




# Expert Clinician Certainty in Diagnosing Autism Spectrum Disorder in 16–30-Month-Olds: A Multi-site Trial Secondary Analysis

Cheryl Klaiman<sup>1,2</sup> · Stormi White<sup>1,2</sup> · Shana Richardson<sup>1</sup> · Emma McQueen<sup>1,2</sup> · Hasse Walum<sup>1,2</sup> · Christa Aoki<sup>1</sup> · Christopher Smith<sup>3</sup> · Mendy Minjarez<sup>4</sup> · Raphael Bernier<sup>4</sup> · Ernest Pedapati<sup>5</sup> · Somer Bishop<sup>6</sup> · Whitney Ence<sup>6</sup> · Allison Wainer<sup>7</sup> · Jennifer Moriuchi<sup>7</sup> · Sew-Wah Tay<sup>8</sup> · Yiming Deng<sup>8</sup> · Warren Jones<sup>1,2,9</sup> · Scott Gillespie<sup>10</sup> · Ami Klin<sup>1,2,9</sup> 

Accepted: 30 October 2022 / Published online: 17 November 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Differential diagnosis of young children with suspected autism spectrum disorder (ASD) is challenging, and clinician uncertainty about a child's diagnosis may contribute to misdiagnosis and subsequent delays in access to early treatment. The current study was designed to replicate and expand a recent report in this Journal (McDonnell et al. in *J Autism Dev Disord* 49:1391–1401, <https://doi.org/10.1080/15374416.2020.1823850>, 2019), in which only 60% of diagnoses were made with complete certainty by clinicians evaluating 478 toddlers and preschool children referred for possible ASD to specialized clinics. In this study, secondary analyses were performed on diagnostic, demographic and clinical data for 496 16–30-month-old children who were consecutive referrals to a 6-site clinical trial executed by specialized centers with experienced clinicians following best-practice procedures for the diagnosis of ASD. Overall, 70.2% of diagnoses were made with complete certainty. The most important factor associated with clinician uncertainty was mid-level autism-related symptomatology. Mid-level verbal age equivalents were also associated with clinician uncertainty, but measures of symptomatology were stronger predictors. None of the socio-demographic variables, including sex of the child, was significantly associated with clinician certainty. Close to one third of early diagnoses of ASD are made with a degree of uncertainty. The delineation of specific ranges on the ADOS-2 most likely to result in clinician uncertainty identified in this study may provide an opportunity to reduce random subjectivity in diagnostic decision-making via calibration of young-child diagnostic thresholds based on later-age longitudinal diagnostic outcome data, and via standardization of decision-making in regard to clinical scenarios frequently encountered by clinicians.

**Keywords** Autism spectrum disorder · Toddlers · Diagnostic certainty · Differential diagnosis

✉ Ami Klin  
ami.klin@emory.edu

<sup>1</sup> Marcus Autism Center, Children's Healthcare of Atlanta, Atlanta, GA, USA

<sup>2</sup> Division of Autism & Related Disorders, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, USA

<sup>3</sup> Southwest Autism Research & Resource Center, Phoenix, AZ, USA

<sup>4</sup> Seattle Children's Hospital & Department of Psychiatry, University of Washington, Seattle, WA, USA

<sup>5</sup> Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

<sup>6</sup> University of California San Francisco, San Francisco, CA, USA

<sup>7</sup> Rush University Medical Center, Chicago, IL, USA

<sup>8</sup> Libra Medical Inc., Minneapolis, MN, USA

<sup>9</sup> Center for Translational Social Neuroscience, Emory University, 954 Gatewood Road, Atlanta, GA, USA

<sup>10</sup> Pediatric Biostatistics Core, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, USA

## Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition diagnosed behaviorally by the presence of early-emerging and persistent differences in social interaction and communication skills, and by the presence of restricted and repetitive patterns of behavior (American Psychiatric Association (APA), 2013). Early identification and treatment are two of the most important factors promoting improved lifetime outcomes for children with ASD (Zwaigenbaum et al., 2015a, 2015b; Zwaigenbaum et al., 2015a, 2015b). Because up to 80% of parents recognize developmental concerns in their children subsequently diagnosed with ASD by age 2 years (Chawarska et al., 2007), the American Academy of Pediatrics recommends universal screening for ASD at 18 and 24 months (Hyman et al., 2020; Johnson et al., 2007). And yet, the median age of ASD diagnosis in the US has been late, between 4 and 5 years of age, and unchanging over various Centers for Disease Control and Prevention (CDC) bi-annual surveillance cohorts (Baio et al., 2018; Christensen et al., 2016; Maenner et al., 2020). This public health challenge moves eligibility and access to intervention services from within the period of maximal neuroplasticity (Johnson, 1999) and of the (US) federal mandate for early intervention services (US Department of Education, 2020)—the period from birth to 3 years—to a point several years hence, when many years of development have already played a large role in shaping the course of a child's condition (Klin et al., 2020). Age of diagnosis is later still in minority, low-income and rural communities (Constantino et al., 2020; Daniels & Mandell, 2014; Daniels et al., 2014; Maenner et al., 2020).

Several factors account for late ages of diagnosis of ASD, including primary care clinicians' time restrictions (Honigfeld et al., 2012), lack of accurate and cost-effective screening and diagnostic tools (Randall et al., 2018), limited number of expert clinicians (US Government Accountability Office (GAO), 2013), and a general "wait and see approach" that results in ASD diagnosis being made only when symptomatic presentation is obvious, typically when the child is older (Daniels et al., 2014). This study focuses on one potential such factor—levels of clinician certainty during diagnostic assessment, which has been understudied in the literature. This is particularly acute in the case of young children, where diagnostic certainty can be challenging because overt ASD symptomatology is still emerging and because other developmental delays or environmental stressors can confound clinical presentation (Zwaigenbaum et al., 2015a, 2015b), even when best-practices are deployed by expert clinicians (Kim & Lord, 2012). The dynamic nature of unfolding

early symptomatology in ASD, particularly, represents a major challenge in early differential diagnosis, with high rates of false negatives having been reported in the literature when children are followed-up from 24 months of age to later years (Ozonoff et al., 2015).

We could identify only one systematic study of clinicians' diagnostic certainty in young children with ASD (McDonnell et al., 2019). It was conducted in three university-based tertiary-care clinics specializing in comprehensive ASD evaluations, in which clinicians were licensed and experienced, all trained to research reliability levels on the *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2)* (Lord et al., 2012). The study included 478 young children with an average age of 42 months (SD 14.7; range 18 to 73 months), 79.1% male, 21.8% non-White, with average IQ of 72.6 (SD 21.3) (global scores on developmental or cognitive assessment tools, depending on child's chronological age). Of the 478 children, 339 (70.9%) were given an expert clinician diagnosis of ASD, and 139 (29.1%) of non-ASD. Most children with non-ASD received another diagnosis (mostly global developmental delay, language disorder or ADHD), with only 7 children (1.5%) receiving no diagnosis. Clinicians rated their certainty in their diagnosis on a 4-point scale as completely uncertain (1.8% of cases), somewhat uncertain (8.2% of cases), somewhat certain (30.7% of cases), and completely certain (59.2% of cases). For purposes of data analysis, "uncertain" included the first 3 categories and "certain" included the last category. Overall, therefore, only approximately 60% of diagnoses were made with certainty. ADOS-2 results were strongly associated with certainty: "uncertain" rating was associated with moderate levels of ASD symptoms as measured by the ADOS-2 (vs high or low).

These are surprisingly low levels of clinician certainty given that the three participating centers are highly specialized for the differential diagnosis of ASD. And yet, it is not possible to estimate what should be considered an optimal level of clinician certainty as there are no other studies available for comparison. In the general community, where practitioners have less ASD-specific training and experience, and best-practice diagnostic assessments are typically not viable, levels of clinician certainty are likely to be even lower. These results have potentially important impact on management of the diagnostic process, public health priorities, and ASD research. For clinical management, clinician uncertainty may protract the diagnostic process, thus contributing to significant stress, time-loss and financial cost for families (Crane et al., 2016; Rutherford et al., 2016). For public health, clinician uncertainty may be an important obstacle to early and efficient diagnosis of ASD and, consequently, to eligibility for services (Rutherford et al., 2018). And in research, clinician uncertainty may contribute to higher

heterogeneity of samples—reducing signal/noise ratio for meaningful analyses—if the “ground truth” for ASD versus Non-ASD diagnostic assignment is certain in some, but not in all, cases (Lijmer et al., 1999).

To address this important gap in the literature, and to attempt a replication of the McConnell and colleagues’ results (McDonnell et al., 2019), the present study analyzed data obtained via a multi-site clinical trial in six US specialized centers to probe clinician certainty in the differential diagnosis of ASD in 16–30-month-old children, a critically important age if we are to advance increased access to early (Birth to Three) intervention services. Specifically, our goals were to measure: (a) overall levels of clinician certainty in Expert Clinician Diagnosis (ECD), as well as certainty of ASD vs Non-ASD ECDs; (b) differences in ECD levels of certainty across the six participating sites; and (c) impact of child demographic and clinical characterization data on clinician certainty. Additionally, we explored clinicians’ stated reasons for their uncertainty when making the ECD of ASD and Non-ASD. This plan was intended to probe the intriguing findings of the only study available on this important question (McDonnell et al., 2019), and if so, to set the stage for future research, designed and executed from the outset to shed light on clinician decision-making in differential diagnosis of toddlers with ASD, with the ultimate goal of improving clinical practice.

