Journal of Autism and Developmental Disorders
https://doi.org/10.1007/s10803-019-04138-2

ORIGINAL PAPER

Stability of Autism Spectrum Disorder in Young Children with Diverse Backgrounds

Ivy Giserman-Kiss[1] · **Alice S. Carter**[2]

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Determining diagnostic stability of ASD, as well stability of functioning in early childhood, is relevant to prevalence, best practices for communicating early ASD diagnoses to caregivers, families' experiences, and developmental trajectories. Generalizability of findings from prior research has been limited by small and homogenous samples, short follow-up time intervals, and inconsistent diagnostic procedures. This report presents follow-up evaluations of 60 children (86.7% male, mean age: 51.3 months) with diverse backgrounds (79.7% racial/ethnic minorities) who received initial ASD diagnoses before 36 months of age (mean age: 27 months). Fifty-three children (88.3%) met diagnostic criteria for ASD at follow-up, a proportion consistent with previous studies. On average, children demonstrated significant cognitive gains and ASD symptom improvement. Clinical implications of findings are discussed.

Keywords Autism spectrum disorder · Autism · Diagnostic stability · Early detection

Introduction

Understanding and documenting the progression of children with early diagnoses of autism spectrum disorder (ASD) is essential for numerous reasons, including framing early diagnoses for families, developing evidence-based screening policies, and understanding the neurobiological substrates and impact of the disorder. Accurate information on the diagnostic stability of ASD is necessary for all stakeholders of the disorder (Woolfenden et al. 2012). Clinicians making the diagnosis of ASD need this information to provide accurate and sensitive guidance to families and to address the short- and long-term outcomes of the condition. Researchers need information on diagnostic stability to understand the prevalence and etiology of the disorder. Most importantly, this information is crucial to the families of children with ASD and the children themselves. A thorough understanding of the progression of the disorder may enable families to feel more comfortable when planning for the future, help them come to terms with a life-altering diagnosis and the emotional, social, and financial challenges that this diagnosis implies, and provide hope for their child's long-term development.

Previous research on the diagnostic stability of ASD has been significantly limited by a number of factors, including relatively small and homogenous samples, short follow-up time periods, and inconsistent diagnostic procedures. In this study, diagnostic stability of ASD was investigated across a one- to three-year follow-up period in a sample of 60 children who received initial ASD diagnoses before 36 months of age. The current study addressed some of the consistent limitations of previous research in this area by (1) including a high percentage of children from traditionally underserved backgrounds (in terms of racial/ethnic background, socioeconomic status, parental education level) in the sample, (2) conducting follow-up evaluations with variability in both age at initial diagnosis and age at follow-up, and (3) using gold-standard diagnostic procedures at both time points.

✉ Ivy Giserman-Kiss
ivygise@gmail.com

[1] Seaver Autism Center for Research and Treatment, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, One Gustave Levy Place, New York, NY 10029, USA

[2] Department of Psychology, University of Massachusetts Boston, 100 Morrissey Blvd, Boston, MA 02125, USA

Published online: 11 July 2019

Overview of Diagnostic Stability of ASD

ASD is typically considered to be a lifelong disorder and a great deal of research supports the stability of the diagnosis over time (Chawarska et al. 2007; Eaves and Ho 2004; Guthrie et al. 2013; Wiggins et al. 2012). Yet, recent studies have reported a small subset of children, sometimes referred to as the "optimal outcome" group, who show improvement over time, often in the context of intervention, and subsequently no longer meet criteria for the diagnosis at follow-up (Fein et al. 2013). Of note, initial diagnoses received in very early childhood (i.e., at or below 30 months of age) have been found to be less stable over time (Helt et al. 2008; Turner and Stone 2007; Turner et al. 2006; van Daalen et al. 2009; Wiggins et al. 2012). The question of stability of ASD, defined as the likelihood that a child will meet criteria for the diagnosis at follow-up (Charman and Baird 2002), has become an increasingly active focus of research and is particularly relevant to scientific and clinical issues regarding prevalence, utility, and cost of early intervention (EI), families' experience of the diagnosis, and developmental trajectories of the symptoms (Cox et al. 1999; Lord and Bishop 2010; Mahli and Singhi 2011; Ozonoff et al. 2015; Woolfenden et al. 2012).

The majority of existing literature investigating the stability of the diagnosis evaluates individuals classified within the category of Pervasive Developmental Disorders (PDDs) from the fourth edition text revision of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; APA 2000). This overarching category included Autistic Disorder (AD), Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS), and Asperger's Syndrome (AS), each of which had different diagnostic criteria than the currently employed DSM-5 Autism Spectrum Disorder (DSM-5; APA 2013). Many researchers examined clinical stability by sub-category and demonstrated that AD was a significantly more stable diagnosis than PDD-NOS and AS (see review: Woolfenden et al. 2012). While this body of literature is informative, the transition to the DSM-5, which contains a single category of ASD with modified diagnostic criteria, has created a lack of clarity regarding the stability of current diagnosis. To account for the modified diagnostic category, some researchers who originally used the DSM-IV-TR to diagnose the disorder dichotomized their outcome variables (i.e., "ASD" and "no ASD," where "ASD" included all three DSM-IV-TR PDD categories); however, this is not analogous to using DSM-5 criteria because of the inconsistencies in the symptoms required for the diagnosis (Eaves and Ho 2004; Fein et al. 2013). In addition to discrepant diagnostic criteria, studies widely vary in the tools used to diagnose the disorder, thus raising questions about the construct validity of ASD across the extant literature. To date, studies have published stability rates of diagnoses made in early childhood that range widely and vary significantly across DSM-IV-TR category (i.e., 63–100% stability for AD, 17–100% for PDD-NOS, and 82–100% for dichotomized ASD outcomes) (Ben Itzchak and Zachor 2009; Bieleninik et al. 2017; Chawarska et al. 2007, 2009; Corsello et al. 2013; Kleinman et al. 2008; Mahli and Singhi 2011). Likewise, estimates of stability of cognitive functioning and symptom severity broadly vary, with a myriad of associated factors, including age of initial diagnosis (e.g., Sutera et al. 2007; Turner and Stone 2007), symptom severity (e.g., Fein et al. 2013; Lord et al. 2006) and cognitive abilities at initial diagnosis (e.g., Turner and Stone 2007; van Daalen et al. 2009), and receipt of EI services (Dawson et al. 2010; Helt et al. 2008).

Demographic Factors

An under-studied element of diagnostic stability is the effect of demographic factors. Few studies have investigated whether or not the stability of the disorder is affected by factors such as socioeconomic status, racial and ethnic minority status, and parental education. This limited body of literature has yielded mixed results, with many studies finding no difference in diagnostic stability rates across groups (Anderson et al. 2014; Ben Itzchak and Zachor 2009; Daniels et al. 2011; Darrou et al. 2010; Turner and Stone 2007), while others report significant differences in diagnostic stability between groups (Fountain et al. 2012; Lord and Bishop 2010), with children from historically underserved groups showing more stability than children from privileged backgrounds. Findings should be interpreted cautiously, however, as these studies are constrained by small sample sizes.

