UC Davis
UC Davis Previously Published Works

Title

A video-based measure to identify autism risk in infancy

Permalink

Journal

ISSN

Authors

Young, Gregory S
Constantino, John N
Dvorak, Simon
et al.

Publication Date

DOI

Peer reviewed

A Video-Based Measure to Identify Autism Risk in Infancy

Gregory S. Young PhD, [a] John N. Constantino MD, [b] Simon Dvorak BS, [c]

Ashleigh Belding MPH, [a] Devon Gangi PhD, [a] Alesha Hill BA, [a] Monique Hill MA, [a]

Meghan Miller PhD, [a] Chandni Parikh PhD, [a] AJ Schwichtenberg PhD, [d] Erika Solis BS, [a]

& Sally Ozonoff PhD [a]

[a] Department of Psychiatry & Behavioral Sciences, MIND Institute, University of California-Davis

[b] Department of Psychiatry, Washington University-St. Louis School of Medicine

[c] Information and Educational Technology, University of California-Davis

[d] Department of Human Development & Family Studies, Purdue University

Word count: 4761 (manuscript), 296 (Abstract)

Running Head: A Video-Based Autism Screener for Infants

Abstract

Background: Signs of autism are present in the first two years of life, but the average age of diagnosis lags far behind. Instruments that improve detection of autism risk in infancy are needed. This study developed and tested the psychometric properties of a novel video-based approach to detecting ASD in infancy.

Methods: A prospective longitudinal study of children at elevated or lower risk for autism spectrum disorder was conducted. Participants were 76 infants with an older sibling with ASD and 37 infants with no known family history of autism. The Video-referenced Infant Rating System for Autism (VIRSA) is a web-based application that presents pairs of videos of parents and infants playing together and requires forced-choice judgments of which video is most similar to the child being rated. Parents rated participants on the VIRSA at 6, 9, 12, and 18 months of age. We examined split-half and test-retest reliability; convergent and discriminant validity; and sensitivity, specificity, and negative and positive predictive value for concurrent and 36-month ASD diagnoses.

Results: The VIRSA demonstrated satisfactory reliability and convergent and discriminant validity. VIRSA ratings were significantly lower for children ultimately diagnosed with ASD than children with typical development by 12 months of age. VIRSA scores at 18 months identified all children diagnosed with ASD at that age, as well as 78% of children diagnosed at 36 months.

Conclusions: This study represents an initial step in the development of a novel video-based approach to detection of ASD in infancy. The VIRSA's psychometric properties were promising when used by parents with an older affected child, but still must be tested in community samples with no family history of ASD. If results are replicated, then the VIRSA's low-burden, web-based format has the potential to reduce disparities in communities with limited access to screening.

Keywords: Autism, Screening, Infancy, Social Development

Abbreviations: ASD = autism spectrum disorder; VIRSA = Video-referenced Infant Rating

System for Autism

Introduction

The developmental course of autism spectrum disorder (ASD) involves the onset of symptoms in the first three years of life. Differences between children who will later receive an ASD diagnosis and those with typical development emerge before the second birthday (Gammer et al., 2015; Landa & Garrett-Mayer, 2006; Ozonoff et al., 2010; Zwaigenbaum et al., 2005), with some studies documenting signs in the first year of life (Maestro et al., 2002; Miller et al., 2017; Werner, Dawson, Osterling & Dinno, 2000), and parents first expressing concerns at an average age of 14 months (Chawarska et al., 2007). Despite advances in knowledge about the earliest presentations of ASD, the mean age of diagnosis has stubbornly remained over 4 years (Baio et al., 2018) and has not declined over the last two decades, squandering years of potential intervention when the brain is most plastic. It is critical that further attempts are made to decrease the age of ASD diagnosis so that it better aligns with the age of first symptom emergence.

One of the identified barriers to more prompt recognition of ASD is measurement (Al Qabandi, Gorter & Rosenbaum, 2011). Over the last decade, much effort has gone into the development of instruments for earlier detection of ASD (Zwaigenbaum et al., 2015). The most feasible method for large-scale screening is parent report and most existing measures use this methodology. However, recent studies have demonstrated low agreement between parent report and more objective measures of ASD symptoms (Ozonoff et al., 2011), as well as lower reliability for screening instruments when used in rural, low income, less educated, and racially diverse samples (Khowaja, Hazzard & Robins, 2015; Scarpa et al., 2013). A population screening study of 10,479 twelve-month-olds (Pierce et al., 2011) using a parent-report measure (Wetherby, Brosnan-Maddox, Peace & Newton, 2008) identified 32 infants with ASD. This represents significant under-identification, even after accounting for cases with later onset (Barger, Campbell & McDonough, 2013), since current prevalence studies estimate that 170 of 10,000 children have ASD (Baio et al., 2018).

