AJSLP

Research Note

Early Observation of Red Flags in 12-Month-Old Infant Siblings Later Diagnosed With Autism Spectrum Disorder

Moira L. Pileggi,[a] Natalie Brane,[a] Jessica Bradshaw,[b] Abigail Delehanty,[c] Taylor Day,[d] Courtney McCracken,[e] Jennifer Stapel-Wax,[e] and Amy M. Wetherby[f]

Purpose: Valid and reliable screening tools are needed to improve early detection and optimize developmental outcomes for toddlers at risk for autism spectrum disorder (ASD). The current study aimed to evaluate the utility of the Systematic Observation of Red Flags (SORF) for ASD at 12 months of age in a sample of high-risk infant siblings of children with ASD.
Method: As part of a prospective, longitudinal study, we examined the sensitivity and specificity of the SORF at 12 months for predicting a diagnosis of ASD at 24 months in a sample of 122 infants, 31 of whom were diagnosed with ASD.
Results: The optimal SORF Composite cutoff score of 18 correctly identified 24 of the 31 twelve-month-olds who were diagnosed with ASD, yielding a sensitivity of .77 and a specificity of .76. The optimal SORF Red Flags cutoff score of 7 correctly identified 20 of the 31 infants, yielding a sensitivity of .65 and a specificity of .75.
Conclusion: This preliminary study demonstrates the potential of the SORF as an effective observational screening measure for 12-month-olds at risk for ASD with good discrimination, sensitivity, and specificity.

With a prevalence of one in 54 (Maenner et al., 2020) and estimated heritability between 50% and 87% (Bai et al., 2019; Sandin et al., 2017), autism spectrum disorder (ASD) is one of the most prevalent and heritable developmental disabilities affecting early childhood. Compared to the general population, younger siblings of children with ASD have a 12-fold increased risk for developing ASD and increased risk rates for related developmental delays, particularly within the domains of social communication and language (Constantino et al., 2006; Ingersoll & Wainer, 2014; Micheletti et al., 2019; Ozonoff et al., 2011).

Prospective studies of younger siblings of children with ASD and children ascertained through general population screening continue to demonstrate that core ASD symptoms (i.e., social interaction deficits, restricted and repetitive behaviors) may be detected as early as 12–14 months old (Bryson et al., 2008; Miller et al., 2017; Pierce et al., 2019; Watson et al., 2013). Delays in verbal and nonverbal communication development may be evident as early as 12–24 months old (Chawarska et al., 2014; Rowberry et al., 2015; Yoder et al., 2009; Zwaigenbaum et al., 2005). Atypical sensory and motor features as well as restricted and repetitive behaviors, such as visual inspection of and repetitive behavior with objects, have been documented in infants between 6 and 12 months old (Zwaigenbaum et al., 2013). These findings suggest that, for at least a subset of high-risk (HR) infants, core features of ASD may be detected by the first birthday; thus, an opportunity exists to screen for the early signs of ASD in infants as young as 12 months old.

[a]Marcus Autism Center, Children's Healthcare of Atlanta, GA
[b]Department of Psychology, University of South Carolina, Columbia
[c]Department of Speech-Language Pathology, Duquesne University, Pittsburgh, PA
[d]Department of Psychology, Florida State University, Tallahassee
[e]Department of Pediatrics, Emory University School of Medicine, Atlanta, GA
[f]Department of Clinical Sciences, Florida State University, Tallahassee
Correspondence to Moira L. Pileggi: moira.lewis@choa.org

Editor-in-Chief: Julie Barkmeier-Kraemer
Editor: Erinn H. Finke
Received June 9, 2020
Revision received September 19, 2020
Accepted December 17, 2020
https://doi.org/10.1044/2020_AJSLP-20-00165

Disclosure: *Amy M. Wetherby is a coauthor of the Communication and Symbolic Behavior Scales and receives royalties for this assessment. All other authors have declared that no competing interests existed at the time of publication.*