Cognitive Abilities

Furthermore, researchers investigating the differences between children who no longer meet ASD criteria at follow-up and those with stable diagnoses have frequently explored the relationship between baseline cognitive scores and diagnostic outcomes. There are particularly mixed findings within this area of research, especially regarding initial language abilities. Several studies have found that children who no longer met criteria for ASD diagnoses were more likely to have higher receptive and expressive language scores at initial evaluation (Turner and Stone 2007; van Daalen et al. 2009), whereas other studies have not found group differences in this area (Sutera et al. 2007; Turner et al. 2006). Additionally, several studies have found group differences in motor skills, measured by both standardized assessment and caregiver-report (Turner and Stone 2007; Sutera et al. 2007), perhaps suggesting that developed motor skills are a sign of a positive prognosis. Moreover,

van Daalen et al. (2009) found that children who "lost" their diagnosis showed a significantly higher increase in cognitive scores at follow-up, compared to those with stable diagnoses.

Early Intervention (EI)

Response to type and intensity of EI services remains a strong focus of research regarding the diagnostic stability of ASD. Lovaas (1987) found that 47% of children who received intense ABA "recovered" from ASD. Lovaas defined "recovery" as achieving success in a regular classroom with average intelligence; however, he did not explicitly measure ASD symptoms at follow-up. While this study has been strongly criticized for poor methodology and questionable construct validity, and studies that attempted to replicate findings did not replicate the initial findings (Cohen et al. 2006), it has sparked an ongoing conversation about the impact of EI on the diagnostic stability of ASD.

Most studies that have investigated type of EI as a predictor of diagnostic stability have found no difference between children who no longer meet diagnostic criteria at follow-up and those with stable diagnoses (Chawarska et al. 2009; Eaves and Ho 2004; Lord et al. 2006; Turner and Stone 2007). Notably, Orinstein et al. (2014) found that children who achieved "optimal outcome" were more likely to have received earlier intervention, more intensive intervention (i.e., significantly more hours), and Applied Behavioral Analysis (ABA) than children with stable diagnoses; however, there were no group differences regarding the number of hours of ABA therapy received. Further, recent research findings support the value of parent-mediated naturalistic developmental behavioral interventions for young children with ASD (see reviews: Oono et al. 2013; Siller and Morgan 2018).

Overall, the paucity of research regarding the stability of the DSM-5 diagnosis of ASD, especially in younger children with diverse backgrounds, has created significant challenges for clinicians in helping caregivers make sense of their child's diagnosis. Caregivers often ask questions about the "permanence" of the disorder and their child's long-term prognosis, but the lack of available information on this topic can leave clinicians unsure how to respond (Woolfenden et al. 2012). The current study attempted to remedy this gap in knowledge by conducting follow-up evaluations of diverse children who were diagnosed before 36 months of age between 1 and 3 years after the original diagnosis.

Methods

All study procedures were approved by the institutional review board and informed consent from participants' legal guardians was documented in writing.

Participants

Participants included 61 children who were previously evaluated and diagnosed with ASD through a university-based, multi-stage screening project embedded within an EI agency. One participant was excluded due to the presence of brain tumors with unknown etiology and impact; thus, all analyses presented below include the remaining 60 participants (86.7% male). Participants and families were diverse with regard to race/ethnicity (79.7% identified as racial/ethnic minorities), annual household income (61.7% of families received government financial assistance within the last year), parental education status (59.3% of primary parents did not graduate from college), and parental employment status (39.0% of primary parents were not employed at the time of follow-up evaluation).

Children received initial ASD diagnoses between 19 and 34 months of age (mean = 27.7 months; SD = 4.3 months) and follow-up evaluations were held between 12 and 50 months (mean = 23.7 months; SD = 7.8 months) after initial evaluations; children ranged in age from 42 to 70 months (mean = 51.3 months; SD = 7.0 months) at follow-up evaluations. All children in this sample received EI services (mean = 14.7 h/week, SD = 11.2 h/week, range = 0–52 h/week). See Table 1 for detailed demographic information for children and families in this sample.

Measures

Initial and follow-up evaluations consisted of a gold-standard behavioral diagnostic measure, developmental/cognitive testing, and caregiver interviews about adaptive behavior and the child's developmental and medical history. ASD diagnoses were assigned by licensed clinical psychologists based on review of all information available within each time-point.

Differential Abilities Scale—Second Edition (DAS-II; Elliott 2007): The DAS-II is a test of cognitive functioning designed for use with preschool- and school-age children and includes an Early Years Battery (for ages 2 years, 6 months to 6 years, 11 months) and a School-Age Battery (for ages 7 years to 17 years, 11 months). Both batteries yield core or diagnostic composite scores, a special nonverbal composite score, and 20 subtest scores. There are three core composites: Nonverbal Reasoning Ability, Verbal Ability, and Spatial Ability, and two diagnostic composites: Working Memory and Processing Speed. The measure is norm-referenced and standardized. Subtests generate T-scores and the composites generate standard scores. At follow-up evaluations, participants were administered the DAS-II Early Years Battery six core subtests (Verbal Comprehension, Picture Similarities, Naming Vocabulary, Pattern Construction, Matrices, Copying) that make up the three core composite

Table 1 Characteristics of sample

Characteristic	N	%
Sex		
Male	52	86.7
Female	8	13.3
Race/Ethnicity		
Arab	2	3.3
Asian	4	6.7
Black Non-Hispanic/Latino	7	11.7
Multiracial Non-Hispanic/Latino	3	5.0
White Non-Hispanic/Latino	12	20
Black Hispanic/Latino	1	1.7
Multiracial Hispanic/Latino	4	6.7
White Hispanic/Latino	26	43.3
Primary Parent Education		
8th grade or less	3	5.0
1–3 years high school	3	5.0
12th grade/high school diploma/GED	16	26.7
Vocational school/Other non-college certificate	4	6.7
1–3 years college/Associate's degree	9	15.0
College degree	14	23.3
Master's degree	7	11.7
Professional degree	3	5.0
Primary Parent Employment Status		
Paid full time job	21	35.0
Paid part time job	15	25.0
No current paid job	23	38.3
Family Household Annual Income		
$0–35,000	18	30.0
$36–65,000	9	15.0
$66,000+	21	35.0
EI Service Receipt Pre-Third Birthday (h/week)		
0–5.0	9	15.0
5.1–10.0	16	26.7
10.1–15.0	13	21.7
15.1–20	6	10.0
20.1–25	3	5.0
25.1+	10	16.7

scores and General Conceptual Ability score by an experienced examiner. If participants were unable to achieve basal scores on at least four of the six subtests, the DAS-II was discontinued and participants were administered the Mullen Scales of Early Learning.

