were included in adjusted analysis.

To assess publication bias, funnel plots and Duval and Tweedie's (2000) trim-and-fill procedure were calculated.

Model estimates were calculated using the trim-and-fill correction when this test indicated significant asymmetry in the funnel plot.

Results

Study characteristics

Table S1 displays sample characteristics of the 46 groups from 33 studies included in the analysis. Participants were predominantly male (81.1%); mean age was 37.4 months. Thirteen studies (40%) reported information on participant race and ethnicity. On average, studies had 32% non-white participants (range: 14.0–72.6%). Seventeen studies (51.5%) reported parental education or income. ASD diagnosis was based on standardized diagnostic criteria (e.g., DSM-IV) and/or standardized diagnostic measures (e.g., Autism Diagnostic Observation Schedule) in 30 studies (90.9%) and was verified by the research team in 21 of the studies (63.6%). Twenty-six groups (56.5%, 868 participants) reported cognitive outcomes, 27 groups (58.7%, 957 participants) reported communication outcomes, 26 groups (56.5%, 945 participants) reported social outcomes, and 28 groups (60.9%, 1,141 participants) reported adaptive behavior outcomes eligible for inclusion.

Table S2 displays intervention characteristics for included groups. Studies took place in the United States (36.4%), Australia (21.2%), United Kingdom (18.2%), Italy (6.1%), Israel (6.1%), Canada (3.0%), Norway (3.0%), Sweden (3.0%), and Taiwan (3.0%). Only 23 groups (50.0%) reported the years over which the intervention occurred (range: 1995–2003 (Cohen, Amerine-Dickens, & Smith, 2006) to 2013–2014 (Whitehouse et al., 2017)). Mean intervention duration was 13.7 months (range: 3–36 months). Forty groups (82.5%) reported intervention intensity. Treatment groups included specified treatment as usual (47.8%), Model programs (28.2%), or unspecified treatment as usual (23.9%). Only six groups (13.0%) from four studies reported tracking fidelity to specific intervention practices.

Uncontrolled effect sizes

Table 1 and Figure 2 present the uncontrolled effect sizes and the results of the random effects models for all outcomes. Studies excluded from specific analyses due to sample overlap or high SAMD value are noted in Table S3. When funnel plots (Figures S1–S4) indicated a need for corrected effect sizes due to publication bias they are reported in the trim-and-fill correction row of Table 1.

Cognitive. Hedges's g ranged from −0.23 to 1.50 for the 26 groups. The average effect size excluding one outlier was small (Table 1, 0.27, 95% CI 0.17–0.37). Trim-and-fill procedures suggested that three studies with effect sizes to the left of mean were missing. The corrected average effect size was 0.24 (95% CI 0.13–0.35).

Communication. Hedges's g ranged from −0.26 to 0.70 for the 27 groups. The average effect size was small (Table, 1, 0.32, 95% CI 0.24–0.40).

Social. Hedges's g ranged from −0.96 to 0.75 for the 26 groups. Thirteen groups (50.0%) demonstrated significant positive effects. The average effect size excluding one outlier was small (Table 1% CI 0.14–0.35, $p < .001$).

Adaptive behavior. Hedges's g ranged from −1.25 to 0.95 for the 28 groups. The average effect size excluding one outlier was small (Table 1% CI 0.13–0.29, $p < .001$).

Moderator analyses

The Q statistic and I^2 index indicated significant effect size heterogeneity for all outcomes (see Table 1); exploratory analyses of potential moderators were conducted to assess whether effect sizes

Table 1 Random effects models

Outcome	Total	k	Hedges's g	95% CI	Q (df)	I^2
Cognitive	All studies	26	0.37***	0.25–0.48	220.55 (25)***	88.66
	Outlier excluded	25	0.32***	0.23–0.42	138.76 (24)***	82.70
	Trim-and-fill correction		0.27	0.17–0.37	202.51	
Communication	All studies	27	0.32***	0.24–0.40	166.82 (26)***	84.42
	Outlier excluded	NA				
	Trim-and-fill correction	NA				
Social	All studies	26	0.20***	0.08–0.32	255.38 (25)***	90.21
	Outlier excluded	25	0.24***	0.14–0.35	198.95 (24)***	87.94
	Trim-and-fill correction	NA				
Adaptive behavior	All studies	28	0.17***	0.08–0.26	328.06 (27)***	91.77
	Outlier excluded	27	0.21***	0.13–0.29	262.56 (26)***	90.10
	Trim-and-fill correction	NA				

CI, confidence interval; k, studies included; NA, Not applicable.
***$p < .001$.