The lower sensitivity of early screening measures may be due to the subtlety of initial ASD symptoms and the difficulty of accurately conveying them to parents through written descriptions. Major sources of error in parent questionnaires include comprehension and interpretation problems (Koriat, Goldsmith & Pansky, 2000; Krosnick & Presser, 2010), such as limited understanding of the queried constructs, inadequate knowledge of developmental milestones, and bias due to post-event information (e.g., eventual diagnosis). The current study moves beyond verbal descriptions by employing video examples to reduce subjective interpretations. The use of videos has been shown to dramatically increase clarity in other fields, from music instruction to motor vehicle repair (Arguel & Jamet, 2009). Recently, video was incorporated in ASD screening by Marrus and colleagues (2015), who had parents complete ratings after watching a video of a socially competent toddler, in order to "reduce discrepant interpretations of items by providing informants with a common naturalistic standard for comparison" (p. 1340).

Here we describe the development of a new instrument, the Video-referenced Infant Rating System for Autism (VIRSA). It extends previous approaches (Marrus et al., 2015) by creating a large library of video clips depicting a wide range of social-communication ability and relying solely on video in the ratings, with no written descriptions of behavior. We hypothesized that the semantic clarity afforded by video would improve early discrimination of infants at highest risk for ASD.

Methods

Instrument Development

The VIRSA was developed using video from participants in a longitudinal infant sibling study and then validated on an independent sample of infants. Videos used in the VIRSA were drawn from an archive of over 300,000 minutes of digitized video recorded in a clinical laboratory setting. Video depicted infants and parents playing together with age-appropriate toys. Segments were selected from a task that used a standardized toy set and instructed

parents to play with their child as they would at home (Schwichtenberg, Kellerman, Miller, Young & Ozonoff, 2019). Video recordings utilized a consistent camera angle facing the child, with the parent in profile. All families gave both informed consent and legal authorization to include their videos in the VIRSA.

Social behaviors, including smiles, vocalizations, and eye contact, were coded by research assistants unaware of participant risk group or outcome, using a previously validated coding scheme that is sensitive to the changes that occur during the onset of ASD symptoms as early as 6 months (Gangi et al., 2019; Ozonoff et al., 2010). In order to include a broad range of behaviors in the VIRSA, candidate videos were ranked by frequencies of the coded behaviors. Twenty-second segments were then excised from the original video files, resulting in a collection of over 3,000 video segments from 100 past participants between 6 and 18 months of age. Next, video segments were rated by 9 clinical research staff on a scale from 1 (least socially competent) to 10 (most socially competent). Each clinician rated a randomly selected set of 39 videos twice to establish test-retest reliability (mean=0.89, range 0.78-0.98). Inter-rater reliability was examined on a larger randomly selected set of 260 video clips rated by all raters using a two-way random ICC model. The average measures ICC for absolute agreement was 0.92, with a lower bound of 0.86, suggesting strong inter-rater reliability of the 10-point scale. Video segments were excluded for poor lighting or audio quality, obscured video angles, or use of the child's name. This resulted in a pool of video comprising 1,132 individual 20-second clips, which was then constrained to insure adequate representation across the 10-point rating scale within each age (6, 9, 12, and 18 months). To limit the software overhead for the VIRSA app, 268 videos were then randomly selected to create the final VIRSA video library. The final pool of video included segments from 11 children with ASD, 23 children with non-ASD developmental concerns (e.g. speech-language delays), and 29 children with typical development, based on 36-month outcome (63 children total). Thirty-eight (60.32%) of the children depicted in VIRSA videos were male and 43 (68.25%) were Non-Hispanic Caucasian.

Analysis of ratings of VIRSA videos, using a generalized linear model with random effects for subjects and age, revealed a significant group effect (X^2=6.76, df=2, p<.05), with the ASD videos rated an average of 4.18 (95% CI = 2.81 to 5.55), the videos of children with non-ASD developmental concerns rated an average of 6.14 (95% CI = 5.19 to 7.09), and the videos of typically developing participants rated an average of 6.35 (95% CI = 5.47 to 7.22). Simple comparisons indicated that the ratings of the ASD videos differed significantly from the videos of both the non-ASD developmental concerns (t=2.31, p=.027) and typically developing cases (t=2.62, p=.013), who did not differ from one another (t=0.31, p=0.76), as expected. Since the VIRSA was designed specifically to detect the social-communication behaviors relevant to ASD, but not broader developmental delays, this pattern of results provided further validation of the final video pool.

