HHS Public Access

Author manuscript

J Am Acad Child Adolesc Psychiatry. Author manuscript; available in PMC 2019 November 01.

Development and Validation of Objective, Quantitative, Eye Tracking-Based Measures of Autism Risk and Symptom Levels

Thomas W. Frazier, PhD,
Center for Autism Cleveland Clinic OH.

Autism Speaks, Independence OH.

statistical experts on this research.

Eric W. Klingemier, BA,
Center for Autism Cleveland Clinic OH.

Sumit Parikh, MD,
Cleveland Clinic, OH.

Leslie Speer, PhD, NCSP,
Center for Autism Cleveland Clinic OH.

Mark S. Strauss, PhD,
University of Pittsburgh, PA.

Charis Eng, MD, PhD,
Genomic Medicine Institute, Cleveland Clinic, OH.

Antonio Y. Hardan, MD, and

Stanford University, CA.

Eric A. Youngstrom, PhD
University of North Carolina at Chapel Hill, NC.

statistical experts on this research.

Abstract

Objective: The primary aim of this study was to develop and validate eye tracking-based measures for estimating autism spectrum disorder (ASD) risk and quantifying autism symptom levels.

Method: Eye tracking data were collected from youth during an initial evaluation visit, with administrators blinded to all clinical information. Consensus diagnoses were given by the multidisciplinary team. Participants viewed a 5- minute video that included 44 dynamic stimuli from 7 distinct paradigms while gaze was recorded. Gaze metrics were computed for temporally-defined regions-of-interest. Autism risk and symptom indices aggregated gaze measures showing significant bivariate relationships with ASD diagnosis and Autism Diagnostic Observation Schedule 2 (ADOS-2) symptom severity levels in a training sample (75%, n=150). Receiver operating characteristic curve analysis and non-parametric correlations were used to cross-validate findings in a test sample (25%; n=51).

Results: Most children (n=201, 92%) completed a valid eye tracking assessment (ages 1.6–17.6; 80% male; ASD n=91, non-ASD n=110). In the test sub-sample, the autism risk index had high accuracy for ASD diagnosis (area under the curve [AUC]=.86, 95%CIs=.75-.95), while the autism symptom index was strongly associated with ADOS-2 total severity scores (r=.41, p<.001). Validity was not substantively attenuated after adjustment for language, non-verbal cognitive ability, or other psychopathology symptoms (r=.40-.67, p>.001).

Conclusion: Eye tracking measures appear to be useful quantitative, objective measures of ASD risk and autism symptom levels. If independently replicated and scaled for clinical use, eye tracking-based measures could be used to inform clinical judgment regarding ASD identification and to track autism symptom levels.

Keywords

Introduction

Numerous studies have identified gaze differences between individuals with autism spectrum disorder (ASD) and controls across a wide range of ages and stimulus paradigms. [1–4] ASD-affected individuals have decreased attention to social information and increased attention to extraneous (non-social) information.[5] Our recent meta-analysis of 122 autism eye tracking studies found largest differences for non-social and face/eye regions.[6] The magnitude of findings was consistent across age and present even when comparing ASD to developmental disability. Motivated by this pattern, we demonstrated, across two samples, that aggregation of looking times to *a priori* regions of interest (ROIs) from a range of

stimuli yielded strong diagnostic validity.[7] These findings, and other recent results,[8,9] suggest that eye gaze patterns, particularly those based on dynamic temporal analysis,[10] may be a promising objective ASD risk marker and a quantitative measure of autism symptoms spanning the full continuum of behavior.

At present, subjective assessment tools are the only methods available for ASD identification.[11,12] Parent- completed questionnaires are used to screen for ASD in primary care settings, while parent interviews and clinical observations are the gold-standard assessment tools used to inform tertiary care ASD evaluations. Developing quantitative, objective measures of ASD is a major ongoing research priority. Several biological measures are being explored (e.g., EEG, MEG, MRI, etc.),[13,14] but these have significant technical, financial, time, and/or ethical limitations (e.g, sedation for MRI in young children). In contrast, remote eye gaze tracking is unobtrusive, can be rapidly collected (5 minutes) across a wide range of ages and cognitive levels, is inexpensive (trackers <$200 are available), and directly assesses the core social attention deficits contributing to ASD. Therefore, the present study's primary purpose was to empirically-identify and cross-validate an autism risk index (ARI) using eye tracking metrics collected from a large sample of clinically-referred patients at-risk for ASD. We expected to identify an ARI with strong diagnostic validity (AUC .80). A secondary purpose was to develop and cross-validate eye tracking measures of autism symptom levels. We expected to find and cross-validate eye tracking measures showing strong relationships (r .30) with clinical observations of autism symptom severity.

Method

Research Participants

Participants were youth referred to a tertiary-care, multi-disciplinary ASD evaluation clinic. Pediatricians made referrals following autism screening, if there was clinical concern of ASD, or if parents or teachers had concerns. Patients were consecutively recruited at the diagnostic evaluation visit (08/25/2015 to 11/30/2016). Gaze data were collected prior to the consensus diagnosis team meeting, the diagnostic team was masked to the eye tracking evaluation; all three research coordinators who administered the eye tracking evaluation were masked to participant diagnosis. The Cleveland Clinic IRB reviewed and approved the protocol.