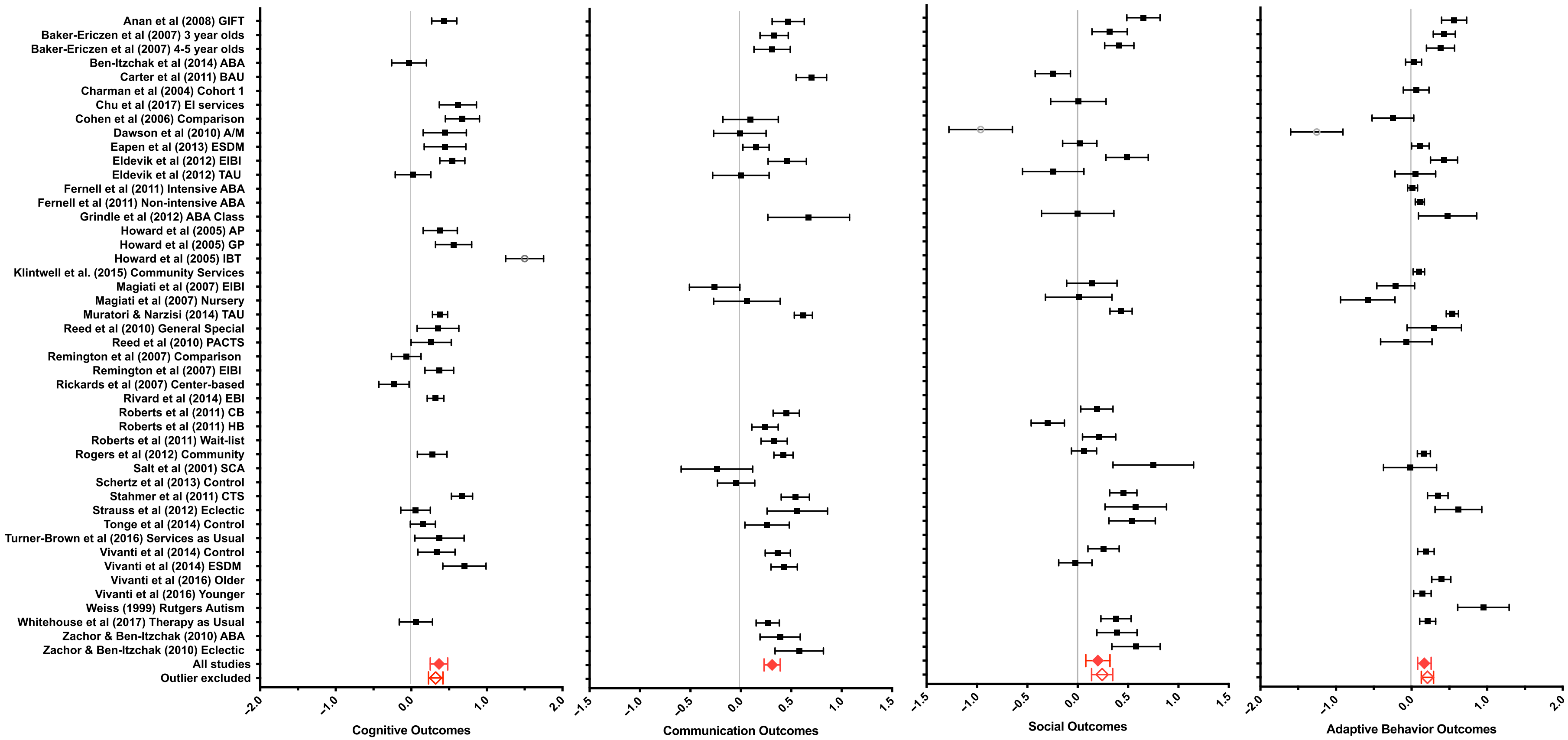

Figure 2 Forest plot of uncontrolled random effects sizes (Hedges's *g*) and 95% confidence intervals. Open circles indicate outliers. Group acronyms: ABA, Applied Behavior Analysis; A/M, Assess and Monitor; AP, autism educational programming; BAU, Business as usual; CB, Center-based Building Blocks; CTS, Children's Toddler School; EI, Early Intervention; EIBI, Early Intensive Behavioral Intervention; ESDM, Early Start Denver Model; GIFT, Group Intensive Family Training Program; GP, generic educational programming; HB, Home-based Building Blocks; IBT, early intensive behavior analytic treatment; PACTS, Parents of Autistic Children Training and Support; SCA, Scottish Centre for Autism; TAU, Treatment as Usual

Table 2 Analyses of moderation for uncontrolled effect sizes

	Cognitive				Communication				Social				Adaptive behavior			
	k	ES	CI	Q (df)	k	ES	CI	Q (df)	k	ES	CI	Q (df)	k	ES	CI	Q (df)
Categorical moderator																
EI Category				**13.94 (2)**				2.98 (2)				2.48 (2)				**13.48 (2)**
MO	6	***0.54***	±0.10		10	***0.39***	±0.13		9	***0.34***	±0.17		11	***0.37***	±0.12	
SPC	13	***0.25***	±0.12		12	***0.30***	±0.14		12	*0.22*	±0.17		12	0.08	±0.15	
USP	6	**0.27**	±0.21		5	**0.22**	±0.16		4	0.11	±0.24		4	***0.14***	±0.06	