Mullen Scales of Early Learning (MSEL; Mullen 1995): The Mullen is a developmental assessment designed for use with infants and young children from birth to 68 months. The MSEL consists of five different scales: Visual Reception, Fine Motor, Gross Motor, Expressive Language, and Receptive Language, and also generates a composite score, the Early Learning Composite (ELC), that indicates a child's overall developmental level (also referred to as Developmental Quotient (DQ)). Subtests generate T-scores and the composite generates a standard score. The measure is norm-referenced and standardized, and has been shown to be an effective psychological assessment for young children with ASD (Akshoomoff 2006). Participants were administered the MSEL by an experienced examiner at initial evaluations.

Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al. 2012): The ADOS-2 is a semi-structured interactive observational tool intended to assess social and communication behaviors of individuals suspected of having an ASD diagnosis. Participants were administered one of five developmental modules, based on language ability, at both initial and follow-up evaluation. The ADOS-2 consists of a number of standardized social presses that are designed to elicit specific behaviors. Modules 1–4 of the ADOS-2 generate a calibrated severity score that is consistent across these four modules, which ranges from 1 to 10, with 10 being the most severe. For the Toddler Module (administered at initial evaluation to children under 31 months of age who were pre-verbal or primarily using single words), a range of concern for ASD is calculated based on an algorithm score; a calibrated severity score is not generated from this module. Methods described in Esler et al. (2015) were used to calculate severity scores for children who received the Toddler Module at their initial evaluation. Developmentally-appropriate modules of the ADOS-2, which were administered by research-reliable assessors, were used as a tool aiding diagnostic evaluation, in collaboration with expert clinical judgment at both time points.

Vineland Adaptive Behavior Scales—Third Edition Domain-Level Interview Form (VABS-3; Sparrow et al. 2016): The VABS-3 is a standardized semi-structured caretaker interview instrument that assesses a child's adaptive functioning. The VABS-3 consists of four domains: Communication, Daily Living, Socialization and Motor. Items are scored on scale for 0 (never) to 2 (usually). Experienced research assistants administered this interview to caregivers during follow-up evaluations. This measure yields standard scores for each of the specific domains as well as an overall Adaptive Behavior Composite score.

Demographic Questionnaire: Caregivers completed a questionnaire in which they reported information regarding marital status, employment status, educational status, and household income. Caregivers also reported the amount and types of services that their child received prior to his/her third birthday, as well as the services that the child received after turning three and aging out of EI services. Specifically, caregivers were asked to report the number and length of monthly sessions of each service that their child received (i.e., autism-specific intervention, developmental specialist, speech/language pathology, occupational therapy, physical

therapy, art therapy, music therapy, play group, social work, and other services).

Results

Follow-up evaluations determined that seven children (11.7% of sample) no longer met DSM-5 diagnostic criteria for ASD. As seen in Table 2, which displays paired *t* tests for cognitive scores and ASD symptom severity between initial and follow-up evaluations, children demonstrated significantly improved cognitive abilities (including nonverbal problem-solving, fine motor, receptive language, and expressive language skills) as well as significantly decreased ASD symptom severity at follow-up. The effect sizes for *t* tests comparing children's initial and follow-up cognitive performance ranged from small to medium, with children making the most gains in fine motor skills ($d = 0.68$). The effect size for the *t* test comparing ASD severity at initial and follow-up evaluations ($d = 0.89$) was found to exceed Cohen's (1988) convention for a large effect. Cognitive scores at initial and follow-up evaluations were significantly, positively correlated ($R = 0.61$, $p < .01$), as were ASD symptom severity scores at initial and follow-up evaluations ($R = 0.44$, $p < .01$), demonstrating cognitive and diagnostic stability over time. In addition, age at initial evaluation was significantly, negatively correlated with ASD symptom severity scores at both initial evaluation ($R = -0.31$, $p < .05$) and at follow-up evaluation ($R = -0.42$, $p < .01$), with younger children evidencing more elevated symptoms at both time points. This suggests that it is possible that children diagnosed at earlier ages were those presenting with more severe symptoms.

Predicting Changes in Child Functioning

A multiple linear regression model was calculated to predict ASD symptom severity (i.e., ADOS-2 symptom severity score) at follow-up, based on the following variables: age at initial evaluation, ASD symptom severity at initial evaluation, and overall DQ at initial evaluation (with each predictor variable entered in different steps sequentially). This regression model was significant ($F(4,52) = 8.45$, $p < .01$) with $R^2 = 0.39$. In the four predictor model, time interval between evaluations, age at initial evaluation, and initial ASD symptom severity each explained unique variance in the prediction of ASD symptom severity at follow-up. The overall DQ score did not explain additional variance in follow-up ASD symptom severity. A second multiple linear regression model, which was calculated to predict change in ASD symptom severity from the initial to the follow-up evaluation based on the same four variables, yielded highly comparable results.

Table 2 Mean changes in cognitive and diagnostic characteristics between initial and follow-up evaluations (T-scores)

	Mean T1	Mean T2	Paired differences					*t*	*df*	Sig. (2-tailed)	Cohen's *d*
			Mean change	SD	SEM	95% CI of the diff.					
						Lower	Upper				
T1 MSEL Visual Recep. and T2 DAS-II Picture Similarities	34.07	39.83	5.466	13.056	1.714	2.033	8.898	3.188	57	0.002**	0.420
T1 MSEL Visual Recep. and T2 DAS-II Matrices	34.07	38.85	5.034	14.410	1.892	1.246	8.823	2.661	57	0.010*	0.403
T1 MSEL Express. Lang. and T2 DAS-II Naming Vocabulary	28.00	35.15	6.931	13.390	1.758	3.410	10.452	3.942	57	0.000**	0.526
T1 MSEL Recep. Lang. and T2 DAS-II Verbal Comprehension	25.29	31.60	6.155	12.768	1.676	2.798	9.512	3.672	57	0.001**	0.477
T1 MSEL Fine Motor and T2 DAS-II Pattern Construction	32.02	41.95	9.948	17.526	2.301	5.340	14.557	4.323	57	0.000**	0.683
T1 ADOS-2 Severity Score and T2 ADOS-2 Severity Score	8.22	6.20	1.966	2.435	0.317	1.322	2.601	6.202	58	0.000**	0.897