The library of video segments was incorporated into a web-based application that presented pairs of videos, depicting differing degrees of social competence, side by side, accompanied by the prompt, "Which video is more like your child's interaction with you on a typical day?" On each trial, the video on the left played automatically, followed by the video on the right, at which point the viewer selected the one most like the child. Presentation of video followed an algorithm that always began with a pair of videos rated as 3 (less social) and 8 (more social) on the 10-point scale. After each choice, the algorithm selected and displayed a second pair of videos with new scale values contingent upon the previously chosen video's ranking. In each subsequent trial, the viewer's video choice dictated the range of sociability represented in the videos on the next trial, analogous to how optometrists help patients select eyeglass prescriptions. In this way, the algorithm presented videos of increasing similarity over subsequent trials until the distance between videos reduced to 1 rating scale point on 2 subsequent trials, at which point the average rating of the last 2 trials was recorded as the final score (see Figure S1 in online Appendix for examples).

The VIRSA web application was designed with a brief introductory video that oriented parents to the concepts and range of social behaviors depicted in the videos and provided rating instructions. The VIRSA app also asked for confirmation of the child's age in order to present videos from a matching age group. Since multiple video exemplars of each scale point (1 to 10) were available, videos were sampled from the pool without replacement.

The UC Davis Institutional Review Board approved the study procedures. Parents signed an informed consent form prior to participation. They completed VIRSA ratings when their child was 6-, 9-, 12-, and 18-months-old and again two weeks later to examine test-retest reliability. An automated email invited parents to the online VIRSA app, which could be accessed by computer or mobile device (e.g., smartphones, tablets). VIRSA ratings were always done prior to in-person assessments, which were conducted at 6, 12, 18, 24, and 36 months by examiners unaware of risk group or previous test results.

Participants

The VIRSA validation sample consisted of 110 infants (73 with an older sibling with ASD, 37 with no known family history of autism), none of whom supplied videos used in instrument development. Twenty-one children in the familial high-risk (HR) group received a diagnosis of ASD, whereas none of the low-risk (LR) children did. ASD diagnoses were made at any age that a child met DSM-5 criteria, based on all information available. One child was diagnosed with ASD at 12 months, 7 were diagnosed at 18 months, 7 at 24 months, and 6 at 36 months. All children diagnosed before 36 months retained the ASD diagnosis at the final visit. The rest of the sample was classified as Non-ASD and then stratified by familial risk to yield the HR Non-ASD and LR Non-ASD comparison groups. Descriptive statistics are shown in Table 1.

Measures

Mullen Scales of Early Learning (MSEL; Mullen, 1995) is an assessment of cognitive, motor, and language development for children aged 1 to 68 months. MSEL scores were used to describe the sample. Scores on the Fine Motor and Visual Reception subtests were used to

evaluate discriminant validity since these developmental domains are theoretically unrelated to the social-communicative focus of the VIRSA.

Autism Diagnostic Observation Schedule, 2nd edition (ADOS-2; Lord et al., 2012) is an observational measure that assesses ASD symptoms through semi-standardized interactions between the clinician and child. It is comprised of five modules appropriate for various ages and language levels; the Toddler module was administered at 18 and 24 months and either module 1 or 2 was used at 36 months, depending upon the child's verbal level. Analyses utilized the overall total algorithm (Social Affect + Restricted Repetitive Behavior or SARRB) score. Scores on the ADOS at 18 months were also used to examine convergent validity with VIRSA ratings at 18 months.

Analysis Plan

Psychometric properties of the VIRSA were examined in several ways. *Split-half reliability* was analyzed by comparing the first half of the ratings within a given session to the second half. *Test-retest reliability* was analyzed by comparing initial VIRSA ratings to those obtained two weeks later. Parents were shown the same series of paired videos they had seen two weeks earlier, instead of video pairs dictated by the VIRSA algorithm, permitting examination of the reliability of individual trial choices. *Convergent validity* was examined through correlations at 18 months with concurrent SARRB algorithm scores on the ADOS-2 Toddler module. *Discriminant validity* was examined by correlating VIRSA scores with concurrent MSEL fine motor and visual reception age equivalents. *Predictive validity* was assessed by examining diagnostic outcome group differences on the VIRSA scores at each age, using a mixed model with group, age, and their interaction included as fixed effects and VIRSA score as the time-varying dependent variable. We also used ROC analysis to assess the VIRSA's *sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV)* in predicting ASD diagnosis. *Area under the curve (AUC)* was computed as a measure of the ability to distinguish between groups. Sample sizes initially projected were 90 high-risk and

45 low-risk infants. With an anticipated ASD outcome rate of approximately 20% in the high-risk group, power was estimated to be .82 to detect an AUC value of .70 on the ROC analyses. All analyses were conducted in R, version 3.5.0.