Country				*23.68 (7)*				***35.11 (5)***				12.16 (5)				***101.22 (6)***
AU	6	0.23	±0.24		8	***0.31***	±0.07		8	0.16	±0.17		5	***0.21***	±0.09	
CA	1	***0.32***	±0.11													
IS	1	−0.03	±0.23		2	***0.47***	±0.19		2	***0.47***	±0.19		1	0.03	±0.10	
IT	2	0.23	±0.32		2	***0.62***	±0.08		2	***0.45***	±0.10		2	***0.55***	±0.08	
NO	2	0.29	±0.51		2	0.24	±0.45		2	0.15	±0.72		2	0.26	±0.37	
SW													2	0.06	±0.09	
TA	1	***0.62***	±0.25													
UK	4	0.22	±0.22		4	0.04	±0.38		4	0.21	±0.30		7	−0.01	±0.22	
US	8	***0.49***	±0.11		9	**0.33**	±0.15		7	*0.24*	±0.22		8	***0.32***	±0.16	
Study Design				6.49 (3)				4.90 (3)				5.45 (3)				*15.41 (3)*
CC	10	***0.34***	±0.15		12	***0.27***	±0.11		12	*0.20*	±0.16		10	0.06	±0.11	
CE	8	*0.22*	±0.19		8	**0.29**	±0.17		6	0.21	±0.25		4	0.18	±0.19	
SI	6	***0.38***	±0.15		6	***0.41***	±0.16		6	***0.38***	±0.15		11	***0.37***	±0.14	
Other	1	***0.62***	±0.25		1	**0.67**	±0.41		1	0.00	±0.36		2	**0.09**	±0.07	
Group Allocation				5.95 (3)				1.66 (3)				6.76 (3)				9.66 (3)
RA	5	0.17	±0.25		7	***0.30***	±0.16		5	0.02	±0.24		2	***0.18***	±0.07	
SA	13	**0.37**	±0.15		12	***0.27***	±0.15		12	**0.24**	±0.16		14	*0.11*	±0.11	
NA	6	**0.36**	±0.12		6	***0.41***	±0.16		6	***0.38***	±0.15		10	***0.34***	±0.15	
Other	1	0.15	±0.17		2	***0.31***	±0.11		2	*0.36*	±0.32		1	*0.10*	±0.07	