*$p < .05$; **$p < .01$

Because only seven children no longer met diagnostic criteria for ASD at follow-up, we lacked sufficient power to use the binary diagnostic outcome variable (i.e., ASD or no ASD) in multivariate analyses. Therefore, an exploratory logistic binary regression model was calculated to predict diagnostic outcome at follow-up (i.e., ASD or no ASD), based on the following variables: age at initial evaluation, ASD symptom severity at initial evaluation, and overall DQ at initial evaluation (each predictor variable entered in different steps, respectively). Time interval between initial and follow-up evaluation was again added as a covariate in the first step. Table 3 shows the logistic regression coefficient, standard error, Wald Chi Square test statistic, and odds ratio (Exp(B)) for each of the predictors. Despite limited power, the four predictor model indicated that time interval between evaluations and age at initial evaluation were significant predictors of diagnostic outcome at follow-up. However, the Wald criterion demonstrated that only age at initial evaluation, not time interval, made a significant contribution to prediction ($p = .01$). Nagelkerke's R^2 (which has a minimum of 0 and maximum of 1.0) of 0.46 indicated a medium to large relationship between prediction variables and binary diagnostic outcome (Cohen 1992). Prediction success overall was 91.2% (98.0% for children who maintained diagnosis and 42.9% for children who no longer met criteria for diagnosis). For every month longer between initial and follow-up evaluations, children showed a 1.3-fold increase in likelihood of losing the diagnosis. In addition, for every month that a child was diagnosed older, there was a 1.6-fold increase in likelihood of the child no longer meeting diagnostic criteria at follow-up.

Table 3 Summary of binary logistic regression analysis for variables predicting diagnostic outcome at follow-up evaluation

Variable	Block 1				Block 2				Block 3				Block 4			
	B	*SE*	Wald	*Exp (B)*	*B*	*SE*	Wald	*Exp (B)*	*B*	*SE*	Wald	*Exp (B)*	*B*	*SE*	Wald	*Exp (B)*
Time interval	−1.03	0.053	3.704	0.902	−0.216	0.076	8.166**	0.806	−0.228	0.081	7.876**	0.796	−0.249	0.090	7.695**	0.779
T1 age					−0.437	0.172	6.423*	0.646	−0.515	0.225	5.226*	0.598	−0.476	0.220	4.665*	0.622
T1 ADOS–02 severity									−0.271	0.419	0.419	0.762	−0.159	0.429	0.137	0.853
T1 MSEL ELC													0.061	0.066	0.850	1.063
Omnibus model sig.			0.044				0.001				0.002				0.003	
Cox and Snell R^2			0.069				0.224				0.230				0.243	
Nagel-kerke R^2			0.131				0.426				0.437				0.462	

*$p < .05$; **$p < .01$

Predicting Cognitive Changes

An additional multiple linear regression model was calculated to predict overall IQ at follow-up (i.e., DAS-II GCA), based on the following variables: age at initial evaluation, ASD symptom severity at initial evaluation, and overall DQ at initial evaluation (each predictor variable entered in different steps, respectively). The overall regression model was significant ($F(3,53) = 10.517$, $p < .01$) with $R^2 = 0.373$. In the three predictor model, only overall initial evaluation DQ was a significant predictor of cognitive abilities at follow-up. When scores from the four MSEL domains were entered individually into the regression model, Expressive Language at initial evaluation was found to be driving the significant relationship between initial and follow-up MSEL scores ($p < .01$).

A final multiple linear regression model was calculated to predict verbal ability at follow-up (i.e., DAS-II Verbal Ability Cluster score), based on the following variables: age at initial evaluation, ASD symptom severity at initial evaluation, and overall DQ at initial evaluation (each predictor

variable entered in different steps, respectively). A significant regression equation was found ($F(3,53) = 14.57$, $p < .01$) with $R^2 = 0.45$. Similar to the previously described regression model predicting overall IQ at follow-up, this three predictor model also showed that only overall DQ at initial evaluation accounted for unique variance in verbal ability at follow-up. When individual MSEL domain scores were entered into the regression model, Expressive Language at initial evaluation was again found to accounting for the largest portion of the variance ($p < .01$).

Moderation Effect

Number of weekly hours of EI services (general and specialty services combined) between the time of a child's initial diagnosis and third birthday was examined as a moderator of the relationship between initial ASD symptom severity and follow-up ASD symptom severity. The analysis operationalized the dependent variable as follow-up ADOS-2 severity score. Total EI hours and initial ADOS-2 severity score were entered in the first step of the regression analysis. In the second step of the regression analysis, the interaction term between total EI hours and initial ADOS-2 severity score were entered. The second step failed to explain an increase in variance in follow-up ADOS-2 severity scores, $\Delta R^2 = 0.02$, $F(1,52) = 1.24$, $p = .27$. Thus, in this model, the number of total EI hours was not a significant predictor or a moderator of the relation between initial and follow-up ASD symptom severity.

Number of weekly hours of EI services between initial diagnosis and third birthday was also examined as a moderator between initial cognitive functioning and follow-up cognitive functioning. Total EI hours and overall cognitive functioning score at initial diagnosis were entered in the first step of the regression analysis. In the second step of the regression analysis, the interaction term between total EI hours and overall cognitive functioning score at initial diagnosis were entered. The second step failed to explain an increase in variance in follow-up overall cognitive functioning, $\Delta R^2 = 0.01$, $F(1,51) = 1.05$, $p = .31$. Thus, in this model, the number of total EI hours was not a significant predictor or a moderator of the relationship between initial and follow-up cognitive functioning.

In addition, number of weekly hours of EI services between initial diagnosis and third birthday was examined as a moderator between initial expressive language skills and follow-up expressive language skills. Total EI hours and expressive language score at initial diagnosis were entered in the first step of the regression analysis. In the second step of the regression analysis, the interaction term between total EI hours and expressive language score at initial diagnosis were entered. The second step failed to explain an increase in variance in follow-up expressive language abilities, $\Delta R^2 = 0.05$, $F(1,51) = 4.46$, $p = .04$. Thus, in this model, the number of total EI hours was not a significant predictor or a moderator of the relationship between initial and follow-up expressive language skills.

Exploring the Relationship Between Sociodemographic Factors and Diagnostic Stability

Independent samples *t* tests as well as ANOVAs were calculated to determine if child/family demographic factors were related to a child's ASD symptom severity at follow-up. As seen in Table 4, *t* tests showed that neither ADOS-2 severity score at follow-up or change in ADOS-2 severity score differed significantly by child's race (dichotomized as white, non-Latino/Hispanic or non-white due to small sample size) nor child's Latino/Hispanic ethnicity (i.e., Latino/Hispanic or not Latino/Hispanic)—analyses yielded small to medium effect sizes. Similarly, ANOVAs displayed in Table 5 show that neither ADOS-2 severity score at follow-up nor change in ADOS-2 severity score differed significantly by a family's household income, primary parental job status, or primary parental education status (i.e., highest grade completed).