Results

VIRSA trials (selections between paired videos) took an average of 56.49 seconds (range=38 to 133, SD=11.49). The average number of trials before a final score was reached was 7.67 (range=6 to 14, SD=1.26), with completion of the VIRSA taking an average of 7.21 minutes (range=3.82 to 15.57, SD=1.61). Split-half reliability was moderate, at r=.48, and test-retest reliability relatively strong, with 72% of the same video choices made two weeks later, which was significantly greater than chance (t=19.58, df=156, p<.0001). Convergent validity correlation with concurrent ADOS-2 SARRB algorithm scores at 18 months was r=-.36, which was significantly stronger than discriminant validity correlations with concurrent MSEL fine motor (r=.05, Fisher's z=2.50, p<.05) and visual reception (r=.10, Fisher's z=2.11, p<.05) age equivalent scores.

Figure 1 shows the modeled parent VIRSA scores for each group between 6 and 18 months. Examination of the model revealed a main effect for group (χ^2=10.32, df=2, p<.01). The main effect for age and the interaction between age and group were not significant. Planned contrasts revealed no significant differences in VIRSA scores between the ASD and comparison groups at 6 months (HR Non-ASD: t=0.71, p=0.48; LR Non-ASD: t=0.89, p=0.38) and 9 months (HR Non-ASD: t=1.75, p=.08; LR Non-ASD: t=1.46, p=0.14). At 12 months, VIRSA scores were significantly lower in the ASD group than both the HR Non-ASD (t=3.15, p=.002) and the LR Non-ASD groups (t=2.11, p=.04). At 18 months, VIRSA scores were significantly lower in the ASD group than in the HR Non-ASD group (t=3.29, p=.002) and marginally lower than the LR Non-ASD group (t=1.73, p=.08).

ROC analyses were conducted on VIRSA scores at each age, predicting a binary 36-month outcome (see Table 2). The threshold/cutoff at which ROC analyses best separated ASD

and Non-ASD cases was defined as the VIRSA score closest to the theoretical limit of maximum specificity and sensitivity. The VIRSA performed best at 18 months. Table 3 presents ROC models that examined how well VIRSA scores at 18 months predicted concurrent 18-month diagnoses (n=8 diagnosed with ASD at that age). Sensitivity was 100% (no false negatives), but specificity and positive predictive value were low.

Discussion

We hypothesized that employing video examples within a screening tool would enable detection of ASD in infancy. Starting at 12 months of age, VIRSA ratings were significantly lower for the group eventually diagnosed with ASD than for the comparison groups. Sensitivity of 18-month VIRSA scores in predicting 36-month diagnosis was approximately 0.80. This compares quite favorably to a study reporting sensitivity of 18-month clinical diagnostic assessment in predicting 36-month diagnosis of only 0.37 (Ozonoff et al., 2015). In fact, VIRSA scores had better sensitivity even at 6-12 months of age than that reported for clinical diagnosis at 18 months in Ozonoff et al. (2015). We hypothesize that the use of video allowed parents to "see" differences in their child that preceded the full onset of symptoms. The VIRSA's sensitivity is especially impressive given the extended time course of development of ASD symptoms. Multiple previous studies have demonstrated that symptoms slowly unfold over the first two years of life and many children who are ultimately diagnosed with ASD do not show overt signs before the first birthday (Gammer et al., 2015; Landa & Garrett-Mayer, 2006; Ozonoff et al., 2010; Zwaigenbaum et al., 2005). Children who are not identified by the VIRSA may not yet be showing signs of ASD for parents to rate.

We also examined the VIRSA's ability to index concurrent symptoms. ROC analyses of VIRSA ratings at 18 months identified all eight children who had been diagnosed with ASD by that age, with no false negatives. A recent meta-analysis of the accuracy of ASD screeners between 14 and 36 months of age (Sanchez-Garcia et al., 2019) reported a pooled sensitivity of 0.72 and specificity of 0.98. The sensitivity of the VIRSA at 18 months is thus comparable to or

better than existing measures and suggests that it may be a useful adjunct in identifying toddlers in need of referral for an ASD evaluation. Its specificity and positive predictive value, however, were lower than recommended standards (Cicchetti et al., 1995), resulting in over-identification of risk. For this reason, the present results do not support use of the VIRSA as a stand-alone ASD screener in infancy yet. Future studies could examine whether using the VIRSA as an initial step in a two-stage screening process improves accuracy.

