



# The stability of the Autism Diagnostic Observation Schedule-2 in children aged 14–36 months with elevated likelihood for autism

Sarah Schaubroeck,<sup>1</sup>  Ellen Demurie,<sup>1</sup> Jannath Begum-Ali,<sup>2</sup> Sven Bölte,<sup>3,4,5</sup> Sofie Boterberg,<sup>1</sup> Jan Buitelaar,<sup>6,7</sup> Tony Charman,<sup>8</sup>  Terje Falck-Ytter,<sup>3,9</sup> Sabine Hunnius,<sup>10</sup> Mark H. Johnson,<sup>2,11</sup> Emily Jones,<sup>2,12,13</sup> Iris Oosterling,<sup>7</sup> Greg Pasco,<sup>2</sup> Mirjam Pijl,<sup>7</sup> Carlijn Van den Boomen,<sup>14</sup> Petra Warreyn,<sup>1</sup> and Herbert Roeyers<sup>1</sup>

<sup>1</sup>Research in Developmental Diversity Lab (RIDDL), UGent, Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium; <sup>2</sup>Centre for Brain and Cognitive Development, School of Psychological Sciences, Faculty of Science, Birkbeck, University of London, London, UK; <sup>3</sup>Center of Neurodevelopmental Disorders (KIND), Centre for Psychiatry Research, Department of Women's and Children's Health, Karolinska Institutet & Stockholm Health Care Services, Region Stockholm, Stockholm, Sweden; <sup>4</sup>Curtin Autism Research Group, School of Allied Health, Curtin University, Perth, WA, Australia; <sup>5</sup>Child and Adolescent Psychiatry, Stockholm Health Care Services, Stockholm, Sweden; <sup>6</sup>Department of Medical Neuroscience, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Centre, Nijmegen, The Netherlands; <sup>7</sup>Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, The Netherlands; <sup>8</sup>Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; <sup>9</sup>Development and Neurodiversity Lab, Department of Psychology, Uppsala University, Uppsala, Sweden; <sup>10</sup>Donders Centre for Cognition, Donders Institute for Brain, Cognition & Behaviour, Radboud University, Nijmegen, The Netherlands; <sup>11</sup>Department of Psychology, University of Cambridge, Cambridge, UK; <sup>12</sup>Centre for Developmental Neurobiology, Institute of Psychiatry, Psychology & Neuroscience, Kings College London, London, UK; <sup>13</sup>Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, Kings College London, London, UK; <sup>14</sup>Experimental Psychology, Helmholtz Institute Utrecht University, Utrecht, The Netherlands

**Background:** This study investigated the stability of Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) classifications in a cohort of 304 siblings at elevated likelihood for autism (EL-siblings). **Methods:** ADOS-2 assessments were conducted at 14, 24 and 36 months, with Clinical Best Estimate (CBE) autism diagnoses determined at 36 months. **Results:** Our findings indicate that while some children have stable ADOS-2 classifications from early on, a significant proportion of the children show inconsistent classifications over time. The overall stability of ADOS-2 autism spectrum classifications increased from 14 to 36 months and agreement with CBE autism clinical diagnosis was moderate and increased with age. **Conclusions:** Caution is warranted when interpreting individual ADOS-2 results, as they should always complement, and can never replace, a comprehensive clinical evaluation. These findings highlight the importance of continued follow-up beyond 14 months in young EL-children, a group for whom early assessment may be both feasible and beneficial and emphasises the need to integrate multiple assessment measures and multiple informants for accurate early autism identification. **Keywords:** Autism; stability; ADOS; diagnosis; EL-siblings.

## Introduction

Autism is a neurodevelopmental condition with a prevalence of about 1%–2% (Maenner et al., 2023; Zeidan et al., 2022). It is characterised by challenges in social communication and interaction, alongside restricted and repetitive behaviours, interests or activities and sensory challenges (American Psychiatric Association [APA], 2022). Children with an older autistic sibling have a significantly elevated likelihood (EL) to be autistic themselves. Sibling studies show that 10%–20% of children with an autistic sibling are diagnosed with autism by age 3 (Ozonoff et al., 2024). This elevated likelihood is further supported by large population studies reporting rates around 10% (Hansen et al., 2019;

Jokiranta-Olkonieni et al., 2016; Sandin, Lichtenstein, Kuja-Halkola, Hultman, & Reichenberg, 2017). Consequently, prospective research in infant EL siblings provides the opportunity to study the early development of autistic children (van 't Hof et al., 2021).

The identification of autism at an early age is challenging due to phenotypic heterogeneity, including differences in the timing of feature onset and behavioural manifestations among autistic children (Kim, Macari, Koller, & Chawarska, 2016). Moreover, co-occurring difficulties or conditions, such as intellectual disability, atypical language development and attention-deficit/hyperactivity disorder (ADHD) complicate the diagnostic process in EL siblings (Charman et al., 2017, 2023; Marrus et al., 2018; Messinger et al., 2013). However, early identification of autism characteristics and

Conflict of interest statement: See Acknowledgements for full disclosures.

subsequent diagnostic classification is crucial for recognising potential challenges and facilitating early support or targeted intervention (Fuller & Kaiser, 2020; Ryberg, 2015). In turn, this might prevent the development of secondary difficulties and improve long term outcomes and quality of life (Turner, Stone, Pozdol, & Coonrod, 2006). Given these potential advantages, timely recognition of autism characteristics in EL-siblings is essential. In clinical practice, administering the Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al., 2012), often combined with the Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003), is standard procedure in assessing autism characteristics. While an autism diagnosis must always be grounded in comprehensive clinical judgement rather than relying solely on the ADOS-2, the instrument provides standardised information to support expert clinician evaluations (Bishop & Lord, 2023; Fombonne, 2023). Considering this, it is important to examine whether ADOS-2 classifications remain stable across development, not because repeated administration should function as a standalone diagnostic pathway, but because such stability can provide insight into how the tool contributes to broader developmental monitoring.

Research by Ozonoff et al. (2015) and Zwaigenbaum et al. (2016) has examined diagnostic stability in EL-siblings and indicated that a reliable and stable clinical diagnosis of autism can be established between 18 and 24 months, although this can vary depending on individual factors, such as developmental level, presence of autistic characteristics or sex (Lai, Lombardo, Auyeung, Chakrabarti, & Baron-Cohen, 2015; Zwaigenbaum et al., 2015). Ozonoff et al. (2015) found that the stability of an autism diagnosis was 93% at 18 months and 82% at 24 months. However, many children with autism outcomes at 36 months were not diagnosed until later, with 63% not identified at 18 months. Similarly, Zwaigenbaum et al. (2015) reported that 82.6% of diagnoses made at 18 months were confirmed at 3 years, but a significant proportion (12.1%) were diagnosed at 3 years who had not been identified earlier. These findings highlight the importance of repeated assessments and ongoing monitoring, as many children may not meet diagnostic criteria at younger ages.