## Method

### Context of the Study

This study is a secondary data analysis of a ClinicalTrials.gov-registered (NCT03469986) multi-site, prospective, double-blind trial of an eye-tracking-based investigational device for the diagnosis of ASD relative to the reference-standard and current best-practice Expert Clinician Diagnosis (ECD) of ASD in 16–30-month-old children. Six sites in the US contributed to the sample, all of which were university-based, specialized centers for the diagnosis of ASD in young children (alphabetically, Cincinnati Children’s Hospital Medical Center, Marcus Autism Center at Emory University and Children’s Healthcare of Atlanta, Rush University Medical Center, Seattle Children’s Hospital and University of Washington, Southwest Autism Research & Resource Center, and University of California San Francisco Center for Autism Spectrum and Neurodevelopmental Disorders). All participating families provided informed consent approved by the Institutional Review Boards of the respective academic health institutions.

## Participants

Children were recruited and enrolled consecutively based on parental or provider’s concern for ASD or in response to advertised recruitment for this study. The relevance of this sample for the subject of study is that the sampling strategy minimized spectrum bias (by recruiting in 6 sites) and selection bias (participants were enrolled consecutively with no ASD-specific screening) (Lijmer et al., 1999), thus representing typical referral patterns to specialized ASD clinics. Nevertheless, given the nature of the pivotal trial, inclusion criteria included normal or corrected-to-normal vision and hearing, and parent or guardian’s ability to communicate with investigators and to provide informed consent in English. Exclusion criteria included known genetic disorders (e.g., neurofibromatosis), vision and/or hearing impairments (e.g., congenital nystagmus), or active and/or acute illness (e.g., fever, uncontrolled seizure disorder). A total of 496 (or 98.2% of the full enrollment of 505) 16–30-month-olds completed the study and had no missing data for Expert Clinician Diagnosis (ECD), Clinician Certainty Rating, and scores on standardized diagnostic and developmental instruments. Of the  $N = 9$  participants not included in the study,  $N = 3$  met exclusion criteria of the trial,  $N = 3$  withdrew from the study, and  $N = 3$  did not complete the clinical protocol in full, and were, therefore, excluded from further analyses.

## Procedures

The experimental and clinical diagnostic procedures were conducted independently on the same day. Clinical procedures included the execution of standardized diagnostic and developmental assessments, parental questionnaires, inventories of developmental history and current presentation, all of which contributed to an expert clinician diagnosis (ECD) following *best-estimate diagnostic procedures* (Leckman et al., 1982) using the totality of information available to the expert clinician (Kim & Lord, 2012), including prior medical records when these were available. This procedure applied to both the diagnoses of ASD and of Non-ASD conditions, including the ECD of (Non-ASD) Developmental Delay (DD), which, although assigned by clinician judgment also typically reflected the rule of  $\geq 1.5$  standard deviations (SDs) below the mean in two subscales of the *Mullen Scales of Early Learning (Mullen)* (Mullen, 1995), or of  $\geq 2.0$  SDs below the mean in one Mullen subscale. Standards for attainment and maintenance of fidelity and reliability on standardized procedures are provided in detail in Supplementary Materials). All procedures were approved by the Institutional Review Board at each site (one site used a central IRB), and all parents/guardians provided informed consent.

## Clinician Certainty in Diagnosis

Clinicians performing the expert clinician diagnosis (ECDs) were required to rate the certainty of their diagnoses for every participant in the study upon completion of the evaluation. They were asked to complete a 5-point scale on their diagnostic certainty levels as follows: “Extremely Uncertain” (0–20%); “Somewhat Uncertain” (21–40%); “On the fence” (41–60%); “Somewhat Certain” (61–80%); and “Certain” (81–100%). They were asked to rate their level of certainty of their ECD regardless of the diagnosis assigned to the child (i.e., ASD or Non-ASD). This procedure was intended to mirror the one adopted in the study targeted for replication (McDonnell et al., 2019). There is currently no available tool for measuring clinician certainty in ASD diagnosis that has been psychometrically validated. Our intent was that a replication of the findings might generate the impetus for future development and study of such a scale. Also, similarly to the targeted study, diagnostic certainty ratings were collapsed for subsequent analyses, with the top level considered “Certain” ECDs, and the other levels considered “Uncertain” ECDs, thus ensuring comparability of findings across the two studies. Finally, clinicians were requested to provide rationale(s) for uncertainty ratings (open-ended narrative).

Fifteen clinicians (2 or 3 per site) conducted all clinical assessments, 14 psychologists and 1 child and adolescent psychiatrist. They all had a minimum of 3 years of relevant clinical experience subsequent to completing post-doctoral fellowships specializing on diagnosis and assessment of toddlers with ASD and related neurodevelopmental conditions. All clinicians had certification to research level on the ADOS-2; also, reliability was demonstrated at the beginning of the study and monitored throughout (see Supplementary Materials for additional details).

## Measures

### Parent Questionnaires

Parents were asked to complete a set of questionnaires that included information on demographic information, developmental history, medical history, and medication and intervention history. These procedures were standardized across all sites.

### Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) (Lord et al., 2012)

The ADOS-2 is a standardized assessment of ASD symptoms using a semi-structured play session consisting of a set of presses for social and communication interaction intended to elicit behaviors relevant to the diagnosis of ASD. These

observations are then coded according to detailed scoring criteria. The ADOS-2 has been extensively validated for use in diagnostic assessment of ASD (Randall et al., 2018). ADOS-2 data were analyzed using total algorithm scores for Social Affect (SA) and Restricted & Repetitive Behaviors (RRB)—as well as Total algorithm scores (SA + RRB scores). Information on ADOS-2 modules and scoring algorithms, their frequency of use in the sample, and rationale for adopting these scores rather than standardized Calibrated Severity Scores are provided in Supplementary Materials. Data collection for this trial ended in 2019, therefore, before COVID-19 necessitated changes and adaptations in ADOS-2 administration.

### Mullen Scales of Early Learning (Mullen) (Mullen, 1995)

The Mullen is a standardized developmental assessment for children 0–68 months of age and is well-validated for measurement of cognitive, language and motor skills, measurements that are considered a required element in best-practice evaluations of children with ASD (Zwaigenbaum et al., 2015a, 2015b). It yields results (T scores and age equivalents) across 5 scales: Receptive and Expressive Language, Visual Reception (nonverbal learning), Fine and Gross Motor skills. In this study, we analyzed age equivalent scores yielding measures of verbal and nonverbal ability. To assign a single Verbal Ability Age Equivalent Score to each child, we averaged the Mullen’s Receptive and Expressive Language age equivalents; Nonverbal Ability Age Equivalent Scores corresponded to the Mullen’s Visual Reception age equivalent.

## Analytic Strategy

All analyses were performed in SAS v.9.4 (Cary, NC) and CRAN R v.4.2 (Vienna, Austria), and statistical significance was evaluated at the 0.05 threshold throughout. Continuous variables are summarized using means with standard deviations and/or medians with ranges, and categorical variables are presented as frequencies with percentages. Any hypothesis testing between two groups (e.g., ECD of ASD versus ECD of Non-ASD) was performed using two-sample *t*-tests for continuous variables and *chi-square* tests of independence for categorical variables; for hypothesis tests with > 2 groups, two-sample tests were replaced with unequal variances ANOVA, and in circumstances when expected frequencies were low, *chi-square* tests were replaced by Fisher’s *exact* tests.

### Overall Levels of Clinician Certainty in Their Expert Clinician Diagnosis (ECD), and by ECDs of ASD vs Non-ASD

Levels of Clinician Certainty were dichotomized as “Certain” (81–100%) versus “Uncertain” (0–80%), as in the



McDonnell and colleagues' study (McDonnell et al., 2019), and tabulated categorically with frequencies and percentages. Hypothesis testing between clinician certainty and ECD (ASD vs Non-ASD) was performed with a *chi-square* test of independence.

### Differences in ECD Levels of Uncertainty Across the Six Participating Sites

Associations between levels of Clinician Certainty and ECD, moderated by site, were assessed using *Cochran–Mantel–Haenszel* (CMH) statistics, including odds ratios with 95% confidence intervals and *p*-values and a *Breslow–Day* test for homogeneity of odds ratios between sites.

### Impact of Child Demographic and Clinical Characterization Data on Levels of Clinician Certainty

Demographic and clinical data were summarized by levels of clinician certainty (“Certain” versus “Uncertain”), and the odds of a certain diagnosis, irrespective of ASD or Non-ASD, were calculated using mixed effects binary logistic regression models. In these models, the demographic and clinical variables were the fixed effects, and the random effects were site-specific intercepts, accounting for random variation in ECD certainty between sites (see Supplementary Materials for additional details). Due to bimodal and/or non-normal distributions of ADOS-2 and Mullen verbal and nonverbal age equivalents, as well as observed non-linearity in the relationships between ADOS-2 and Mullen sub-scores and the probability of diagnostic certainty, ADOS-2 and Mullen sub-scores were first assessed as absolute differences from their respective means. As a secondary analysis, continuous ADOS-2 and Mullen sub-scores were analyzed as natural cubic splines within the mixed effects logistic regression models, employing a truncated power function basis for the spline expansion and knots established at the 5th, 27.5th, 50th, 72.5th, and 95th percentiles, as recommended by Harrell (Harrell, 2015). Probabilities of certainty results from these natural cubic splines models helped inform thresholds for clinically “Mild”, “Moderate”, and “Severe” score ranges. These score ranges were then utilized as linear piecewise covariates within the mixed effects binary logistic regression models. Results from all logistic models are presented either as odds ratios with 95% confidence intervals and *p*-values or estimated probabilities. Predictors in the logistic regression models are considered both individually (univariable) and together with all other covariates (multivariable), i.e., we performed no variable reduction. Model fits and validation were evaluated using AICC and Chi-square/DF statistics. During the model-building and data visualization processes, Mullen scores were found to be weaker predictors of diagnostic certainty than

ADOS-2; thus, only ADOS-2 was considered in secondary, linear piecewise regression analyses.

