9-5-2023

Development and replication of objective measurements of social visual engagement to aid in early diagnosis and assessment of autism

Warren Jones

Cheryl Klaiman

Shana Richardson

Meena Lambha

Morganne Reid

See next page for additional authors

Authors

Warren Jones, Cheryl Klaiman, Shana Richardson, Meena Lambha, Morganne Reid, Taralee Hamner, Chloe Beacham, Peter Lewis, Jose Paredes, Laura Edwards, Natasha Marrus, John N Constantino, Sarah Shultz, and Ami Klin

JAMA Network Open

Original Investigation | Psychiatry

Development and Replication of Objective Measurements of Social Visual Engagement to Aid in Early Diagnosis and Assessment of Autism

Warren Jones, PhD; Cheryl Klaiman, PhD; Shana Richardson, PhD; Meena Lambha, PhD; Morganne Reid, PhD; Taralee Hamner, PhD; Chloe Beacham, PhD; Peter Lewis, MBA; Jose Paredes, MS; Laura Edwards, PhD; Natasha Marrus, MD, PhD; John N. Constantino, MD; Sarah Shultz, PhD; Ami Klin, PhD

Abstract

IMPORTANCE Autism spectrum disorder is a common and early-emerging neurodevelopmental condition. While 80% of parents report having had concerns for their child's development before age 2 years, many children are not diagnosed until ages 4 to 5 years or later.

OBJECTIVE To develop an objective performance-based tool to aid in early diagnosis and assessment of autism in children younger than 3 years.

DESIGN, SETTING, AND PARTICIPANTS In 2 prospective, consecutively enrolled, broad-spectrum, double-blind studies, we developed an objective eye-tracking–based index test for children aged 16 to 30 months, compared its performance with best-practice reference standard diagnosis of autism (discovery study), and then replicated findings in an independent sample (replication study). Discovery and replication studies were conducted in specialty centers for autism diagnosis and treatment. Reference standard diagnoses were made using best-practice standardized protocols by specialists blind to eye-tracking results. Eye-tracking tests were administered by staff blind to clinical results. Children were enrolled from April 27, 2013, until September 26, 2017. Data were analyzed from March 28, 2018, to January 3, 2019.

MAIN OUTCOMES AND MEASURES Prespecified primary end points were the sensitivity and specificity of the eye-tracking–based index test compared with the reference standard. Prespecified secondary end points measured convergent validity between eye-tracking–based indices and reference standard assessments of social disability, verbal ability, and nonverbal ability.

RESULTS Data were collected from 1089 children: 719 children (mean [SD] age, 22.4 [3.6] months) in the discovery study, and 370 children (mean [SD] age, 25.4 [6.0] months) in the replication study. In discovery, 224 (31.2%) were female and 495 (68.8%) male; in replication, 120 (32.4%) were female and 250 (67.6%) male. Based on reference standard expert clinical diagnosis, there were 386 participants (53.7%) with nonautism diagnoses and 333 (46.3%) with autism diagnoses in discovery, and 184 participants (49.7%) with nonautism diagnoses and 186 (50.3%) with autism diagnoses in replication. In the discovery study, the area under the receiver operating characteristic curve was 0.90 (95% CI, 0.88-0.92), sensitivity was 81.9% (95% CI, 77.3%-85.7%), and specificity was 89.9% (95% CI, 86.4%-92.5%). In the replication study, the area under the receiver operating characteristic curve was 0.89 (95% CI, 0.86-0.93), sensitivity was 80.6% (95% CI, 74.1%-85.7%), and specificity was 82.3% (95% CI, 76.1%-87.2%). Eye-tracking test results correlated with expert clinical assessments of children's individual levels of ability, explaining 68.6% (95% CI, 58.3%-78.6%), 63.4% (95% CI, 47.9%-79.2%), and 49.0% (95% CI, 33.8%-65.4%) of variance in reference standard assessments of social disability, verbal ability, and nonverbal cognitive ability, respectively.

(continued)

Key Points

Question Can objective measurements of social visual engagement be developed and replicated to aid in early diagnosis and assessment of autism before age 3 years?

Findings In 2 prospective double-blind studies of diagnostic performance in 1089 children aged 16 to 30 months, 719 in discovery and 370 in replication, eye-tracking–based measurements of social visual engagement relative to expert clinical diagnosis had area under the receiver operating characteristic curve of 0.90, sensitivity of 81.9%, and specificity of 89.9% in discovery; and area under the curve of 0.89, sensitivity of 80.6%, and specificity of 82.3% in replication.

Meaning These results offer the prospect of an objective biomarker to aid in autism diagnosis and assessment.

+ Supplemental content

+ **Related article** at jama.com

Author affiliations and article information are listed at the end of this article.

Abstract (continued)

CONCLUSIONS AND RELEVANCE In two diagnostic studies of children younger than 3 years, objective eye-tracking–based measurements of social visual engagement quantified diagnostic status as well as individual levels of social disability, verbal ability, and nonverbal ability in autism. These findings suggest that objective measurements of social visual engagement can be used to aid in autism diagnosis and assessment.

JAMA Network Open. 2023;6(9):e2330145. doi:10.1001/jamanetworkopen.2023.30145

Introduction

Approximately 1 in 36 US children is affected by autism.[1] Thirty percent of parents of children with autism had concerns for their child's development before age 12 months, 50% of parents had concerns by age 18 months, and 80% had concerns by age 2 years.[2-4] Despite these early concerns and the manifest behaviors that elicited these concerns, the median age of US diagnosis remains delayed until the age of 4 to 5 years.[5,6] The age of diagnosis is even later among those who lack resources or lack access to expert clinicians: diagnoses for US racial minority families, families with low income, and families residing in rural areas lag further.[1,6-9]

The goal of diagnosis in autism is to facilitate timely and targeted support to help a child and family as needed. To that end, there may be an important role for new tools and objective biomarkers that can accurately and efficiently aid in diagnosing children as well as aid in quantifying individual strengths and vulnerabilities.[10] Such tools could enhance health care system capacity and help facilitate timely access to individually appropriate services.[10,11]

In current best practice, autism is diagnosed behaviorally by symptomatic deficits in social interaction and communication and by the presence of restricted and repetitive behaviors.[12] Current gold (reference) standard[13] diagnostic instruments are standardized validated assessments that measure the presence of autistic social disability through both behavioral observation and parent interview.[14,15] Best-practice guidelines also call for standardized assessments of a child's cognitive and language skills.[16]

Unfortunately, there are often long wait lists to access expert clinicians using gold standard instruments (a situation now described as a crisis)[17] and community use of gold standard instruments is limited.[18,19] Consequently, many children experience delayed diagnosis, and most receive diagnostic labels without receiving comprehensive evaluations and standardized assessments.[18,19]

In the present studies, we tested the performance of eye-tracking–based measurements of social visual engagement to accurately predict autism diagnoses and to objectively quantify individual levels of social disability, verbal ability, and nonverbal cognitive ability. In primary analyses, we measured the sensitivity and specificity of eye-tracking assays in comparison with clinician best-estimate diagnosis by expert clinicians. In secondary analyses, we quantified convergent validity between eye-tracking–based indices of social disability, verbal ability, and nonverbal cognitive ability in comparison with standardized assessments thereof as administered by expert clinicians.[20,21]

The present studies build on prior research using eye tracking to quantify social visual engagement, defined as how children look at and learn from their surrounding social environment. The prior research found that social visual engagement is strongly influenced by individual genetic variation (with monozygotic twin-twin concordance of approximately 0.9),[22] is highly biologically conserved,[23] and is atypical in autism[24,25] from very early ages in development (ie, 2-6 months).[26] Here we test the hypothesis that measurements of social visual engagement collected via eye tracking can serve as a robust biomarker to enable early diagnosis and assessment of autism.

Methods

The goal of the current studies was to evaluate performance of eye-tracking–based assays to accurately assess categorical presence of autism and to measure dimensional levels of ability or disability. In design, terminology, and reporting, this research followed the Standards for Reporting of Diagnostic Accuracy (STARD) guidelines,[27-29] with eye-tracking assays referred to as the "index test," and expert clinical diagnosis using standardized assessments referred to as the "reference standard." Two observational studies were conducted: a discovery study that was used to develop the data collection tool and algorithms, and a replication study that was used to test performance in an independent sample. Children were consecutively enrolled from April 27, 2013, until September 26, 2017. Data were analyzed from March 28, 2018, to January 3, 2019. The research protocol was approved by the institutional review boards of Emory University and Washington University in St Louis. Written informed consent was obtained from all parents or legal guardians.