Diagnosis

Consensus diagnosis was based on a parent interview conducted by a psychologist, psychosocial history confirmed by the psychologist, medical evaluation and developmental history confirmed by a physician, cognitive testing administered by a speech language pathologist or psychometrist, and the Autism Diagnostic Observation Schedule- Second Edition (ADOS-2) administered by a reliable administrator. Within two weeks of the initial visit, a multidisciplinary team met to confirm the presence/absence of ASD using DSM-5 criteria and document any other psychiatric diagnoses.

Clinical Assessments

The ADOS-2 measured autism symptom levels.[15,16] The ADOS-2 modules 1–4 and the toddler module are considered the gold-standard clinical observation measures for assessing autism symptom severity. This study used ADOS-2 total, social affect sub-scale, and restricted/repetitive behavior calibrated severity scores.[17–20] The Social Responsiveness Scale Second Edition (SRS-2) measured parent-reported autism trait levels.[21] The SRS-2 is a 65- item, ordinally-scaled (1= "not true" to 4= "almost always true") quantitative assessment of autism traits. The SRS sex-adjusted total T-score has been extensively validated and distinguishes youth with autism from other psychiatric conditions,[22,23] but has low specificity in at-risk or referred populations.[24] Receptive and expressive language was collected as part of the clinical evaluation using the Mullen Scales of Early Learning,[25] the Clinical Evaluation of Language Fundamentals - Fourth Edition[26] or Preschool Version - Second Edition,[27] or the Preschool Language Scales - Fifth Edition.[28] Non-verbal ability was collected using the Mullen visual reception subtest. For Mullen scales, T-scores were converted to standard scores (M=100, SD=15). Other psychopathological symptoms were collected using the Child Behavior Checklist (CBCL) - ages 1.5 to 5 and 6–18 parent-report versions.[29] Internalizing, externalizing, and total problems T-scores were used to describe the sample and examine the impact of behavior problems on eye gaze measures.

Eye Tracking Acquisition and Processing

Eye tracking data were collected in a quiet room adjacent to the diagnostic clinic. Data were recorded using an SMI Red250 remote eye tracker (sampling at 60Hz) attached to the frame of a 1280 horizontal X 1024 vertical 19-inch LCD stimulus presentation monitor. Maximum spatial resolution was 0.1° with gaze position accuracy of 0.5°. The system allows for head movement (32 X 21 X 30 for Red250) at a maximum distance of 75cm. Two 5-point calibrations were obtained at fixed times throughout the experiment, once at the beginning and once approximately 2 minutes into the stimulus presentation. Additional calibrations could be inserted if the examiner judged that gaze capture accuracy had been lost when observing the participant's gaze as depicted on the experimental control monitor. Gaze metrics to each ROI were derived using SMI BeGaze software and included 5 measures that initial analyses indicated provided unique information within temporal ROI: glances, fixation count, fixation duration percent, first fixation duration, and average fixation duration. A glance was defined as any entry to an ROI that includes at least one fixation. Glance count reflects the number of entries to an ROI with at least one fixation. A fixation was defined as at least 80ms of samples within a 100 pixel dispersion area.

Forty-four stimuli were presented using SMI Experiment Center, selected to represent 7 distinct stimulus paradigms previously used in the eye gaze literature,[6] including single person facial affect, two-person facial affect discrimination, gaze following and joint attention, dyadic bids toward the participant involving humor, side-by-side abstract shape movement and human activity, high autism interest images mixed with social stimuli, and naturalistic social interaction scenes (See Figure S1, available online). Stimuli were presented in a single order, intermixed with attention grabbing center fixation stimuli and stimuli designed to evaluate receptive language (not examined). Temporal ROIs were identified by the first author (who was not involved in the clinical assessment and did not

participate in the eye tracking evaluation) using data from six healthy control participants (ages 3–15, 3 males) not included in the present study. Healthy control participants were recruited from families of children with autism with the intention of identifying "typical" gaze patterns. Parents had no concerns regarding development for these children, there was no evidence of elevated autism traits (SRS-2 T-score<60), and all available clinical information indicated no neuropsychiatric disorders. The first author chose time periods for temporal ROIs based on the observed gaze patterns of these 6 individuals. For example, if at least half of the healthy controls began looking at the eye region at ~800 ms into the stimulus and then gaze moved to other regions at ~1500ms, a temporal ROI was set for 800ms-1500ms. Spatial aspects of ROIs were chosen by the first author to capture all major elements within each stimulus, including theoretically important social (facial expressions, body movements, target objects) and non-social stimulus elements (abstract shapes, non-target distractor objects). Eye gaze metrics were computed for temporal ROIs and the total stimulus period. Across the 44 stimuli, 1,592 temporal ROIs were defined and gaze metrics were collected (See Table S1, available online).

Eye tracking data collection followed recommendations from Sasson and Elison.[30] Children were seated alone or in their parent's lap approximately 65cm from the LCD display and viewed stimuli subtending a visual angle of ~18.8°. Standard room lighting was used and the room was sparse, with visual barriers used to reduce distraction. After calibration, children (who were of sufficient age and cognitive level) were told, "You will see some pictures and videos; pay attention, but look however you want." Eye tracking evaluations were considered invalid and data were excluded if gaze to the screen during the entire experiment was tracked <40% of the time, if more than two unplanned re-calibrations had to be inserted during the evaluation, or participants had <15 stimuli with adequate looking time (defined as >60% fixation duration percent to the stimulus).