Research specifically examining the stability of ADOS-2 classifications is limited. Although its psychometric properties are well established (De Bildt et al., 2009; Gotham et al., 2008; Gray, Bruce, Ae, & Sweeney, 2007; Kamp-Becker et al., 2011; Le Couteur, Haden, Hammal, & McConachie, 2008; Oosterling et al., 2010), the longitudinal stability of ADOS-2 classifications in EL-siblings across early childhood has not been systematically studied. It is important to recognise that diagnostic decisions are based on comprehensive assessments that integrate

multiple data sources, with the ADOS-2 representing only one component (Bishop & Lord, 2023). Clarifying the stability of ADOS-2 classifications over time contributes to the broader understanding of diagnostic stability and informs the clinical interpretation of ADOS-2 classifications relative to other assessment modalities and tools. It is also important to note that this study was conducted with volunteer families with an older autistic child participating in a research study, not children who were clinically referred for developmental concerns. The assessment context therefore differs from clinical samples, where parents actively seek clarity on a diagnosis, which may influence diagnostic decisions and the agreement between the ADOS-2 and Clinical Best Estimate (CBE) diagnosis. Research diagnoses were not communicated to families, but were used solely for scientific purposes, which further distinguishes the research context from clinical practice. With ADOS-2 classifications, we refer specifically to the classifications of autism spectrum according to ADOS-2 threshold.

The first research question addresses the stability of ADOS-2 classifications over time (defined as being either consistently above or consistently below the ADOS-2 autism spectrum cutoff at both time points). Although no prior research is available, we expected increasing stability in ADOS-2 classifications with age. Second, we investigated whether sex and developmental levels differed between the ADOS-2 classifications at each time point and between stable and unstable ADOS-2 classification patterns over time. Based on previous literature, we expected boys and children with lower developmental levels to show more ADOS-2 positive classifications (Loomes, Hull, & Mandy, 2017; Messinger et al., 2015). We also examined the agreement between ADOS-2 classifications at 14, 24 and 36 m and CBE diagnosis at 36 m and analysed the ADOS-2 classification patterns of children with an autism outcome. We expected that a subset of children would display stable ADOS-2 classifications up to the age of 3. These children may already show a clear picture early on, with more developmental difficulties, evident by cognitive delay or more pronounced autism characteristics compared to children with unstable ADOS-2 classifications (Chawarska et al., 2014; Zwaigenbaum et al., 2016).

## Method

### Participants

The sample included 304 EL-siblings drawn from a prospective, longitudinal multi-site study conducted by the EuroSibs Autism Research Network (Jones et al., 2019). Assessments were conducted at 5, 10, 14, 24 and 36 months (m, for months). Ethical approval was given by local or national ethics committees, following the Declaration of Helsinki and APA guidelines. Parents gave written consent. EL-children had at least one older sibling with a community diagnosis of autism.

Families were recruited through well-baby clinics, child-care and diagnostic centers, autism services and parent events. Exclusion criteria included epilepsy, preterm birth (i.e., <36 weeks gestational age) and genetic syndromes related to autism. EL-data from four sites were used: the Early Autism Sweden (EASE) project ( $n = 221$ , Sweden), Babystudy project ( $n = 60$ , Belgium), Siblings of Children with Autism (ZEBRA) project ( $n = 52$ , the Netherlands) and Studying Autism and ADHD in the Early Years (STAARS) project ( $n = 101$ , UK). Participants were excluded if ADOS-2 data were missing at two or more timepoints ( $n = 128$ ). Missing data resulted from missed appointments, dropouts and unadministered instruments (e.g., ADOS-2 at 14 m was excluded in the UK protocol and largely unavailable for Sweden). Missing data were not imputed, as this was deemed inappropriate for the study's aims. Analyses were based on available data; for an overview, see Table S1. Excluded participants did not differ significantly from the included sample in terms of sex or developmental level at 14 and 24 m but at 36 m excluded children had a lower developmental level than included children; however, only 24 excluded children had Mullen Scales of Early Learning (MSEL) data available at this age (see Table S2). The final sample of 304 EL-siblings (46.7% girls) was distributed as follows: Sweden ( $n = 140$ ), UK ( $n = 77$ ), Belgium ( $n = 52$ ), the Netherlands ( $n = 35$ ). The number of children with a CBE autism diagnosis at 36 m per site was: Sweden ( $n = 36$ ), UK ( $n = 12$ ), Belgium ( $n = 13$ ), the Netherlands ( $n = 10$ ). For 95 participants, data were available at all three time points, 97 at 14 m and 24 m, 296 at 24 m and 36 m, and 101 at 14 m and 36 m.

## Measures

**Autism Diagnostic Observation Schedule-2.** The ADOS-2 (Lord et al., 2012) is a standardised, semi-structured observational measure designed to assess autism-related features in communication, social interaction, play and restricted, repetitive or sensory behaviours and interests. In this study, the Toddler Module, Module 1 or Module 2 were administered based on the standard age and language ability criteria specified in the manual (see Table S3). The exception to this was that at 14 m the ADOS Toddler Module was administered if the child could move independently; but 27 children had not yet achieved independent walking at that time. Observations were performed by ADOS-trained researchers, who met standardised administrations and scoring reliability requirements. ADOS-2 scores include the Calibrated Severity Score (CSS), Social Affect Score (CSS-SA) and Restricted and Repetitive Behaviour Score (CSS-RRB), ranging from 1 to 10. CSS-values of  $\geq 4$  indicate an ADOS-positive classification for autism spectrum (Gotham, Pickles, & Lord, 2009). CSS-values for the ADOS-T were calculated per Esler et al. (2015), while those for Module 1 and 2 per Hus, Gotham, and Lord (2014).

**Social Responsiveness Scale (SRS).** The SRS (Constantino & Gruber, 2012) is a 65-item parent-report questionnaire assessing autistic characteristics over the past 6 months. Items are rated on a 4-point scale, with higher scores indicating greater social communication difficulties. SRS raw scores were used, with higher scores reflecting more severe social impairments.

**Mullen Scales of Early Learning.** The MSEL (Mullen, 1995) assesses verbal and non-verbal skills in children aged 1–68 m across four domains: fine motor, visual perception, receptive language and expressive language. *T*-scores and percentiles were calculated, with the Early Learning Composite (ELC;  $M = 100$ ,  $SD = 15$ ) reflecting overall cognitive ability. Trained researchers, meeting 90% scoring reliability through

role-play and video scoring, conducted the assessments. Experienced clinicians validated all assessments.

**Clinical Best Estimate Diagnosis of Autism.** CBE outcome, based on DSM-5 criteria (APA, 2013), was determined at 36 m through clinical consensus, the gold standard for diagnosing autism (Bishop & Lord, 2023). Decisions integrated ADOS-2 results, diagnostic interviews (i.e., ADI-R) and parent-reported screeners but were also informed by item-level or qualitative information, developmental assessments, researcher observations made during the assessments and additional parent-reported information. All diagnoses were made by a team including at least one experienced senior clinical psychologist and a researcher involved in the assessments, both with expertise in young children with developmental disorders. At least one team member remained consistent across all CBE assessments.

**Maternal education.** Maternal education, coded from 1 (primary) to 4 (university) was used to characterise the sample. It approximates socioeconomic status but reflects only part of this broad construct (Duncan & Magnuson, 2005).

Sample descriptives are shown in Table 1.