Associations were calculated in the full sample using mixed effects binary logistic regression, modeling the outcome dichotomously as certain (81–100%) versus not certain (0–80%), and accounting for random variation in diagnostic certainty between sites.

### Clinicians' Stated Reasons for Their Uncertainty When Making the ECD of ASD and of Non-ASD

We explored clinicians' uncertainty and decision-making by reviewing narratives offered as rationale for diagnostic uncertainty. Initially, one expert clinician extracted discrete reasons for uncertainty contained in the narratives, and then two expert clinicians rated each case using these reasons (whether a given case contained a given reason or not), while documenting the number of cases that lacked any rationale for the uncertainty rating. Our goal was to summarize the most frequent reasons for diagnostic uncertainty mentioned by clinicians, separately for ECDs of ASD and Non-ASD. This descriptive summary was important because it covers domains of the evaluation that were not quantitative in nature, but which were likely to influence clinician decision-making, particularly developmental and family histories.

## Results

### Sample Characterization

Site contribution to the finalized sample was uneven, with specific contributions being  $N = 32$  (6.4%),  $N = 148$  (29.8%),  $N = 68$  (13.7%),  $N = 121$  (24.4%),  $N = 68$  (13.7%), and  $N = 59$  (11.9%), respectively, for Sites 1, 2, 3, 4, 5 and 6, to a total of  $N = 496$  valid cases with outcome data. Of the  $N = 496$  cases,  $N = 16$  had some missing data (race, ethnicity or maternal education), but all  $N = 496$  cases were kept in the analysis. Overall, the sample was 60.7% Male, had a mean age of 24.2 months (SD 4.4; Median 25.0; Range 16–30 months), 7.7% Black (and 17.0% with race other than White), 14.6% Hispanic/Latino, and maternal education was higher than high school in 89.1% of participating families that provided these data. There were broad distributions of verbal and nonverbal age equivalents (AE) and ADOS-2 scores. Demographic and Clinical data for the full sample, by Expert Clinician Diagnosis (ECD) (ASD and Non-ASD), are provided in Table 1. The total sample included 234 (47.2%) children with ASD and 262 (52.8%) children with Non-ASD ECDs. Of the  $N = 262$  children with Non-ASD,  $N = 216$  (82.4%) had one or more Non-ASD developmental delay(s) (DD) or condition(s) (most commonly speech-language and

**Table 1** Demographic and clinical characterization of the sample (N = 496)

	Full sample	Expert clinician diagnosis (ECD)		<i>p</i> values <sup>a</sup> ASD—Non- ASD compari- son
		ASD	Non-ASD	
N	496	234	262	
% Of full sample		47.2%	52.8%	
<b>Demographics</b>				
Sex: Male n/N (%)	301/496 (60.7%)	172/234 (73.5%)	129/262 (49.2%)	< 0.0001*
<b>Child age (months)</b>				
Mean (SD)	24.2 (4.4)	25.0 (4.2)	23.4 (4.5)	< 0.0001*
Median	25.0	26.0	24.0	
Range (Min–Max) <sup>b</sup>	(16.0–30.0)	(16.0–30.0)	(16.0–30.0)	
<b>Race—n/N (%)</b>				
White	372/494 (75.3%)	160/233 (68.7%)	212/261 (81.2%)	0.0052*
<b>Other categories</b>				
Black	38/494 (7.7%)	22/233 (9.4%)	16/261 (6.1%)	
Other	84/494 (17.0%)	51/233 (21.9%)	33/261 (12.6%)	
Ethnicity: Hispanic/Latino—n/N (%)	72/493 (14.6%)	47/234 (20.1%)	25/259 (9.7%)	0.0011*
Maternal Education > high school—n/N (%)	434/487 (89.1%)	190/228 (83.3%)	244/259 (94.2%)	< 0.0001*
<b>Clinical characteristics</b>				
Mullen Verbal Age Equivalent <sup>c</sup> Mean (SD)	18.9 (9.7)	12.7 (7.1)	24.6 (8.2)	< 0.0001*
Mullen Non-Verbal Age Equivalent <sup>d</sup> Mean (SD)	22.8 (8.5)	18.6 (5.9)	26.5 (8.7)	< 0.0001*
ADOS-2 Total Score Mean (SD)	11.3 (8.9)	19.4 (5.2)	4.1 (3.5)	< 0.0001*
ADOS-2 SA Score Mean (SD)	8.4 (6.9)	14.5 (4.3)	2.9 (3.0)	< 0.0001*
ADOS-2 RRB Score Mean (SD)	2.9 (2.5)	4.9 (2.0)	1.2 (1.3)	< 0.0001*
<b>ADOS-2 Form Performed</b>				
Toddler Module—n/N (%)	437/496 (88.1%)	227/234 (97.0%)	210/262 (80.2%)	< 0.0001*
Module 2—n/N (%)	59/496 (11.9%)	7/234 (3.0%)	52/262 (19.8%)	
Presence of one or more (Non-ASD) DDs <sup>e</sup> —n/N (%)	345/496 (69.6%)	129/234 (55.1%)	216/262 (82.4%)	< 0.0001*
No ASD or DD diagnosis—n/N (%)	46/496 (9.2%)	0/234	46/262 (17.6%)	

<sup>a</sup>*p* values correspond to  $\chi^2$  or *t* tests. Significance established at *p* < 0.05 or lower, denoted by \*. Non-significant comparison denoted by “ns”

<sup>b</sup>Inclusion criteria for enrollment in trial included age range of 16 to 30 months

<sup>c</sup>In months. Mullen Verbal Age Equivalents were calculated as the Average of Mullen Expressive Language Age Equivalent and the Mullen Receptive Language Age Equivalent for each child

<sup>d</sup>In months. Mullen Nonverbal Age Equivalents corresponds to Mullen Visual Reception Age Equivalents

<sup>e</sup>Expert Clinician Diagnosis of one or more (Non-ASD) Developmental Delays (DDs) or conditions

global developmental delays), with N = 46 (17.6%) having no diagnosis (i.e., being unaffected). Of the N = 234 children with ASD, N = 129 (55.1%) also had one or more DDs (most commonly speech-language and cognitive). In ASD vs Non-ASD comparisons, the samples differed significantly on every demographic variable: Sex Ratio (ASD more males), Age (ASD slightly older), Race (ASD more Non-White), Ethnicity (ASD more Hispanic/Latino), and Maternal Education (ASD fewer families with maternal education greater than high school). The samples also differed significantly on every clinical variable: as expected, the samples differed on verbal and nonverbal abilities (ASD more delayed) and on social disability scores (ASD more elevated), and fewer children with ASD received the

ADOS-2 Module 2 (rather than the ADOS-2 Toddler Module) than Non-ASD children.

### Overall Levels of Clinician Certainty in Their Expert Clinician Diagnosis (ECD), and by ECDs of ASD and Non-ASD

For the whole sample, clinicians made an ECD with certainty in 348 of 496 cases (70.2%), with a trend for higher levels of certainty for ECD diagnosis of Non-ASD relative to ASD, although this trend did not reach statistical significance (“Certain” ECD for ASD of 67.1% and for Non-ASD of 72.9%,  $\chi^2$  (df) = 1.991 (1), *p* = 0.158; see below for a more detailed analysis of this trend). In terms of the five

levels of Uncertainty, most “Uncertain” cases fell in the 61–80% category (20.2% of all cases, and 67.6% of total Uncertain cases), followed by the 41–60% category (6.7% of all cases, and 22.3% of total Uncertain cases), with categories of <40% capturing only a minority of cases (3.0% of all cases, and 10.1% of total Uncertain cases). This distribution applied similarly to both ECD diagnoses of ASD and Non-ASD ( $\chi^2$  (df) = 7.422 (4),  $p = 0.115$ ).

### Differences in Certainty Between ECD Levels (ASD and Non-ASD) Across the Six Participating Sites

Rates of “Certain” diagnoses for all cases by site ranged from 56.3 to 88.2% (Mean 69.1%, SD 13.3%). The cross-site variability of rates of Certain diagnoses applied to both diagnoses of ASD (Mean 61.8%, SD 19.6%) and Non-ASD (Mean 72.3%, SD 13.0%). For all but one site, there was more certainty in the ECD of Non-ASD than ASD. Table 2 provides the odds ratio, by site, of a Certain diagnosis for ASD relative to Non-ASD. As can be seen, odds ratios varied markedly by site, ranging from a low of 0.22 (i.e., the odds of a Certain diagnosis for ASD were 78% lower than the odds of a Certain diagnosis for Non-ASD) to 1.83 (i.e., the odds of a Certain diagnosis for ASD were 83% higher than the odds of a Certain diagnosis for Non-ASD). While there was marked variability in levels of ECD clinician certainty across the six sites, these differences did not reach statistical significance (Breslow-Day test for homogeneity of odds ratios  $p = 0.069$ ). Across all sites, the pooled Mantel–Haenszel odds ratio for certainty in an ASD diagnosis relative to certainty in a Non-ASD diagnosis was 0.86 (0.57, 1.28), indicating a trend of overall higher certainty in Non-ASD diagnoses. Nevertheless, this trend is not statistically significant, and, for practical clinical inference, it does not indicate impactful difference in levels of clinician certainty between the ECDs of ASD vs. Non-ASD.