Statistical Analyses

The study design and analyses followed recommendations for evaluating test validity (STARD; See Table S2, available online)[31] and reporting the results of a multivariate prediction model (TRIPOD; See Table S3, available online).[32] Univariate and bivariate distributions identified outliers and high leverage cases. Analyses were computed separately with and without high leverage cases, but no substantive differences in results were observed. For this reason, analyses on all available data are reported. Descriptive statistics were presented separately for patients with ASD and patients without ASD to characterize the sample. Comparisons between ASD and non-ASD groups were made across demographic and clinical measures using independent samples t-tests, Chi- square statistics, or non-parametric alternatives, where appropriate.

To create the autism risk index, the sample was randomly split into train (75%, n=150) and test (n=51, 25%) sub-samples using the random variable compute function in SPSS v24 where 150 rows were randomly selected as the training sub-sample and the remainder served as the test sub-sample. All indicator selection occurred in the training sample, consistent with recommended prediction cross-validation procedures.[33] In the training sample, bivariate bootstrap correlations (k=100) were computed between ASD diagnosis and each

eye gaze metric for all temporal ROIs. Positive correlations indicated greater gaze to the temporal ROI in youth with ASD and negative correlations indicated less gaze in youth with ASD. Gaze metrics showing significant correlations (non-zero bootstrap 95%CI) were selected, standardized by computing z-scores for each temporal ROI using the full sample mean and SD, and averaged. Indicators with significantly less gaze in ASD-affected children were multiplied by −1 and averaged with indicators showing significantly more gaze in ASD-affected individuals. Linear averaging was chosen because significant indicators tended to have similar relationships with ASD (.17 ≤ r ≤ .30) in the training sample; patterns of missing data were highly variable across cases and stimuli; and this unit weighting procedure has been shown to produce robust prediction under a range of circumstances,[34] permitting straightforward computation of Cronbach's alpha. Linear averaging also would be a reasonably efficient and practical approach for future software implementation. Similar procedures created indicators for the autism symptom index (ASI) and ASI sub-domain measures for social communication/interaction (SCI) (ASI:SCI) and restricted/repetitive behavior (RRB) (ASI:RRB) using ADOS-2 total and sub-scale severity scores as the criteria. To create the final ARI and ASI measures used in all future analyses, Age, age^2, and sex were included as covariates in linear regression models predicting raw ARI/ASI measures. Residuals from these models were saved and normed using the non-ASD sample mean and SD so that scores could be interpreted with reference to this distribution; where higher ARI scores are associated with greater likelihood of ASD. Internal consistency reliability of the ARI and ASI measures was estimated by computing Cronbach's α using z-transformed temporal ROI values that composed each measure.

To cross-validate the ARI in the test sub-sample, receiver operating characteristic curve analysis was computed with consensus ASD diagnosis as the state variable. Bivariate non-parametric Spearman's rank-order correlations between ASI measures and ADOS-2 severity scores quantified cross-validation. Bivariate Pearson correlations assessed the relationships between eye tracking risk and symptom indices, eye tracking validity measures, and clinical measures. Parametric and non-parametric partial correlations were computed between eye gaze measures and ASD diagnosis or ADOS-2 severity scores adjusting for cognitive or other psychopathology measures to evaluate whether autism-eye gaze relationships were specific or conflated with other factors.

Data preparation and descriptives used SPSS v24. ROC analyses were computed using pROC[35] and bivariate Spearman's rank-order partial correlations were computed using the ppcor program[36] in R. Statistical significance was set at α=.05, one-tailed, given the explicit directionality of predictions. Power to detect a significant AUC ≥ .80 was excellent (>.99) with a test sub-sample of 51 cases. Power was at least adequate (>.71) for detecting significant positive bivariate correlations of r ≥ .30.

Results

Participant Characteristics

Most participants (n=201 of 219; 92%) completed a valid eye tracking assessment (See Figure S2, available online). Cases with invalid eye tracking evaluations had high externalizing behavior problems and low cognitive ability (See Table S4, available online).

The clinically-realistic and diagnostically-challenging nature of the valid case sample is best demonstrated by two observations: 1) SRS-2 scores did not differentiate ASD and non-ASD cases and 2) non- ASD cases had significantly higher levels of behavior problems than ASD cases (Table 1). Not surprisingly, ASD cases had lower language and non-verbal ability scores. There were no significant differences between ASD and non-ASD cases on eye tracking validity measures after case exclusion. As expected, the randomly divided train and test sub-samples were highly similar across demographic and clinical characteristics (See Table S5, available online).

ASD Diagnosis

ARI classification accuracy for ASD diagnosis was excellent in the train sub-sample (AUC=. 92, 95%CI=.88-.96) and very good in the test sub-sample (AUC=.86, 95%CI=.75-.95; Figure 1), with only modest validity attenuation despite slightly lower symptom severity and higher cognitive ability in the test sample (See Table S5, available online). In the full sample, classification accuracy was comparable in children <4 years old (AUC=.925) and children 4+ (AUC=.931). Although only 17 children age 2.5 or younger were available, classification accuracy was also strong in this age group (AUC=.921). Intriguingly, the ARI had incremental validity for predicting ASD diagnosis after accounting for ADOS-2 severity scores R^2=.22, p<.001), even though ADOS-2 scores were available to the clinicians.