Discussion

The current study had three main aims: 1) to examine diagnostic stability in a diverse sample of children diagnosed under the age of 36 months; 2) to examine the initial child evaluation factors associated with ASD symptom and cognitive improvement at follow-up, including age at diagnosis, initial symptom severity, initial cognitive abilities, as well as the amount of EI services received between the child's initial diagnosis and third birthday; and 3) to explore the relation between diagnostic stability and child/family demographic factors, including race, ethnicity, household income, parental education, and parental employment status. This study differed from previous studies of the diagnostic stability of ASD in several ways. Specifically, the sample was markedly more diverse (79.7% of participants identified as racial/ethnic minorities; 30.0% of participants reported family household annual income of $35,000 or less) and diagnoses were assigned following DSM-5 diagnostic criteria for ASD. Of the 60 children diagnosed with ASD under 36 months, seven children did not meet ASD diagnostic criteria at follow-up, revealing an overall ASD diagnostic stability rate of 88.3% in the sample. This stability rate is comparable to those found in previous studies of early-assigned diagnoses (i.e., before age 3 years), specifically of DSM-IV-TR AD. This consistent finding suggests that DSM-5 ASD diagnoses assigned in early childhood are valid and likely to be retained over a 1–3 year period. While several of the children who no longer met diagnostic criteria

Table 4 Independent samples *t* tests examining relationships between child/family demographic factors and follow-up ASD symptom severity

	Mean	SD	Follow-up ADOS-2 Severity Score		
			t	*df*	Cohen's *d*
Race					
White, Non-Latino/Hispanic	5.78	2.82	− 1.093	53.260	0.286
Non-White	6.55	2.56			
Ethnicity					
Latino/Hispanic	6.06	2.51	0.417	53.806	0.110
Not Latino/Hispanic	6.36	2.91			
Government assistance in last year					
Yes	5.68	2.58	1.945	45.364	0.516
No	7.04	2.69			
	Mean	SD	Change in ADOS-2 Severity Score		
			t	*df*	Cohen's *d*
Race					
White	2.44	2.52	1.388	53.635	0.363
Non-White	1.56	2.33			
Ethnicity					
Latino/Hispanic	1.58	2.17	1.273	52.141	0.337
Not Latino/Hispanic	2.39	2.67			
Government assistance in last year					
Yes	2.46	2.46	− 2.132*	48.050	0.565
No	1.14	2.21			

*$p < .05$; **$p < .01$

for ASD at follow-up continued to show "residual" signs of ASD (e.g., strong interest in a specific topic, inconsistent eye contact), the majority of their social-communication skills were developmentally appropriate and were not interfering with their learning or peer/family relationships, and none demonstrated pervasive restricted/repetitive behaviors or interests. Of the seven children that did not meet ASD criteria at follow-up, one fit diagnostic criteria for Social (Pragmatic) Communication Disorder. It is possible that as the social demands of childhood and adolescence increase (e.g., navigating more sophisticated peer groups, developing friendships, managing social pressures), the children who did not meet diagnostic criteria at follow-up will show a re-emergence of ASD social-communication symptoms. Basic social-communication skills expected of preschool children, such as sharing interests and engaging in brief conversations, differ from the more nuanced skills typical in older children's relationships, such as reading subtle facial expressions and body language. Future research should examine trajectories of social-communication skills and impairments over time in a large sample of early-diagnosed children.

On average, children demonstrated meaningful reductions in ASD symptomatology and cognitive gains between initial and follow-up evaluations. Specifically, on average, participants showed significantly decreased ASD symptomology at follow-up, suggesting that while diagnoses remained mostly stable, intervention as well as the passage of time allowed for the development of additional social-communication skills and/or compensatory strategies. In addition, participants exhibited substantial improvements on intellectual assessment subtests related to nonverbal problem-solving, fine motor, receptive language, and expressive language skills with effect sizes that range from 0.403 to 0.683, indicating moderate to large gains in these areas. At the same time, there was moderate stability in relation to relative ranking within this sample. Children with more well-developed abilities continued to evidence greater skills over time. This finding is consistent with recent research examining the stability of cognitive abilities over time in young children with ASD (e.g., Flanagan et al. 2015). The cognitive gains observed in this report suggest that early cognitive abilities often under-estimate later cognitive abilities in children with ASD; once ASD symptoms are directly targeted in treatment, children may develop skills that promote more general acquisition of language and nonverbal problem solving skills. In addition, the ability to sit and attend for an extended period of

Table 5 One-way analysis of variance of follow up ASD severity by household income, primary parent employment status, and primary parent education status

	Mean	SD	Follow up ADOS-2 Severity Score				
			SS	*df*	*MS*	*F*	*p*
Primary Parent Education							
No HS degree	6.50	2.168					
HS degree	5.94	2.594					
Some college	6.00	2.677					
College degree or higher	6.54	2.949					
Between Groups			4.791	3	1.597	0.214	.886
Within Groups			410.396	55	7.462		
Total			415.186	58			
Annual Household Income							
$0–35,000	5.44	2.281					
$36–65,000	6.67	2.550					
$66,000+	7.05	2.837					
Between Groups			25.853	2	12.927	1.930	.157
Within Groups			301.397	45	6.698		
Total			327.250	47			
Primary Parent Employment Status							
Paid full time job	6.48	2.943	8.609	2	4.305	0.593	.556
Paid part time job	5.60	2.414	406.577	56	7.260		
No paid job	6.48	2.626	415.186	58			
	Mean	SD	Change in ADOS-2 Severity Score				
			SS	*df*	*MS*	*F*	*p*
Primary Parent Education							
No HS degree	1.83	1.94					
HS degree	2.25	2.46					
Some college	1.46	0.240					
College degree or higher	2.04	2.69					
Between Groups			4.824	3	1.608	0.257	.856
Within Groups			338.021	54	6.260		
Total			415.186	58			
Annual Household Income							
$0–35,000	2.83	2.12					
$36–65,000	1.22	1.64					
$66,000+	1.85	2.54					
Between Groups			17.863	2	8.931	1.781	.180
Within Groups			220.606	44	5.014		
Total			238.468	46			
Primary Parent Employment Status							
Paid full time job	1.95	2.46					
Paid part time job	2.80	2.04					
No paid job	1.39	2.62					
Between Groups			18.017	2	9.008	1.525	.227
Within Groups			324.828	55	5.906		
Total			342.845	57			

time may have allowed children to demonstrate their skills on structured testing at follow-up with more ease.