Study Design

To eliminate or minimize design-related bias (as highlighted by Lijmer et al[30]), data were collected prospectively; participants were enrolled consecutively; enrolled participants had a broad spectrum of case presentation (spanning the full spectrum of symptom severity and absence of symptoms); clinical assessments were blind to eye tracking, eye tracking was blind to clinical assessments; and the index test and reference standard diagnosis were performed with all participants. Only the results of best-practice standardized assessments and reference standard clinician best-estimate diagnosis[31] were used clinically or communicated to parents. In this way, best-practice standard of care was maintained for all participants, blind to eye-tracking results; neither a child's parents nor expert clinical staff were informed of a child's eye-tracking results.

Participants

A total of 1089 children participated: 719 participated in the discovery study, and 370 participated in the replication study (**Figure 1**). Eligible participants were identified on the basis of chronological age and were recruited through placement of advertising materials in local media, specialty clinics, and pediatric practices. The studies were designed to develop and test a tool to aid in the diagnosis and assessment of autism, not to test the tool's utility as a screening instrument. To that end, children for whom there were concerns about autism were recruited (ie, children typically evaluated in specialized clinics for the diagnosis of autism). To be eligible for participation, children could not have clinically meaningful hearing or visual impairments (eg, congenital deafness, blindness, or nystagmus); could not have previously diagnosed genetic conditions associated with autism-related symptoms (eg, not known to have fragile X or Rett syndromes); had to be generally healthy at the time of testing, with no acute illness; had to be either born at or after 37 weeks' gestational age (discovery study) or born at or after 32 weeks' gestational age (replication study); and had to be either between the ages of 16 and 30 months (discovery study) or between the ages of 16 and 45 months (replication study). In both studies, the age of participants was guided by future intended use (ie, to align in time with ages that would ideally enable diagnosed children to be referred to early intervention before age 36 months). In the replication study, to test performance among a broader range of children, increased prematurity at birth and older age at enrollment was allowed. For the purpose of sample characterization, patient demographic data (including race, ethnicity, and maternal educational level) were collected by parents' selection of fixed categories. Race and ethnicity data were collected to enable evaluation of whether test performance varied based on these characteristics. Further details are available in the eMethods in Supplement 1.

Reference Standard Diagnosis

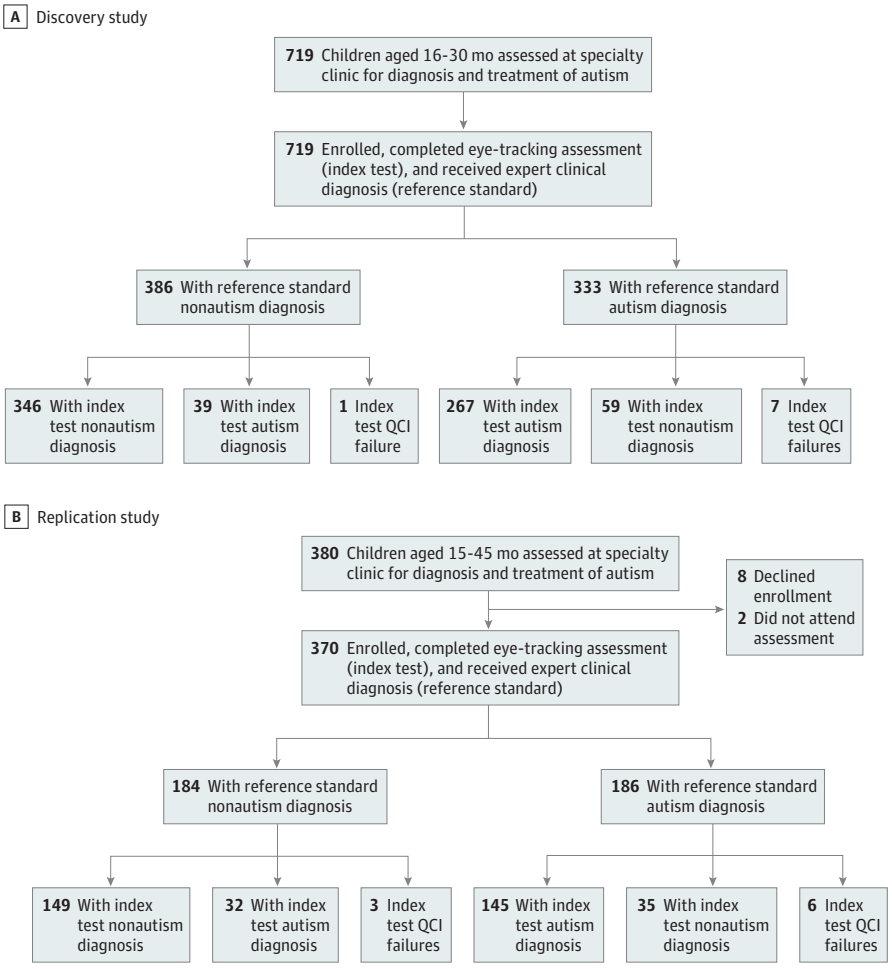
Reference standard diagnosis consisted of clinician best-estimate diagnosis[32-35] by experienced licensed clinicians using standardized diagnostic protocols and developmental assessments.[20]

Reference standard diagnosis was assigned based on all available clinical information, including developmental assessments as well as medical and developmental history gathered in clinical interviews. At young ages, clinician best-estimate diagnosis (ie, experienced clinicians' judgments using the totality of information available) is a more stable predictor of later diagnosis than strict reliance on cutoff scores.[20,36] For example, while scores on the Autism Diagnostic Observation Schedule, second edition (ADOS-2), may vary during the first 2 to 3 years of life, clinician best estimate is more stable.[32,36] The standardized diagnostic protocol was sequential so that a child's developmental history and scores on screening tests dictated subsequent assessments. A complete description of this protocol is available in the Reference Standard Diagnostic Assessment Procedures subsection of the eMethods in Supplement 1.

Index Test Eye Tracking

For index test measurements of social visual engagement, eye-tracking data were collected while participants watched video scenes of social interaction (examples are provided in eFigure 1 in Supplement 1). Fourteen video scenes were presented, each with a mean (SD) duration of approximately 54.0 (21.5) seconds (range, 21.7 seconds to 1 minute, 29.7 seconds; sum, 12 minutes, 35.5 seconds). Experimental procedures, data collection, and data processing were performed as described in prior studies[22] and in the Experimental Procedures and Data Collection subsection of the eMethods in Supplement 1. Data collection for the discovery study was performed in an academic

Figure 1. Participant Enrollment and Outcomes for Comparing Objective Measurements of Social Visual Engagement With Expert Clinical Diagnosis of Autism in Discovery and Replication Studies



A, Participant flow for the discovery study. B, Participant flow for the replication study. In both studies, during a single visit at the testing site, enrolled participants received expert clinical diagnosis using standardized assessments (reference standard diagnosis) as well as eye-tracking–based measurement of social visual engagement (index test). Index test quality control indicator (QCIs) failures occurred when participants' data failed to meet automated preset data QCIs (additional details are available in eTables 2 and 3 in Supplement 1).

medical center laboratory setting. Data collection for the replication study was performed in both an academic medical center laboratory setting and a community clinic using a standalone investigational eye-tracking device (eFigure 2 in Supplement 1). Eye-tracking data were collected using near-infrared video-based measurements of eye movements using specialized cameras and hardware (additional details are provided in the Experimental Procedures and Data Collection subsection of the eMethods in Supplement 1).

All collected data underwent automated quality control analyses measuring calibration accuracy, integrity of eye movements, duration of data collected, and time spent fixating on video scenes. Data that met or exceeded predefined automated static quality control thresholds proceeded to analysis (Quality Control Indicators subsection of the eMethods in Supplement 1). All steps in data processing and analysis were automated, with no manual human review or analysis required.