The ARI had high internal consistency reliability (α=.92) and wide quantitative range, with 95% of non-ASD cases falling from z=−2.3 to 1.6 and 95% of ASD cases falling from z= −0.1 to 5.0 (Figure 2). Most missed cases (68%) fell within +/−0.75 SD of the optimal cut point z=0.74.

While the ARI had significant negative relationships with measures of language and non-verbal ability (Table 2), controlling for these did not substantively attenuate correlation with ASD diagnosis. Furthermore, the ARI was not significantly associated with eye tracking evaluation validity indicators (See Table S6, available online). Missed cases (false positives and false negatives based on Youden's J cut point=.74) were not significantly related to most demographic, clinical, or eye tracking validity variables. However, false negative cases had lower SRS-2 and CBCL scores (See Supplement 1, available online). False positive cases required a larger number of calibrations inserted during the experiment (M=1.00, SD=1.05) relative to correctly identified (M=0.46, SD=0.79) and false negative (M=0.58, SD=1.17; p=. 031) cases. Missed cases were not significantly associated with any other eye tracking validity indicator.

Autism Symptom Severity

Cross-validated correlations between ASI measures and ADOS-2 total and sub-scale severity scores were very strong (r=.71-.74) in the train sub-sample. These correlations attenuated but remained significant in the test sub-sample (Figure 3; ASI r=.41, p=.002; ASI:SCI r=.26, p=.040; ASI:RRB r=.30, p=.022). In the full sample, the ASI had significant negative relationships with internalizing behavior, language, and non-verbal ability measures (Table 2), but maintained a strong relationship with ADOS-2 total severity scores (r=.58-.67) after adjusting for these associations. Intriguingly, in a post-hoc analysis suggested by a

reviewer, the ASI showed significant negative relationships with CBCL anxiety and oppositional defiant problem scales (r=−.17 and −.21) but not affective or ADHD problem scales (r=−.10 and −.07). ASI measures had high internal consistency reliability (ASI α=.93, ASI:SCI α=.93, ASI:RRB α=.90).

Small, but significant, relationships (r=.17 and .20) were present between the ASI and gaze deviation in the X and Y-axes (See Table S6, available online). However, adjustment for these and other validity indicators did not diminish associations between the ASI and ADOS-2 total severity scores (r=.63-.64). ASI:SCI and ASI:RRB measures correlated more strongly with their respective ADOS-2 sub-scales than with the opposite ADOS-2 sub-scale (smallest z=4.48, p<.001).

Discussion

This is the first study to empirically-derive and cross-validate objective, quantitative measures of ASD risk and autism symptom levels in a large, clinically realistic sample. The ARI showed very good diagnostic accuracy, particularly given the imperfect reliability of clinical ASD diagnoses,[37] and the ASI had a large correlation with clinical observations of autism symptom severity. ASI sub-domain indices also showed strong and specific relationships with clinical observations of social communication/interaction and restricted/ repetitive behavior. ARI and ASI measures were not influenced by demographic factors and maintained large relationships with ASD diagnosis and autism symptom severity levels after accounting for measures of cognitive ability and other psychopathology. Future work may produce even greater validity by using more efficient variable selection methods at earlier stages of ARI and ASI creation and by including other gaze measurements (e.g., saccades, blink, pattern analysis, etc.). This work may also consider using more sophisticated methods for identifying temporal and spatial aspects of gaze patterns to include as inputs to ARI and ASI measures. Regardless, the present study suggests that high cross-validation classification and prediction accuracy can be achieved after averaging within and across stimuli and gaze metrics.

Use of non-ASD cases that were clinically-referred for ASD further strengthens findings. The non-ASD group represented a challenging comparison cohort of children referred for evaluation of ASD. This group was screened prior to referral (often by a pediatrician using the M-CHAT) and as a result had elevated parent-reported autism traits (SRS-2 scores) comparable to the ASD group. Children in the non-ASD group also had ADOS-2 scores overlapping the ASD group and almost all non-ASD children had some form of neuropsychiatric diagnosis, including intellectual disability. Further increasing the challenging nature of the comparison, the non-ASD group had elevated levels of other behavior problems and highly overlapping cognitive and language abilities relative to the ASD group. Thus, diagnostic discrimination was not inflated and should reflect values seen in high prevalence tertiary care settings - an important consideration for test evaluation studies.[38] Future validation studies should examine whether other psychopathological conditions, such as anxiety and oppositional behavior, influence ARI/ASI scores via engagement in the eye tracking paradigm or independently.

Objective, quantitatively-scaled eye tracking measures of ASD risk and autism symptom levels could be a major advance for clinical practice. Beyond risk assessment, eye tracking measures may enhance patient management by providing an unbiased method for tracking treatment response. Intriguingly, our experience with healthy controls indicates that they score considerably lower on the ARI, suggesting that the ARI might have excellent screening utility. If demonstrated in a population sample, the ARI might be a cost-effective tool for primary care settings, with limited time and training requirements, similar to new vision screening methods that have seen growing adoption.