	Cognitive			Communication			Social			Adaptive behavior		
	Coeff	CI	Q (df)	Coeff	CI	Q (df)	Coeff	CI	Q (df)	Coeff	CI	Q (df)
Continuous moderator												
Age	−0.00	±0.01	0.51 (1)	−0.00	±0.01	0.63 (1)	−0.00	±0.01	0.05 (1)	0.00	±0.01	0.13 (1)
Year	−0.00	±0.03	0.01 (1)	0.03	±0.03	4.30 (1)	−0.01	±0.03	0.50 (1)	0.00	±0.02	0.00 (1)
Duration	0.00	±0.01	0.11 (1)	*−0.01*	±0.01	8.23 (1)	−0.01	±0.01	2.60 (1)	**−0.02**	±0.01	11.72 (1)
Hours	0.00	±0.00	1.03 (1)	−0.00	±0.00	3.64 (1)	−0.00	±0.00	0.94 (1)	−0.00	±0.00	−4.61 (1)

AU = Australia; CA = Canada; CC = group design comparing community EI programs; CE = group design comparing a community EI program to an experimental EI program; CI = 95% Confidence Interval; CO = Country; Coeff = Coefficient; ES = Hedges's g; IT = Italy; IS = Israel; MO = "model" EI program; NA = Not applicable (single treatment group); NO = Norway; RA = random assignment to treatment group; SA = assignment to treatment group using standard community practice; SI = study of one community program; SPC = specified treatment as usual program; SW = Sweden; TA = Taiwan; UK = United Kingdom; US = United States; USP = unspecified treatment as usual program.
Italic = significant at Bonferroni-corrected $p < .05$; Bold = significant at Bonferroni-corrected $p < .01$; Bold & Italic = significant at Bonferroni-corrected $p < .001$.

differed based on group and intervention characteristics (see Table 2). Only results significant at Bonferroni-corrected $p < .05$ within each outcome measure (corrected critical value of 0.00625) are reported below.

EI category. Children in Model EI programs made moderate cognitive gains and small communication, social, and adaptive behavior gains (see Table 2). Children receiving specified and unspecified TAU made small gains across all four outcomes. Differences among the three EI categories were statistically significant for cognitive and adaptive behavior outcomes ($p < .05$). Model programs had significantly higher cognitive and adaptive behavior scores than the other two program types ($ps < .05$).

Out of concern that Model programs would only publish with positive results, a post hoc publication bias analysis was conducted. Publication bias was identified for adaptive behavior outcomes. Duval and Tweedie's trim and fill adjusted value was 0.30.

Country. Differences in cognitive, communication, and adaptive behavior outcomes by country were statistically significant (Table 2). Interventions conducted in the United States had small to medium effects cognitive, communication, and adaptive behavior outcomes. Interventions conducted in Italy had medium effects on communication and adaptive behavior outcomes. Interventions conducted in Australia had small effects on communication and adaptive behavior outcomes. Interventions conducted in Israel had medium effects on communication outcomes. Interventions conducted in Taiwan had medium effects on cognitive outcomes. Those conducted in Canada had small effects.

Study design. Only adaptive behavior outcomes differed significantly by study design (Table 2). Post-hoc analyses indicated that studies using "Single" designs had significantly larger effects on adaptive behavior than those using "Other" or "Comparison to other community groups" designs (p-values $< .05$).

Group allocation. Group allocation method was not associated with any outcome (Table 2).

Age. Average age of the sample at intake was not associated with any outcome (Table 2).

Year. Year of publication was not associated with any outcome (Table 2).

Intervention duration. EI duration was significantly negatively associated with communication and adaptive behavior outcomes (Table 2).

Intervention hours. Approximate EI hours were not associated with any outcome (Table 2).