## Data analysis

Descriptive analyses calculated the number of EL-siblings with ADOS-positive scores at the different time points. Changes in ADOS-positive proportions (14–36 m) were assessed using Cochran's *Q*-test, with post-hoc pairwise comparisons using Dunn's procedure and Bonferroni correction (Abdi, 2007; Dunn, 1964). Stability coefficients measured the proportion of children maintaining an ADOS-2 autism (spectrum) classification over time. Mann-Whitney *U* tests compared ADOS-2 CSS (total/SA/RRB) between children with and without a CBE autism diagnosis, with Holm-Bonferroni correction for multiple testing (Holm, 1979). Chi-squared test analysed sex and site differences in ADOS-2 classification and stability. Given variation in CBE diagnosis across sites, site differences in ADOS-2 classification were analysed separately for the CBE autism and CBE non-autism groups. Mann-Whitney *U* tests were used to compare MSEL scores across ADOS-2 classifications at each timepoint. Group differences in MSEL- and SRS-scores between children with stable positive, stable negative and unstable ADOS-2 classifications were examined using a Kruskal-Wallis *H* test, followed by post-hoc pairwise comparisons with Bonferroni correction. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of ADOS-2 classifications at 14, 24 and 36 m compared to CBE autism diagnosis at 36 m were calculated. A sensitivity analysis excluding non-walking children at 14 m is reported in Table S4. Analyses were conducted using SPSS version 28 (IBM Corp., 2021).

## Results

At 36 m 71 from 304 EL-siblings (23.4%) met DSM-5 CBE diagnostic criteria for autism. This comprised 45 from 161 boys (28.0%) and 26 from 142 (18.1%) girls, a marginally significant sex difference ( $\chi^2(1) = 3.79$ ,  $p = .052$ ).

### ADOS-2 classification at 14, 24 and 36 months

A total of 40 children (38.8%; 95% CI [29%, 48%]) received an ADOS-positive classification at 14 m. No significant sex ( $\chi^2(1) = 1.36$ ,  $p = .244$ ) or site (CBE

**Table 1** EL-children characteristics

	CBE autism <i>n</i> = 71 (23.4%) Frequency (%)		CBE non-autism <i>n</i> = 233 (76.6%) Frequency (%)		Test	<i>p</i> -Value	$\alpha^*$
Sex	26 female (36.6%) 45 male (63.4%)		116 female (49.8%) 117 male (50.2%)		$\chi^2 = 3.79$	.052	
Maternal education					$\chi^2 = 3.06$	.383	
1: Primary	2 (2.8%)		4 (1.7%)				
2: Secondary	21 (29.6%)		48 (20.6%)				
3: Tertiary-undergraduate	18 (25.4%)		67 (28.8%)				
4: Tertiary-postgraduate	21 (29.6%)		83 (35.6%)				
Missing	9 (12.6%)		31 (13.3%)				
	<b>Mean (SD)</b>	<b>Range</b>	<b>Mean (SD)</b>	<b>Range</b>			
Age (months)							
14-month visit	14.34 (0.53)	13.34–15.93	14.40 (0.67)	12.98–16.95	<i>U</i> = 7,244	.694	
24-month visit	24.87 (1.12)	23.20–29.47	24.93 (1.21)	23.16–30.63	<i>U</i> = 7,946	.870	
36-month visit	38.20 (2.59)	35.57–47.28	37.73 (2.25)	35.53–48.79	<i>U</i> = 6,982.50	.233	
ELC (MSEL)							
14-month visit	86.15 (13.93)		91.57 (14.79)		<i>U</i> = 5,708	<b>.007</b>	.05
24-month visit	85.64 (19.20)		99.42 (16.54)		<i>U</i> = 3,928.50	<b>&lt;.001</b>	.017
36-month visit	91.34 (23.22)		105.37 (17.58)		<i>U</i> = 4,376	<b>&lt;.001</b>	.025

Maternal education = highest achieved diploma of participant's mother,  $\alpha^*$ : adjusted alpha level determining significance after Holm-Bonferroni correction was applied; Significant differences are marked in bold. CBE, Clinical Best Estimate; ELC, Early Learning Composite; MSEL, Mullen Scales of Early Learning; SD, standard deviation.

no autism:  $\chi^2(2) = 1.97, p = .37$ ; CBE autism: ( $\chi^2(2) = 3.88, p = .144$ ) differences were found, see Table 2 and S4. Children with ADOS-2 positive classifications had significantly lower MSEL-ELC scores ( $U = 637.5, p < .001$ ), see Table S5.

A total of 122 children (40.9%; 95% CI [35%, 46%]) received an ADOS-positive classification at 24 m. No significant sex differences were found ( $\chi^2(1) = 1.69, p = .194$ ), see Table S5. In the CBE no autism group site differences were significant ( $\chi^2(3) = 12.13, p = .007$ ), but not in the autism group ( $\chi^2(3) = 5.83, p = .120$ ), see Table 2. Children with ADOS-2 positive classifications had significantly lower MSEL-ELC scores ( $U = 6,619, p < .001$ ), see Table S5.

A total of 132 children (43.7%; 95% CI [38%, 49%]) received an ADOS-positive classification at 36 m. No significant sex differences were found ( $\chi^2(1) = 3.15, p = .082$ ). In both the CBE no autism and autism groups, significant site differences were found (respectively:  $\chi^2(3) = 18.66, p < .001$ ;  $\chi^2(3) = 25.64, p < .001$ ). See Table 2 and S4. Children with ADOS-2 positive classifications had significantly lower MSEL-ELC scores ( $U = 8,146, p < .001$ ); see Table S5.

Cochran's *Q* test showed a significant difference in ADOS-positive proportions across different time points ( $\chi^2(2) = 9.529, p = .009$ ). Post-hoc Dunn's tests indicated no significant increase from 14 to 24 m ( $p = .368$ ) and from 24 to 36 m ( $p = .368$ );

**Table 2** Site differences in ADOS-2 classifications

Site	CBE autism		<i>p</i> -Value	CBE non-autism		<i>p</i> -Value	
	ADOS-positive <i>n</i>	ADOS-negative <i>n</i>		ADOS-positive <i>n</i>	ADOS-negative <i>n</i>		
14 months			$\chi^2 = 3.88$			$\chi^2 = 1.97$	.37
Belgium	6	7		10	27		
Netherlands	0	2		1	7		
Sweden	10	3		11	19		
24 months			$\chi^2 = 5.83$			$\chi^2 = 12.13$	<b>.007</b>
Belgium	11	1		8	30		
Netherlands	6	4		3	22		
Sweden	26	9		43	58		
UK	6	6		19	46		
36 months			$\chi^2 = 25.64$			$\chi^2 = 18.66$	<b>&lt;.001</b>
Belgium	13	0		15	24		
Netherlands	8	2		3	22		
Sweden	35	1		43	59		
UK	5	7		10	55		

Significant differences are marked in bold. CBE, Clinical Best Estimate.

however, there was a significant increase from 14 to 36 m ( $p = .006$ ). The relationship between ADOS-2 scores and CBE diagnosis of autism is shown in Table S6.

### Stability of ADOS-2 classification from 14 to 36 months

**Stability and change from 14 to 24 months.** Of the 35 siblings with an ADOS-positive result at 14 m, 21 retained this classification at 24 m, which resulted in a stability coefficient of 60%. 38.7% of the siblings with no ADOS-positive classification at 14 m met the cut-off at the 24 m assessment (see Figure 1a).