Time-varying kernel density estimation was used to quantify social visual engagement[37] (eFigure 3 in Supplement 1). Probability density functions of visual fixation and scanning were calculated during each moment of collected eye-tracking data (eFigure 4 in Supplement 1). Moments in time when the majority of participants with nonautism diagnoses in the discovery study fixated on approximately the same location(s) at the same moments at levels greater than expected by chance were identified by permutation testing.[38] Discovery study data were then mined to identify time points when the majority of participants with autism fixated on alternate locations (defining a classification index). Data were also mined to identify time points when autism and nonautism discovery study data were correlated with measurements of (1) social disability (correlated with ADOS-2 total scores), (2) verbal ability (correlated with verbal age-equivalent scores from the Mullen Scales of Early Learning, hereinafter, Mullen), or (3) nonverbal cognitive ability (correlated with visual reception age-equivalent scores from the Mullen). Data mining for these associations thereby defined 3 indices of individual variability in levels of disability and ability. Further details are provided in the Data Processing subsection of the eMethods in Supplement 1. Discovery study results were tested by leave-one-out cross validation,[39] with each participant tested as an independent comparison relative to the rest of the sample. All parameters were fixed and then tested again in the independent replication study.

Statistical Analysis

Primary effectiveness analyses were planned as a comparison between the eye-tracking index test results and the reference standard diagnosis results (either autism or nonautism). Sensitivity and specificity were calculated according to standard practice: sensitivity was calculated as the proportion of participants with reference standard autism diagnoses who had eye-tracking results that also indicated autism; specificity was calculated as the proportion of participants with reference standard nonautism diagnoses who had eye-tracking results that also indicated nonautism. The test positivity threshold was derived in the discovery study using the Youden index[40]; the threshold was then fixed for testing in the independent replication study. Receiver operating characteristic curves, area under the curve, positive predictive value (PPV), negative predictive value (NPV), and accuracy of the eye-tracking index test were also calculated, all with 95% CIs.[41] Primary end point analyses were tested at a 1-sided significance level of $\alpha = .025$.

Secondary effectiveness analyses were planned as measurements of correlation between eye-tracking–based severity indices and their respective expert clinician–administered reference standard assessments, including the ADOS-2 total score for social disability, the mean of Mullen receptive and expressive language age-equivalent scores for verbal ability, and the Mullen visual reception age-equivalent score for nonverbal ability. For social disability, the correlation was expected to be negative because higher scores on the ADOS-2 denote greater social disability, whereas for the eye-tracking test, lower scores denote greater social disability. Deming regression[42,43] was used to quantify the relationships between eye-tracking–based indices and their respective reference standards. Standard regression diagnostics (including Cook distance and

difference-in-fits),[44,45] Pearson correlation coefficients, and adjusted R^2 coefficients[46-48] together with 95% CIs were calculated. Secondary outcome analyses were tested at a 1-sided significance level of α = .025. Data analyses were performed in Matlab R2016a (Mathworks, Inc).

Results

Participants

A total of 719 children (mean [SD] age, 22.4 [3.6] months; 224 [31.2%] female and 495 [68.8%] male) were enrolled in the discovery study. A total of 370 children (mean [SD] age, 25.4 [6.0] months; 120 [32.4%] female and 250 [67.6%] male) were enrolled in the replication study. Based on reference standard diagnosis, the discovery study comprised 386 participants (53.7%) with nonautism diagnoses and 333 (46.3%) with autism diagnoses, while the replication study comprised 184 participants (49.7%) with nonautism diagnoses and 186 (50.3%) with autism diagnoses.

Participant characteristics and demographic data are shown in the **Table**. In both studies, participants with autism had higher ADOS-2 domain and total scores (all $t > 24.0$, all $P < .001$). Participants with autism also had lower Mullen verbal age-equivalent scores (both $t > 7.4$, $P < .001$) and lower Mullen nonverbal age-equivalent scores (both $t > 5.1$, $P < .001$). The ADOS-2 scores in both studies indicated that participants with autism represented the full spectrum of autism symptom severity. Likewise, the Mullen scores in both studies indicated that participants with nonautism and autism diagnoses represented a broad range of verbal and nonverbal abilities, extending from substantially delayed to age-appropriate to advanced abilities. In each study, mean (SD) age of the sample with autism diagnoses was significantly older than the sample with nonautism diagnoses (discovery: 23.1 [3.7] months vs 21.7 [3.4] months; $t = 5.2$, $P < .001$; replication: 28.1 [5.8] months vs 22.7 [4.9] months; $t = 9.7$, $P < .001$). Sex differences were as expected,[49] with a higher number of boys diagnosed with autism in both studies (both $\chi^2 > 16.6$; $P < .001$).

Quality Control Indicators

Average calibration accuracy was within 1 degree of visual angle and did not differ significantly between diagnostic groups or study samples (eFigure 5 in Supplement 1). There were no significant between-group differences in duration of data collected (discovery: $t = 1.48$; $P = .14$; replication: $t = 0.81$, $P = .42$). Children with nonautism diagnoses did fixate (discovery: $t = 4.97$, $P < .001$; replication: $t = 8.51$, $P < .001$) and saccade (discovery: $t = 6.75$, $P < .001$; replication: $t = 8.10$, $P < .001$) significantly more, and blink less (discovery: $t = 4.61$, $P < .001$; replication: $t = 4.08$, $P < .001$), than children with autism, which was consistent with expected diagnostic differences in attention to and engagement with social cues in the environment[50] that have been commonly noted in autism.[36]

Primary End Points: Estimates of Diagnostic Accuracy

Prespecified primary end point analyses measured the diagnostic accuracy of eye-tracking–based index test results in comparison with reference standard diagnosis. Results are shown in **Figure 2** as receiver operating characteristic curves (panels A and B) and diagnostic cross-tabulations with performance measure estimates (panels C and D); underlying score distributions are plotted in eFigure 6 in Supplement 1. Index test performance had area under the curve statistics equal to 0.90 (95% CI, 0.88-0.92) in the discovery study and 0.89 (95% CI, 0.86-0.93) in the replication study. The test positivity threshold in the discovery study was selected to match the Youden index (represented by the diamond in Figure 2A). After discovery study determination, the test positivity threshold was fixed and applied in the replication study. Achieved sensitivity and specificity in the replication study are shown in Figure 2B (represented by the diamond, which corresponds to the cross-tabulation results and performance measure estimates in Figure 2D). Eye-tracking–based index test results predicted expert clinician reference standard diagnosis with sensitivity equal to 81.9% (95% CI, 77.3%-85.7%) and specificity equal to 89.9% (95% CI, 86.4%-92.5%) in the discovery study, and sensitivity equal to 80.6% (95% CI, 74.1%-85.7%) and specificity equal to 82.3% (95% CI,

Table. Participant Characteristics

Reference standard diagnosis	Participants, No. (%)[a]			
	Discovery study (n = 719)		Replication study (n = 370)	
	Nonautism (n = 386)	Autism (n = 333)	Nonautism (n = 184)	Autism (n = 186)
Age, mo				
Mean (SD)	21.7 (3.4)	23.1 (3.7)	22.7 (4.9)	28.1 (5.8)
Median (IQR)	23 (18-24)	24 (20-26)	21 (21-25)	28 (24-31)
Sex				
Female	154 (39.9)	70 (21.0)	78 (42.4)	42 (22.6)
Male	232 (60.1)	263 (79.0)	106 (57.6)	144 (77.4)
Race				
Asian	5 (1.3)	10 (3.0)	1 (0.5)	23 (12.4)
Black or African American	21 (5.4)	67 (20.1)	22 (12.0)	38 (20.4)
Native Hawaiian or Pacific Islander	4 (1.0)	3 (0.9)	2 (1.1)	0
White	281 (72.8)	179 (53.8)	139 (75.6)	106 (57.0)
>1 Race	28 (7.3)	41 (12.3)	19 (10.3)	16 (8.6)
Prefer not to answer or unknown	47 (12.2)	33 (9.9)	1 (0.5)	3 (1.6)
Ethnicity				
Hispanic	24 (6.2)	23 (6.9)	12 (6.5)	20 (10.8)
Non-Hispanic	309 (80.1)	268 (80.5)	166 (90.2)	154 (82.8)
Prefer not to answer or unknown	53 (13.7)	42 (12.6)	6 (3.3)	12 (6.5)
Household income, $				
≤20 000	5 (1.3)	13 (3.9)	14 (7.6)	2 (1.1)
20 001-40 000	17 (4.4)	29 (8.7)	21 (11.4)	16 (8.6)
40 001-60 000	32 (8.3)	48 (14.4)	35 (19.0)	42 (22.6)
60 001-80 000	37 (9.6)	51 (15.3)	31 (16.8)	57 (30.6)
80 001-100 000	51 (13.2)	33 (9.9)	29 (15.8)	29 (15.6)
100 001-125 000	56 (14.5)	26 (7.8)	21 (11.4)	17 (9.1)
125 001-150 000	26 (6.7)	13 (3.9)	10 (5.4)	11 (5.9)
150 001-200 000	40 (10.4)	12 (3.6)	8 (4.3)	5 (2.7)
≥200 000	33 (8.5)	6 (1.8)	5 (2.7)	0
Prefer not to answer or unknown	89 (23.1)	102 (30.6)	10 (5.4)	7 (3.8)
Maternal educational level				
Some high school	0	4 (1.2)	1 (0.5)	4 (2.2)
High school diploma or GED certificate	8 (2.1)	20 (6.0)	19 (10.3)	29 (15.6)
Some college	15 (3.9)	55 (16.5)	31 (16.8)	21 (11.3)
Vocational school certificate	1 (0.3)	12 (3.6)	6 (3.3)	2 (1.1)
Associate's degree	4 (1.0)	14 (4.2)	14 (7.6)	13 (7.0)
Bachelor's degree	114 (29.5)	98 (29.4)	74 (40.2)	76 (40.9)
Master's degree	135 (35.0)	55 (16.5)	31 (16.8)	29 (15.6)
Professional or doctoral degree	51 (13.2)	14 (4.2)	5 (2.7)	8 (4.3)
Prefer not to answer or unknown	58 (15.0)	61 (18.3)	3 (1.6)	4 (2.2)