Adjusted EI category moderator analyses. For cognitive outcomes, when controlling for country, the effect size of Model programs remained significantly higher than unspecified TAU programs ($p < .01$), but not specified TAU programs ($p = .10$). For communication outcomes, when controlling for country and intervention duration, Model programs had significantly higher effect sizes that unspecified TAU ($p < .05$) but not specified TAU ($p = .25$). No adjusted models were estimated for social outcomes because no statistically significant moderators were identified. For adaptive behavior outcomes, the difference in effect sizes between Model and the other two program types were no longer significant when controlling for country, study design, and intervention duration.

Discussion

The present study is one of the first syntheses of child outcomes in community-based early intervention, and, to our knowledge, the only to specifically focus on children with ASD. Consistent with prior findings regarding early childhood special education (Cook & Odom, 2013) we found little evidence to suggest that the gap between research and practice has been meaningfully been reduced. Effect sizes associated with community-based early intervention for cognitive, communication, social, and adaptive behavior outcomes for children with autism were small, ranging from 0.21 for adaptive behavior to 0.32 for communication. These results starkly contrast with those reported in prior meta-analyses of university-based clinical trials, which find effect sizes of 0.42–0.76 for these domains (Reichow et al., 2012). It should be noted that the effect sizes from the university-based clinical trials represent the difference between the treatment and control groups, instead of the total effect size over time, which makes the difference even greater between clinical trials and community-based interventions.

Despite the low average effect sizes, several programs showed stronger outcomes. These programs were developed in association with universities, hospitals, and researchers, and offer potentially replicable and sustainable models (Stahmer & Aarons, 2009). Although not examined systematically in this study, they also tended to report using intervention strategies aligned with published, evidence-based interventions, including Early Start Denver Model, Pivotal Response Training, and interventions in the family of applied behavior analysis.

A surprising finding was that, at the group level, intervention duration was negatively associated with communication and adaptive behavior outcomes. We think that the most likely reason for this finding relates to the use of standard scores. The developmental trajectory of autism is delayed compared to typically developing peers (Green & Carter, 2014; Iverson et al., 2018; Lord, Risi, & Pickles, 2004), therefore the longer the duration of intervention, the greater the observed gap will be between the scores of children with autism and age-adjusted norms. It is also possible that children may not make progress when receiving a long duration of the same ineffective intervention or the parent-report instruments used to measure communication and adaptive behavior are not sensitive to gains or to clinically important gains made in other domains more broadly. Regardless of the reason for the negative associations, the results highlight the importance of monitoring treatment progress using systematic assessment with measures sensitive to intervention targets, so clinicians can make appropriate clinical decision regarding changing treatment strategies or treatments if benefit is not observed (Kazdin, 2016; National Research Council, 2001). As one possibility, curriculum-based assessments have been used to measure progress on proximal intervention goals and assist with treatment individualization and planning for toddlers with ASD (Bacon et al., 2014).

Improved results were not observed in any outcome domain in more recent studies relative to older studies. Evidence-based practices may not be making their way into standard community care or are not being implemented well in community settings. This finding also may result from restricted range; most studies published prior to 2004 did not meet our inclusion criteria. Year of publication may also have been a poor proxy for the year that data were collected (e.g., conducting secondary data analysis, publishing many years after the completion of treatment, see Table S2).

Programs associated with universities and hospitals had significantly better cognitive and adaptive behavior outcomes than other community programs, suggesting that academic involvement may bolster effectiveness in these domains. We were limited in our ability to investigate this question further, as limited details regarding treatment strategies, fidelity, or other potential metrics of quality were not consistently provided. However, these types of studies reflect an important step in studying treatments in "progressively more genuine circumstances" (Chorpita, 2004; Southam-Gerow, Silverman, & Kendall, 2006; Weisz, 2004). Public-academic collaborations may offer an important step towards improving community practice. Academic involvement may lead to greater access to experts in ASD treatment (Stahmer, Carter, Baker, & Miwa, 2003) which could increase use of evidence-based practices and access to supervision. This is a rich area for further exploration. Other explanations are that these programs are more likely to measure an outcome more proximal to their intervention target or to publish results because they have positive findings. In support of the latter hypothesis, publication bias was identified for communication and adaptive behavior, although not cognitive or social outcomes.