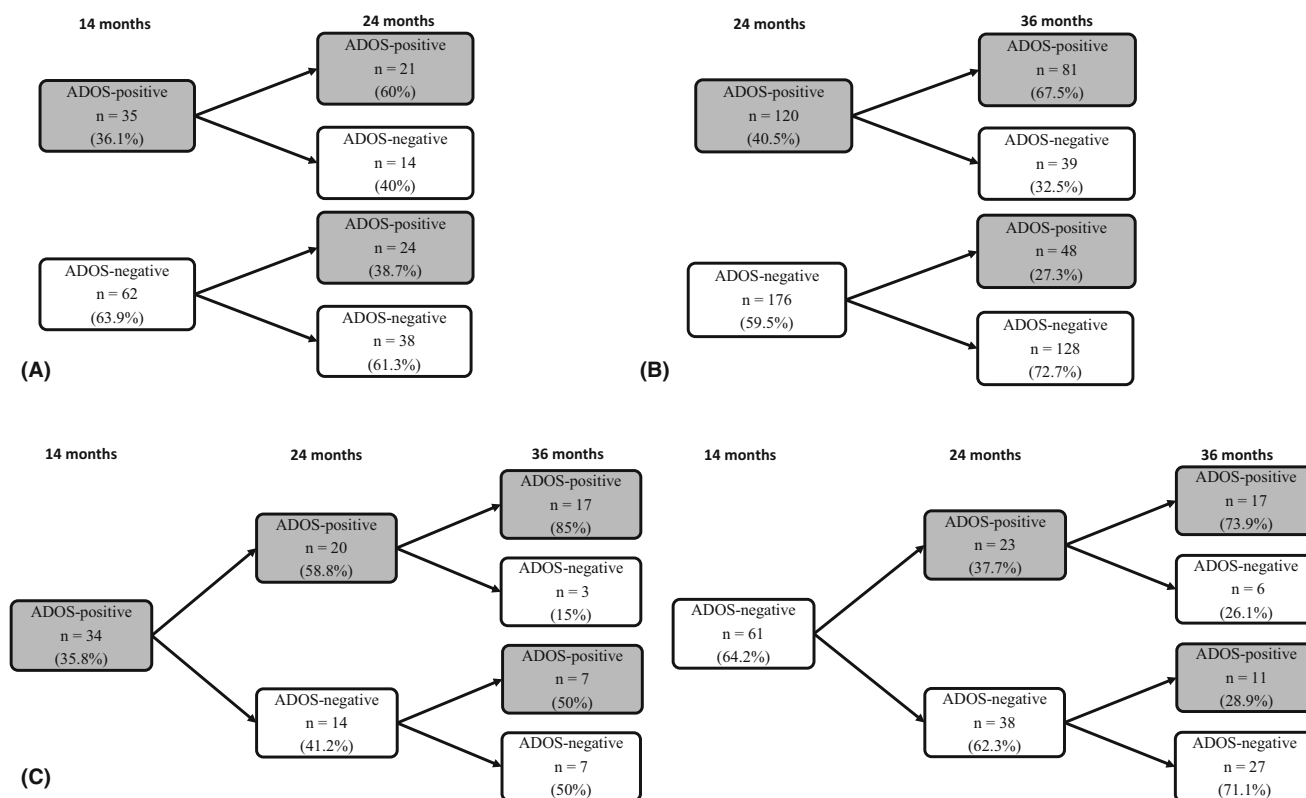
**Stability and change from 24 to 36 months.** Of the 120 siblings with an ADOS-positive result at 24 m, 81 retained this classification at 36 m, which resulted in a stability coefficient of 67.5%. 27.3% of the children with no ADOS-positive classification at 24 m met the cut-off at the 36 m assessment (see Figure 1b).

**Stability and change from 14 to 36 months.** As shown in Figure 1c, the overall stability of an ADOS-positive classification from 14 to 36 m was 70.6%, with 24 from 34 siblings retaining their classification over this period. Of these, 17 had stable ADOS-positive classifications across 14, 24 and 36 m, while 7 had an unstable pattern, being

ADOS-negative at 24 m. Of those siblings who were ADOS-negative at 14 m, 28 (45.9%) received an ADOS-positive classification at 36 m.

### Stability patterns of children with a CBE outcome of autism

Table 3 gives an overview of the ADOS-classification patterns in relation to CBE outcome. At 36 m, out of the 95 EL-siblings with ADOS-2 data at all three timepoints 26 (27.4%) were given a CBE clinical diagnosis of autism according to DSM-5 criteria. More than half (16 out of 26; 61.5%) of these children already had an ADOS-positive result at 14 m. Furthermore, out of these 16 children, 13 had ADOS-positive results from 14 m onwards, although three of them had instable ADOS-classifications, with an ADOS-positive result at 14 m, but not at 24 m (see Table 3). At 24 m, almost all children with a CBE diagnosis of autism had an ADOS-positive result (23 out of 26; 88.5%). All children with a CBE of autism were ADOS-positive at 36 m. However, 60.9% (42 out of 69) of the children with no CBE of autism had an ADOS-positive result at one or more time points by the age of 3. An overview of the ADOS-classification patterns at 24 and 36 m in relation to CBE per site can be found in Tables S7–S11. No significant sex differences were found between children with stable-positive, stable-negative and unstable ADOS-2 classifications over



**Figure 1** Stability of ADOS-2 classifications across visits. (A) stability from 14 to 24 months, (B) stability from 24 to 36 months, (C) stability from 14 to 36 months

ADOS-classification			Clinical Best Estimate		
14 months	24 months	36 months	CBE autism n = 26	CBE no autism n = 69	Total n = 95
			13 (50%)	4 (5.8%)	17
			0	3 (4.3%)	3
			0	7 (10.2%)	7
			3 (11.5%)	4 (5.8%)	7
			0	6 (8.7%)	6
			10 (38.5%)	7 (10.2%)	17
			0	11 (15.9%)	11
			0	27 (39.1%)	27

Classification, green ADOS-positive, red ADOS-negative. CBE, Clinical Best Estimate.

**Table 3** ADOS-classification patterns in relation to Clinical Best Estimate diagnosis of autism

time ( $\chi^2(2) = 3.083$ ,  $p = .214$ ). Mullen ELC scores were significantly higher in children without CBE of autism at all timepoints ( $U = 5,708$ ,  $p = .007$ ;  $U = 5928.5$ ,  $p < .001$ ;  $U = 4,376$ ,  $p < .001$  respectively, see also Table 1). There were significant differences in MSEL scores between children with stable-positive, stable-negative and unstable ADOS-2 classification patterns at all time points (14 m:  $H(2) = 11.93$ ,  $p = .003$ ; 24 m:  $H(2) = 13.54$ ,  $p = .001$ ; 36 m:  $H(2) = 16.93$ ,  $p < .001$ ). Post-hoc comparisons with Bonferroni correction showed that children with stable-positive classifications had significantly lower scores than those with stable-negative classifications (14 m:  $p = .004$ , 24 m:  $p = .001$ , 36 m:  $p < .001$ ) and unstable ADOS-2 classifications (14 m:  $p = .005$ , 24 m:  $p = .004$ , 36 m:  $p = .006$ ) at all timepoints. Children with stable-negative classifications did not differ from children with unstable classifications at any timepoint (14 m:  $p = 1$ , 24 m:  $p = 1$ , 36 m:  $p = .296$ ), see Table S12. There were no significant differences in SRS scores between children with stable-positive, stable-negative and unstable ADOS-2 classification patterns ( $H(2) = 4.03$ ,  $p = .133$ ), see Table S12. To further explore the variability between ADOS-2 classification and CBE diagnosis, we have plotted the ADOS

CSS-trajectories of children with fluctuating ADOS-2 classifications per child, split by CBE outcome (autism vs. no autism; see Figure S1). This includes children with ADOS-positive classifications at 36 m but who showed inconsistent classifications at 14 and 24 m.