With regard to identifying factors associated with diagnostic stability in this sample, time interval between initial and follow-up evaluations, age at initial diagnosis, and initial symptom severity were associated with both follow-up symptom severity as well as overall change in symptom severity. Specifically, children with fewer ASD symptoms at initial diagnosis demonstrated fewer symptoms at follow-up, and were more likely to show overall symptom improvement. This finding is consistent with previous studies of ASD diagnostic stability and "optimal outcome" in early childhood (Eaves and Ho 2004; Fein et al. 2013; Sutera et al. 2007; Turner and Stone 2007). Interestingly, the younger a child was at initial evaluation, the more ASD symptoms they showed at follow-up and the less likely they were to demonstrate overall symptom improvement. While this finding is inconsistent with previous research (e.g., Helt et al. 2008; van Daalen et al. 2009; Turner et al., 2006; Wiggins et al., 2012), this finding may be related to the universal targeted ASD screening process through which participants in the current study were originally detected and diagnosed. Although screening and diagnostic assessment were equally accessible to all children in the EI agency, when caregivers and/or clinicians were concerned about ASD (possibly due to more severe symptom presentations), evaluations were pursued at faster rates and children moved through the diagnostic process with greater speed (Sheldrick et al. 2019). In addition, neither cognitive functioning nor language abilities at initial evaluation were related to autism symptom severity characteristics at follow-up. Moreover, time interval between initial and follow-up evaluations was found to be a significant predictor of both symptom severity at follow-up as well as symptom improvement over time. Interestingly, regardless of the age at which children were initially diagnosed, those with more time between evaluations were more likely to show overall improvement. This finding raises questions about the impact of the passage of time as well as service receipt.

Analyses also examined predictors of cognitive and verbal abilities at follow-up. Cognitive ability, specifically expressive language skills, at initial evaluation was shown to significantly predict cognitive and verbal abilities at follow-up. Children in the current sample with stronger expressive language skills at initial evaluation showed higher overall cognitive and verbal scores at follow-up. This finding highlights the critical need to closely monitor early language milestones and quickly intervene when development is delayed or atypical, in order to ensure that a child can reach his or her full potential.

Counter to our expectations, the number of EI hours (general and specialty services combined) that children received was not associated with improvements in autism symptoms and did not moderate the relationship between initial symptoms and follow-up symptom severity. This finding is consistent with multiple previous research studies that fail to show an association between number of community-delivered EI (mostly ABA) hours and diagnostic outcome (e.g., Orinstein et al. 2014). The majority of the children in the current sample received ABA services, a mostly provider-led intervention, to target ASD symptoms. Further, recent research has described the effectiveness of parent-mediated intervention programs, including sustained symptom improvements (Pickles et al. 2016). In addition, information around quality of services (i.e., professionalism of providers, consistency, family satisfaction) was not collected and thus the impact of these factors on the child's diagnostic outcome is unknown. While the amount of EI did not predict diagnostic outcome in the current sample, overall, children gained cognitive skills and showed clinically significant symptom improvement over time. Moreover, the amount of EI that a child received also did not moderate the relationship between initial and follow-up cognitive skills or initial and follow-up expressive language skills in this sample. This suggests that children with the greatest diagnostic clarity at the initial assessment, based on higher symptoms severity, continued to present with the most impairment at follow-up.

Despite high sociodemographic diversity in our sample, sociodemographic factors were not related to changes in children's symptom severity over time: race, ethnicity, family household income, primary parental education status, and primary parental job status. In other words, neither ASD symptom severity scores nor overall symptom improvement scores differed by demographic groups. Previous studies examining ASD diagnostic stability have yielded similar findings (e.g., Anderson et al. 2014; Ben Itzchak and Zachor 2009; Turner and Stone 2007); however, the majority of previous studies were considerably limited by homogenous samples. In contrast, the current sample, though small, was highly diverse in relation to these factors. Thus, this finding suggests that when detected early through targeted screening of families who have already accessed EI services and are provided with accessible evaluation services, changes in children's symptom severity do not appear to be associated with sociodemographic factors, in contrast to data presented by Fountain et al. (2012). It is possible that without the families' previously-established connection with EI and the provision of accessible evaluation services, differences related to sociodemographic factors may have been seen in this sample.

Clinical Implications

The current study, which was a longitudinal extension of an ongoing project that implemented a multi-stage, universal, targeted ASD screening process in EI agencies, provides

support for the effectiveness of targeted screening in detecting ASD in young children across demographic groups. While previous research documented racial and ethnic disparities regarding the age of initial diagnosis, and reduced opportunity for intensive intervention for children from racial and ethnic minority groups (Mandell et al. 2002), the current study demonstrated ASD symptom outcomes that were unrelated to sociodemographic factors (i.e., race, ethnicity, household income, parental education status, parental job status), when ASD is initially detected at an early age among a group of families already connected to EI services. Moreover, this study provides evidence for the effectiveness of gold-standard diagnostic procedures in early childhood and validity of early-assigned diagnoses, as the majority of participants retained their diagnoses at follow-up.

In addition, findings from the current study can aid clinicians in difficult discussions with families regarding the stability of early-assigned ASD diagnoses, especially those that were made using DSM-5 diagnostic criteria. Findings, including the 88.3% diagnostic stability rate (consistent with previous studies) and the benefits of EI and the passage of time (i.e., on average participants gained cognitive skills and showed overall ASD symptom improvement at follow-up), could provide families with additional information about a diagnosis that is often perceived as overwhelming, confusing, and anxiety-producing (Wachtel and Carter 2008).

Strengths and Limitations

The current study included a particularly diverse sample, with regard to all sociodemographic factors measured and consisted of a longitudinal design in which the same cohort of children was followed over a period of 1 to 3 years. In addition, gold-standard ASD diagnostic procedures, as well as DSM-5 criteria, were employed at both time points, maximizing the validity and long-term generalizability of diagnoses. The design and diverse sample made this study well-suited for addressing hypotheses regarding the diagnostic stability of ASD in children from historically underserved statuses.

However, the findings of the current study must be interpreted in the context of some potential limitations. While the sample is quite diverse, especially compared to previous studies of diagnostic stability, it is limited in size, reducing the statistical power of the analyses. The small percentage of the overall sample that did not meet diagnostic criteria at follow-up restricted the use binary statistical analyses, thus continuous variables (i.e., symptom severity at follow-up, change in overall severity) were analyzed instead of a binary diagnostic outcome. In addition, many caregivers had difficulty reporting specific information regarding receipt of EI services. While caregivers frequently recalled the total number of hours their child worked with therapists, it was often challenging for them to report the specific breakdown of the services that their child received between initial diagnosis and third birthday. Thus, the retrospective nature of information about EI hours and the lack of information about the quality of these hours, may have introduced bias. Finally, as children were only followed into preschool, the ultimate diagnostic outcome for children in this study is unknown at the current time. As discussed above, it is possible that ASD symptoms will wax and wane as social demands increase, thus potentially limiting the long-term generalizability of these findings.