(continued)

Table. Participant Characteristics (continued)

Reference standard diagnosis	Participants, No. (%)[a]			
	Discovery study (n = 719)		Replication study (n = 370)	
	Nonautism (n = 386)	Autism (n = 333)	Nonautism (n = 184)	Autism (n = 186)
ADOS-2[b]				
Social affect score				
Mean (SD)	2.3 (2.3)	13.6 (4.1)	3.1 (2.6)	13.8 (4.4)
Median (IQR)	2 (1-3)	14 (10-17)	3 (1-5)	14 (10-17)
Restricted and repetitive behavior score				
Mean (SD)	1.0 (0.9)	4.3 (1.8)	2.4 (1.6)	5.6 (1.4)
Median (IQR)	1 (0-2)	4 (3-6)	2 (1-4)	6 (5-7)
Total score				
Mean (SD)	3.3 (2.6)	17.9 (5.1)	5.5 (3.2)	19.4 (5.0)
Median (IQR)	3 (2-5)	18 (14-22)	5 (3-7)	20 (15-24)
Mullen Scales of Early Learning[c]				
Verbal ability age-equivalent score				
Mean (SD)	24.2 (5.6)	13.0 (6.2)	23.1 (8.0)	14.8 (7.7)
Median (IQR)	24 (20-28)	12 (8-16)	23 (16-28)	12 (10-18)
Nonverbal ability age-equivalent score				
Mean (SD)	24.8 (6.1)	19.0 (5.2)	27.3 (9.8)	20.7 (6.8)
Median (IQR)	24 (20-29)	19 (16-23)	25 (19-32)	20 (16-24)

Abbreviations: ADOS-2, Autism Diagnostic Observation Schedule, second edition; GED, general educational development.

[a] Category percentages may sum to less than or greater than 100 due to rounding.

[b] Includes 564 children (329 with autism diagnoses and 235 with nonautism diagnoses) in the discovery study and 255 children (186 with autism diagnoses and 69 with nonautism diagnoses) in the replication study. The standardized diagnostic protocol was sequential so that a child's developmental history and scores on screenings dictated subsequent assessments. Further details provided in Reference Standard Diagnostic Assessment Procedures subsection of the eMethods in Supplement 1.

[c] Includes 620 children (322 with autism diagnoses and 298 with nonautism diagnoses; 10 with missing nonverbal scores) in the discovery study and 251 children (183 with autism diagnoses and 68 with nonautism diagnoses) in the replication study. The verbal ability age-equivalent score, in months, was calculated as the mean of expressive and receptive language age-equivalent scores. The nonverbal ability age-equivalent score, in months, was calculated as the visual reception age-equivalent score. Further details provided in Reference Standard Diagnostic Assessment Procedures subsection of the eMethods in Supplement 1.

76.1%-87.2%) in the replication study. Sensitivity, specificity, PPV, NPV, and accuracy did not differ significantly by sex (all with overlapping 95% CIs of performance estimates). Additional information regarding clinical outcomes is provided in the Clinical Outcomes of False Positives and Negatives subsection in eResults in Supplement 1.

Secondary End Points: Measurements of Symptom Severity

Prespecified secondary end point analyses measured the strength of association between eye-tracking–based indices and reference standard behavioral assessments of social disability, verbal ability, and nonverbal ability. Results are shown in **Figure 3** as scatter plots with Deming regression fitted functions, Pearson *R* values, and adjusted R^2 coefficients of determination (also summarized in eTable 1 in Supplement 1).

In the discovery study (Figure 3A-C), the correlation between the index test social disability index and the ADOS-2 total score was −0.74 (95% CI, −0.78 to −0.70), the correlation between the index test verbal ability index and the Mullen verbal age-equivalent score was 0.71 (95% CI, 0.67-0.75), and the correlation between the nonverbal ability index and the Mullen nonverbal age-equivalent score was 0.66 (95% CI, 0.61-0.70). In the replication study (Figure 3D-F), the correlation between the index test social disability index and the ADOS-2 total score was −0.72 (95% CI, −0.78 to −0.65), the correlation between the index test verbal ability index and the Mullen verbal age-equivalent score was 0.59 (95% CI, 0.50-0.77), and the correlation between the index test nonverbal ability index and the Mullen nonverbal age-equivalent score was 0.53 (95% CI, 0.43-0.62).

From the replication study, adjusted for reference standard measurement error, the eye-tracking–based social disability index accounted for 68.6% (adjusted R^2 = 0.69; 95% CI, 0.58-0.79) of variance in ADOS-2 total scores. The verbal ability index accounted for 63.4% (adjusted R^2 = 0.63; 95% CI, 0.48-0.79) of variance in Mullen verbal age-equivalent scores. The nonverbal

ability index accounted for 49.0% (adjusted R^2 = 0.49; 95% CI, 0.34-0.65) of variance in Mullen nonverbal age-equivalent scores.

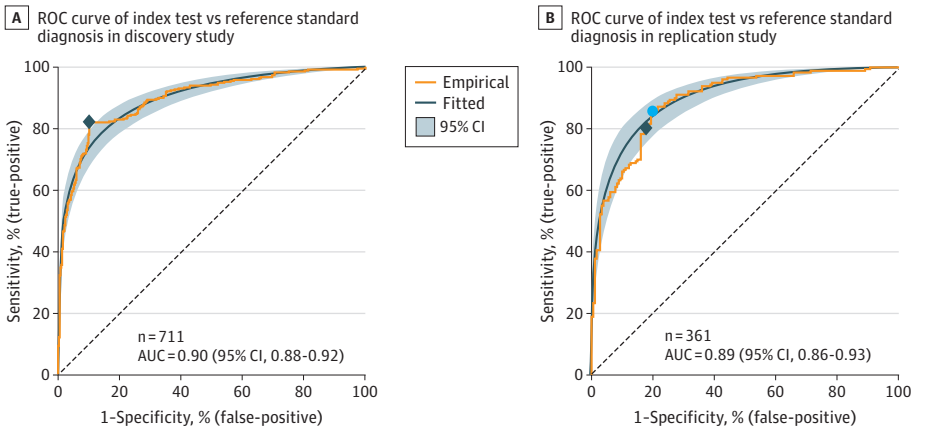
In all comparisons, the strength of association between the eye-tracking–based indices and their respective expert clinician–administered assessments was high (R > 0.5), suggesting strong convergent validity between index and reference standard measures for social disability, verbal ability, and nonverbal ability. There were no significant differences in strength of association by sex. Participants with eye-tracking quality control indicator failures (8 of 719 in the discovery study and 9 of 370 in the replication study), with no results returned, are described further in eTables 2 and 3 in Supplement 1. Additional details and comparisons are available in the eResults in Supplement 1.

Discussion

In 2 prospective double-blind diagnostic studies, the first a discovery study and the second a replication study, 1089 children were tested to measure the diagnostic performance of index test measurements of social visual engagement relative to reference standard expert clinical diagnosis of autism. Children aged 16 to 30 months (discovery study) and 16 to 45 months (replication study) were assessed by expert clinicians to test whether measurements of social visual engagement could accurately predict categorical diagnosis as well as dimensional levels of social disability, verbal ability, and nonverbal ability.