The psychometric properties of the ADOS-2 classifications at 14, 24 and 36 m are presented in Table 4.<sup>1</sup> Sensitivity increased significantly with age, from 64% at 14 m to 86% at 36 m ( $\chi^2(2) = 12.154$ ,  $p = .002$ ), with post-hoc McNemar tests showing a significant increase between 14 and 36 m, and additionally between 24 and 36 m using Holm correction (Holm, 1979). Specificity and PPV remained relatively stable across time points, while NPV showed a descriptive increase (from 84% at 14 m to 94% at 36 m), though not statistically significant.

## Discussion

In this prospective longitudinal study, we examined the stability of ADOS-2 autism spectrum classifications across three early childhood timepoints in EL-siblings. EL-siblings are an important population to study, as the presence of an autism diagnosis in an

**Table 4** Sensitivity, specificity PPV and NPV of the ADOS-2 as compared to CBE diagnosis

Instrument	No. of positive cases		No. of negative cases		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
	True positive	False positive	True negative	False negative				
ADOS 14 m <sup>a</sup>	18	22	53	10	64.3% (45.9–80.2)	70.7% (59.8–80.2)	45% (30.3–60.4)	84.1% (73.8–91.7)
ADOS 24 m <sup>a</sup>	49	73	156	20	71% (59.7–80.8)	68.1% (61.9–73.9)	40.2% (31.7–49)	88.6% (83.4–92.8)
ADOS 36 m <sup>a</sup>	61	71	160	10	85.9% (76.6–92.7)	69.3% (63.1–75)	46.2% (37.8–54.7)	94.1% (89.9–97)

CBE, Clinical Best Estimate; NPV, negative predictive value; PPV, positive predictive value; No. total number. 14 m timepoint at 14 months, 24 m timepoint at 24 months, 36 m timepoint at 36 months.

<sup>a</sup>An ADOS-2 classification of autism requires meeting or exceeding the autism spectrum cut-point. The cut-point for the Toddler Module is a total score  $\geq 10$  for children between 10 and 20 months old or between 21 and 30 months with little-to-no words or  $\geq 8$  for children between 21 and 30 months with some words. The cut-point for Module 1 is a total score of  $\geq 11$  (no-to-few words) or  $\geq 8$  (some words). The cut-point for Module 2 is a total score of  $\geq 7$ .

older sibling provides a natural context to monitor younger siblings from an early age using standardised tools such as the ADOS-2. To our knowledge, this is the first investigation to examine whether the ADOS-2, a well-established observational diagnostic instrument, provides a stable and reliable identification of autism in EL-siblings prior to the age of three.

The recurrence rate of 23.4% and the (marginally) higher rate in males than females are consistent with prior reports in EL-sibling samples (Ozonoff et al., 2024). Our results show that for some EL-children, ADOS-2 classifications are stable from early on. The stability of ADOS-positive classifications at 14 and 24 m relative to the classification at 36 m was moderate (70.6% and 67.5%, respectively). The majority of the ADOS-2 classifications made before 36 m, were confirmed at 36 m (see Figure 1). For a subgroup of children however, a classification at one point in time was inconsistent with their classification at subsequent timepoints. Approximately 30% of the sample with unstable patterns received ADOS-positive classifications at 14 m and/or 24 m, and ADOS-negative classifications at age 3. Several factors may account for this discrepancy; one possibility is that EL-siblings exhibit transient behaviours in early childhood, which may not persist over time (Gamliel, Yirmiya, & Sigman, 2007). Another possibility is misclassification due to measurement error, especially near the diagnostic threshold, as well as overlap with other behaviours, such as shyness and anxiety or characteristics of developmental delay or ADHD that can impact ADOS score (Greene, Vasile, Bradbury, Olsen, & Duvall, 2021; Hamrick, Ros-Demarize, Kanne, & Carpenter, 2024; Hayashi et al., 2022). Module selection may also affect stability. Administering higher modules with more demanding tasks, when children only just qualify for them, may result in higher ADOS-2 scores, whereas administering lower modules may result in lower scores (Klein-Tasman, Risi, & Lord, 2007). Differences in classifications at 24 m may partly reflect differences between the Toddler Module and Module 2, since the five children without a CBE autism diagnosis who met the ADOS-2 cutoff on Module 2, only did so at this timepoint. These variations likely capture developmental shifts rather than measurement error; see Table S3.

The stability of an ADOS-negative classification was lower, approximately 55% at 14 m and about 73% at 24 m relative to the ADOS-2 assessment at 36 m. As PPV and specificity remained relatively stable across timepoints (see Table 4), this pattern likely reflects increasing diagnostic clarity with age rather than a rise in false positives. Sensitivity and specificity were generally moderate (Glascoe, 2005). Sensitivity increased with age, indicating that the ADOS-2 better identifies children with autism as they grow older, although the good sensitivity at

36 m may be partly reflective of the incorporation of the ADOS-2 in the CBE. Specificity remained relatively stable across time. These estimates are somewhat lower than those reported in clinical samples (e.g., Lebersfeld, Swanson, Clesi, & O'Kelley, 2021), likely due to differences in referral context and autism prevalence. NPV was consistently high, while PPV remained low across ages, suggesting limited predictive value of a positive ADOS-2 classification in EL-siblings up to 3 years. As few studies have reported PPV in this population, direct comparisons are limited. This underscores the complexity of early identification and the importance of continued follow-up, even in children who do not yet display clear autism characteristics, as some may not meet ADOS-2 classification thresholds, or receive a CBE, until or after age 3 (Bazelmans et al., 2024; Brian et al., 2016; Ozonoff et al., 2018).

Direct comparison is challenging given the absence of prior research examining the stability of ADOS-2 classification at such a young age. The most comparable studies are those assessing diagnostic stability over time in EL-siblings (Ozonoff et al., 2015; Zwaigenbaum et al., 2016), albeit with a significant distinction; these studies established a diagnosis at each age with a more comprehensive evaluation, whereas an ADOS-positive classification was not a prerequisite for diagnosis. As might be expected, these studies that rely on multiple instruments and sources of information and expert consensus diagnostic opinion show higher stability (Ozonoff et al. (2015): 82%–93%; Zwaigenbaum et al. (2015): 82.6%) than we found based on a single observational instrument the ADOS-2. ADOS-2 assessments capture a momentary snapshot and are highly dependent on circumstances and the child's response during a time-limited assessment often in an unfamiliar context. They provide only limited information on restricted and repetitive and sensory behaviours, necessitating complementary information to determine a CBE diagnosis (Kim & Lord, 2010). Therefore, the higher stability observed in studies that assess diagnostic stability, compared to our results is not surprising.

Moreover, an autism diagnosis requires more than scoring on a standardised instrument. Best practices emphasise the need for comprehensive assessment combining information from multiple sources including parents and direct observations, with instruments like the ADOS-2 serving as informative, but not definitive components (Bishop & Lord, 2023; Fombonne, 2023). While it may be possible to identify developmental concerns early in some EL-siblings, ongoing monitoring and surveillance remain essential, particularly when characteristics are ambiguous (Ozonoff et al., 2011; Zwaigenbaum et al., 2021). This process may include repeated assessments, such as the ADOS-2, alongside parental interviews and other clinical information (Zwaigenbaum et al., 2015).