Eye tracking measures of ASD risk and symptoms may also greatly accelerate autism research. Etiologic studies would benefit from diagnoses and symptom severity estimates based on reliable, objective measurements. Clinical trials have been limited by the lack of easily-acquired objective measures that link closely to the autism phenotype. Increasing identification of molecular pathway abnormalities leading to ASD within genetic syndromes has led to the initiation of clinical trials of targeted therapeutics. Additionally, trials of novel behavioral intervention packages for idiopathic ASD are working towards cost-effective population care models. Reliable, quantitative, objective measures of autism symptom levels have the potential to increase sensitivity to treatment effects, thereby increasing statistical power, decreasing the sample size needed to detect treatment effects, and improving the ability to identify efficacious treatments. Integrating eye tracking measures into longitudinal studies may improve understanding of ASD trajectories and clarify underlying etiology and outcomes. Test-retest reliability and sensitivity to change need to be evaluated prior to inclusion in these investigations, but existing data suggest good stability of eye gaze measures.[39]

Although this was one of the largest autism eye-tracking samples collected to date, a very large (N 500) multisite validation study is needed to replicate diagnostic classification accuracy and symptom severity prediction, determine generalizability across sampling variation,[40,41] and ensure resistance to minor procedural variation in gaze collection. Future research may also benefit from matching the ASD and non-ASD groups on key demographic and clinical factors that may influence accuracy and predictive value, such as age and sex. This work should also consider recruiting a larger number of children less than 2 years old to examine whether accuracy is maintained near the time of initial primary care autism screening. The stimulus battery was longer than desired for very young children (~5min). Results suggested that battery length might be decreased without loss of validity by including only those stimuli containing ROIs with large (r .25) gaze metrics bivariate correlations with ASD diagnosis or ADOS-2 symptom severity. Future investigations should validate an abbreviated battery for young children, explore methods for increasing gaze capture in the youngest and most challenging children, and evaluate whether repeated testing might improve accuracy. Additional work needs to create a streamlined package that automates scoring and reporting for clinical adoption and to examine how clinicians might integrate these measures into decision-making.

Eye tracking measures, such as the ARI and ASI, may be useful quantitative, objective measures of ASD risk and autism symptom levels. If replicated in a large multi-site study, and scaled for routine clinical and research use, the present eye tracking-based indices could

inform clinical judgment regarding ASD diagnosis in tertiary care settings and track autism symptom levels during standard treatments, longitudinal studies, and clinical trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Chita-Tegmark M Social attention in ASD: A review and meta-analysis of eye-tracking studies. Res Dev Disabil. 2016;48:79–93. [PubMed: 26547134]

2. Chita-Tegmark M Attention allocation in ASD: A review and meta-analysis of eye-tracking studies. Review Journal of Autism and Developmental Disorders. 2016;3:209.

3. Papagiannopoulou EA, Chitty KM, Hermens DF, Hickie IB, Lagopoulos J. A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. Soc Neurosci. 2014;9(6):610–632. [PubMed: 24988218]

4. Klin A, Jones W. Altered face scanning and impaired recognition of biological motion in a 15-month-old infant with autism. Dev Sci. 2008;11(1):40–46. [PubMed: 18171365]

5. 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. [PubMed: 19329996]

6. Frazier TW, Strauss M, Klingemier EW, et al. A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. J Am Acad Child Adolesc Psychiatry. 2017;56(7):546–555. [PubMed: 28647006]

7. Frazier TW, Klingemier EW, Beukemann M, et al. Development of an objective autism risk index using remote eye tracking. J Am Acad Child Adolesc Psychiatry. 2016;55(4):301–309. [PubMed: 27015721]

8. Pierce K, Marinero S, Hazin R, McKenna B, Barnes CC, Malige A. Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder

subtype associated with increased symptom severity. Biol Psychiatry. 2016;79(8):657–666. [PubMed: 25981170]

9. Chevallier C, Parish-Morris J, McVey A, et al. Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters. Autism Res. 2015;8(5):620–628. [PubMed: 26069030]

10. Guillon Q, Hadjikhani N, Baduel S, Roge B. Visual social attention in autism spectrum disorder: insights from eye tracking studies. Neurosci Biobehav Rev. 2014;42:279–297. [PubMed: 24694721]

11. 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. [PubMed: 21883205]

12. 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. [PubMed: 16926617]

13. Ingalhalikar M, Parker WA, Bloy L, Roberts TP, Verma R. Creating multimodal predictors using missing data: classifying and subtyping autism spectrum disorder. J Neurosci Methods. 2014;235:1–9. [PubMed: 24983132]

14. Hazlett HC, Gu H, Munsell BC, et al. Early brain development in infants at high risk for autism spectrum disorder. Nature. 2017;542(7641):348–351. [PubMed: 28202961]

15. Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Bishop SL. Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part 1): Modules 1–4. Torrance, CA: Western Psychological Services; 2012.

16. Luyster R, Gotham K, Guthrie W, et al. The Autism Diagnostic Observation Schedule-toddler module: a new module of a standardized diagnostic measure for autism spectrum disorders. J Autism Dev Disord. 2009;39(9):1305–1320. [PubMed: 19415479]

17. Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. J Autism Dev Disord. 2009;39(5):693–705. [PubMed: 19082876]

18. Hus V, Gotham K, Lord C. Standardizing ADOS domain scores: separating severity of social affect and restricted and repetitive behaviors. J Autism Dev Disord. 2014;44(10):2400–2412. [PubMed: 23143131]

19. Hus V, Lord C. The autism diagnostic observation schedule, module 4: revised algorithm and standardized severity scores. J Autism Dev Disord.44(8):1996–2012.