Observational measures administered by trained clinicians may provide certain advantages over parent report tools for identifying overt and subtle risk factors for ASD at younger ages, including joint attention, gaze shifts, and repetitive vocalizations (Bryson et al., 2008; Stone et al., 2008). At present, however, there are limited interactive observational screening tools for children at HR for ASD with adequate psychometric properties (Bryson & Zwaigenbaum, 2014; Zwaigenbaum et al., 2015). Furthermore, very few exist for differentiating ASD from typical development in infants prior to the second birthday (Barton et al., 2012). To fill a gap in existing early detection measures, Wetherby et al. (2004, 2016) developed the Systematic Observation of Red Flags (SORF) for ASD, an observational tool designed to detect and quantify ASD symptoms in young children with the goal of identifying children who should be referred for a comprehensive diagnostic evaluation. Although the SORF was developed for use with young children at risk for ASD and other developmental delays, it has not yet been studied in an infant sample.

The current study aimed to evaluate the utility of the SORF at 12 months in a population of HR infant siblings of children with ASD and a low-risk (LR) comparison group. As part of a larger prospective, longitudinal study of infant siblings of children with ASD, we examined the sensitivity and specificity of the SORF at 12 months for predicting a diagnosis of ASD at 24 months and evaluated the association between 12-month SORF scores and 24-month phenotypic outcomes with regard to developmental level, social communication, and autism symptoms. By screening for early signs of ASD in HR infants, we may "push the needle" toward earlier identification, lowering the age of detection and onset of treatment in this group.

Method

Participants

Participants in this study were enrolled in a longitudinal, prospective study examining developmental trajectories of infant siblings (National Institute of Mental Health Autism Centers of Excellence study, Principal Investigator: A. Klin). Study procedures were approved by the Emory University Institutional Review Board, and parents gave written, informed consent prior to their participation. Participants were either HR or LR for ASD. HR infants had an older full-biological sibling with ASD, confirmed through clinical review of a diagnostic evaluation report signed by a licensed clinical or school psychologist or medical provider; had scores within the ASD range on the Social Responsiveness Scale–Second Edition (Constantino, 2013); and met the cutoff on the Social Communication Questionnaire (Rutter et al., 2003). LR participants had no familial history of ASD in first- or second-degree relatives. Exclusion criteria for both HR and LR infants included gestational age below 35 weeks, significant hearing and/or visual impairment, a nonfebrile seizure disorder, a known genetic syndrome, and significant pre- or perinatal complications.

All participants who were enrolled in the larger longitudinal study and had completed a communication and developmental assessment at 12 months and a comprehensive diagnostic evaluation at 24 months (N = 156) were further evaluated for diagnostic inclusion criteria. Only participants who were diagnosed with ASD (n = 31) or determined to be typically developing (TD; n = 91) were included in this study. Participants were excluded if they exhibited subclinical features of ASD (n = 18), non-ASD developmental delay (n = 14), or significant clinical features consistent with anxiety (n = 2; see Table 1).

Measures

Communication and Symbolic Behavior Scales–Developmental Profile

The Communication and Symbolic Behavior Scales (CSBS)–Developmental Profile Behavior Sample (Wetherby & Prizant, 2002) was administered by licensed speech-language pathologists (SLPs) with expertise in infant development and ASD at 12 and 24 months. The CSBS is a standardized, norm-referenced instrument designed to measure early communication development. The CSBS takes approximately 20–30 min to administer and consists of object temptation activities and a symbolic play probe. The measure is designed to create interactive opportunities for toddlers to direct communication to the examiner and caregiver, thereby yielding a sample of child communicative behavior within a semistructured setting. Interrater reliability of the SLPs' coding was established with the completion of 25 CSBS training videos. Generalizability (g) coefficients or intraclass correlations measure the source and magnitude of variance accounted for by the subjects and the raters and were used to calculate interrater reliability, as used in similar research (McCathren et al., 2000; McWilliam & Ware, 1994; Morgan et al., 2020; Wetherby & Prizant, 2002). Compared to expert raters, SLPs' independent rating of the 25 training videos demonstrated g coefficients above 0.90 on the Social, Speech, and Symbolic composites and the total score, which are in the excellent range. The CSBS is often utilized for infants and toddlers with suspected language or social communication delays in clinic settings, allowing for the SORF (described below) to be coded without requiring an additional evaluation.