The results, reported in accordance with the STARD initiative,[27-29] found that measurements of social visual engagement had 81.9% sensitivity and 89.9% specificity relative to expert clinical diagnosis of autism in the discovery study and 80.6% sensitivity and 82.3% specificity in the

Figure 2. Diagnostic Performance Comparing Eye-Tracking–Based Measurement of Social Visual Engagement (Index Test) With Expert Clinical Diagnosis of Autism (Reference Standard)



C Cross-tabulation of index test vs reference standard diagnosis in discovery study

		Reference standard diagnosis: Autism	Nonautism	
Index test	Autism	267	39	306
	Nonautism	59	346	405
		326	385	711

	%	95% CI
Sensitivity	81.9	(77.3-85.7)
Specificity	89.9	(86.4-92.5)
PPV	87.3	(83.0-90.6)
NPV	85.4	(81.6-88.6)
Accuracy	86.2	(83.5-88.6)

D Cross-tabulation of index test vs reference standard diagnosis in replication study

		Reference standard diagnosis: Autism	Nonautism	
Index test	Autism	145	32	177
	Nonautism	35	149	184
		180	181	361

	%	95% CI
Sensitivity	80.6	(74.1-85.7)
Specificity	82.3	(76.1-87.2)
PPV	81.9	(75.6-86.9)
NPV	81.0	(74.7-86.0)
Accuracy	81.4	(77.1-85.1)

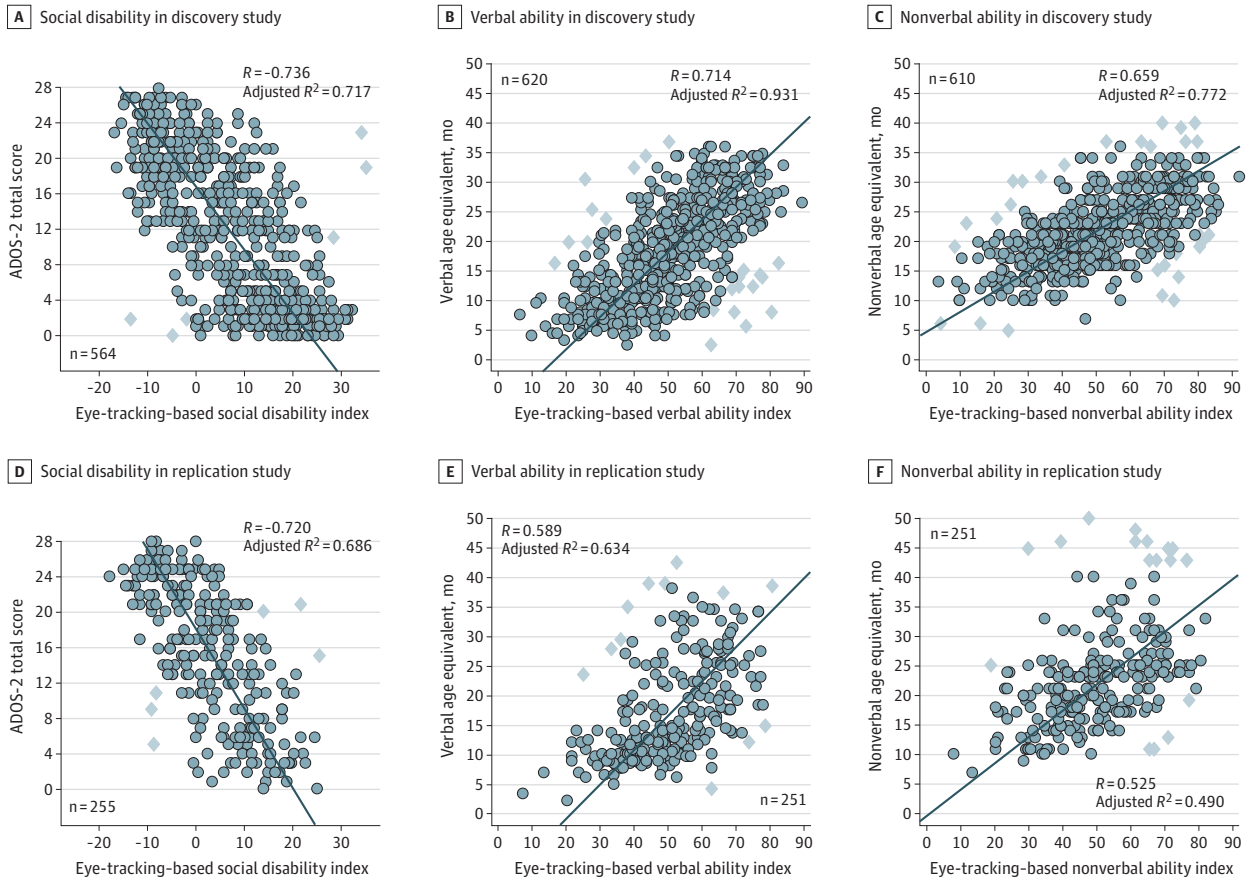
Performance among 711 children in the discovery study and 361 children in the replication study. A, The diamond represents the optimal test positivity threshold for the discovery study (Youden index). B, The test positivity threshold determined in the discovery study was fixed and applied independently in the replication study. The diamond represents the achieved sensitivity and specificity in the replication study using the test positivity threshold from the discovery study. The solid blue circle represents the post hoc theoretical optimal threshold. C, Tabulation corresponds to the diamond in panel A. D, Tabulation corresponds to the diamond in panel B. AUC indicates area under the curve; ROC, receiver operating characteristic. Negative predictive value (NPV) and positive predictive value (PPV) estimates reported here are calculated based on study sample prevalence.

replication study. Sensitivity, specificity, PPV, NPV, and accuracy did not differ significantly between the discovery and replication studies, suggesting robust and replicable performance. In addition, measurements of social visual engagement were also predictive of children's individual scores on gold standard behavioral assessments: measurements of social visual engagement effectively explained 68.6% of variance in individual levels of social disability (ADOS-2 total scores), 63.4% of variance in verbal ability (Mullen verbal age-equivalent scores), and 49.0% of variance in nonverbal cognitive ability (Mullen nonverbal age-equivalent scores).

These results suggest high convergent validity with reference standard assessments that otherwise require highly trained experts to spend multiple hours of assessment time per child. In contrast, for measurements of social visual engagement, biomarker data collection consisted of children watching videos (eFigure 1 in Supplement 1), with data collected on a standalone mobile eye-tracking device that was deployed in a clinic and operated by technicians with no required clinical or technical expertise (eFigure 2 in Supplement 1).