Examining patterns of EL-siblings with a CBE outcome of autism at 36 m also revealed that the stability of ADOS-positive classifications increased with age suggesting that for some children the ADOS-2 is already informative early on. In the subgroup with ADOS-2 data at all timepoints ( $n = 95$ ), all EL-siblings with an autism CBE were ADOS-positive at 36 m. However, the ADOS-2 also classified some non-autistic children as positive, underscoring the need for clinical judgement (see Table 3). The agreement between ADOS-2 classifications and CBE diagnosis varied across sites. This variability may reflect differences in sample characteristics, ADOS-2 scoring or diagnostic differences, factors we could not formally assess. Given language, practical and resource constraints of this multi-national study we did not attempt any cross-site diagnostic consensus process. Although variability in ADOS-2 classification patterns was observed in both groups (autism vs. no autism) in our sample, no clear pattern emerged (see Figure S1). No sex differences were found in ADOS-2 classification patterns. Despite efforts to minimise site differences in ADOS-2 scoring through training and reliability meetings, some time point-specific variability was observed though it was not consistent over time. These findings underscore the need to interpret early ADOS-2 classifications within a broader clinical context. Additionally, it is important to consider that the ADOS-2 at 36 m was taken into account when determining the CBE diagnosis, which introduces bias into the assessment of agreement. Our results suggest that, for some children, ADOS-2 classifications are already stable and accurate before age 3 and may be informative as early as 14 m of age. Although we found no evidence for more pronounced autism characteristics in this group, children with stable positive ADOS-2 classification consistently showed lower developmental levels across all timepoints.

However, caution is needed, as approximately 60% of those without a CBE of autism scored above the ADOS autism spectrum cut-off at least once up to age 3. It is challenging to determine why these children score above the ADOS cut-off yet do not receive a CBE diagnosis of autism. One possible explanation is the relationship between ADOS-2 score and developmental delay, which could lead to children with delayed development being incorrectly identified by the ADOS-2. In our study children without a CBE diagnosis but with a positive ADOS-2 classification exhibited statistically significantly lower MSEL scores at 14 and 24 m (see Table S5). However, by 36 m, this difference was no longer significant. Another potential explanation is that certain children receive positive ADOS-2 results driven by social communication difficulties, yet an autism diagnosis requires the presence of both social communicative difficulties and restricted, repetitive behaviours and interests and sensory differences.

The ADOS-2 provides only limited information on repetitive and sensory behaviours, necessitating complementary information to determine a CBE diagnosis (Kim & Lord, 2010). Furthermore, observed behaviours may not cause clinically significant challenges in everyday functioning, which is a requirement for diagnosis (APA, 2022). Lastly, these children may not meet the criteria for an autism diagnosis at age 3 but may do so later in their development (Bazelmans et al., 2024; Brian et al., 2016).

### Limitations and directions for future research

While this study provides valuable insights into the stability of the ADOS-2 classifications in EL-siblings, several limitations and future directions should be noted. First, the sample of highly educated mothers limits generalisability. Second, a substantial number of children ( $n = 128$ ) were excluded due to missing or insufficient ADOS-2 data across time points, potentially impacting representativeness. Third, there was a considerable amount of missing data at 14 m, and 27 siblings were not yet walking independently at their ADOS-2 assessment, which deviates from standard administration guidelines. However, of these siblings, only two without CBE of autism exceeded the cutoff solely at this point, suggesting minimal impact of motor development on early classifications in our sample. Fourth, potential interventions were not accounted for. Furthermore, site differences may exist between CBE autism diagnostic decision-making in this multi-site study, despite being conducted by experienced teams led by expert clinicians using similar methods. Importantly, participants in this study were assessed in the context of a longitudinal research project, rather than referred for diagnostic evaluation. As such, expectations surrounding diagnosis may differ from clinical practice and may also be influenced by parent's knowledge of their older autistic child, potentially influencing the variability between CBE diagnostic thresholds and ADOS scores across sites. Moreover, the relatively small number of children with a CBE autism diagnosis, particularly when broken down by site, limits generalisability. Finally, the absence of agreed criteria for broader autism phenotype (BAP) precluded further subdivision of the non-autism CBE group into BAP and typically developing subgroups. Future research could (1) examine ADOS trajectories within these subgroups and (2) explore whether CSS scores over time reveal distinct developmental profiles.

### Conclusion

This study demonstrates that, while the ADOS-2 can offer valuable information to the early monitoring of EL siblings, its stability before age 3 shows notable variability between individuals. A significant

proportion of children without a CBE of autism received positive ADOS-2 classifications at one or more timepoints, highlighting the potential for false positives. These findings underscore the importance of interpreting early ADOS-2 outcomes with caution and always within the context of broader clinical assessment. Diagnostic decisions should not be based solely on ADOS-2 results, particularly in very young children. Continued developmental surveillance, incorporating multiple sources of information over time, remains essential for accurate identification. Further research with longer follow-up periods is needed to clarify which early behavioural features are associated with false positive or false negative ADOS-2 outcomes and to advance our understanding of the developmental trajectories of autism.

### Supporting information

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

**Table S1.** Available data for the ADOS-2 per time point.

**Table S2.** MSEL and sex of included and excluded children.

**Table S3.** ADOS-2 modules and classifications per age and per CBE outcome.

**Table S4.** Sensitivity, specificity PPV and NPV of the ADOS-2 as compared to CBE diagnosis excluding non-walking children.

**Table S5.** Sex and developmental differences in ADOS-2 classifications.

**Table S6.** Difference in ADOS-2 scores between EL-children with and without a CBE diagnosis of autism.

**Table S7.** ADOS-classification patterns at 24 and 36 months in relation to Clinical Best Estimate diagnosis of autism.

**Table S8.** ADOS-classification patterns at 24 and 36 months in relation to Clinical Best Estimate diagnosis of autism (Belgium).

**Table S9.** ADOS-classification patterns at 24 and 36 months in relation to Clinical Best Estimate diagnosis of autism (The Netherlands).

**Table S10.** ADOS-classification patterns at 24 and 36 months in relation to Clinical Best Estimate diagnosis of autism (Sweden).

**Table S11.** ADOS-classification patterns at 24 and 36 months in relation to Clinical Best Estimate diagnosis of autism (UK).

**Table S12.** Sex, Developmental and SRS Differences in Stable-Positive, Stable-Negative and unstable ADOS-2 Classification patterns.

**Figure S1.** ADOS CSS trajectories of children with fluctuating ADOS-2 classifications and ADOS-positive classification at 36 months.

### Acknowledgements

This research was supported by the Medical Research Council (G0701484, MR/K021389/1, MR/T003057/1), BASIS (led by Autistica), the Innovative Medicines Initiative 2 Joint Undertaking (grants No. 777394 and

115300, AIMS-2-TRIALS), the EU's Horizon 2020 programme (grant No. 642990, BRAINVIEW) and the Wellcome Institutional Strategic Support Fund (Birkbeck).

Additional funding was provided by the Research Foundation Flanders (including a postdoctoral Fellowship 1215425N for S.B.), Ghent University, the Special Research Fund and the Support Fund Marguerite-Marie Delacroix. This work was further supported by the EU Seventh Framework Programme (FP7/2007–2013), EFPIA, Autism Speaks and Simons Foundation Autism Research Initiative (SFARI).

Any views expressed are those of the authors and do not necessarily reflect those of the funders (IHI-JU2).

The authors thank the families who participated in the Eurosibs study and the STAARS, EASE, ZEBRA and Babystudie teams for their support in data collection.

T.C. has served as a paid consultant to F. Hoffmann-La Roche Ltd. and Servier and receives royalties from Sage Publications and Guilford Publications.

S.B. discloses that he has in the last 3 years acted as an author, consultant or lecturer for Medice and LinusBio. He receives royalties for textbooks and diagnostic tools from Hogrefe, UTB, Ernst Reinhardt, Kohlhammer and Liber. S.B. is a partner at NeuroSupportSolutions International AB. The remaining authors have declared that they have no competing or potential conflicts of interest.