20. Esler AN, Bal VH, Guthrie W, Wetherby A, Ellis Weismer S, Lord C. The Autism Diagnostic Observation Schedule, Toddler Module: Standardized Severity Scores. J Autism Dev Disord. 2015;45(9):2704–2720. [PubMed: 25832801]

21. Constantino JN, Gruber CP. The social responsiveness scale manual, second edition(SRS-2). Los Angeles, CA: Western Psychological Services; 2012.

22. Virkud YV, Todd RD, Abbacchi AM, Zhang Y, Constantino JN. Familial aggregation of quantitative autistic traits in multiplex versus simplex autism. Am J Med Genet B Neuropsychiatr Genet. 2009;150B(3):328–334. [PubMed: 18618672]

23. Constantino JN, Gruber CP. Social Responsiveness Scale: Manual. Los Angeles, CA: Western Psychological Services; 2005.

24. Aldridge FJ, Gibbs VM, Schmidhofer K, Williams M. Investigating the clinical usefulness of the Social Responsiveness Scale (SRS) in a tertiary level, autism spectrum disorder specific assessment clinic. J Autism Dev Disord. 2012;42(2):294–300. [PubMed: 21516433]

25. Mullen EM. Mullen Scales of Early Learning. Circle Pines, MN: American Guidance Service Inc; 1995.

26. Semel E, Wiig EH, Secord WA. Clinical Evlauation of Language Fundamentals, fourth edition(CELF-4). Toronto, Canada: The Psychological Corporation / A Harcourt Assessment Company; 2003.

27. Wiig EH, Secord WA, Semel E. Clinical evaluation of language fundamentals - Preschool, second edition(CELF-2). Toronto, Canada: The Psychological Corporation / A Harcourt Assessment Company; 2004.

28. Zimmerman IL, Steiner V,G, Pond E Preschool Language Scales-Fifth Edition (PLS-5). San Antonio, TX: Pearson; 2011.

29. Achenbach TM, Rescorla LA. Manual for the ASEBA school-age forms and profiles. Burlington, VT: University of Vermont, Department of Psychiatry; 2001.

30. Sasson NJ, Elison JT. Eye tracking young children with autism. J Vis Exp. 2012(61):3675. [PubMed: 22491039]

31. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. Ann Intern Med. 2003;138:40–44. [PubMed: 12513043]

32. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55–63. [PubMed: 25560714]

33. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning with applications in R. New York: Springer; 2013.

34. Bobko P, Roth PL, Buster MA. The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. Organizational Research Methods. 2007;10(4):689–709.

35. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77. [PubMed: 21414208]

36. Kim S ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. Communications for Statistical Applications and Methods. 2015;22(6):665–674. [PubMed: 26688802]

37. Regier DA, Narrow WE, Clarke DE, et al. DSM-5 field trials in the United States and Canada, Part II: test- retest reliability of selected categorical diagnoses. Am J Psychiatry. 2013;170(1):59–70. [PubMed: 23111466]

38. Youngstrom EA, Genzlinger J, Egerton G, Van Meter AR. Multivariate meta-analysis of the discriminative validity of caregiver, youth, and teacher rating scales for pediatric bipolar disorder: mother knows best about mania. Archives of Scientific Psychology. 2015;3(1):112–137.

39. Farzin F, Scaggs F, Hervey C, Berry-Kravis E, Hessl D. Reliability of eye tracking and pupillometry measures in individuals with fragile X syndrome. J Autism Dev Disord. 2011;41(11): 1515–1522. [PubMed: 21267642]

40. Konig I, Malley J, Weimar C, Diener H, Ziegler A. German Stroke Study C. Practical experiences on the necessity of external validation. .Stat Med. 2007;26(30):5499–5511. [PubMed: 17907249]