In addition to the predictive validity of the VIRSA, we examined a number of other psychometric properties. Test-retest reliability was strong, with parents selecting over 70% of the same videos when they retook the VIRSA two weeks later. VIRSA scores at 18 months were significantly correlated with ADOS-2 scores and correlations were significantly higher than with measures of divergent abilities (e.g., fine motor and visual reception skills).

The majority of participants, and all the children who developed ASD, came from the high-risk group (e.g., had older siblings with ASD). It is imperative, prior to recommending the VIRSA for clinical use, to examine its psychometric properties when used by parents who are naïve to ASD. The positive predictive value of an instrument is dependent upon the base rate of the condition in the population (Clark & Harrington, 1999; Grimes & Schulz, 2002) and thus the VIRSA's predictive ability may be reduced in a community-based sample that has a lower prevalence of ASD than in high-risk families. Studies are currently underway in our laboratory to determine whether the present results generalize to low-risk samples.

Despite these limitations, the VIRSA makes several contributions to the literature. First, it demonstrates that it is possible to develop a parent report instrument capable of identifying ASD risk in the first year of life. Second, it demonstrates that video can be used to clarify developmental phenomena and improve parent reporting of early development. Finally, an innovation of the VIRSA is its web-based, mobile-optimized application. Over 90% of American adults of childbearing age own a smartphone, with rates over 65% even in lower income, rural, and minority communities (Pew Research Center, 2018). It is vital that screening procedures

keep pace with such advances in technology and society's increasingly internet-based preferences for information acquisition and communication. Thus, this study provides an initial step in the proof of principle of video- and web-based screening for ASD. With further development, the VIRSA, with its low-burden, quick, online ratings, has potential to reduce disparities in communities with limited access to screening and provide the possibility of initiating intervention before the full symptom set of ASD has emerged.

Key Points

1. Signs of ASD are present in the first two years of life, but the average age of diagnosis lags far behind. Instruments that improve detection of autism risk in infancy are needed.

2. We hypothesized that employing video examples within a screening tool would improve detection of ASD in infancy.

3. A newly developed video-based screening tool had high sensitivity at 18 months in concurrently identifying the toddlers diagnosed with ASD at that age, as well as predicting ASD at 36 months.

4. Employing video examples within a screening tool may be helpful in identifying ASD in infancy. A brief, low-burden, web-based screening tool could help reduce disparities in communities with limited access to care.

Acknowledgments

Correspondence

Sally Ozonoff, MIND Institute, UC Davis Health, 2825 50th Street, Sacramento CA 95817; 916-703-0259; sozonoff@ucdavis.edu

References

Al-Qabandi M, Gorter JW, Rosenbaum P. (2011). Early autism detection: Are we ready for routine screening? *Pediatrics, 128(1),* e211-e217.

Arguel A, Jamet E. (2009). Using video and static pictures to improve learning of procedural contents. *Computers in Human Behavior, 25,* 354-359.

Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, Kurzius-Spencer M, Zahorodny W, Robinson-Rosenberg C, White T, Durkin MS, Imm P, Nikolaou L, Yeargin-Allsopp M, Lee LC, Harrington R, Lopez M, Fitzgerald RT, Hewitt A, Pettygrove S, Constantino JN, Vehorn A, Shenouda , Hall-Lande J, Van Naarden-Braun K, Dowling NF. (2018). Prevalence of autism spectrum disorder among children aged 8 years: autism and developmental disabilities monitoring network, 11 sites, United States 2014. *MMWR Surveillance Summary, 67(6),* 1-23.

Barger BD, Campbell JM, McDonough JD. (2013). Prevalence and onset of regression within autism spectrum disorders: A meta-analytic review. *Journal of Autism and Developmental Disorders, 43(4),* 817-828.

Chawarska K, Paul R, Klin A, Hannigen S, Dichtel LE, Volkmar F. (2007). Parental recognition of developmental problems in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37(1),* 62-72.

Cicchetti DV, Volkmar F, Klin A, Showalter D. (1995). Diagnosing autism using ICD-10 criteria: A comparison of neural networks and standard multivariate procedures. *Child Neuropsychology, 1,* 26–37.