### Ethical considerations

Ethical approval was obtained from local ethics committees in the participating countries. Specifically, approval was granted by the National Health Service REC London Central (reference 13/LO/0751, 29 July 2013) for the UK site; the Regional Ethics Board in Stockholm (reference 2010/2085-31/3, 23 March 2011) for the Swedish site; the CMO Region Arnhem-Nijmegen (reference NL42726.091.13, 3 September 2013) for the Dutch sites in Nijmegen and Utrecht; and the ethical board of the Faculty of Psychology and Educational Sciences of Ghent University Ghent (reference: 2011/41, 2014/33, 2015/10 and 2015/81; 24 August 2011, 4 July 2014, 27 February 2015, 29 October 2015). The study was conducted in accordance with the Declaration of Helsinki and the ethical principles of the American Psychological Association. Written informed consent was obtained from parents prior to participation.

### Data availability statement

The data sets generated and analysed during this study are available from the corresponding author upon reasonable request.

### Correspondence

Sarah Schaubroeck, Research in Developmental Diversity Lab (RIDDL), Department of Experimental Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium; Email: [sarah.schaubroeck@ugent.be](mailto:sarah.schaubroeck@ugent.be)

## Key points

### What's known?

- The ADOS-2 is widely used to assess autism characteristics in young children, including siblings at elevated likelihood for autism. However, the stability of ADOS-2 classifications across early childhood remains unclear.

### What's new?

- ADOS-2 classification stability increases with age but remains moderate at 14 and 24 months compared to 36 months. A notable proportion of children without clinical best estimate diagnosis of autism showed positive ADOS-2 results at one or more early time points.

### What's relevant?

- These findings inform clinical practice by reinforcing the importance of integrating ADOS-2 results with broader clinical evaluations and longitudinal follow-up of elevated likelihood siblings to reduce the over-reliance on single-time-point ADOS-2 results.

## Endnote

1. Sensitivity, specificity, PPV and NPV for the Toddler Module did not substantially differ when non-walking children were excluded (see Table S4).

## References

- Abdi, H. (2007). Bonferroni and Sidak corrections for multiple comparisons. In N.J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 103–107). Thousand Oaks, CA: Sage.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th edn). Washington, DC: American Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890425596>
- American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th edn, text rev.; DSM-5-TR). Washington, DC: American Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890425787>
- Bazelmans, T., Arthur, R., Pasco, G., Shephard, E., Milosavljevic, B., Ali, J.B., ... & Team, T.B. (2024). Mid-childhood autism sibling recurrence in infants with a family history of autism. *Autism Research, 17*, 1501–1514.
- Bishop, S.L., & Lord, C. (2023). Commentary: Best practices and processes for assessment of autism spectrum disorder – The intended role of standardized diagnostic instruments. *Journal of Child Psychology and Psychiatry, 64*, 834–838.
- Brian, J., Bryson, S.E., Smith, I.M., Roberts, W., Roncadin, C., Szatmari, P., & Zwaigenbaum, L. (2016). Stability and change in autism spectrum disorder diagnosis from age 3 to middle childhood in a high-risk sibling cohort. *Autism, 20*, 888–892.
- Charman, T., Pasco, G., Hendry, A., Bazelmans, T., Narvekar, N., Goodwin, A., ... & The STAARS Team. (2023). Three year outcomes in infants with a family history of autism and/or attention deficit hyperactivity disorder. *JCPP Advances, 3*, e12189.
- Charman, T., Young, G.S., Brian, J., Carter, A., Carver, L.J., Chawarska, K., ... & Zwaigenbaum, L. (2017). Non-ASD outcomes at 36 months in siblings at familial risk for autism spectrum disorder (ASD): A Baby Siblings Research Consortium (BSRC) study. *Autism Research, 10*, 169–178.
- Chawarska, K., Shic, F., Macari, S., Campbell, D.J., Brian, J., Landa, R., ... & Bryson, S. (2014). 18-month predictors of later outcomes in younger siblings of children with autism spectrum disorder: A Baby Siblings Research Consortium Study. *Journal of the American Academy of Child & Adolescent Psychiatry, 53*, 1317–1327.e1.
- Constantino, J.N., & Gruber, C.P. (2012). *Social Responsiveness Scale, second edition (SRS-2)*. Los Angeles, CA: Western Psychological Services.
- De Bildt, A., Sytema, S., Van Lang, N.D.J., Minderaa, R.B., Van Engeland, H., & De Jonge, M.V. (2009). Evaluation of the ADOS revised algorithm: The applicability in 558 Dutch children and adolescents. *Journal of Autism and Developmental Disorders, 39*, 1350–1358.
- Duncan, G.J., & Magnuson, K.A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children, 15*, 35–54.
- Dunn, O.J. (1964). Multiple comparisons using rank sums. *Technometrics, 6*, 241–252.
- 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*, 2704–2720.
- Fombonne, E. (2023). Editorial: Is autism overdiagnosed? *Journal of Child Psychology and Psychiatry, 64*, 711–714.
- Fuller, E.A., & Kaiser, A.P. (2020). The effects of early intervention on social communication outcomes for children with autism spectrum disorder: A meta-analysis. *Journal of Autism and Developmental Disorders, 50*, 1683–1700.
- Gamliel, I., Yirmiya, N., & Sigman, M. (2007). The development of young siblings of children with autism from 4 to 54 months. *Journal of Autism and Developmental Disorders, 37*, 171–183.
- Glascoe, F.P. (2005). Screening for developmental and behavioral problems. *Mental Retardation and Developmental Disabilities Research Reviews, 11*, 173–179.

- Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39, 693–705.
- Gotham, K., Risi, S., Dawson, G., Tager-Flusberg, H., Joseph, R., Carter, A., ... & Pickles, A. (2008). A replication of the Autism Diagnostic Observation Schedule (ADOS) revised algorithms. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47, 642–651.
- Gray, K.M., Bruce, A.E., Ae, J.T., & Sweeney, D.J. (2007). Using the autism diagnostic interview-revised and the Autism Diagnostic Observation Schedule with young children with developmental delay: Evaluating diagnostic validity. *Journal of Autism and Developmental Disorders*, 38(4), 657–667.
- Greene, R.K., Vasile, I., Bradbury, K.R., Olsen, A., & Duvall, S.W. (2021). Autism Diagnostic Observation Schedule (ADOS-2) elevations in a clinical sample of children and adolescents who do not have autism: Phenotypic profiles of false positives. *The Clinical Neuropsychologist*, 36, 943–959.
- Hamrick, L.R., Ros-Demarize, R., Kanne, S., & Carpenter, L.A. (2024). Profiles of nonverbal skills used by young pre-verbal children with autism on the ADOS-2: Relation to screening disposition and outcomes. *Autism Research*, 17, 2370–2385.
- Hansen, S.N., Schendel, D.E., Francis, R.W., Windham, G.C., Bresnahan, M., Levine, S.Z., ... & Parner, E.T. (2019). Recurrence risk of autism in siblings and cousins: A multinational, population-based study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 58, 866–875.
- Hayashi, W., Hanawa, Y., Yuriko, I., Aoyagi, K., Saga, N., Nakamura, D., & Iwanami, A. (2022). ASD symptoms in adults with ADHD: A preliminary study using the ADOS-2. *European Archives of Psychiatry and Clinical Neuroscience*, 272, 217–232.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Hus, V., Gotham, K., & Lord, C. (2014). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders*, 44, 2400–2412.
- IBM Corp. (2021). *IBM SPSS Statistics for Windows (Version Released 2021) [Computer software]*. Armonk, NY: IBM Corp.
- Jokiranta-Olkonieni, E., Cheslack-Postava, K., Joelsson, P., Suominen, A., Brown, A.S., & Sourander, A. (2016). Risk of psychiatric and neurodevelopmental disorders among siblings of probands with autism spectrum disorders. *JAMA Psychiatry*, 73, 622–629.
- Jones, E.J.H., Mason, L., Begum Ali, J., Van Den Boomen, C., Braukmann, R., Cauvet, E., ... & Johnson, M.H. (2019). Eurosibs: Towards robust measurement of infant neurocognitive predictors of autism across Europe. *Infant Behavior and Development*, 57, 101316.
- Kamp-Becker, I., Ghahreman, M., Heinzl-Gutenbrunner, M., Peters, M., Remschmidt, H., & Becker, K. (2011). Evaluation of the revised algorithm of Autism Diagnostic Observation Schedule (ADOS) in the diagnostic investigation of high-functioning children and adolescents with autism spectrum disorders. *Autism*, 17, 87–102.
- Kim, S.H., & Lord, C. (2010). Restricted and repetitive behaviors in toddlers and preschoolers with autism spectrum disorders based on the Autism Diagnostic Observation Schedule (ADOS). *Autism Research*, 3, 162–173.
- Kim, S.H., Macari, S., Köller, J., & Chawarska, K. (2016). Examining the phenotypic heterogeneity of early autism spectrum disorder: Subtypes and short-term outcomes. *Journal of Child Psychology and Psychiatry*, 57, 93–102.
- Klein-Tasman, B.P., Risi, S., & Lord, C.E. (2007). Effect of language and task demands on the diagnostic effectiveness of the Autism Diagnostic Observation Schedule: The impact of module choice. *Journal of Autism and Developmental Disorders*, 37, 1224–1234.
- Lai, M.C., Lombardo, M.V., Auyeung, B., Chakrabarti, B., & Baron-Cohen, S. (2015). Sex/gender differences and autism: Setting the scene for future research. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54, 11–24.
- Le Couteur, A., Haden, G., Hammal, D., & McConachie, H. (2008). Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: The ADI-R and the ADOS. *Journal of Autism and Developmental Disorders*, 38, 362–372.
- Lebersfeld, J.B., Swanson, M., Clesi, C.D., & O'Kelley, S.E. (2021). Systematic review and meta-analysis of the clinical utility of the ADOS-2 and the ADI-R in diagnosing autism spectrum disorders in children. *Journal of Autism and Developmental Disorders*, 51, 4101–4114.
- Loomes, R., Hull, L., & Mandy, W.P.L. (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56, 466–474.
- Lord, C., Rutter, M., DiLavore, P.C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule (2nd edn)*. Los Angeles, CA: Western Psychological Services.
- Maenner, M.J., Warren, Z., Williams, A.R., Amoakohene, E., Bakian, A.V., Bilder, D.A., ... & Shaw, K.A. (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. *MMWR Surveillance Summaries*, 72, 1–14.
- Marrus, N., Hall, L.P., Paterson, S.J., Elison, J.T., Wolff, J.J., Swanson, M.R., ... & IBIS Network. (2018). Language delay aggregates in toddler siblings of children with autism spectrum disorder. *Journal of Neurodevelopmental Disorders*, 10, 29.
- Messinger, D., Young, G.S., Ozonoff, S., Dobkins, K., Carter, A., Zwaigenbaum, L., ... & Sigman, M. (2013). Beyond autism: A Baby Siblings Research Consortium Study of high-risk children at three years of age. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52, 300–308.e1.
- Messinger, D.S., Young, G.S., Webb, S.J., Ozonoff, S., Bryson, S.E., Carter, A., ... & Zwaigenbaum, L. (2015). Early sex differences are not autism-specific: A baby siblings research consortium (BSRC) study. *Molecular Autism*, 6, 32.
- Mullen, E.M. (1995). *Mullen scales of early learning*. Circle Pines, MN: American Guidance Service Inc.
- Oosterling, I.J., Wensing, M., Swinkels, S.H., Van Der Gaag, R.J., Visser, J.C., Woudenberg, T., ... & Buitelaar, J.K. (2010). Advancing early detection of autism spectrum disorder by applying an integrated two-stage screening approach. *Journal of Child Psychology and Psychiatry*, 51, 250–258.
- Ozonoff, S., Young, G.S., Bradshaw, J., Charman, T., Chawarska, K., Iverson, J.M., ... & Zwaigenbaum, L. (2024). Familial recurrence of autism: Updates from the baby siblings research consortium. *Pediatrics*, 154, e2023065297.
- Ozonoff, S., Young, G.S., Brian, J., Charman, T., Shephard, E., Solish, A., & Zwaigenbaum, L. (2018). Diagnosis of autism spectrum disorder after age 5 in children evaluated longitudinally since infancy. *Journal of the American Academy of Child & Adolescent Psychiatry*, 57, 849–857.e2.
- Ozonoff, S., Young, G.S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., ... & Stone, W.L. (2011). Recurrence risk for autism spectrum disorders: A baby siblings research consortium study. *Pediatrics*, 128, e488–e495.
- Ozonoff, S., Young, G.S., Landa, R.J., Brian, J., Bryson, S., Charman, T., ... & Iosif, A. (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, 988–998.
- Rutter, M., Le Couteur, A., & Lord, C. (2003). *ADI-R: Autism Diagnostic Interview-Revised (ADI-R)*. Los Angeles, CA: Western Psychological Services.
- Ryberg, K.H. (2015). Evidence for the implementation of the Early Start Denver Model for young children with autism spectrum disorder. *Journal of the American Psychiatric Nurses Association*, 21, 327–337.
- Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Hultman, C.M., & Reichenberg, A. (2017). The heritability of autism spectrum disorder. *JAMA*, 318, 1182–1184.
- 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, 243–265.
- van 't Hof, M., Tisseur, C., van Berckeleer-Onnes, I., van Nieuwenhuyzen, A., Daniels, A.M., Deen, M., ... & Ester, W.A. (2021). Age at autism spectrum disorder diagnosis: A systematic review and meta-analysis from 2012 to 2019. *Autism*, 25, 862–873.
- Zeidan, J., Fombonne, E., Scolah, J., Ibrahim, A., Durkin, M.S., Saxena, S., ... & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism Research*, 15, 778–790.
- Zwaigenbaum, L., Bauman, M.L., Fein, D., Pierce, K., Buie, T., Davis, P.A., ... & Wagner, S. (2015). Early screening of autism spectrum disorder: Recommendations for practice and research. *Pediatrics*, 136, S41–S59.
- Zwaigenbaum, L., Brian, J., Smith, I.M., Sacrey, L.R., Franchini, M., Bryson, S.E., ... & Roncadin, C. (2021). Symptom trajectories in the first 18 months and autism risk in a prospective high-risk cohort. *Journal of Child Psychology and Psychiatry*, 62, 1435–1443.
- Zwaigenbaum, L., Bryson, S.E., Brian, J., Smith, I.M., Roberts, W., Szatmari, P., ... & Vaillancourt, T. (2016). Stability of diagnostic assessment for autism spectrum disorder between 18 and 36 months in a high-risk cohort. *Autism Research*, 9, 790–800.

Accepted for publication: 1 October 2025