41. Youngstrom EA, Meyers OI, Youngstrom JK, Calabrese JR, Findling RL. Comparing the effects of sampling designs on the diagnostic accuracy of eight promising screening instruments for pediatric bipolar disorder. Biol Psychiatry. 2006;60:1013–1019. [PubMed: 17056395]
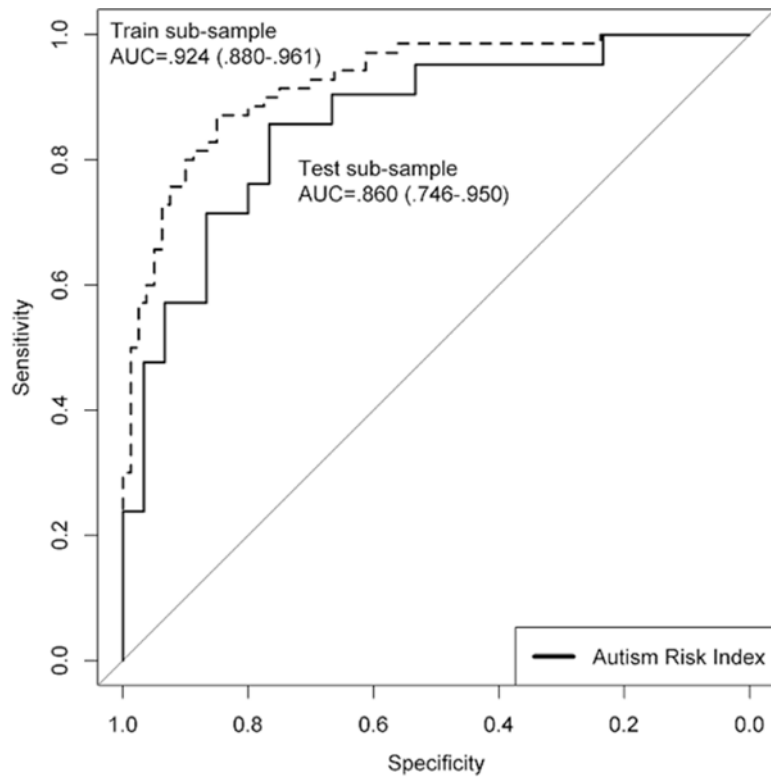
Figure 1.
Receiver Operating Characteristic Curves in the Training and Test Sub-Samples.
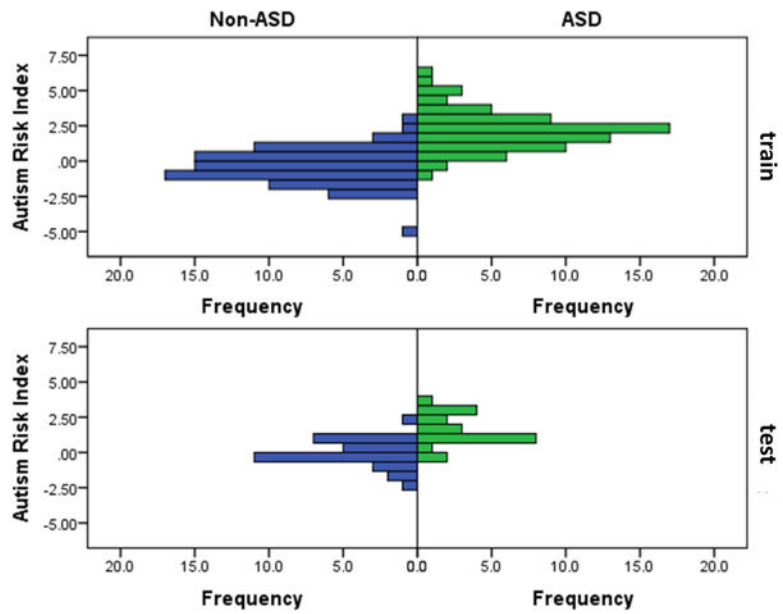
Figure 2. Autism Risk Index (ARI) Distribution Across Youth With and Without an Autism Spectrum Disorder (ASD) Diagnosis in the Training and Test Sub-Samples.
Note: ARI values are re-scaled so that non-ASD participants have a mean score of 0 with a standard deviation of 1.
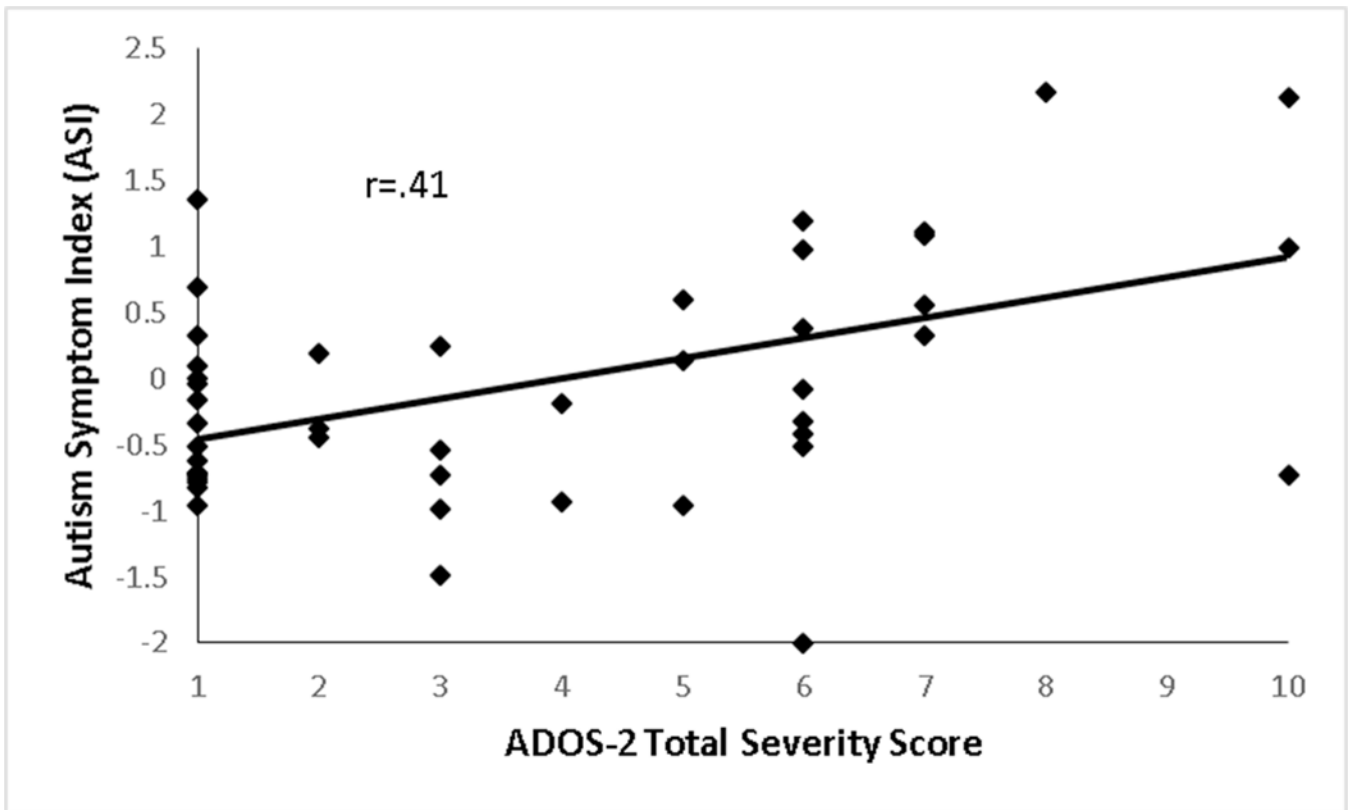
Figure 3.
Relationship Between the Autism Symptom Index (ASI) and Autism Diagnostic
Observation Schedule 2 (ADOS-2) Total Severity Scores in the Test Sub-Sample.