While statistical testing of ECD clinician certainty between sites did not reach statistical significance, our analytic plan still robustly accounted for site-specific random variation in ECD certainty (due to both unmeasured and measured sources of variation), via incorporation of site-specific random effects in all regression models (Feaster et al., 2011). Possible sources of cross-site variability arising from measured and unmeasured site-specific characteristics in this sample are discussed further in the Supplementary Materials (See Supplementary Table S1 for demographic and clinical characterization of the sample by participating site). By incorporating random effects, fixed effects regression associations are calculated using as much information as possible.

**Table 2** Clinician certainty (categorical: certain or uncertain) in expert clinician diagnosis (ECD) of ASD and non-ASD by site (x6)

N (row %)	Not certain	Certain	ECD—certainty OR (95% CI) <sup>a</sup>	<i>p</i> values
<b>Site 1 (N = 32)</b>				
ECD—ASD	5 (71.4%)	2 (28.6%)	0.23 (0.04, 1.41)	0.215
ECD—Non-ASD	9 (36%)	16 (64%)	Reference	
<b>Site 2 (N = 148)</b>				
ECD—ASD	13 (18.6%)	57 (81.4%)	1.83 (0.85, 3.98)	0.176
ECD—Non-ASD	23 (29.5%)	55 (70.5%)	Reference	
<b>Site 3 (N = 68)</b>				
ECD—ASD	5 (23.8%)	16 (76.2%)	0.22 (0.05, 1.02)	0.098
ECD—Non-ASD	3 (6.4%)	44 (93.6%)	Reference	
<b>Site 4 (N = 121)</b>				
ECD—ASD	29 (41.4%)	41 (58.6%)	0.99 (0.48, 2.06)	1.000
ECD—Non-ASD	21 (41.2%)	30 (58.8%)	Reference	
<b>Site 5 (N = 68)</b>				
ECD—ASD	8 (26.7%)	22 (73.3%)	0.62 (0.20, 1.97)	0.603
ECD—Non-ASD	7 (18.4%)	31 (81.6%)	Reference	
<b>Site 6 (N = 59)</b>				
ECD—ASD	17 (47.2%)	19 (52.8%)	0.60 (0.20, 1.75)	0.501
ECD—Non-ASD	8 (34.8%)	15 (65.2%)	Reference	

<sup>a</sup>Breslow-Day test for homogeneity of odds ratios  $p = 0.069$ , indicating a statistically insignificant difference in odds ratios between strata (i.e., sites); Pooled Mantel–Haenszel OR (95% CI) = 0.86 (0.57, 1.28)

### The Impact of Demographic and Clinical Characterization Data on Levels of Clinician Certainty

Univariable and multivariable binary logistic regression associations between demographic and clinical variables with dichotomous ECD certainty [i.e., certain (81–100%) versus not certain (0–80%)], irrespective of an ASD or Non-ASD diagnosis, are presented in Table 3. Due to bimodal and/or non-normal distributions of ADOS-2 and Mullen sub-scores (see Fig. 1), as well as observed non-linearity in the relationships between ADOS-2 and Mullen sub-scores and the probability of diagnostic certainty, Social Affect (SA) and Restricted and Repetitive Behavior (RRB),

**Table 3** Univariable and multivariable binary logistic regression models evaluating the impact of demographic and clinical variables on levels of clinician certainty, accounting for random variation in diagnostic certainty between sites, with Mullen and ADOS-2 scales modeled as absolute differences from their respective means

Characteristic	Uncertain, N = 148		Certain, N = 348		p	Mullen		ADOS		Mullen + ADOS	
	Mean (SD) or N (row %)	Mean (SD) or N (row %)	Mean (SD) or N (row %)	Mean (SD) or N (row %)		Multivariable OR (95% CI)	p	Multivariable OR (95% CI)	p	Multivariable OR (95% CI)	p
Age (months) <sup>a</sup>	24.4 (4.5)	24.1 (4.4)	0.89 (0.71, 1.11)	0.305	0.77 (0.60, 0.99)	<b>0.044</b>	0.79 (0.60, 1.04)	0.096	0.78 (0.59, 1.04)	0.095	
Sex											
Female	48 (24.6%)	147 (75.4%)	1.46 (0.97, 2.22)	0.073	1.39 (0.90, 2.16)	0.142	1.51 (0.93, 2.44)	0.096	1.52 (0.93, 2.46)	0.092	
Male	100 (33.2%)	201 (66.8%)	Reference		Reference		Reference		Reference		
Race											
Black/Other	29 (23.8%)	93 (76.2%)	1.42 (0.87, 2.32)	0.165	1.45 (0.86, 2.42)	0.159	1.24 (0.70, 2.21)	0.458	1.24 (0.70, 2.21)	0.458	
White	119 (31.8%)	255 (68.2%)	Reference		Reference		Reference		Reference		
Maternal education											
Post-high school	130 (29.9%)	304 (70.1%)	1.29 (0.68, 2.45)	0.443	1.06 (0.52, 2.12)	0.881	0.91 (0.41, 2.05)	0.822	0.91 (0.40, 2.05)	0.818	
High school or less	17 (32.1%)	36 (67.9%)	Reference		Reference		Reference		Reference		
Non-ASD DDS											
Yes	91 (26.4%)	254 (73.6%)	1.69 (1.08, 2.63)	<b>0.021</b>	1.60 (1.01, 2.56)	<b>0.048</b>	1.28 (0.76, 2.15)	0.358	1.30 (0.77, 2.20)	0.330	
No	57 (37.8%)	94 (62.2%)	Reference		Reference		Reference		Reference		
Ethnicity											
Not Hispanic or Latino	117 (27.8%)	304 (72.2%)	1.36 (0.78, 2.37)	0.282	1.33 (0.73, 2.44)	0.352	1.19 (0.58, 2.42)	0.633	1.21 (0.59, 2.46)	0.600	
Hispanic or Latino	30 (41.7%)	42 (58.3%)	Reference		Reference		Reference		Reference		
Module											
Module 2	13 (22%)	46 (78%)	1.31 (0.67, 2.58)	0.427	0.64 (0.28, 1.50)	0.307	1.71 (0.77, 3.82)	0.190	1.70 (0.66, 4.35)	0.270	
Module T	135 (30.9%)	302 (69.1%)	Reference		Reference		Reference		Reference		
Abs. Diff. from Mean <sup>b,c</sup>											
Mullen											
Verbal AE	6.8 (4.4)	8.7 (5.5)	1.08 (1.03, 1.12)	<b>0.001</b>	1.11 (1.05, 1.18)	<b>0.001</b>	–	–	1.02 (0.95, 1.09)	0.570	
Non-verbal AE	5.6 (4.8)	6.9 (5.6)	1.05 (1.01, 1.10)	<b>0.014</b>	0.99 (0.94, 1.05)	0.776	–	–	0.98 (0.92, 1.04)	0.471	
ADOS-2											
SA	4.1 (3.0)	7.0 (2.7)	1.42 (1.31, 1.54)	< <b>0.001</b>	–	–	1.38 (1.27, 1.50)	< <b>0.001</b>	1.38 (1.27, 1.51)	< <b>0.001</b>	
RRB	1.7 (1.2)	2.3 (1.3)	1.58 (1.33, 1.89)	< <b>0.001</b>	–	–	1.43 (1.17, 1.74)	< <b>0.001</b>	1.43 (1.17, 1.74)	<b>0.001</b>	

Statistically significant *p* values are given in bold

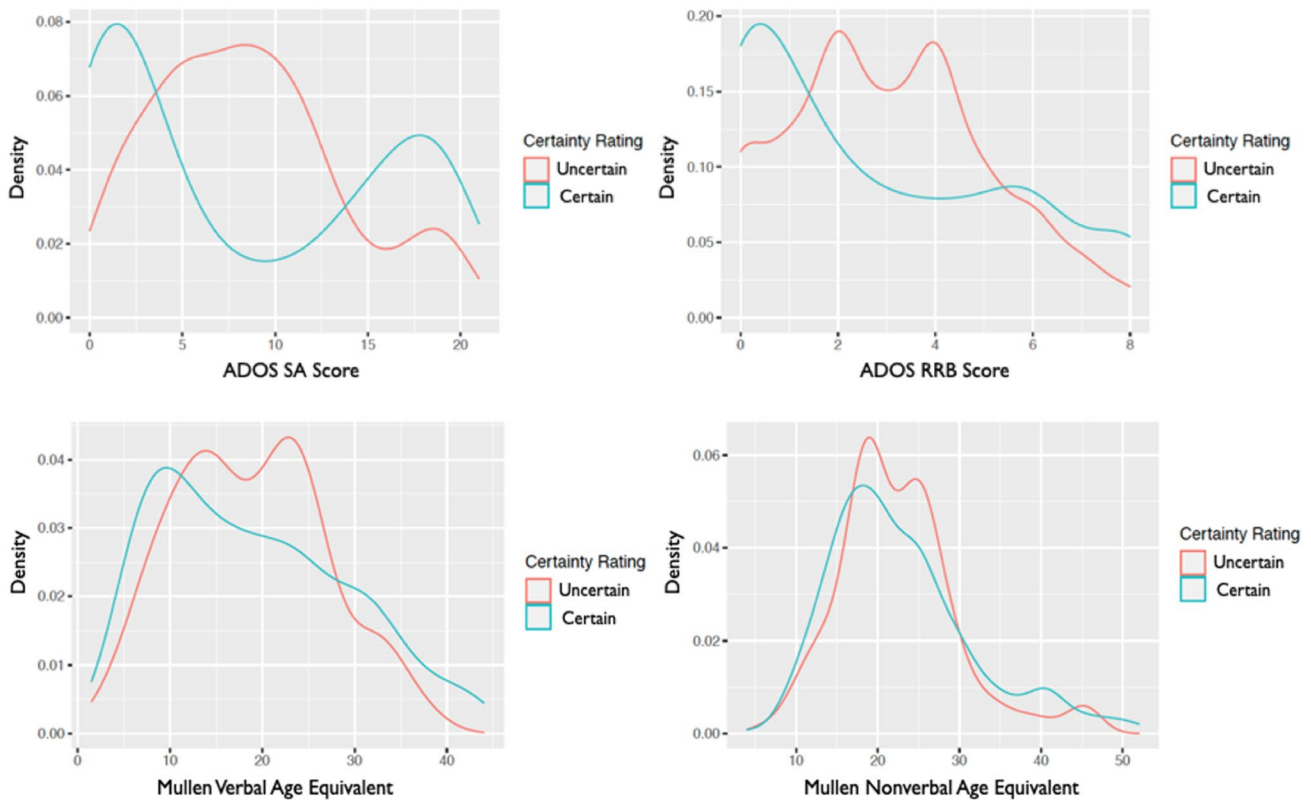
Models include random effects accounting for random variation in certainty between centers