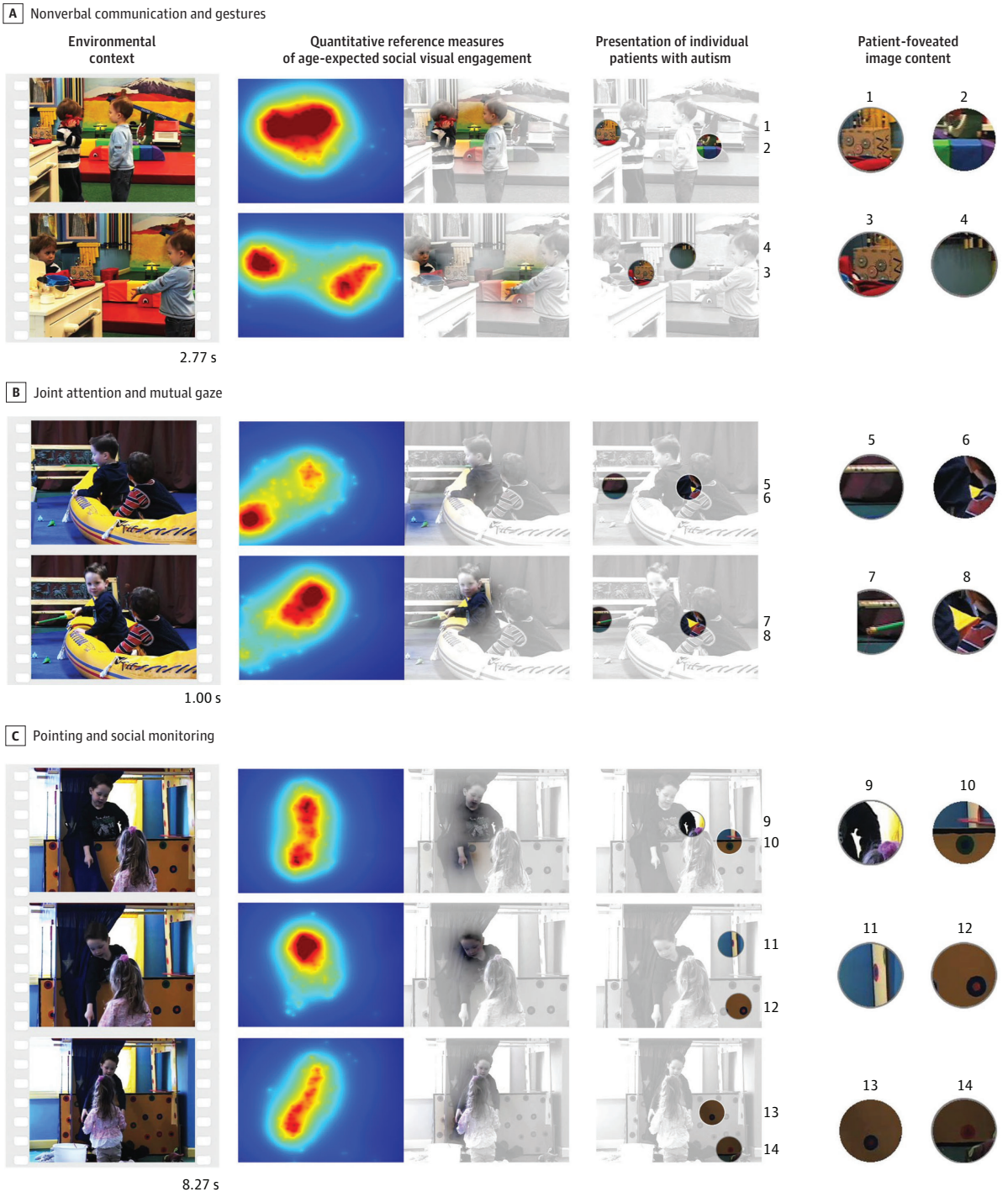
Once social visual engagement data are collected, although data processing and analysis are computationally intensive, they are also automated, deployed on cloud-based servers, and capable of returning a results report in less than 30 minutes. The index test is objective and quantitative and

Figure 3. Convergent Validity Between Eye-Tracking–Based Measurement of Social Visual Engagement (Index Test) and Expert Clinician–Administered Standardized Assessments of Social Disability, Verbal Ability, and Nonverbal Cognitive Ability



A, Discovery study correlation between eye-tracking–based indices of social disability versus children's total scores on the Autism Diagnostic Observation Schedule, second edition (ADOS-2). B, Discovery study correlation between eye-tracking–based indices of verbal ability versus children's verbal age equivalent scores as measured by the Mullen Scales of Early Learning (Mullen). C, Discovery study correlation between eye-tracking–based indices of nonverbal cognitive ability versus children's nonverbal age equivalent scores as measured by the Mullen. D, E, F, Replication study correlations between eye-tracking–based indices and reference standard assessments. In all scatterplots, circles represent individual data and diamonds represent regression outliers (bivariate outliers identified using Cook distance and difference-in-fits regression diagnostic assessment). The adjusted R^2 values were adjusted for measurement error variance of the reference standard (yielding percentage of reference standard nonerror variance explained by the index test). Additional information is provided in the Secondary End Point Analyses subsection of the eMethods in Supplement 1.

Figure 4. Performance-Based Measures of Children's Individual Strengths, Vulnerabilities, and Opportunities for Skill Development



Measurement of social visual engagement quantifies how a child engages with social and nonsocial cues occurring continuously within naturalistic environmental contexts (left column, shown as still frames from testing videos). In relation to those contexts, normative reference measures provide objective quantification of nonautism age-expected visual engagement (middle columns, shown as density distributions in both pseudocolor format and as color to grayscale fades overlaid on corresponding still frames). The age-expected reference measures can be used to measure and visualize patient comparisons, revealing individual strengths, vulnerabilities, and opportunities for skill building (right columns, sample patient data shown as overlaid circular apertures that encompass the portion of video foveated by each patient [each aperture spans the central 5.2 degrees of a patient's visual field]). Children with autism present as engaging with toys of interest (1, 3, 5, and 7), color and contrast cues (2, 6, 8, and 9), and objects and background elements not directly relevant to social context (4 and 10-14). Elapsed times at the bottom right of still frames highlight the rapidly changing nature of social interaction in which many hundreds of verbal and nonverbal communicative cues are presented, each eliciting age-expected patterns of engagement and offering corresponding opportunities for objective quantitative comparisons of patient behavior.

directly measures thousands of instances of children's behavior for comparison with age-expected norms (examples are shown in eFigure 4 in Supplement 1). Data processing and analysis to derive diagnostic classification and indices of symptom severity are entirely automated, requiring no special expertise or eye-tracking knowledge on the part of clinicians.

It is important to note that the test results derived from measurements of social visual engagement are not intended to replace clinicians with expertise in developmental disabilities; to the contrary, a tool like this could be used by expert clinicians to aid in accurately and efficiently diagnosing autism as well as quantifying children's strengths and vulnerabilities. Therefore, these results offer important opportunities to enhance health care system capacity and facilitate more rapid progress from the time of first concern to the start of individually appropriate services.[10] While empirically supported services have their own access challenges, those challenges are not a reason to delay diagnosis or to delay initiation of supports for children and families.[51]

Finally, the meaning of these measurements resides in what they quantify: repeated divergence from shared social experience with rapid accrual of atypical experience (**Figure 4**). Shared experience is the foundation for communication and social development. By quantifying the number, extent, and timing of divergence from shared experience, measurements of social visual engagement provide a transactional biomarker: direct objective measurements of a child's unique biology transacting with specific environmental contexts. Those transactions are the building blocks of learning and brain development.[52]

Recognizing the transactional nature of this developmental process is a reminder that the emergence of disability is itself transactional, driven by genetic liabilities but also by atypical learning experiences that are correlated with those liabilities.[53] Recognizing this provides 2 notable reasons for optimism. First, it reminds us that disability is a cocreation, a consequence of individual vulnerabilities transacting with particular environmental contexts, severely disabling in some contexts but less so or not at all in others.[54] Fostering early intervention approaches and contexts that embrace difference and diversity while also augmenting individual adaptive skills is important to reducing disability and optimizing outcomes for all. Second, knowing that the unfolding of disability is transactional means that it can be measured as such to (1) identify children in need of support; (2) monitor specific behaviors and contexts that may exacerbate or ameliorate disability over time; and (3), ideally, intervene more successfully and treat specific individual manifestations and vulnerabilities for disability.

Limitations

This study has several limitations. Clinical procedures were performed by a relatively small group of expert clinicians, and eye-tracking procedures were implemented under well-controlled laboratory conditions in the discovery study or with a single prototype standalone eye-tracking device in the replication study. The efforts in this study should be complemented by studies collecting reference standard and index test data at multiple different sites with multiple different clinical teams and eye-tracking devices.[55] The results of this study should also be complemented by data quantifying repeatability and reproducibility variance in eye-tracking–based measurements.[56] Previous studies[57,58] have also noted expert clinical uncertainty in the reference standard diagnosis of autism in some children. Uncertainty in the reference standard sets an upper limit on the performance measures of any comparison test (eFigure 3 in Jones et al[55]). In the current study, we did not prospectively track expert clinician certainty of diagnosis in all children. Consequently, we were unable to analyze the effects of clinician certainty in the discovery or replication studies. This limitation was improved in a subsequent multisite study.[55]

Conclusions

In 2 diagnostic studies of children aged 16 to 30 months with and without autism, objective measurements of social visual engagement were able to quantify diagnostic status and assess

individual levels of social disability, verbal ability, and nonverbal ability. These findings suggest that objective measurements of social visual engagement can be used to aid in autism diagnosis and assessment.

ARTICLE INFORMATION

Accepted for Publication: July 11, 2023.

Published: September 5, 2023. doi:10.1001/jamanetworkopen.2023.30145

Corresponding Author: Warren Jones, PhD, Marcus Autism Center, Children's Healthcare of Atlanta, 1920 Briarcliff Rd NE, Atlanta, GA 30329 (warren.jones@emory.edu).