SORF

The SORF (Wetherby et al., 2016) is a 22-item observational measure designed to detect 22 specific early indicators for ASD in toddlers based on current diagnostic criteria (*Diagnostic and Statistical Manual of Mental Disorders* [5th ed.; *DSM-5*; American Psychiatric Association, 2013]). Eleven items from each diagnostic domain —Impairment in Social Communication and Social Interaction (SC) and Restricted and Repetitive Behaviors (RRB) —are included. Items are organized by symptom subdomains, for example, with deficits in social–emotional reciprocity encompassing SORF items associated with sharing smiles, facial expressions, sharing interests, and response to name (see Appendix for a list of the SORF 22 red flag items).

Table 1. Participant demographics.

Variable	ASD (n = 31)[a,b]	TD (n = 91)[a,c]	Statistic	Effect size[d]
Sex			$\chi^2(1) = 3.3$	0.33
Male	24 (77%)	54 (59%)		
Female	7 (23%)	37 (41%)		
Race/ethnicity[e]			$\chi^2(1) = 5.7^*$	0.44
White	22 (71%)	81 (89%)		
Black/African American	7 (23%)	4 (4%)		
Asian	0 (0%)	1 (1%)		
More than one race	2 (7%)	5 (5%)		
Maternal education[f]			$\chi^2(2) = 17.2^{***}$	0.81
Less than college	6 (20%)	2 (2%)		
College graduate	16 (53%)	34 (37%)		
Postcollege	8 (27%)	55 (60%)		
Mullen Scales of Early Learning				
Expressive Language	38.29 (12.78)	57.84 (11.26)	$t(120) = -8.06^{***}$	1.69
Receptive Language	36.55 (15.32)	58.21 (6.35)	$t(120) = -11.05^{***}$	2.32
Fine Motor	46.68 (10.07)	55.32 (9.05)	$t(120) = -4.46^{***}$	0.94
Visual Reception	49.03 (13.04)	61.20 (10.43)	$t(120) = -5.25^{***}$	1.10
ADOS calibrated severity score				
Social Affect	4.27 (0.91)	1.92 (1.09)	$Z = 7.21^{***}$	2.27
Restricted/Repetitive Behavior	7.52 (1.72)	3.95 (2.33)	$Z = 6.92^{***}$	1.65
Total	6.70 (2.38)	1.65 (1.03)	$Z = 8.25^{***}$	3.47

Note. ASD = autism spectrum disorder; TD = typically developing; ADOS = Autism Diagnostic Observation Schedule.

[a]Diagnosis determined using clinical best estimate diagnosis at 24 months. [b]Includes n = 28 high-risk (HR) infants and n = 3 low-risk (LR) infants. [c]Includes n = 17 HR infants and n = 74 LR infants. [d]Cohen's d. [e]Minority groups (Black/African American, Asian, and more than one race) combined for between-group comparisons. [f]One participant in the ASD group did not report maternal education.

$^*p < .05$. $^{***}p < .001$.