Clark, A., & Harrington, R. (1999). On diagnosing rare disorders rarely: Appropriate use of screening instruments. *Journal of Child Psychology and Psychiatry, 40(2),* 287-290.

Gammer I, Bedford R, Elsabbagh M, Garwood H, Pasco G, Tucker L, Volein A, Johnson MH, Charman T, BASIS Team. (2015). Behavioral markers for autism in infancy: Scores on the

Autism Observational Scale for Infants in a prospective study of at-risk siblings. *Infant Behavior and Development, 38,* 107-115.

Gangi, D.N., Boterberg, S., Schwichtenberg, A.J., Solis, E., Young, G.S., Iosif, A., & Ozonoff, S. (2019). Use of prospective longitudinal gaze measurements in defining regression. Paper presented at the annual meeting of the International Society for Autism Research, Montreal, May.

Grimes, D.A., & Schulz, K.F. (2002). Uses and abuses of screening tests. *Lancet, 359,* 881-884.

Khowaja MK, Hazzard AP, Robins DL. (2015). Sociodemographic barriers to early detection of autism: Screening and evaluation using the M-CHAT, M-CHAT-R, and follow-up. *Journal of Autism and Developmental Disorders, 45(6),* 1797-1808.

Koriat A, Goldsmith M, Pansky A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51,* 481-537.

Krosnick JA, Presser S. (2010). Question and questionnaire design. In PV Marsden, JD Wright (Eds.), *Handbook of Survey Research, 2nd edition* (pp. 263-314). London: Emerald.

Landa R, Garrett-Mayer E. (2006). Development in infants with autism spectrum disorders: A prospective study. *Journal of Child Psychology and Psychiatry, 47(6),* 629-638.

Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Bishop SL. (2012). *Autism Diagnostic Observation Schedule Manual, 2nd edition.* Torrance, CA: Western Psychological Services.

Maestro S, Muratori F, Cavallaro MC, Pei F, Stern D, Golse B, Palacio-Espasa F. (2002). Attentional skills during the first six months of age in autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 41(10),* 1239-1245.

Marrus N, Glowinski AL, Jacob T, Klin A, Jones W, Drain CE, Holzhauer KE, Hariprasad V, Fitzgerald RT, Mortensen EL, Sant SM, Cole L, Siegel SA, Zhang Y, Agrawal A, Heath AC, Constantino JN. (2015). Rapid video-referenced ratings of reciprocal social behavior in toddlers: A twin study. *Journal of Child Psychology and Psychiatry, 56(12),* 1338-1346.

Miller M, Iosif A, Hill M, Young GS, Schwichtenberg AJ, Ozonoff S. (2017). Response to name in infants developing autism spectrum disorder: A prospective study. *Journal of Pediatrics, 183,* 141-146.

Mullen EM. (1995). *Mullen Scales of Early Learning*. Circle Pines, MN: AGS Publishing.

Ozonoff S, Iosif AM, Baguio F, Cook IC, Hill MM, Hutman T, Rogers SJ, Rozga A, Sangha S, Sigman M, Steinfeld MB, Young GS. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child and Adolescent Psychiatry, 49(3),* 256-266.

Ozonoff S, Iosif A, Young GS, Hepburn S, Thompson M, Colombi C, Cook IC, Werner E, Goldring S, Baguio F, Rogers S. (2011). Onset patterns in autism: Correspondence between home video and parent report. *Journal of the American Academy of Child and Adolescent Psychiatry, 50(8),* 796-806.

Ozonoff S, Young GS, Landa RJ, Brian J, Bryson S, Charman T. Chawarska K, Macari SL, Messinger D, Stone WL, Zwaigenbaum L, Iosif AM. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: A baby siblings research consortium study. *Journal of Child Psychology and Psychiatry, 56(9),* 988-998.

Pew Research Center. (2018). Demographics of mobile device ownership and adoption in the United States. http://www.pewinternet.org/fact-sheet/mobile/. Accessed March 15, 2019.

Pierce K, Carter C, Weinfeld M, Desmond J, Hazin R, Bjork R, Gallagher N. (2011). Detecting, studying, and treating autism early: The one-year well-baby check-up approach. *Journal of Pediatrics, 159(3),* 458-465.

Sanchez-Garcia, A.B., Galindo-Villardon, P., Nieto-Librero, A.B., Martin-Rodero, H., & Robins, D.L. (2019). Toddler screening accuracy for autism spectrum disorder: A meta-analysis of diagnostic accuracy. *Journal of Autism and Developmental Disorders, 49(5),* 1837-1852.