Author Affiliations: Marcus Autism Center, Children's Healthcare of Atlanta, Atlanta, Georgia (Jones, Klaiman, Richardson, Lambha, Reid, Hamner, Beacham, Lewis, Paredes, Edwards, Shultz, Klin); Division of Autism and Related Disabilities, Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia (Jones, Klaiman, Richardson, Lambha, Paredes, Edwards, Shultz, Klin); Center for Translational Social Neuroscience, Emory University, Atlanta, Georgia (Jones, Klin); Department of Psychiatry, Washington University in St Louis School of Medicine, St Louis, Missouri (Marrus); Intellectual and Developmental Disabilities Research Center, Washington University in St Louis School of Medicine, St Louis, Missouri (Marrus, Constantino); Now with Department of Psychiatry, Emory University School of Medicine, Atlanta, Georgia (Constantino); Now with Division of Behavioral and Mental Health, Children's Healthcare of Atlanta, Atlanta, Georgia (Constantino).

Author Contributions: Dr Jones had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Jones, Klaiman, Lewis, Shultz, Klin.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Jones, Klaiman, Edwards, Klin.

Critical review of the manuscript for important intellectual content: Hamner, Paredes.

Statistical analysis: Jones, Edwards.

Obtained funding: Jones, Lewis, Edwards, Klin.

Administrative, technical, or material support: Jones, Klaiman, Richardson, Lambha, Hamner, Beacham, Lewis, Paredes, Edwards, Marrus, Constantino, Shultz, Klin.

Supervision: Jones, Klaiman, Lewis, Edwards, Marrus, Shultz, Klin.

Conflict of Interest Disclosures: Dr Jones reported receiving grants from the Georgia Research Alliance, the Joseph B. Whitehead Foundation, the Marcus Foundation, and the National Institute of Mental Health (NIMH) during the conduct of the study (all via his institution); being a scientific cofounder of and owning equity in EarliTec Diagnostics; holding patents licensed to EarliTec Diagnostics; and receiving personal fees for consulting from EarliTec Diagnostics and a lecture honorarium from Washington University in St Louis School of Medicine outside the submitted work. Dr Klaiman reported receiving grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), the Georgia Research Alliance, the Joseph B. Whitehead Foundation, the Marcus Foundation, and the NIMH (all via her institution) during the conduct of the study and personal fees from ABA Centers of America, Dekalb County School District, Cherokee County School District, Fulton County School District, Beaming Health, and EarliTec Diagnostics outside the submitted work. Dr Richardson reported receiving funding and/or equipment from the Georgia Research Alliance, the John B. Whitehead Foundation, and the Marcus Foundation (all via her institution) during the conduct of the study. Dr Lambha reported receiving funding and/or equipment from the Georgia Research Alliance, the Joseph B. Whitehead Foundation, the Marcus Foundation, the NICHD, and the NIMH (all via her institution) during the conduct of the study. Dr Beacham reported receiving funding and/or equipment from the John B. Whitehead Foundation and the Marcus Foundation during the conduct of the study. Mr Lewis reported receiving grants from the Georgia Research Alliance, the Joseph B. Whitehead Foundation, and the Marcus Foundation (all via his institution) during the conduct of the study; being a scientific cofounder of and owning equity in EarliTec Diagnostics; and holding patents licensed to EarliTec Diagnostics outside the submitted work. Mr Paredes reported receiving funding and/or equipment from the Joseph B. Whitehead Foundation and the Marcus Foundation (both via his institution) during the conduct of the study. Dr Marrus reported receiving grants from the NIMH and from Washington University in St Louis School of Medicine during the conduct of the study. Dr Constantino reported receiving grants from the NICHD during the conduct of the study and royalties from Western Psychological

Services outside the submitted work. Dr Klin reported receiving grants from the Georgia Research Alliance, the Joseph B. Whitehead Foundation, the Marcus Foundation, the NICHD, and the NIMH (all via his institution) during the conduct of the study; being a scientific cofounder of and owning equity in EarliTec Diagnostics; holding patents licensed to EarliTec Diagnostics; and receiving personal fees from the Alliance for Early Success, the Brazilian Society of Speech Therapy, EarliTec Diagnostics, the McKnight Endowment Fund for Neuroscience, the National Autism Conference, and Washington University in St Louis School of Medicine outside the submitted work. No other disclosures were reported.

Funding/Support: This study was supported by funding from the Marcus Foundation (Drs Jones and Klin), the Joseph B. Whitehead Foundation (Drs Jones and Klin), and the Georgia Research Alliance (Drs Jones and Klin and Mr Lewis); grants HD068479 and U54 HD087011 from the NICHD (Dr Constantino); and grants MH100029 (Drs Jones and Klin) and MH100019 (Dr Marrus) from the NIMH.

Role of the Funder/Sponsor: The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Sharing Statement: See Supplement 2.

Additional Contributions: We thank the families and children for their participation. We also thank the fellows of the Donald J. Cohen Fellowship in Developmental Social Neuroscience, the Simons Fellowship in Computational Neuroscience, and the Simons Fellowship in Design Engineering for providing help with data collection and processing at Marcus Autism Center, Children's Healthcare of Atlanta, Emory University. We thank Erika Mortenson, BA, Sayli Sant, MD, Teddi Gray, BA, Yi Zhang, MS, Laura Campbell, BA, and Leena Malik, BA, of Washington University in St Louis, for assisting with data collection and processing. We thank Steve Kovar, BFA, of Marcus Autism Center, Children's Healthcare of Atlanta, for providing help with building the standalone eye-tracking device used in the replication study and Robin Sifre, PhD, of EarliTec Diagnostics, for providing comments on the manuscript. The named individuals were not compensated outside of their normal salary.

REFERENCES

1. Maenner MJ, Warren Z, Williams AR, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. *MMWR Surveill Summ*. 2023;72(2):1-14. doi:10.15585/mmwr.ss7202a1

2. Chawarska K, Paul R, Klin A, Hannigen S, Dichtel LE, Volkmar F. Parental recognition of developmental problems in toddlers with autism spectrum disorders. *J Autism Dev Disord*. 2007;37(1):62-72. doi:10.1007/s10803-006-0330-8

3. Young RL, Brewer N, Pattison C. Parental identification of early behavioural abnormalities in children with autistic disorder. *Autism*. 2003;7(2):125-143. doi:10.1177/1362361303007002002

4. Wetherby AM, Brosnan-Maddox S, Peace V, Newton L. Validation of the Infant-Toddler Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. *Autism*. 2008;12(5):487-511. doi:10.1177/1362361308094501

5. Christensen DL, Baio J, Van Naarden Braun K, et al; Centers for Disease Control and Prevention. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2012. *MMWR Surveill Summ*. 2016;65(3):1-23. doi:10.15585/mmwr.ss6503a1

6. Shaw KA, McArthur D, Hughes MM, et al. Progress and disparities in early identification of autism spectrum disorder: Autism and Developmental Disabilities Monitoring Network, 2002-2016. *J Am Acad Child Adolesc Psychiatry*. 2022;61(7):905-914. doi:10.1016/j.jaac.2021.11.019

7. Mandell DS, Novak MM, Zubritsky CD. Factors associated with age of diagnosis among children with autism spectrum disorders. *Pediatrics*. 2005;116(6):1480-1486. doi:10.1542/peds.2005-0185

8. Mandell DS, Wiggins LD, Carpenter LA, et al. Racial/ethnic disparities in the identification of children with autism spectrum disorders. *Am J Public Health*. 2009;99(3):493-498. doi:10.2105/AJPH.2007.131243

9. Constantino JN, Abbacchi AM, Saulnier C, et al. Timing of the diagnosis of autism in African American children. *Pediatrics*. 2020;146(3):e20193629. doi:10.1542/peds.2019-3629

10. Zwaigenbaum L, Warren Z. Commentary: embracing innovation is necessary to improve assessment and care for individuals with ASD: a reflection on Kanne and Bishop (2020). *J Child Psychol Psychiatry*. 2021;62(2):143-145. doi:10.1111/jcpp.13271

11. MacLachlan M. Commentary: challenges and opportunities in autism assessment—a commentary on Kanne and Bishop (2020). *J Child Psychol Psychiatry*. 2021;62(2):146-148. doi:10.1111/jcpp.13360

12. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

13. Falkmer T, Anderson K, Falkmer M, Horlin C. Diagnostic procedures in autism spectrum disorders: a systematic literature review. *Eur Child Adolesc Psychiatry*. 2013;22(6):329-340. doi:10.1007/s00787-013-0375-0

14. Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Bishop S. *Autism Diagnostic Observation Schedule*. 2nd ed. Western Psychological Services; 2012.

15. Rutter M, Le Couteur A, Lord C. *ADI-R: Autism Diagnostic Interview—Revised (ADI-R)*. Western Psychological Services; 2003.