<sup>a</sup>Age increments in units of 5 for interpretation purposes

<sup>b</sup>Mullen scales modeled as absolute differences from their respective means (Mean Verbal AE = 19, Mean Non-Verbal AE = 22.8)

<sup>c</sup>ADOS-2 scales modeled as absolute differences from their respective means (Mean SA = 8.4, Mean RRB = 2.9)





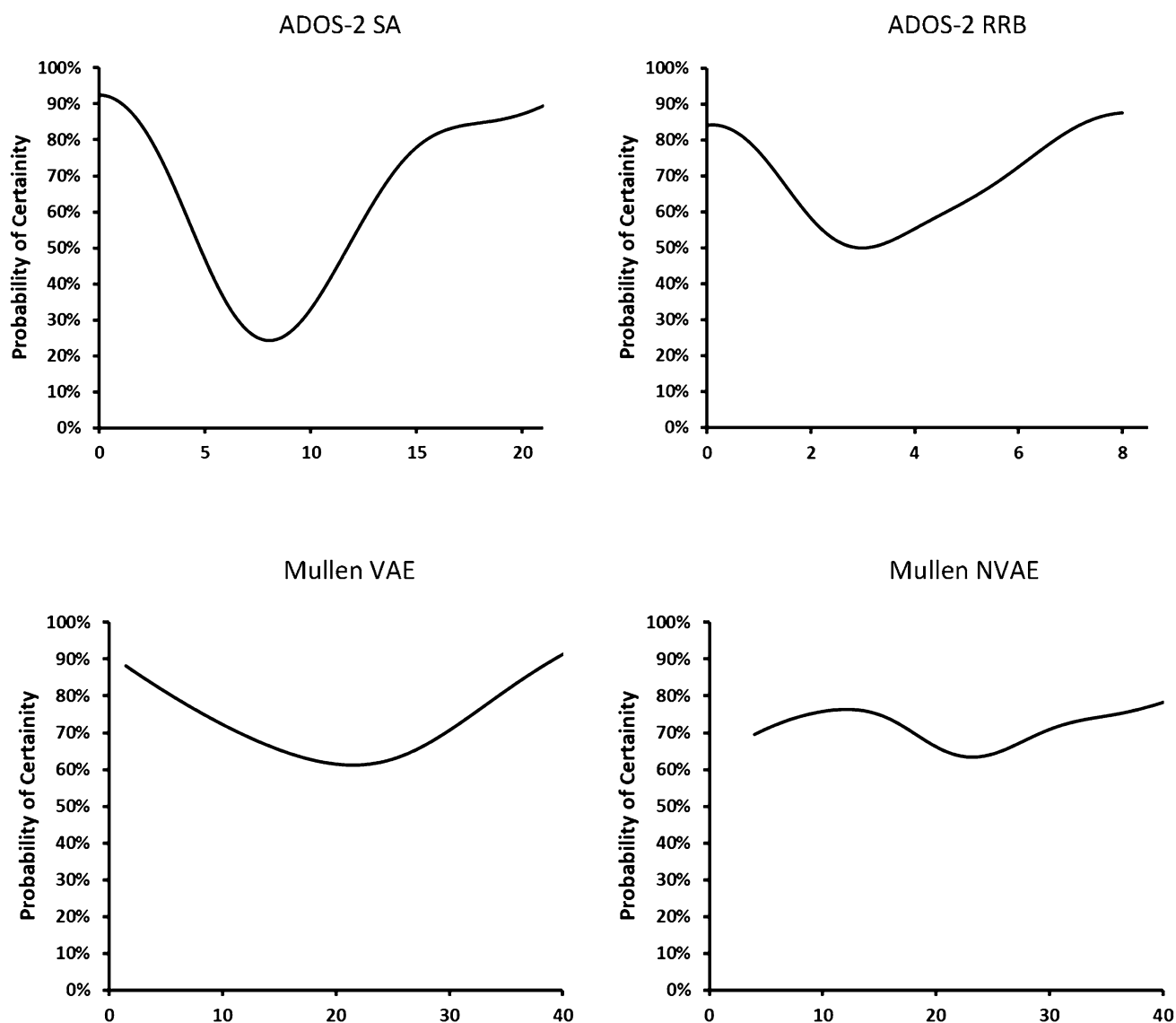
**Fig. 1** Density Plots for ADOS-2 subscores—Social Affect (SA) and Restricted & Repetitive Behavior (RRB) and Mullen Verbal and Nonverbal Age Equivalents—by Expert Clinician Diagnosis (ECD)

Certain vs Uncertain Cases<sup>a</sup> (Density of 1.0 = 100% of the sample). Curves represent the number of Certain and of Uncertain Cases (in percentage) with given ADOS-2 and Mullen scores

and Mullen Verbal Age Equivalents (VAE) and Nonverbal Age Equivalents (NVAE), were first modeled as absolute differences from their respective mean values. As univariable effects, participants classified as diagnostically Certain had higher odds for a presence of Non-ASD developmental delays (DDs) (OR 1.69, 95% CI 1.08, 2.63,  $p=0.021$ ), had absolute ADOS-2 sub-scores further from their respective means (SA OR 1.42, 95% CI 1.31, 1.54,  $p<0.001$ ; RRB OR 1.58, 95% CI 1.33, 1.89,  $p<0.001$ ), and Mullen sub-scores further from their respective means (VAE OR 1.08, 95% CI 1.03, 1.12,  $p=0.001$ ; NVAE OR 1.05, 95% CI 1.01, 1.10,  $p=0.014$ ). In other words, for every absolute one-unit increase away from the mean value for SA and RRB, the odds of ECD certainty increased by 42% and 58%, respectively. Implications for the ADOS-2 sub-score results point to higher diagnostic certainty when a child's scores are particularly high (more likely ASD) or particularly low (less likely ASD). While Mullen scores were significant in univariable analyses, their association with clinician certainty was much weaker: for every absolute one-month increase away from the mean value for VAE and NVAE, the odds of ECD certainty increased by 8% and 5% only, respectively. These trends were further clarified on multivariable analyses. In a model excluding the ADOS-2, only the Mullen VAE

remained statistically significant predictors of outcome, with absolute values further from the mean indicating higher odds of diagnostic certainty; in contrast, a model excluding the Mullen, both the ADOS-2 SA and RRB remained statistically significant predictors of diagnostic certainty. Finally, when both ADOS-2 and Mullen were entered into the model, only ADOS-2 SA and RRB remained statistically significant, with the stronger effect of the ADOS-2 seemingly nullifying Mullen sub-scores associations with ECD certainty (when treated as absolute differences from their respective means). In summary, the Mullen VAE played a minor role in predicting clinician certainty in comparison to much stronger role of both the ADOS-2 SA and RRB scores.

Secondary analysis employed natural cubic splines to model the non-linear relationships between ADOS-2 and Mullen sub-scores and the probability of diagnostic certainty (see Fig. 2). For increasing values of ADOS-2 SA, estimated probabilities of certainty follow a roughly parabolic pattern, with highest certainty of diagnosis (> 81%) occurring when SA scores are between 0–2 and 16–21. Estimated probabilities of diagnostic certainty fall (< 70%) when SA scores are between 4 and 13. For increasing values of RRB, estimated probabilities of diagnostic certainty are highest when RRB is either 0 or between 7 and 8; and similarly, estimated



**Fig. 2** Natural cubic splines model estimates for the probabilities of diagnostic certainty over increasing values of ADOS-2—Social Affect (SA) and Restricted & Repetitive Behavior (RRB)—and Mullen sub-scores—Verbal (VAE) and Nonverbal (NVAE) Age Equivalents

probabilities of diagnostic certainty fall (<70%) when RRB scores are between 2 and 5. For the Mullen, only the VAE sub-score had a statistically significant spline ( $p=0.003$ ), whereas the NVAE sub-score was not statistically significant ( $p=0.100$ ). For increasing values of Mullen VAE, estimated probability of certainty follow a roughly parabolic pattern, with highest certainty of diagnosis (>81%) occurring when VAE are <5 months or >35 months; estimated probabilities of diagnostic certainty fall (<70%) when VAE scores are between 12 and 30 months. This broad range corresponds to approximately 2.2 SDs of the Mullen VAE scores obtained in this sample, or about 72% of children whose chronological age range was 16–30 months. Therefore, consistent with previous analyses, the ADOS-2 sub-scores were more clinically meaningful as predictors of diagnostic certainty than

the Mullen VAE, whereas the Mullen NVAE was much less helpful.