Scarpa A, Reyes NM, Patriquin MA, Lorenzi J, Hassenfeldt TA, Desai VJ Kerkering KW. (2013). The modified checklist for autism in toddlers: Reliability in a diverse rural American sample. *Journal of Autism and Developmental Disorders, 43(10),* 2269-2279.

Schwichtenberg AJ, Kellerman A, Miller M, Young GS, Ozonoff S. (2019). Mothers of children with autism spectrum disorder: Play behaviors with infant siblings and social responsiveness. *Autism* [epub ahead of print]. doi: 10.1177/1362361318782220.

Werner E, Dawson G, Osterling J, Dinno N. (2000). Recognition of autism spectrum disorder before one year of age: A retrospective study based on home videotapes. *Journal of Autism and Developmental Disorders, 30(2),*157-162.

Wetherby AM, Brosnan-Maddox S, Peace V, Newton L. (2008). Validation of the infant-toddler checklist as a broadbrand screener for autism spectrum disorders from 9 to 24 months of age. *Autism, 12(5),* 487-511.

Zwaigenbaum L, Bryson S, Rogers T, Roberts W, Brian J, Szatmari P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience, 23(2-3),* 143-152.

Zwaigenbaum L, Bauman ML, Fein D, Pierce K, Buie T, Davis PA, Newschaffer C, Robins DL, Wetherby A, Choueiri R, Kasari C, Stone WL, Yirmiya N, Estes A, Hansen RL, McPartland JC, Natowicz MR, Carter A, Granpeesheh D, Mailloux Z, Smith-Roley S, Wagner S. (2015). Early screening of autism spectrum disorder: Recommendations for practice and research. *Pediatrics, 136(Suppl 1),* S41-S59.

Table 1: Sample Descriptives.

	ASD	HR Non-ASD	LR Non-ASD
Sample size	21	52	37
Sex (% male)	61.91%[a]	40.39%[a]	59.46%[a]
Race/Ethnicity			
% Non-Hispanic Caucasian	36.84%[a]	42.00%[a]	62.16%[a]
% Non-White Race or Multiracial	21.05%[a]	38.00%[a]	27.03%[a]
% Hispanic	42.11%[a]	20.00%[a]	10.81%[b]
Maternal education			
% Graduate degree	14.29%[a]	48.08%[b]	35.14%[ab]
% College degree	57.14%[a]	40.39%[a]	59.46%[a]
% High school or Vocational training	19.05%[a]	9.62%[a]	5.41%[a]
Household income			
% $60k or less	19.05%[a]	7.69%[a]	16.22%[a]
% $61k to $100k	33.33%[a]	25.00%[a]	18.92%[a]
% $101k or higher	23.81%[a]	46.15%[a]	51.35%[a]
Age at outcome (months)	36.39 (0.72)[a]	36.92 (1.82)[a]	36.61 (0.84)[a]
ADOS-2 SARRB[*] algorithm score at 36 months	14.95 (6.25)[a]	3.08 (2.31)[b]	2.24 (2.02)[b]
MSEL outcome fine motor age eq	26.35 (6.76)[a]	33.86 (4.79)[b]	37.11 (5.48)[c]
MSEL outcome visual reception age eq	29.45 (8.93)[a]	41.58 (7.58)[b]	42.89 (6.57)[b]
MSEL outcome expressive language age	26.25 (9.53)[a]	38.14 (4.97)[b]	39.54 (4.60)[b]
MSEL outcome receptive language age eq	25.60 (9.55)[a]	35.66 (5.68)[b]	37.51 (4.78)[b]

Note:

Values with different subscripts are significantly different at $p<.05$

* Social Affect + Restrictive Repetitive Behavior overall total

MSEL outcome = Mullen Scales of Early Learning at 36 months of age

Table 2: ROC analyses with 36-month ASD diagnostic classification.

	6 months	9 months	12 months	18 months
True positives	7	10	8	14
False positives	11	37	31	31
True negatives	36	29	51	35
False negatives	6	6	8	4
AUC	.62	.42	.59	.71
Specificity (95% CI)	.77 (.65 to .89)	.44 (.32 to .56)	.62 (.52 to .73)	.53 (.41 to .65)
Sensitivity (95% CI)	.54 (.26 to .81)	.63 (.39 to .86)	.50 (.26 to .75)	.78 (.59 to .97)
Negative Predictive Value (95% CI)	.86 (.75 to .96)	.83 (.70 to .95)	.86 (.78 to .95)	.90 (.80 to .99)
Positive Predictive Value (95% CI)	.39 (.16 to .61)	.21 (.10 to .33)	.21 (.08 to .33)	.31 (.18 to .45)
Threshold	5.25	8.25	6.75	8.25

Table 3: ROC analyses for 18-month VIRSA with concurrent 18-month diagnosis.