Table 1.

Participant Characteristics.

	Non-ASD M (SD)	ASD M (SD)	Cohen's d (p)
N	110	91	
Age (SD, range)	6.8 (3.3, 1.8–17.6)	5.7 (3.6, 1.6–15.8)	.31 (.001)
Male (n, %)	86 (78.2%)	75 (82.5%)	−.11 (.454)
White Non-Hispanic (n, %)	80 (72.7%)	62 (68.1%)	.10 (.476)
ADOS-2 Total Severity (SD, range)	2.4 (2.0, 1–10)	6.3 (2.4, 1–10)	−1.78 (<.001)
ADOS-2 Social Affect Severity (SD, range)	3.0 (2.3, 1–10)	6.4 (2.3, 1–10)	−1.48 (<.001)
ADOS-2 Repetitive Behavior Severity (SD, range)	2.3 (2.2, 1–9)	6.4 (2.5, 1–10)	−1.75 (<.001)
Social Responsiveness Scale-2 (T-Score; SD, range)	66.7 (12.7, 39–86)	68.1 (10.9, 41–89)	−.11 (.616)
Child Behavior Checklist - Total Problems (T-Score)	67.0 (11.2)	61.4 (11.4)	.50 (<.001)
Child Behavior Checklist - Internalizing (T-Score)	64.6 (10.4)	60.1 (11.5)	.41 (.003)
Child Behavior Checklist - Externalizing (T-Score)	64.2 (12.8)	57.4 (12.2)	.55 (<.001)
Receptive Language (SS)	85.5 (23.9)	69.9 (20.7)	.70 (<.001)
Expressive Language (SS)	88.1 (24.7)	74.2 (23.6)	.57 (<.001)
Total Language (SS)	87.9 (23.8)	74.4 (23.5)	.57 (<.001)
Non-Verbal Ability (SS)	88.6 (19.2)	65.7 (13.9)	1.33 (<.001)
Tracking Ratio %	81.5 (13.4)	77.7 (16.1)	.26 (.127)
Number of Stimuli Tracked (out of 44)	39.6 (5.6)	37.9 (6.9)	.26 (.112)
Calibration Accuracy - X Deviation°	1.8 (2.1)	2.1 (2.9)	−.20 (.531)
Calibration Accuracy - Y Deviation°	1.4(1.6)	1.8 (2.5)	−.14 (.756)

Note: Cohen's d is based on conversion of phi coefficient or parametric independent samples t-test. *P* values are derived from X^2 or non-parametric Mann-Whitney U. Sample sizes for Autism Diagnostic Observation Schedule 2 (ADOS-2), Social Responsiveness Scale 2nd Editiion (SRS-2), and Child Behavior Checklist (CBCL) were 195, 184, and 192. Sample sizes for receptive language, expressive language, total language, and non-verbal ability estimates were 138, 168, 172, and 54. Non-verbal ability was only collected for children ages 6 and under. A subset of 149 participants completed gaze calibration at the end of the experiment. ASD = autism spectrum disorder. SS = standard score

Table 2.

Bivariate and Partial Correlations Between the Autism Risk Index (ARI), Autism Symptom Index (ASI), and Clinical Factors in the Full Sample.

	ARI	ASI
Bivariate Correlations		
ASD Diagnosis	.70***	.49***
ADOS Total Severity	.46***	.66***
SRS Total T-Score	.22*	.14
CBCL Internalizing	−.06	−.14*
CBCL Externalizing	−.09	−.16*
CBCL Total	−.07	−.14*
Receptive Language	−.23*	−.34***
Expressive Language	−.15	−.22*
Total Language	−.16*	−.24***
Non-Verbal Ability	−.57***	−.52***
Partial Correlations (adjusting for cognitive variables)		
ASD Diagnosis (receptive language)	.63***	.40***
ASD Diagnosis (non-verbal ability)	.52***	.41***
ADOS Total Severity (receptive language)	.40***	.59***
ADOS Total Severity (non-verbal ability)	.18	.58***
Partial Correlations (adjusting for behavior problems)		
ASD Diagnosis (CBCL total problems)	.67***	.49***
ADOS Total Severity (CBCL total problems)	.47***	.67***

Note. All correlations with Autism Diagnostic Observation Schedule (ADOS) Total Severity are non-parametric (Spearman's) bivariate or partial correlations. ASD = autism spectrum disorder; CBCL = Child Behavior Checklist; SRS = Social Responsiveness Scale

*
$p<.05$

***
$p<.001$