Several study limitations should be noted. We conducted multiple comparisons to examine potential moderators and did not pre-register a protocol for the study, which can inflate the risk of type I error. We have attempted to mitigate this risk by using Bonferroni-corrected *p*-values, but results should be interpreted with caution. We were limited in the characteristics of the intervention models and the participants that we could include in our analysis. Communication, social, and adaptive behavior outcomes were all measured using the same parent reported outcome measure (i.e., VABS) and parent reported outcomes may be biased towards programs that include parent training/model programs. We also required standardized scores, which may have resulted in important studies being excluded. The restrictions on inclusion criteria may have contributed to type II error. Additionally, we included only study design and group allocation methods as indicators of study quality. Other methodological factors (e.g., diagnostic processes; assessment procedures) may contribute to variable quality in these studies.

Despite these limitations, these findings have important implications, especially if the small effects of community-based interventions resulted from poor translation of research into practice. For example, community providers may lack high quality training and supervision. Differences also could result from fewer resources to implement complex, resource intensive programs. Differences in characteristics of children and families between community settings and research trials also may lead to different outcomes. Lord et al. (2005) point out that in treatment studies that report demographic characteristics of participants, the overwhelming majority are white and of high socio-economic status. Families enrolling in studies may have more resources, fewer obstacles, and more motivation/skill. Unlike research trials, community sites often must accept all children without exclusion. Thus, community programs have children that differ from those in clinical trials in functioning, at-home support, family resources, native language, and complex comorbidities. We largely did not see differences in the results based on study design and group allocation method, which suggests that the small effect sizes for community programs are more likely due to intervention as opposed to participant characteristics.

Any of these reasons for the observed differences point to the need for closer and bidirectional partnerships between academic and community sites, and the urgency of moving from tests of efficacy, or even effectiveness, to testing strategies that increase the successful implementation of demonstrated-effective interventions for young children with autism in community settings.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. PsycINFO search strategy.
Table S1. Sample characteristics.
Table S2. Intervention characteristics.
Table S3. Random weighted uncontrolled effect sizes and SAMDs.
Figure S1. Funnel plot of standard error by Hedges's *g* for cognitive results.
Figure S2. Funnel plot of standard error by Hedges's *g* for communication results.
Figure S3. Funnel plot of standard error by Hedges's *g* for social results.
Figure S4. Funnel plot of standard error by Hedges's *g* for adaptive behavior results.

Acknowledgements

The research reported here was supported in part by Autism Speaks grant #8874 and the Institute of Education Sciences, U.S. Department of Education, through Grant #R305B090015 to the University of Pennsylvania. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education. The authors are also grateful to Maya Matheis and Joseph Pereira for their help with reliability coding. The authors have declared that they have no competing or potential conflicts of interest.

Correspondence

Allison S. Nahmias, Department of Psychiatry and Behavioral Sciences, MIND Institute, University of California, Davis, 2825 50th St., Sacramento, CA 95817, USA; Email: asnahmias@ucdavis.edu

Key points

- Children with Autism Spectrum Disorder can make moderate to large gains in cognitive, communication, social, and adaptive behavior domains when receiving effective early intervention (EI).
- Less is known about the overall effectiveness of EI delivered in community settings.
- Findings show that children with ASD make small gains when receiving community-based EI.
- Model programs, associated with universities, are particularly promising for cognitive and adaptive behavior outcomes.
- Additional research is warranted to bring effective EI into standard community practice.

References

Bacon, E.C., Dufek, S., Schreibman, L., Stahmer, A.C., Pierce, K., & Courchesne, E. (2014). Measuring outcome in an early intervention program for toddlers with autism spectrum disorder: Use of a curriculum-based assessment. *Autism Research and Treatment*, *2014*, 1–9.

Balas, E. A., & Boren, S. A. (2000). Managing clinical knowledge for health care improvement. Yearbook of Medical Informatics 2000: Patient-Centered Systems. Available from: https://augusta.openrepository.com/augusta/handle/10675.2/617990. [last accessed 19 November 2018].