Informed by these natural cubic splines models, as well as by clinically meaningful interpretations of the ADOS-2 sub-scores characteristic of this sample, clinically “Mild”, “Moderate”, and “Severe” ranges of SA and RRB sub-scores were specified and resulting univariable and multivariable linear piecewise regression associations are presented in Supplementary Table S2. Univariable associations for SA found that for every one-unit increase in the Clinically Mild range (scores 0–7), the odds of certainty decrease by 14% (OR 0.86, 95% CI 0.80, 0.93,  $p<0.001$ ); Clinically Moderate scores (8–13) did not associate with significant change in the odds of diagnostic certainty, but the Clinically Severe score range (14–21) showed the odds of diagnostic certainty

increase by 79% (OR 1.79, 95% CI 1.39, 2.30,  $p < 0.001$ ) for every one-unit increase in SA. Similar significance was found for RRB, with odds ratios in the Clinically Mild (0–1) and Severe ranges (5–8) demonstrating increasing odds of diagnostic certainty, respectively. These linear piecewise regression results mirror those from the absolute mean difference analysis and the natural cubic splines models, showing highest diagnostic certainties in the tails of the ADOS-2 sub-score distributions and lowest diagnostic certainties in the mid-range values. On multivariable analysis, SA retained significance for Clinically Mild and Severe scores; whereas the effect of RRB was attenuated, only showing marginal statistical significance for scores in the Clinically Severe range (OR 1.79, 95% CI 0.92, 3.51,  $p = 0.087$ ).

### Clinicians' Stated Reasons for Their Uncertainty When Making the ECD of ASD and of Non-ASD

Of children receiving a diagnosis of ASD, one or more rationales for diagnostic uncertainty were provided for 78.7% of cases. Of these, the following were reasons cited in more than 10% of cases, ordered by frequency: "Mild symptom presentation and inconsistent use of skills (40.0%); "Parent had no ASD concerns", "Medical complexity (e.g., prematurity, in utero drug exposure, feeding/sleeping challenges)", and "Cognitive/global developmental delay" (each cited in 13.3% of cases). Additionally, only "Presence of language delay" and "Too young to be certain of diagnosis" were closer to the threshold of 10% (each, 9.3% of cases). Of children receiving a diagnosis of Non-ASD, one or more rationales for diagnostic uncertainty were provided for 82.2% of cases. Of these, the following were reasons cited in more than 10% of cases, ordered by frequency: "Restrictive and Repetitive Behaviors (RRBs) present but alongside strengths in social and communicative behavior" (37.3%); "Mild autism-related symptoms that do not quite meet threshold for ASD diagnosis" (32.9%); "Language delay" (28.8%); and "Family history of ASD, or presentation consistent with broader autism phenotype (BAP)" (23.3%). Additionally, only "Medical complexity (e.g., prematurity, in utero drug exposure, feeding/sleeping challenges)" was close to the threshold of 10% (9.6% of cases).

These results are consistent with previous analyses. Across ECDs of ASD and Non-ASD, the most commonly cited reason for diagnostic uncertainty is "mild symptomatology", i.e., the presence of ASD symptoms that cluster around the perceived threshold for a diagnosis of ASD vs Non-ASD (when Non-ASD cases have other diagnoses). Specific to the diagnosis of Non-ASD, the most commonly cited reason for diagnostic uncertainty is the presence of RRB symptomatology alongside strengths in social-communicative behavior. Other common complicating factors in the diagnosis of Non-ASD were the presence of language delay

and BAP presentation. Interestingly, the young age of the child (16–30 months) was uncommonly cited by clinicians for their diagnostic uncertainty in ECDs of ASD, and not at all in ECDs of Non-ASD. A comprehensive description of reasons for clinician uncertainty for ECDs of ASD and Non-ASD, separately, and the percentage of clinicians endorsing them are provided in Table 4.

## Discussion

In secondary analyses of a large, multi-site clinical trial involving best-practices diagnostic and developmental assessments of 16–30-month-old children referred to specialized ASD centers, 29.8% of cases received an ASD or Non-ASD diagnosis with variable levels of clinician uncertainty. This result replicates the finding of the only other ASD study of clinician certainty (McDonnell et al., 2019), also a multi-site study conducted in specialized ASD centers. Because both studies included only experienced clinicians, who were monitored for standardized and reliable execution of reference standard diagnostic and developmental instruments, it is very likely that clinician certainty is lower still in community-based, less-specialized centers with less expertise in ASD diagnosis. Given the importance of early ASD diagnosis to access appropriate early treatment services, and given the absence of a psychometrically and clinically validated tool to document and measure clinician decision-making in this context, there is an urgent need for much more detailed, standardized and quantitative protocols than those used in these two studies to elicit more comprehensive data on clinician certainty. Such a tool should likely deploy a rating system reflecting explicit clinical criteria and scenarios that mirror the everyday experiences of clinicians conducting diagnostic evaluations of young children suspected to have ASD. In turn, these criteria will likely evolve from qualitative research eliciting challenges in the diagnostic process and in diagnostic decision-making, for which the findings of the current study provide a starting point. The need for such a tool was made all the more obvious by this study's finding that there was marked variability in levels of clinician certainty across the six participating sites, despite steps taken to reduce uncertainty, which included a planning phase to achieve uniformity of clinical procedures and reliability on administration and scoring of standardized tools across all sites, as well as periodic monitoring to prevent procedural or reliability drift.

Unsurprisingly, the most potent predictor of clinician uncertainty was mid-level autism-related symptomatology as measured by the ADOS-2, falling around the thresholds for the differential diagnosis of ASD vs Non-ASD. This effect was measured precisely: for every absolute one-unit increase away from the sample mean value of the ADOS-2 Social

**Table 4** Clinicians' stated reasons for their uncertainty when making the expert clinician diagnosis (ECD) of ASD and Non-ASD

Rank order	Clinicians' stated reasons for diagnostic uncertainty	Percentage of cases (%)
ECD of ASD		
1	Mild symptom presentation/inconsistent use of skills	40.0
2	Parent had no ASD concerns	13.3
2	Medical complexity (e.g., prematurity, drug exposure in utero, feeding/sleeping challenges)	13.3
2	Cognitive/global developmental delay	13.3
5	Language delay	9.3
5	Too young to be certain of diagnosis	9.3
7	Family history of ASD or other mental health concerns	6.7
8	Lack of language delay	4.0
9	Bilingual	2.7
10	Slow to warm up/anxious	1.3
	Clinician did not indicate reason for uncertainty	21.3
ECD of Non-ASD		
1	Restricted and Repetitive Behaviors (RRBs) present but child shows social and communicative strengths	37.3
2	Mild autism-related symptoms that do not quite reach threshold for ASD diagnosis	32.9
3	Language delay	28.8
4	Family history of ASD, or presentation consistent with broader autism phenotype (BAP)	23.3
5	Medical complexity (e.g., prematurity, drug exposure in utero, feeding/sleeping challenges)	9.6
6	Cognitive/global developmental delay	8.2
7	Slow to warm up/anxious	6.8
8	Emotional dysregulation (e.g., tantrums, unwillingness to participate, high activity level)	5.5
9	Parent had ASD concerns	2.7
10	Too young to be certain of diagnosis	0.0
	Clinician did not indicate reason for uncertainty	17.8

Affect (SA) and Restricted and Repetitive Behavior (RRB) scores, the odds of a certain diagnosis of ASD or Non-ASD increased by 42% and 58%, respectively. Mid-level verbal age equivalents measured by the Mullen were also predictors of clinician uncertainty, albeit much less strongly than the ADOS-2 scores. These quantitative results were consistent with clinicians' stated reasons for uncertainty: for the diagnosis of ASD, "mild" ASD symptomatology was the most common reason cited (40% of uncertain ASD cases), followed by much less common reasons (~13%) such as global developmental delay, medical complexities or lack of parental concern; for the diagnosis of Non-ASD developmental conditions, "Mild" ASD symptomatology (~33% of uncertain Non-ASD cases), and specifically, the presence of restricted and repetitive behaviors (RRBs) in the presence of social-communicative strengths (~37%) were the most common reasons for clinician uncertainty, followed by language delays (~29%) and ASD-related family history (or presentation consistent with broader autism phenotype) (~23%). Collectively, this emerging clinical profile of children at risk of encountering clinician uncertainty should be highlighted in clinical practice for special attention as these

children are at higher risk of misdiagnosis or the "wait and see" approach, possibly delaying both eligibility and access to early intervention services (Smith-Young et al., 2020). In research, diagnostic uncertainty is likely to be associated with higher heterogeneity and diagnostic "noise" (Thomas et al., 2021). Of importance, none of the sociodemographic variables, including age, sex, race, ethnicity or maternal education were found to be significant predictors of clinician uncertainty, in either the quantitative analyses or in clinicians' stated reasons for their diagnostic uncertainty. Importantly, however, there was a (a non-significant) trend for the sex variable to be associated with clinician uncertainty, although it went in the opposite direction than suggested by the literature (McDonnell et al., 2021), with a higher percentage of boys rather than girls falling among the cases deemed by clinicians as uncertain. Given the importance of sociodemographic variables for actionable steps that could be taken to improve equitable access to early diagnosis, and given that we were not able to replicate some of the sociodemographic findings in the McDonnell and colleagues' study (McDonnell et al., 2019), additional research is warranted.