16. Hyman SL, Levy SE, Myers SM; Council on Children With Disabilities, Section on Developmental and Behavioral Pediatrics. Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics*. 2020;145(1):e20193447. doi:10.1542/peds.2019-3447

17. Kanne SM, Bishop SL. Editorial perspective: the autism waitlist crisis and remembering what families need. *J Child Psychol Psychiatry*. 2021;62(2):140-142. doi:10.1111/jcpp.13254

18. Wiggins LD, Baio J, Rice C. Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. *J Dev Behav Pediatr*. 2006;27(2)(suppl):S79-S87. doi:10.1097/00004703-200604002-00005

19. Rice CE, Baio J, Morrier MJ, et al. Changes in diagnostic testing practice for the autism spectrum disorders (ASDs) in four US populations. Paper presented at: International Meeting for Autism Research; May 9, 2009; Chicago, IL. Accessed December 6, 2022. https://imfar.confex.com/imfar/2009/webprogram/Paper3602.html

20. Kim SH, Lord C. Combining information from multiple sources for the diagnosis of autism spectrum disorders for toddlers and young preschoolers from 12 to 47 months of age. *J Child Psychol Psychiatry*. 2012;53(2):143-151. doi:10.1111/j.1469-7610.2011.02458.x

21. Mullen EM. *Mullen Scales of Early Learning*. American Guidance Service; 1995.

22. Constantino JN, Kennon-McGill S, Weichselbaum C, et al. Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature*. 2017;547(7663):340-344. doi:10.1038/nature22999

23. Wang A, Payne C, Moss S, Jones WR, Bachevalier J. Early developmental changes in visual social engagement in infant rhesus monkeys. *Dev Cogn Neurosci*. 2020;43:100778. doi:10.1016/j.dcn.2020.100778

24. Klin A, Jones W, Schultz R, Volkmar F, Cohen D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch Gen Psychiatry*. 2002;59(9):809-816. doi:10.1001/archpsyc.59.9.809

25. Klin A, Lin DJ, Gorrindo P, Ramsay G, Jones W. Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*. 2009;459(7244):257-261. doi:10.1038/nature07868

26. Jones W, Klin A. Attention to eyes is present but in decline in 2-6–month-old infants later diagnosed with autism. *Nature*. 2013;504(7480):427-431. doi:10.1038/nature12715

27. Bossuyt PM, Reitsma JB, Bruns DE, et al; Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ*. 2003;326(7379):41-44. doi:10.1136/bmj.326.7379.41

28. Korevaar DA, Cohen JF, Reitsma JB, et al. Updating standards for reporting diagnostic accuracy: the development of STARD 2015. *Res Integr Peer Rev*. 2016;1:7. doi:10.1186/s41073-016-0014-7

29. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799. doi:10.1136/bmjopen-2016-012799

30. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282(11):1061-1066. doi:10.1001/jama.282.11.1061

31. Leckman JF, Sholomskas D, Thompson WD, Belanger A, Weissman MM. Best estimate of lifetime psychiatric diagnosis: a methodological study. *Arch Gen Psychiatry*. 1982;39(8):879-883. doi:10.1001/archpsyc.1982.04290080001001

32. Chawarska K, Klin A, Paul R, Macari S, Volkmar F. A prospective study of toddlers with ASD: short-term diagnostic and cognitive outcomes. *J Child Psychol Psychiatry*. 2009;50(10):1235-1245. doi:10.1111/j.1469-7610.2009.02101.x

33. Macari SL, Campbell D, Gengoux GW, Saulnier CA, Klin AJ, Chawarska K. Predicting developmental status from 12 to 24 months in infants at risk for autism spectrum disorder: a preliminary report. *J Autism Dev Disord*. 2012;42(12):2636-2647. doi:10.1007/s10803-012-1521-0

34. Zwaigenbaum L, Bryson S, Lord C, et al. Clinical assessment and management of toddlers with suspected autism spectrum disorder: insights from studies of high-risk infants. *Pediatrics*. 2009;123(5):1383-1391. doi:10.1542/peds.2008-1606

35. Risi S, Lord C, Gotham K, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry*. 2006;45(9):1094-1103. doi:10.1097/01.chi.0000227880.42780.0e

36. Chawarska K, Klin A, Paul R, Volkmar F. Autism spectrum disorder in the second year: stability and change in syndrome expression. *J Child Psychol Psychiatry*. 2007;48(2):128-138. doi:10.1111/j.1469-7610.2006.01685.x

37. Silverman BW. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall; 1986.

38. Good P. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. 3rd ed. Springer; 2000.

39. Stone M. Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc Ser B Methodol*. 1974;36(2):111-133. doi:10.1111/j.2517-6161.1974.tb00994.x

40. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-35. doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

41. Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. *Am Stat*. 1998;52(2):119-126. doi:10.1080/00031305.1998.10480550

42. Linnet K. Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. *Clin Chem*. 1998;44(5):1024-1031. doi:10.1093/clinchem/44.5.1024

43. Martin RF. General Deming regression for estimating systematic bias and its confidence interval in method-comparison studies. *Clin Chem*. 2000;46(1):100-104. doi:10.1093/clinchem/46.1.100

44. Bollen KA, Jackman RW. Regression diagnostics: an expository treatment of outliers and influential cases. *Sociol Methods Res*. 1985;13(4):510-542. doi:10.1177/0049124185013004004

45. Kutner MH, Nachtsheim CJ, Neter J, Li W. *Applied Linear Statistical Models*. 5th ed. McGraw-Hill/Irwin; 2005.

46. Spearman C. The proof and measurement of association between two things. by C. Spearman, 1904. *Am J Psychol*. 1987;100(3-4):441-471. doi:10.2307/1422689

47. Law MR, Wald NJ, Wu T, Hackshaw A, Bailey A. Systematic underestimation of association between serum cholesterol concentration and ischaemic heart disease in observational studies: data from the BUPA study. *BMJ*. 1994;308(6925):363-366. doi:10.1136/bmj.308.6925.363

48. Charles EP. The correction for attenuation due to measurement error: clarifying concepts and creating confidence sets. *Psychol Methods*. 2005;10(2):206-226. doi:10.1037/1082-989X.10.2.206

49. Loomes R, Hull L, Mandy WPL. What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis. *J Am Acad Child Adolesc Psychiatry*. 2017;56(6):466-474. doi:10.1016/j.jaac.2017.03.013

50. Shultz S, Klin A, Jones W. Inhibition of eye blinking reveals subjective perceptions of stimulus salience. *Proc Natl Acad Sci U S A*. 2011;108(52):21270-21275. doi:10.1073/pnas.1109304108

51. Zwaigenbaum L, Bishop S, Stone WL, et al. Rethinking autism spectrum disorder assessment for children during COVID-19 and beyond. *Autism Res*. 2021;14(11):2251-2259. doi:10.1002/aur.2615

52. LeDoux J. *Synaptic Self: How Brains Become Who We Are*. Viking Adult; 2002.

53. Scarr S, McCartney K. How people make their own environments: a theory of genotype → environment effects. *Child Dev*. 1983;54(2):424-435. doi:10.1111/j.1467-8624.1983.tb03884.x

54. Ellis BJ, Boyce WT, Belsky J, Bakermans-Kranenburg MJ, van Ijzendoorn MH. Differential susceptibility to the environment: an evolutionary–neurodevelopmental theory. *Dev Psychopathol*. 2011;23(1):7-28. doi:10.1017/S0954579410000611

55. Jones W, Klaiman C, Richardson S, et al. Eye-tracking–based measurement of social visual engagement compared with expert clinical diagnosis of autism. *JAMA*. Published online September 5, 2023. doi:10.1001/jama.2023.13295

56. Burdick RK, Larsen GA. Confidence intervals on measures of variability in R&R studies. *J Qual Technol*. 1997;29(3):261-273. doi:10.1080/00224065.1997.11979768

57. McDonnell CG, Bradley CC, Kanne SM, Lajonchere C, Warren Z, Carpenter LA. When are we sure? predictors of clinician certainty in the diagnosis of autism spectrum disorder. *J Autism Dev Disord*. 2019;49(4):1391-1401. doi:10.1007/s10803-018-3831-3

58. Klaiman C, White S, Richardson S, et al. Expert clinician certainty in diagnosing autism spectrum disorder in 16-30–month-olds: a multi-site trial secondary analysis. *J Autism Dev Disord*. 2022:1-16. doi:10.1007/s10803-022-05812-8

SUPPLEMENT 1.

SUPPLEMENT 2.