Beal, D.J., Corey, D.M., & Dunlap, W.P. (2002). On the bias of Huffcutt and Arthur's (1995) procedure for identifying outliers in the meta-analysis of correlations. *Journal of Applied Psychology*, *87*, 583–589.

Ben-Itzchak, E., Watson, L.R., & Zachor, D.A. (2014). Cognitive ability is associated with different outcome trajectories in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*, 2221–2229.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. https://doi.org/10.1002/9780470743386.refs

Borenstein, M., Hedges, L., Higgins, J.P.T., & Rothstein, H. (2014). *Comprehensive meta-analysis version 3*. Englewood, NJ: Biostat.

Chorpita, B.F. (2004). Four considerations for dissemination of intervention innovations. *Clinical Psychology: Science and Practice*, *11*, 364–367.

Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Journal of Developmental & Behavioral Pediatrics*, *27*, S145–S155.

Cook, B.G., Landrum, T.J., Tankersley, M., & Kauffman, J.M. (2003). Bringing research to bear on practice: Effecting evidence-based instruction for students with emotional or behavioral disorders. *Education and Treatment of Children*, *26*, 345–361.

Cook, B.G., & Odom, S.L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children*, *79*, 135–144.

Council, National Research (Ed.) (2001). *Educating children with autism*. Washington, DC: National Academy Press.

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*, 455–463.

Elliott, C.D. (1990). *Differential abilities scale*. San Antonio, TX: Psychological Corporation.

Green, S.A., & Carter, A.S. (2014). Predictors and course of daily living skills development in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*, 256–263.

Greenwood, C.R., & Abbott, M. (2001). The research to practice gap in special education. *Teacher Education and Special Education*, *24*, 276–289.