While watching the CSBS via video review, the SORF can be scored by rating each SORF item based upon social communication and behavior observed in each of six CSBS activities. For example, the SLP collects data on a worksheet to code the frequency and variety of facial expressions, noting and/or tallying behaviors such as shared smiles used during the CSBS or occurrence of hand/finger posturing. For each item, a 4-point (0–3) rating system is applied, in which a zero indicates either the presence of typical behavior (e.g., frequent shared smiles) or the absence of atypical behavior (e.g., no clear repetitive behaviors). Some items capture an absence or limited frequency of typical behavior (e.g., vocalizations with consonants), whereas others measure the presence of an unusual or atypical behavior (e.g., repetitive movements with objects). For example, for Item 1, *Limited sharing warm, joyful expressions,* a rating of 0 indicates joyful smiles with directed gaze were shared during at least four activities, whereas a rating of 1 indicates joyful smiles occurred during three activities, 2 indicates joyful smiles occurred in one or two activities, and 3 indicates no clear and joyful smiles shared. The SORF allows for calculation of SC and RRB Domain scores, which are the sums of item scores in each diagnostic domain, and a Total Composite score. Higher domain and Total Composite scores indicate more early indicators and a greater level of ASD concern. Additionally, SORF codes that indicate clear symptom presence, based on an item-level code of 2 or 3, are collapsed to yield a count of the total number of red flags, which is informative for describing and quantifying clinically significant behaviors observed. Dow et al. (2017) validated the SORF in a large community-based sample of toddlers (N = 247) aged 16–24 months who received early communication assessments. Results suggested that the SORF has good discrimination, sensitivity, and specificity for early identification of ASD as young as 16 months of age. They found 17 of the 22 SORF items to have medium-to-large effect sizes and area under the curve (AUC) values of at least .60 when differentiating ASD from non-ASD. Subsequently, an algorithm was developed to include only those items to compute the Total Composite and SC and RRB Domain scores, while all 22 items were utilized to derive the number of red flags. Building on these results, the current study followed the same scoring algorithm.

In the current study, at the conclusion of the assessment, the SLP who administered the CSBS immediately scored and coded both the CSBS and the SORF via video recording of the CSBS to rate the presence of red flag items across the SC and RRB Domains. All SLPs were trained by an author of the CSBS and SORF (A. M. Wetherby) and trained, reliable coders at the Florida State University (FSU) FIRST WORDS Project. Training in administration and scoring was done via weekly videoconference meetings and within the first year of the study and via ongoing consultation with trained coders at the FSU FIRST WORDS Project as needed. Additionally, SLPs were previously trained on the CSBS and Autism Diagnostic Observation Schedule (ADOS), as well as the core diagnostic features of ASD and early detection, including through the Autism Navigator About Autism in Toddlers and ASD Video Glossary, which are web-based tools on the early signs of ASD with video illustrations of toddlers with ASD. SLPs first rated 22 CSBS training videos on children

16–24 months of age to establish interrater agreement at 85% or higher on 10 consecutive videos. This study was the first to code the SORF on children 12 months of age; therefore, for this study, consensus coding of 12-month-old videos was used after independent rating of the SORF by the evaluating SLP and an independent, research-reliable SLP at the FSU FIRST WORDS Project to allow discussion of ratings at this young age. Independent SLPs at FSU were blind to risk status of all participants during this subsequent CSBS video observation and SORF scoring process, allowing for blind rating of 12-month SORFs.

Mullen Scales of Early Learning

The Mullen Scales of Early Learning (or Mullen; Mullen, 1995) was administered at 12 and 24 months. The Mullen is considered a gold standard measure of early nonverbal cognitive, motor, and language development and is normed from birth through 68 months old. It provides *T* scores and age equivalencies for five domains based on child performance relative to the norming sample: Gross Motor, Visual Receptive, Fine Motor, Receptive Language, and Expressive Language. *T* scores have a mean of 50 and an *SD* of 10. Corrected age was used to calculate *T* scores for infants born younger than 37 weeks of gestation.

ADOS–Second Edition

The ADOS–Second Edition is a semistructured, standardized assessment of communication, social interaction, and restricted and repetitive behaviors designed for identification of behaviors consistent with ASD. The ADOS-Toddler Module was designed specifically for children 12–30 months old with limited language and was administered at the 24-month diagnostic evaluation. The calibrated severity score was calculated in an effort to reduce the effect of verbal language abilities when capturing a metric of ASD symptom severity at this young age (Esler et al., 2015).