Specific to the association between Mid-level ADOS-2 scores and clinician uncertainty, our results provide an opportunity to constrain and calibrate diagnostic decision-making based on greater understanding of the eventual diagnostic outcomes of children with this profile. Diagnostic thresholds of the ADOS-2 modules for young children were developed by maximizing sensitivity and specificity relative to *concurrent* expert clinician diagnosis (ECD), a “gold standard” which, however, includes diagnostic uncertainty for a sizable portion of these children. These thresholds might require adjustments that could come from follow-up data on uncertain cases to probe rate of conversion of false positives to true positives and false negatives to true negatives, as well as cost–benefit, to the family and to society, of a child being a false positive or false negative (for example, a false negative diagnosis is likely to correspond to the costliest scenario to both family and society, since the child would be denied needed services with potential lifetime consequences) (see below). The need for longitudinal data goes beyond data on diagnostic stability; it should also include documentation of families’ “diagnostic odyssey” or protracted diagnostic pathways, including attainment of eligibility for and access to adequate services, as this is ultimately the most important goal of the diagnostic process. Such a research line might provide the data necessary for re-calibration of thresholds for early ASD diagnosis and for the establishment of a standard, rule-governed framework of directives for clinician decision-making using standardized diagnostic instruments such as the ADOS-2.

To contextualize this discussion, it is important to highlight the fact that, from a population perspective, there is a large and unacceptable number of ASD false negatives, as indicated by lower prevalence rates of ASD in younger rather than older children, and a late median age of diagnosis in the US, still hovering over 4 years of age (Maenner et al., 2021; Shaw et al., 2021). Adding to the population high rates of false negatives is the low sensitivity exhibited by widespread population ASD screeners (Guthrie et al., 2019), which typically act as gatekeepers for further assessment. All of these challenges are further exacerbated for minority and traditionally under-served populations (Constantino et al., 2020). Therefore, given the importance of early treatment to optimize outcomes, the cost to family and society of false negatives is high. False positives also carry costs in the form of alarm and investments associated with an ASD misdiagnosis. Nevertheless, false positive cases within the age range included in this study would likely be children with one or more actionable developmental delays, who would also benefit from the kinds of early treatment intended for young children with ASD, which, typically, focus on social-communication skill acquisition via cost-efficient caregiver-mediated interventions (Klin et al., 2020). Therefore, the cost of false positives may be mitigated by the fact that such

children are likely to benefit from identification, particularly in the current system of early intervention services which typically requires a diagnosis to access adequate levels of treatment. Collectively, these considerations suggest the need to consider and research the cost–benefit implications, to child and family, of diagnostic decision-making in situations of clinician uncertainty. More empirical knowledge to inform this process might be an important element in any attempt to optimize the diagnostic process for young children with ASD.

## Limitations

This study has several limitations. First, there was less diversity in race, ethnicity, and SES in our sample than in McConnell and colleagues’ study, a factor that may have impeded our ability to replicate their sociodemographic findings pointing to factors such as race/ethnicity and SES as predictors of clinician uncertainty. Second, as noted, there is a need for a much more detailed, standardized, quantitative and qualitative protocol to elicit data on clinician certainty, with a rating system that reflects explicit clinical criteria and scenarios that mirror the everyday experiences of clinicians conducting ASD-related diagnostic evaluations of young children. Third, there is a need for systematic data collection on clinician variables such as level of training, professional discipline, experience (e.g., number of clinical assessments completed) and attitudes relevant to decision-making (e.g., hesitation to make an ASD diagnosis in less obvious cases or in very young children). And finally, as noted, this research line needs to be informed by longitudinal data on the evolving diagnostic status of children at later ages as well as by data on their families’ “diagnostic odyssey”, documenting and measuring eligibility for and attainment of early intervention services.

## Conclusions

Differential diagnosis of ASD in young children is challenging, and uncertainty regarding a child’s diagnosis may result in under-identification or prolonged diagnostic pathways. In the current multi-site study, and in the only other study probing clinician certainty in the diagnosis of ASD (McDonnell et al., 2019), both involving large cohorts of young children adequately representing referrals to specialized ASD centers, suboptimal diagnostic certainty was documented for 30–40% of cases. These results warrant future research addressing areas of limitation in both studies, including the deployment of especially designed and validated tools and procedures for the study of clinician certainty. Unsurprisingly, clinician uncertainty was more likely to involve cases with mid-range levels of ASD symptoms (for both ASD and Non-ASD diagnoses), and with language delays and ASD-related family



history (in the case of Non-ASD developmental diagnoses). Also, given the implications of these results for clinical practice, public health and research, future studies should include less-specialized community settings, and should map, codify and quantify clinician decision-making with a view to ultimately optimize the diagnostic process of young children and meet the needs of their families.

In essence, the diagnostic process in very young children has primarily two goals: to operationalize a child's needs at present for access to immediate early treatment; and to predict a child's future diagnostic and developmental outcomes. Both goals depend on the differential diagnosis of ASD. For a sizable proportion of cases seen in clinics, this process can be fraught with uncertainty, which seems to be primarily due to “noise” (intuitive judgments that may be applied inconsistently), a factor that is difficult to avoid in subjective decision-making in all areas of clinical medicine, but particularly in developmental and behavioral health. Several decades of social and clinical science research of clinical decision-making have shown that in such situations, rules, including standardized, algorithmic decision-making, consistently outperform subjective decision-making (Kahneman et al., 2021). With the advent of longitudinal follow-up data, of both stability of diagnosis and of families' “diagnostic odyssey”, diagnostic thresholds in standardized diagnostic instruments can be re-calibrated and algorithmic directives mapping on specific clinical scenarios can be created to optimize the diagnostic process for young children and their families.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10803-022-05812-8>.

**Acknowledgments** The authors would like to thank the children and families for their participation in this study. The study was funded by a Grant from the Marcus Foundation, Atlanta, Georgia, USA.

**Author Contributions** Cheryl Klaiman contributed to clinical components of trial design, directed clinical execution of trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Stormi White contributed participants to trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Shana Richardson contributed participants to trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Emma McQueen contributed participants to trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Hasse Walum contributed to statistical analysis plan of trial. Christa Aoki contributed participants to trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Christopher Smith contributed participants to trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Mendy Minjarez contributed participants to trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Raphael Bernier contributed participants to trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Ernest Pedapati contributed participants to trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Somer Bishop contributed participants to trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Whitney Ence contributed participants to trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Allison Wainer

contributed participants to trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Jennifer Moriuchi contributed participants to trial, and reviewed manuscript for accuracy and clinical and scientific integrity. Sew-Wah Tay contributed to data processing, management and analysis of the trial, and reviewed manuscript for accuracy and scientific integrity. Yiming Deng contributed to data processing, management and analysis of the trial, and reviewed manuscript for accuracy and scientific integrity. Warren Jones designed the trial, provided overall scientific leadership, and reviewed manuscript for accuracy and scientific integrity. Scott Gillespie designed statistical analysis plan, conducted data analysis, drew statistical inferences, and reviewed manuscript for accuracy and scientific integrity. Ami Klin contributed to trial design, statistical analysis plan, and interpretation of results, wrote and reviewed the manuscript for accuracy and clinical and scientific integrity.

## Declarations

**Conflict of interest** WJ and AK are scientific consultants to and minority shareholders in EarliTec Diagnostics, Inc. EarliTec develops technology for early identification of autism and gives revenue to support treatment of children with autism. The activity has been reviewed and approved by Emory University's Conflict of Interest Review Office.

## References

- American Psychiatric Association (APA). (2013). *The diagnostic and statistical manual of mental disorders, fifth edition (DSM-5)*. American Psychiatric Association.
- Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Robinson Rosenberg, C., White, T., Durkin, M. S., Imm, P., Nikolaou, L., Yeargin-Allsopp, M., Lee, L. C., Harrington, R., Lopez, M., Fitzgerald, R. T., Hewitt, A., ... Dowling, N. F. (2018). Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *MMWR Surveillance Summaries*, 67(6), 1–23. <https://doi.org/10.15585/mmwr.ss6706a1>
- Chawarska, K., Paul, R., Klin, A., Hannigen, S., Dichtel, L. E., & Volkmar, F. (2007). Parental recognition of developmental problems in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 37(1), 62–72. <https://doi.org/10.1007/s10803-006-0330-8>
- Christensen, D. L., Baio, J., Van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J. N., Daniels, J., Durkin, M. S., Fitzgerald, R. T., Kurzius-Spencer, M., Lee, L. C., Pettygrove, S., Robinson, C., Schulz, E., Wells, C., Wingate, M. S., Zahorodny, W., Yeargin-Allsopp, M., Centers for Disease, C., & Prevention. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *MMWR Surveillance Summaries*, 65(3), 1–23. <https://doi.org/10.15585/mmwr.ss6503a1>
- Constantino, J. N., Abbacchi, A. M., Saulnier, C., Klaiman, C., Mandell, D. S., Zhang, Y., Hawks, Z., Bates, J., Klin, A., Shattuck, P., Molholm, S., Fitzgerald, R., Roux, A., Lowe, J. K., & Geschwind, D. H. (2020). Timing of the diagnosis of autism in African American children. *Pediatrics*. <https://doi.org/10.1542/peds.2019-3629>
- Crane, L., Chester, J. W., Goddard, L., Henry, L. A., & Hill, E. (2016). Experiences of autism diagnosis: A survey of over 1000 parents in the United Kingdom. *Autism*, 20(2), 153–162. <https://doi.org/10.1177/1362361315573636>