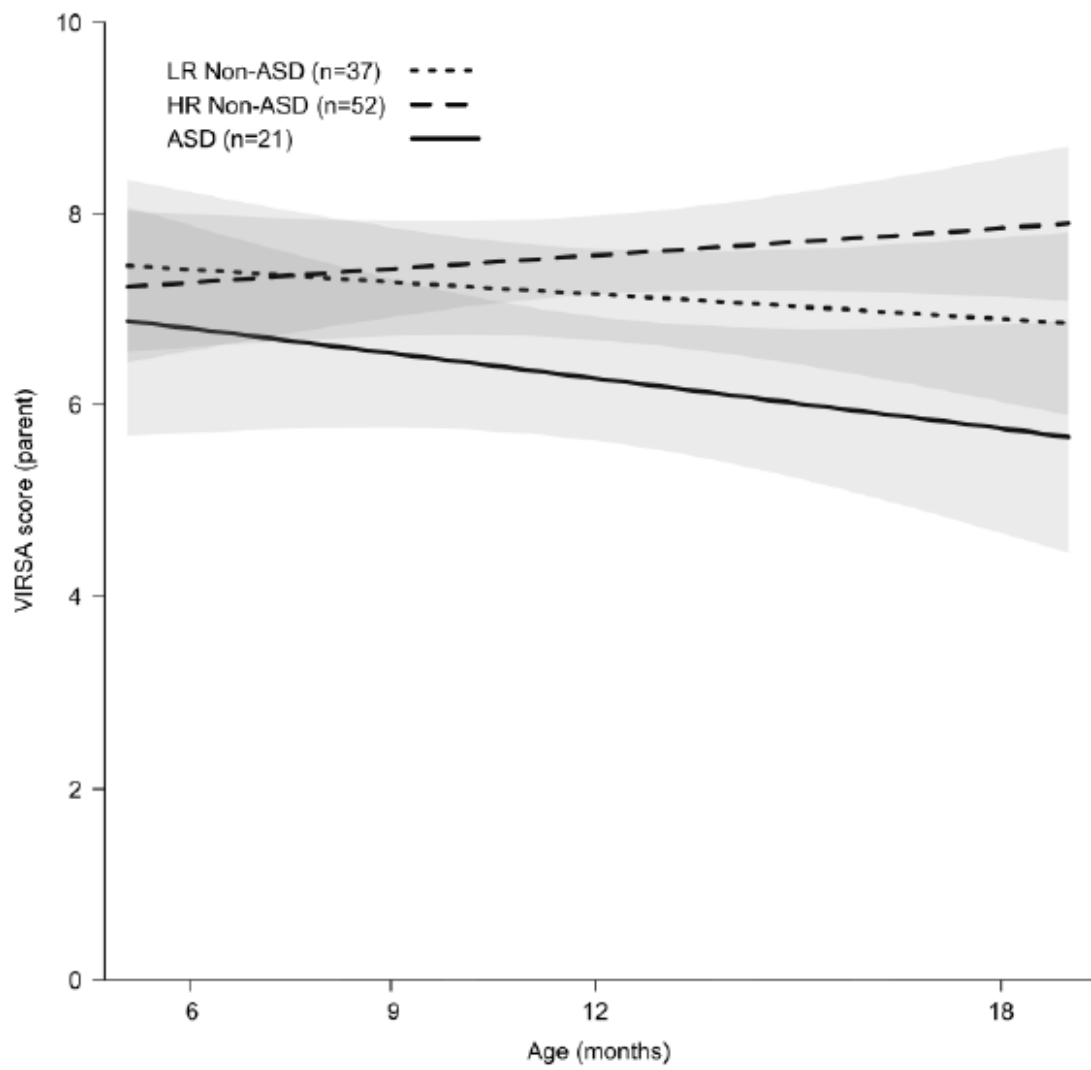|	VIRSA (18 months)
True positives	8
False positives	34
True negatives	38
False negatives	0
AUC	.78
Specificity	.53
Sensitivity	1.00
Negative Predictive Value	1.00
Positive Predictive Value	.19
Threshold	7.75

Figure 1: VIRSA ratings by group from 6 to 18 months.

Figure S1: VIRSA search example.