The current study contributes to the growing body of literature regarding the stability of early-assigned ASD diagnoses, and is among the first to include a sample of children with widely diverse sociodemographic characteristics. The findings of this study raise important questions that warrant further investigation. As discussed above, future studies should aim to follow children for longer periods of time, including across developmental stages that require complex social skills. Studies with such parameters would allow researchers to carefully document the trajectories of ASD symptoms over an extended period of time, thus providing clinically useful information to both families and interventionists (i.e., for treatment-planning). It is possible that children who do not show sufficient symptoms for a diagnosis in preschool, will show additional symptoms when social demands increase later in childhood, leading them to receive a diagnosis of ASD or Social (Pragmatic) Communication Disorder. As recently described by Georgiades et al. (2017), a deeper understanding of the "turning points" at which children change developmental trajectories could inform research on the etiologies of the disorder as well as individual treatment planning.

Overall, the current study offers optimism to families of young children diagnosed with ASD. Though outcomes for each child vary, participants in this study on average demonstrated significant improvements in language and cognitive abilities, as well as ASD symptom presentation. While it is possible that children will continue to show cognitive and behavioral gains over time, previous studies have shown stability of preschool intellectual and diagnostic profiles into middle childhood and adolescence (Woolfenden et al. 2012), although further research in this area is critically needed. Though few participants in the current study "lost" their diagnoses, a majority of children showed clear developmental progress. Thus, the findings of this study can provide families with concrete evidence that young children with ASD can gain cognitive and social-communication skills over time, which can offer hope when faced with a difficult diagnosis.

Acknowledgements This paper is based on the doctoral dissertation of Dr. Ivy Giserman-Kiss, for which Dr. Carter was the

committee chairperson. Funding for this project was provided by HRSA (R40MC26195), NIMH (R01MH104400), Autism Speaks, the UMB Office of Graduate Studies, and the UMB Graduate Student Assembly. The authors would like to acknowledge the children and families who participated in this project, the Early Intervention agency staff, as well as the graduate students and staff of the UMB Child Development Lab who were integral in data collection. The authors would also like to specifically thank Melanie Feldman for her support with all aspects of this project.

Author Contributions IGK and ASC conceived of and designed the project together. IGK collected the data, performed data analysis, and wrote the paper, with significant input, feedback, and support from ASC.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflicts of interest.

Informed Consent All study procedures were approved by the institutional review board and informed consent from participants' legal guardians was documented in writing.

References

Akshoomoff, N. (2006). Use of the Mullen Scales of Early Learning for the assessment of young children with autism spectrum disorders. *Child Neuropsychology, 12*(4–5), 269–277. https://doi.org/10.1080/09297040500473714.

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

American Psychological Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed. text revision). Washington, D.C.: Author.

Anderson, D. K., Liang, J. W., & Lord, C. (2014). Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. *Journal of Child Psychology and Psychiatry, 55*(5), 485–494. https://doi.org/10.1111/jcpp.12178.

Ben Itzchak, E., & Zachor, D. A. (2009). Change in autism classification with early intervention: Predictors and outcomes. *Research in Autism Spectrum Disorders, 3*(4), 967–976. https://doi.org/10.1016/j.rasd.2009.05.001.

Bieleninik, Ł., Posserud, M. B., Geretsegger, M., Thompson, G., Elefant, C., & Gold, C. (2017). Tracing the temporal stability of autism spectrum diagnosis and severity as measured by the Autism Diagnostic Observation Schedule: A systematic review and meta-analysis. *PLoS ONE, 12*(9), e0183160. https://doi.org/10.1371/journal.pone.0183160.

Charman, T., & Baird, G. (2002). Practitioner review: Diagnosis of autism spectrum disorder in 2- and 3-year-old children. *Journal of Child Psychology and Psychiatry, 43*(3), 289–305. https://doi.org/10.1111/1469-7610.00022.

Chawarska, K., Klin, A., Paul, R., Macari, S., & Volkmar, F. (2009). A prospective study of toddlers with ASD: Short-term diagnostic and cognitive outcomes. *Journal of Child Psychology and Psychiatry, 50*(10), 1235–1245. https://doi.org/10.1111/j.1469-7610.2009.02101.x.

Chawarska, K., Klin, A., Paul, R., & Volkmar, F. (2007). Autism spectrum disorder in the second year: Stability and change in syndrome expression. *Journal of Child Psychology and Psychiatry, 48*(2), 128–138. https://doi.org/10.1111/j.1469-7610.2006.01685.x.

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

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. https://doi.org/10.1037/0033-2909.112.1.155.

Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Journal of Developmental and Behavioral Pediatrics, 27*(Suppl2), S145–S155. https://doi.org/10.1097/00004703-200604002-00013.

Corsello, C. M., Akshoomoff, N., & Stahmer, A. C. (2013). Diagnosis of autism spectrum disorders in 2-year-olds: A study of community practice. *Journal of Child Psychology and Psychiatry, 54*(2), 178–185. https://doi.org/10.1111/j.1469-7610.2012.02607.x.

Cox, A., Klein, K., Charman, T., Baird, G., Baron-Cohen, S., Swettenham, J.,… Wheelwright, S. (1999). Autism spectrum disorders at 20 and 42 months of age: Stability of clinical and ADI-R diagnosis. *Journal of Child Psychology and Psychiatry*, *40*(5), 719–732. https://doi.org/10.1111/1469-7610.00488.

Daniels, A. M., Rosenberg, R. E., Law, J. K., Lord, C., Kaufmann, W. E., & Law, P. A. (2011). Stability of initial autism spectrum disorder diagnoses in community settings. *Journal of Autism and Developmental Disorders, 41*(1), 110–121. https://doi.org/10.1007/s10803-010-1031-x.

Darrou, C., Pry, R., Pernon, E., Michelon, C., Aussilloux, C., & Baghdadli, A. (2010). Outcome of young children with autism: Does the amount of intervention influence developmental trajectories? *Autism, 14*(6), 663–677.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J.,… Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics*, *125*(1), e17–e23. https://doi.org/10.1542/peds.2009-0958.

Eaves, L. C., & Ho, H. H. (2004). The very early identification of autism: Outcome to age 4 1/2-5. *Journal of Autism and Developmental Disorders, 34*(4), 367–378. https://doi.org/10.1023/B:JADD.0000037414.33270.a8.

Elliott, C. (2007). *Differential Ability Scales* (2nd ed.). San Antonio, TX: Harcourt Assessment.

Esler, A. N., Bal, V. H., Guthrie, W., Wetherby, A., Weismer, S. E., & Lord, C. (2015). The Autism Diagnostic Observation Schedule, toddler module: Standardized severity scores. *Journal of Autism and Developmental Disorders, 45*(9), 2704–2720. https://doi.org/10.1007/s10803-015-2432-7.

Fein, D., Barton, M., Eigsti, I., Kelley, E., Naigles, L., Schultz, R. T.,… Tyson, K. (2013). Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry*, *54*(2), 195–205. https://doi.org/10.1111/jcpp.12037.