- Daniels, A. M., Halladay, A. K., Shih, A., Elder, L. M., & Dawson, G. (2014). Approaches to enhancing the early detection of autism spectrum disorders: A systematic review of the literature. *Journal of the American Academy of Child and Adolescent Psychiatry*, 53(2), 141–152. <https://doi.org/10.1016/j.jaac.2013.11.002>
- Daniels, A. M., & Mandell, D. S. (2014). Explaining differences in age at autism spectrum disorder diagnosis: A critical review. *Autism*, 18(5), 583–597. <https://doi.org/10.1177/1362361313480277>
- Feaster, D. J., Mikulich-Gilbertson, S., & Brincks, A. M. (2011). Modeling site effects in the design and analysis of multi-site trials. *American Journal of Drug and Alcohol Abuse*, 37(5), 383–391. <https://doi.org/10.3109/00952990.2011.600386>
- Guthrie, W., Wallis, K., Bennett, A., Brooks, E., Dudley, J., Gerdes, M., Pandey, J., Levy, S. E., Schultz, R. T., & Miller, J. S. (2019). Accuracy of autism screening in a large pediatric network. *Pediatrics*. <https://doi.org/10.1542/peds.2018-3963>
- Harrell, F. E. (2015). *Regression modeling strategies—With applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.
- Honigfeld, L., Chandhok, L., & Spiegelman, K. (2012). Engaging pediatricians in developmental screening: The effectiveness of academic detailing. *Journal of Autism and Developmental Disorders*, 42(6), 1175–1182. <https://doi.org/10.1007/s10803-011-1344-4>
- Hyman, S. L., Levy, S. E., Myers, S. M., Council on Children with Disabilities, Section on Developmental and Behavioral Pediatrics. (2020). Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics*. <https://doi.org/10.1542/peds.2019-3447>
- Johnson, C. P., Myers, S. M., American Academy of Pediatrics Council on Children With Disabilities. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics*, 120(5), 1183–1215. <https://doi.org/10.1542/peds.2007-2361>
- Johnson, M. H. (1999). Cortical plasticity in normal and abnormal cognitive development: Evidence and working hypotheses. *Development and Psychopathology*, 11(3), 419–437. <https://doi.org/10.1017/s0954579499002138>
- Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown Spark.
- Kim, S. H., & Lord, C. (2012). Combining information from multiple sources for the diagnosis of autism spectrum disorders for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Child Psychology and Psychiatry*, 53(2), 143–151. <https://doi.org/10.1111/j.1469-7610.2011.02458.x>
- Klin, A., Micheletti, M., Klaiman, C., Shultz, S., Constantino, J. N., & Jones, W. (2020). Affording autism an early brain development redefinition. *Development and Psychopathology*, 32(4), 1175–1189. <https://doi.org/10.1017/S0954579420000802>
- Leckman, J. F., Sholomskas, D., Thompson, W. D., Belanger, A., & Weissman, M. M. (1982). Best estimate of lifetime psychiatric diagnosis: A methodological study. *Archives of General Psychiatry*, 39(8), 879–883. <https://doi.org/10.1001/archpsyc.1982.04290080001001>
- Lijmer, J. G., Mol, B. W., Heisterkamp, S., Bossel, G. J., Prins, M. H., van der Meulen, J. H., & Bossuyt, P. M. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*, 282(11), 1061–1066. <https://doi.org/10.1001/jama.282.11.1061>
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule, second edition (ADOS-2)*. Western Psychological Services.
- Maenner, M. J., Graves, S. J., Peacock, G., Honein, M. A., Boyle, C. A., & Dietz, P. M. (2021). Comparison of 2 case definitions for ascertaining the prevalence of autism spectrum disorder among 8-year-old children. *American Journal of Epidemiology*, 190(10), 2198–2207. <https://doi.org/10.1093/aje/kwab106>
- Maenner, M. J., Shaw, K. A., Baio, J., EdS, Washington, A., Patrick, M., DiRienzo, M., Christensen, D. L., Wiggins, L. D., Pettygrove, S., Andrews, J. G., Lopez, M., Hudson, A., Baroud, T., Schwenk, Y., White, T., Rosenberg, C. R., Lee, L. C., Harrington, R. A., ... Dietz, P. M. (2020). Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveillance Summaries*, 69(4), 1–12. <https://doi.org/10.15585/mmwr.ss6904a1>
- McDonnell, C. G., Bradley, C. C., Kanne, S. M., Lajonchere, C., Warren, Z., & Carpenter, L. A. (2019). When are we sure? Predictors of clinician certainty in the diagnosis of autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 49(4), 1391–1401. <https://doi.org/10.1007/s10803-018-3831-3>
- McDonnell, C. G., DeLucia, E. A., Hayden, E. P., Penner, M., Curcin, K., Anagnostou, E., Nicolson, R., Kelley, E., Georgiades, S., Liu, X., & Stevenson, R. A. (2021). Sex differences in age of diagnosis and first concern among children with autism spectrum disorder. *Journal of Clinical Child and Adolescent Psychology*, 50(5), 645–655. <https://doi.org/10.1080/15374416.2020.1823850>
- Mullen, E. M. (1995). *Mullen scales of early learning*. American Guidance Services.
- Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., Chawarska, K., Macari, S. L., Messinger, D., Stone, W. L., Zwaigenbaum, L., & 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*, 56(9), 988–998. <https://doi.org/10.1111/jcpp.12421>
- Randall, M., Egberts, K. J., Samtani, A., Scholten, R. J., Hooft, L., Livingstone, N., Sterling-Levis, K., Woolfenden, S., & Williams, K. (2018). Diagnostic tests for autism spectrum disorder (ASD) in preschool children. *Cochrane Database of Systematic Reviews*. <https://doi.org/10.1002/14651858.cd009044.pub2>
- Rutherford, M., Burns, M., Gray, D., Bremner, L., Clegg, S., Russell, L., Smith, C., & O'Hare, A. (2018). Improving efficiency and quality of the children's ASD diagnostic pathway: Lessons learned from practice. *Journal of Autism and Developmental Disorders*, 48(5), 1579–1595. <https://doi.org/10.1007/s10803-017-3415-7>
- Rutherford, M., McKenzie, K., Forsyth, K., McCartney, D., O'Hare, A., McClure, I., & Irvine, L. (2016). Why are they waiting? Exploring professional perspectives and developing solutions to delayed diagnosis of autism spectrum disorders in adults and children. *Research in Autism Spectrum Disorders*, 31, 53–65.
- Shaw, K. A., Maenner, M. J., Bakian, A. V., Bilder, D. A., Durkin, M. S., Fumier, S. M., Hughes, M. M., Patrick, M., Pierce, K., Salinas, A., Shenouda, J., Vehorn, A., Warren, Z., Zahorodny, W., Constantino, J. N., DiRienzo, M., Esler, A., Fitzgerald, R. T., Grzybowski, A., ... Cogswell, M. E. (2021). Early identification of autism spectrum disorder among children aged 4 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2018. *MMWR Surveillance Summaries*, 70(10), 1–14. <https://doi.org/10.15585/mmwr.ss7010a1>
- Smith-Young, J., Chafe, R., & Audas, R. (2020). “Managing the Wait”: Parents' experiences in accessing diagnostic and treatment services for children and adolescents diagnosed with autism spectrum disorder. *Health Services Insights*, 13, 1178632920902141. <https://doi.org/10.1177/1178632920902141>
- Thomas, H. R., Rooney, T., Cohen, M., Bishop, S. L., Lord, C., & Kim, S. H. (2021). Spontaneous expressive language profiles in a clinically ascertained sample of children with autism spectrum disorder. *Autism Research*, 14(4), 720–732. <https://doi.org/10.1002/aur.2408>
- US Department of Education. (2020). *2020 individuals with disabilities education act annual report to congress*. <https://sites.ed.gov/idea/2020-annual-report-congress-idea/>
- US Government Accountability Office (GAO). (2013). *Combating autism act: HHS agencies responded with new and continuing*

activities, including oversight. <https://www.gao.gov/products/gao-13-232>

- Zwaigenbaum, L., Bauman, M. L., Choueiri, R., Kasari, C., Carter, A., Granpeesheh, D., Mailloux, Z., Smith Roley, S., Wagner, S., Fein, D., Pierce, K., Buie, T., Davis, P. A., Newschaffer, C., Robins, D., Wetherby, A., Stone, W. L., Yirmiya, N., Estes, A., ... Natowicz, M. R. (2015a). Early intervention for children with autism spectrum disorder under 3 years of age: Recommendations for practice and research. *Pediatrics*, *136*(Suppl 1), S60-81. <https://doi.org/10.1542/peds.2014-3667E>
- Zwaigenbaum, L., Bauman, M. L., Stone, W. L., Yirmiya, N., Estes, A., Hansen, R. L., McPartland, J. C., Natowicz, M. R., Choueiri, R., Fein, D., Kasari, C., Pierce, K., Buie, T., Carter, A., Davis, P. A., Granpeesheh, D., Mailloux, Z., Newschaffer, C., Robins, D.,

... Wetherby, A. (2015b). Early identification of autism spectrum disorder: Recommendations for practice and research. *Pediatrics*, *136*(Suppl 1), S10-40. <https://doi.org/10.1542/peds.2014-3667C>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.