Vineland II

The Vineland Adaptive Behavior Scales–Survey Interview Form (Sparrow et al., 2005) was administered to parents at the 12- and 24-month evaluations. The Socialization, Communication, Daily Living Skills, and Motor domains, as well as the overall Adaptive Behavior composite score, are included in analyses.

Clinical Best Estimate Diagnosis

At the 24-month visit, a clinical best estimate (CBE) diagnosis was determined by the diagnostic team (clinical psychologist and SLP) based on the results of the gold standard ASD diagnostic evaluation described above and according to *DSM-5* criteria. Based on the 24-month CBE, children classified as having ASD and as TD were included in this sample. A judgment of being TD was made if the child did not show early signs for ASD during evaluation and if all scores from the Mullen were in the average range. Clinical psychologists who administered the 24-month diagnostic battery and made the CBE were blind to participant risk status at the outcome assessment; however, SLPs who administered the CSBS and rated the SORF could not be blinded as they needed to facilitate clinical care across longitudinal visits.

Statistical Analyses

Descriptive statistics were calculated for all variables of interest and included means and standard deviations or counts and percentages, as appropriate. Chi-square tests of independence for categorical variables and independent two-sample *t* tests for continuous variables were utilized to determine if there were significant differences between participants with ASD and TD participants on demographic and outcome variables. Multivariable regression models were used to determine if SORF scores at 12 months differed between participants with ASD and TD participants while adjusting for maternal education. Results are presented as model-based means with associated 95% confidence intervals adjusted for the covariates in the model. Associations between 12-month SORF scores and 12- and 24-month developmental scores were evaluated using Pearson correlations for normally distributed variables or Spearman correlations for nonnormally distributed variables. To determine the diagnostic utility of the 12-month SORF scores, sensitivity (true-positive rate), specificity (true-negative rate), and the AUC were estimated for all SORF domain and total scores as well as domain and total red flags, and 95% confidence intervals are provided for each metric. AUC is a common metric of classification to determine how well a model distinguishes between classes, for example, ASD versus TD. An AUC of about .5 indicates no discrimination capacity, and 1 indicates perfect discriminability. Optimal cutoff scores were selected by prioritizing sensitivity while maintaining an adequate level of specificity. False discovery rate–adjusted *p* values were used to account for multiple comparisons, when applicable. Effect sizes were calculated to aid in clinical interpretation and are presented as standardized differences using Cohen's *d* as the common effect size. Significance was assessed at the .05 level, unless otherwise noted, and all tests were two-sided. Analyses were conducted using SAS 9.4.

Results

Descriptive statistics for participants are presented in Table 1. Significant differences with regard to race/ethnicity and maternal education were found between ASD and TD; therefore, maternal education was included as a covariate in subsequent general linear models. On average, infants who were later diagnosed with ASD scored significantly lower than the TD group on measures of verbal and nonverbal skills and significantly higher on a measure of autism symptomology.

Model-based estimates of SORF total scores from the 12-month evaluation, controlling for maternal education, are displayed in Table 2. Infants with ASD scored significantly higher than TD infants on all six SORF scores. The SC Domain score and SC Red Flags showed the largest

Table 2. Twelve-month Systematic Observation of Red Flags model-based means.

Domain	ASD (n = 31) Mean [95% CI]	TD (n = 91) Mean [95% CI]	Test statistic $F(1, 120)$	Effect size[a]
Composite score	21.46 [18.87, 24.05]	11.43 [9.25, 13.62]	41.70***	1.19
SC Domain score	17.62 [15.64, 19.60]	9.36 [7.69, 11.04]	48.16***	1.28
RRB Domain score	3.84 [2.84, 4.84]	2.07 [1.22, 2.91]	8.77**	0.55
Total Red Flags	8.09 [7.07, 9.11]	4.01 [3.14, 4.87]	44.45***	1.24
SC Red Flags	6.53 [5.77, 7.29]	3.13 [2.49, 3.78]	55.23***	1.37
RRB Red Flags	1.57 [1.07, 2.06]	0.87 [0.45, 1.19]	5.50*	0.43

Note. ASD = autism spectrum disorder; TD = typically developing; CI = confidence interval; SC = Social Communication and Social Interaction; RRB = Restricted and Repetitive Behaviors.