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

Hess, K.L., Morrier, M.J., Heflin, L.J., & Ivey, M.L. (2008). Autism treatment survey: Services received by children with autism spectrum disorders in public school classrooms. *Journal of Autism and Developmental Disorders*, *38*, 961–971.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*, 557–560.
Huffcutt, A.I., & Arthur, W. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology*, *80*, 327–334.
Iverson, J.M., Northrup, J.B., Leezenbaum, N.B., Parladé, M.V., Koterba, E.A., & West, K.L. (2018). Early gesture and vocabulary development in infant siblings of children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *48*, 55–71.
Kasari, C., & Smith, T. (2016). Forest for the trees: Evidence-based practices in ASD. *Clinical Psychology: Science and Practice*, *23*, 260–264.
Kazdin, A.E. (2016). Closing the research–practice gap: How, why, and whether. *Clinical Psychology: Science and Practice*, *23*, 201–206.
Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P.A., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, *6*, e1000100.
Lipsey, M.W., & Wilson, D.B. (2001). Practical meta-analysis. In L. Bickman & D.J. Rog (Eds), *Applied social research methods series: Vol. 49* (pp. 247). Thousand Oaks, CA: Sage Publications.
Lord, C., Risi, S., & Pickles, A. (2004). Trajectory of language development in autistic spectrum disorders. In M.L. Rice & S.F. Warren (Eds), *Developmental language disorders: From phenotypes to etiologies* (pp. 7–29). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., Charman, T., ... & Yoder, P. (2005). Challenges in evaluating psychosocial interventions for autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*, 695–708.
Magiati, I., Charman, T., & Howlin, P. (2007). A two-year prospective follow-up study of community-based early intensive behavioural intervention and specialist nursery provision for children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *48*, 803–812.
Mandell, D. S., Stahmer, A. C., Shin, S., Xie, M., Reisinger, E., & Marcus, S. C. (2013). The role of treatment fidelity on outcomes during a randomized field trial of an autism intervention. *Autism: The International Journal of Research and Practice*, *17*, 281–295.
Mullen, E.M. (1995). *Mullen scales of early learning manual*. Circle Pines, MN: American Guidance Service.
Murza, K.A., Schwartz, J.B., Hahs-Vaughn, D.L., & Nye, C. (2016). Joint attention interventions for children with autism spectrum disorder: A systematic review and meta-analysis: Joint attention meta-analysis. *International Journal of Language & Communication Disorders*, *51*, 236–251.
Nevill, R.E., Lecavalier, L., & Stratis, E.A. (2018). Meta-analysis of parent-mediated interventions for young children with autism spectrum disorder. *Autism*, *22*, 84–98.
Ospina, M.B., Krebs Seida, J., Clark, B., Karkhaneh, M., Hartling, L., Tjosvold, L., ... & Smith, V. (2008). Behavioural and Developmental Interventions for Autism Spectrum Disorder: A Clinical Systematic Review. *PLoS ONE*, *3*, e3755.
Reichow, B. (2012). Overview of meta-analyses on early intensive behavioral intervention for young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *42*, 512–520.
Reichow, B., Barton, E. E., Boyd, B. A., & Hume, K. (2012). 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
Roid, G.H., & Sampers, J.L. (2004). *Merrill-palmer-revised: Scales of development*. Wood Dale, IL: Stoelting.
Shire, Stephanie Y., Chang, Y.-C., Shih, W., Bracaglia, S., Kodjoe, M., & Kasari, C. (2017). Hybrid implementation model of community-partnered early intervention for toddlers with autism: A randomized trial. *Journal of Child Psychology and Psychiatry*, *58*, 612–622.
Shire, Stephanie Yoshiko, & Kasari, C. (2014). Train the trainer effectiveness trials of behavioral intervention for individuals with autism: A systematic review. *American Journal on Intellectual and Developmental Disabilities*, *119*, 436–451.
Southam-Gerow, M.A., Silverman, W.K., & Kendall, P.C. (2006). Client similarities and differences in two childhood anxiety disorders research clinics. *Journal of Clinical Child & Adolescent Psychology*, *35*, 528–538.
Sparrow, S.S., Balla, D.A., & Cicchetti, D.V. (1985). *Vineland: Adaptive behavior scales: Interview edition: Survey form manual*. Circle Pines, MN: American Guidance Service.
Sparrow, S.S., Cicchetti, D.V., & Balla, D.A. (2005). *Vineland-II: Vineland adaptive behavior scales, survey forms manual*. Circle Pines, MN: American Guidance Service.
Stahmer, A.C., & Aarons, G.A. (2009). Attitudes toward adoption of evidence-based practices: A comparison of autism early intervention providers and children's mental health providers. *Psychological Services*, *6*, 223–234.
Stahmer, A., Carter, C., Baker, M., & Miwa, K. (2003). Parent perspectives on their toddlers' development: comparison of regular and inclusion childcare. *Early Child Development and Care*, *173*, 477–488.
Stahmer, A.C., Collings, N.M., & Palinkas, L.A. (2005). Early intervention practices for children with autism: Descriptions from community providers. *Focus on Autism and Other Developmental Disabilities*, *20*, 66–79.
Suhrheinrich, J., Stahmer, A.C., Reed, S., Schreibman, L., Reisinger, E., & Mandell, D. (2013). Implementation challenges in translating pivotal response training into community settings. *Journal of Autism and Developmental Disorders*, *43*, 2970–2976.
Weisz, J.R. (2004). Treatment dissemination and evidence-based practice: Strengthening intervention through clinician-researcher collaboration. *Clinical Psychology: Science and Practice*, *11*, 300–307.
What Works Clearinghouse. (2014). Procedures and standards handbook version 3.0. Available from: https://ies.ed.gov/ncee/wwc/Handbooks [last accessed 19 November 2018].
Whitehouse, A.J.O., Granich, J., Alvares, G., Busacca, M., Cooper, M.N., Dass, A., ... & Anderson, A. (2017). A randomised controlled trial of an iPad-based application to complement early behavioural intervention in Autism Spectrum Disorder. *Journal of Child Psychology and Psychiatry*, *58*, 1042–1052.
Williams, N.J., & Beidas, R.S. (2018). Annual Research Review: The state of implementation science in child psychology and psychiatry: A review and suggestions to advance the field. *Journal of Child Psychology and Psychiatry*, *60*, 430–450.

Accepted for publication: 15 April 2019