Flanagan, H. E., Smith, I. M., Vaillancourt, T., Duku, E., Szatmari, P., Bryson, S.,… Waddell, C. (2015). Stability and change in the cognitive and adaptive behaviour scores of preschoolers with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 45*(9), 2691–2703. https://doi.org/10.1007/s10803-015-2433-6.

Fountain, C., Winter, A. S., & Bearman, P. S. (2012). Six developmental trajectories characterize children with autism. *Pediatrics, 129*(5), e1112–e1120. https://doi.org/10.1542/peds.2011-1601.

Georgiades, S., Bishop, S. L., & Frazier, T. (2017). Editorial perspective: Longitudinal research in autism—introducing the concept of 'chronogeneity'. *Journal of Child Psychology and Psychiatry, 58*(5), 634–636. https://doi.org/10.1111/jcpp.12690.

Guthrie, W., Swineford, L. B., Nottke, C., & Wetherby, A. M. (2013). Early diagnosis of autism spectrum disorder: Stability and change in clinical diagnosis and symptom presentation. *Journal*

of Child Psychology and Psychiatry, 54(5), 582–590. https://doi.org/10.1111/jcpp.12008.

Helt, M., Kelley, E., Kinsbourne, M., Pandey, J., Boorstein, H., Herbert, M., et al. (2008). Can children with autism recover? If so, how? *Neuropsychology Review, 18*(4), 339–366. https://doi.org/10.1007/s11065-008-9075-9.

Kleinman, J. M., Ventola, P. E., Pandey, J., Verbalis, A. D., Barton, M., Hodgson, S.,… Fein, D. (2008). Diagnostic stability in very young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(4), 606–615. https://doi.org/10.1007/s10803-007-0427-8.

Lord, C., & Bishop, S. L. (2010). Autism spectrum disorders: Diagnosis, prevalence, and services for children and families. *Social Policy Report, 24*(2), 1–23.

Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of General Psychiatry, 63*(6), 694–701. https://doi.org/10.1001/archpsyc.63.6.694.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule: ADOS-2*. Los Angeles, CA: Western Psychological Services.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*(1), 3–9. https://doi.org/10.1037/0022-006X.55.1.3.

Mahli, P., & Singhi, P. (2011). Follow-up of children with autism spectrum disorders: Stability and change in diagnosis. *Indian Journal of Pediatrics, 78*(8), 941–945. https://doi.org/10.1007/s12098-011-0370-8.

Mandell, D. S., Listerud, J., Levy, S. E., & Pinto-Martin, J. A. (2002). Race differences in the age at diagnosis among Medicaid-eligible children with autism. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*(12), 1447–1453. https://doi.org/10.1097/00004583-200212000-00016.

Mullen, E. M. (1995). *Mullen Scales of Early Learning* (AGS ed.). Circle Pines, MN: American Guidance Service, Incorporated.

Oono, I. P., Honey, E. J., & McConachie, H. (2013). Parent-mediated early intervention for young children with autism spectrum disorders (ASD). *Cochrane Database of Systematic Reviews, 4*, Art. No. CD009774. https://doi.org/10.1002/14651858.cd009974.pub2.

Orinstein, A. J., Helt, M., Troyb, E., Tyson, K. E., Barton, M. L., Eigsti, I.,… Fein, D. A. (2014). Intervention for optimal outcome in children and adolescents with a history of autism. *Journal of Developmental and Behavioral Pediatrics, 35*(4), 247–256. https://doi.org/10.1097/dbp.0000000000000037.

Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T.,… Iosif, A. M. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: A baby siblings research consortium study. *Journal of Child Psychology and Psychiatry.* Advance online publication. https://doi.org/10.1111/jcpp.12421.

Pickles, A., Le Couteur, A., Leadbitter, K., Salomone, E., Cole-Fletcher, R., Tobin, H.,… Green, J. (2016). Parent-mediated social communication therapy for young children with autism (PACT): Long-term follow-up of a randomised controlled trial. *The Lancet, 388*(10059), 2501–2509. https://doi.org/10.1016/s0140-6736(16)31229-6.

Sheldrick, R. C., Frenette, E., Vera, J. D., Mackie, T. I., Martinez-Pedraza, F., Hoch, N., … Carter, A. S. (2019). What drives detection and diagnosis of autism spectrum disorder? Looking under the hood of a multi-stage screening process in early intervention. *Journal of Autism and Developmental Disorders, 49*(6):2304–2319. https://doi.org/10.1007/s10803-019-03913-5.

Siller, M., & Morgan, L. (2018). Systematic review of research evaluating parent-mediated interventions for young children with autism: Years 2013 to 2015. In M. Siller & L. Morgran (Eds.), *Handbook of parent-mediated interventions for very young children with autism* (pp. 1–21). Cham: Springer.

Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). *Vineland Adaptive Behavior Scales, Third Edition (Vineland-3)*. San Antonio, TX: Pearson.

Sutera, S., Pandey, J., Esser, E. L., Rosenthal, M. A., Wilson, L. B., Barton, M.,… Fein, D. (2007). Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*(1), 98–107. https://doi.org/10.1007/s10803-006-0340-6.

Turner, L. M., & Stone, W. L. (2007). Variability in outcome for children with an ASD diagnosis at age 2. *Journal of Child Psychology and Psychiatry, 48*(8), 793–802. https://doi.org/10.1111/j.1469-7610.2007.01744.x.

Turner, L. M., Stone, W. L., Pozdol, S. L., & Coonrod, E. E. (2006). Follow-up of children with autism spectrum disorders from age 2 to age 9. *Autism, 10*(3), 243–265. https://doi.org/10.1177/1362361306063296.

van Daalen, E., Kemner, C., Dietz, C., Swinkels, S. N., Buitelaar, J. K., & van Engeland, H. (2009). Inter-rater reliability and stability of diagnoses of autism spectrum disorder in children identified through screening at a very young age. *European Child and Adolescent Psychiatry, 18*(11), 663–674. https://doi.org/10.1007/s00787-009-0025-8.

Wachtel, K., & Carter, A. S. (2008). Reaction to diagnosis and parenting styles among mothers of young children with ASDs. *Autism, 12*(5), 575–594. https://doi.org/10.1177/1362361308094505.

Wiggins, L. D., Baio, J., Schieve, L., Lee, L., Nicholas, J., & Rice, C. E. (2012). Retention of autism spectrum diagnoses by community professionals: Findings from the Autism and Developmental Disabilities Monitoring Network, 2000 and 2006. *Journal of Developmental and Behavioral Pediatrics, 33*(5), 387–395. https://doi.org/10.1097/DBP.0b013e3182560b2f.

Woolfenden, S., Sarkozy, V., Ridley, G., & Williams, K. (2012). A systematic review of the diagnostic stability of autism spectrum disorder. *Research in Autism Spectrum Disorders, 6*(1), 345–354. https://doi.org/10.1016/j.rasd.2011.06.008.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.