[a]Cohen's *d*.

*$p < .05$. **$p < .01$. ***$p < .001$.

effects, $F(3, 120) = 48.16$, $p < .0001$, and $F(3, 117) = 55.23$, $p < .0001$, respectively. With regard to associations among 12-month SORF scores and concurrent developmental measures (see Table 3), significant, large correlations ($r > .60$) were observed between the SORF (SC Domain, Composite, SC Red Flags, Total Red Flags) and the CSBS Total Score and CSBS Social composite. Significant, medium correlations (all rs $> .40$) were observed between the SORF (SC Domain, Composite, SC Red Flags, Total Red Flags) and Mullen Expressive Language and CSBS Speech. Small but significant correlations were found among the SORF RRB Domain score ($r > .20$–$.33$) and Mullen Receptive Language, all CSBS scores, and Vineland Communication. Significant associations were not observed between the SORF and the Mullen Visual Reception, Vineland Socialization, or Vineland Motor Skills scores.

Optimal cutoff scores for the SORF at 12 months were determined based on the psychometric features relative to the ASD diagnosis at 24 months and are displayed in Table 4. The highest AUC, sensitivity, and specificity was observed for the SORF Composite cutoff of 18, which correctly identified 24 of the 31 infants who were later diagnosed with ASD and yielded a sensitivity of .77 and a specificity of .76. The optimal cutoff of SORF Red Flags was 7, with 20 of the 31 infants having seven or more red flags at 12 months, yielding a sensitivity of .65 and a specificity of .75. The cutoff for the SC Domain score was 15, with a sensitivity of .74 and a specificity of .73, and that for SC Red Flags was 6, with a sensitivity of .68 and a specificity of .82. Finally, the optimal cutoff was 3 for the RRB Domain score (sensitivity of .68, specificity of .55) and 2 for RRB Red Flags (sensitivity of .45, specificity of .67).

Associations between 12-month SORF scores and 24-month developmental and autism severity outcome are presented in Table 5. SORF Total and SC Domain scores were significantly associated with most outcome measures.

Table 3. Correlations among 12-month Systematic Observation of Red Flags and 12-month developmental measures.

Measure	Scores			Red Flags		
	SC Domain	RRB Domain	Composite	SC Domain	RRB Domain	Composite
Mullen Scales of Early Learning						
Expressive Language	–.51***	–.15	–.45***	–.54***	–.10	–.46***
Receptive Language	–.33***	–.22*	–.34***	–.37***	–.12	–.34***
Visual Reception	–.16	–.10	–.16	–.15	–.02	–.12
Fine Motor	–.24*	–.03	–.20*	–.23*	.08	–.15
Communication and Symbolic Behavior Scales						
Social	–.75***	–.32***	–.70***	–.73***	–.22*	–.65***
Speech	–.58***	–.24*	–.53***	–.57***	–.13	–.50***
Symbolic	–.27**	–.20*	–.28**	–.28**	–.18	–.29**
Total	–.74***	–.33***	–.69***	–.72***	–.22*	–.65***
Vineland Adaptive Behavior Scales						
Socialization	–.18	–.01	–.14	–.18	–.02	–.15
Communication	–.40***	–.20*	–.38***	–.40***	–.14	–.37***
Daily Living Skills	–.18	–.09	–.17	–.20*	–.17	–.23*
Motor	–.06	.03	–.04	–.10	–.05	–.09
Composite	–.30**	–.11	–.27**	–.32***	–.14	–.31**

Note. SC = Social Communication and Social Interaction; RRB = Restricted and Repetitive Behaviors.

*$p < .05$, using false discovery rate–adjusted p values. **$p < .01$. ***$p < .001$.

Table 4. Sensitivity, specificity, and area under the curve (AUC) for the 12-month Systematic Observation of Red Flags.

Domain	Cutoff score[a]	AUC [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]
Composite score	18	.77 [.68, .85]	.77 [.59, .90]	.76 [.66, .84]
SC Domain score	15	.73 [.64, .82]	.74 [.55, .88]	.73 [.62, .81]
RRB Domain score	3	.61 [.52, .71]	.68 [.49, .83]	.55 [.44, .65]
Total Red Flags	7	.70 [.60, .79]	.65 [.45, .81]	.75 [.65, .83]
SC Red Flags	6	.75 [.66, .84]	.68 [.48, .83]	.82 [.73, .90]
RRB Red Flags	2	.56 [.46, .66]	.45 [.27, .64]	.67 [.56, .77]

Note. CI = confidence interval; SC = Social Communication and Social Interaction; RRB = Restricted and Repetitive Behaviors.
[a]Cutoff score is inclusive (e.g., a cutoff score of 6 includes participants who have six or more red flags).

Specifically, medium and large, significant correlations were observed between the ADOS–Second Edition total calibrated severity score and the SORF SC Domain (r = .52), SORF Composite (r = .46), SORF SC Red Flags (r = .51), and SORF Total Red Flags (r = .42). SORF SC Domain scores and SC Red Flags were significantly associated with all four subscales of the Mullen (all $|rs| > .20$). Statistically significant associations were not found among the SORF RRB Domain score and RRB Red Flags and any of the 24-month measures.

Discussion

Using the SORF observational measure rated from a video-recorded CSBS Behavior Sample, a standardized communication assessment, we found that the core features of ASD were detected as early as 12 months in a sample of HR infant siblings who were diagnosed with ASD at 24 months, suggesting that, for some children, symptoms of ASD may unfold by the end of the first year of life. Our findings reaffirm research with other HR sibling samples showing that delays in the development of robust social communication skills, including eye contact, response to name, and sharing of positive affect and interests, may be present from a very early age (Landa & Garrett-Mayer, 2006; Ozonoff et al., 2010; Rowberry et al., 2015; Zwaigenbaum et al., 2005). Our findings also suggest that early signs of ASD in the social interaction domain were more powerful in predicting ASD in 12-month-olds compared to the RRB Domain. We speculate that RRBs are just beginning to emerge in this young sample and, while it is telling that the SORF identified RRBs in these infants, these behaviors were not as robust or predictive as diminished social interactive behaviors.

A serious consideration in attempting to identify early signs for ASD in very young children is adequate sensitivity and specificity. This concern is echoed in the report of the U.S. Preventive Services Task Force (Siu, 2016), which suggested that evidence is not yet sufficiently compelling to adequately evaluate the harms and benefits of screening young children with ASD in the absence of caregiver concern. Yet, compelling neuroscience research and promising studies of very early intervention continue to indicate the importance of optimizing development within the first 2 years of life (Bradshaw et al., 2015; Wolff et al., 2018). Results of this study suggested a screening cutoff score that resulted in acceptable sensitivity and specificity for detecting ASD, despite the young age of this sample (Dumont-Mathieu & Fein, 2005).

Our results also show significant predictive value of the tool regarding clinical phenotypes at 24 months of age

Table 5. Correlations among 12-month Systematic Observation of Red Flags and 24-month developmental outcomes.

	Scores			Red Flags		
Measure	SC Domain	RRB Domain	Composite	SC Domain	RRB Domain	Composite
Mullen Scales of Early Learning						
Expressive Language	−.29**	−.12	−.27**	−.31**	−.06	−.27**
Receptive Language	−.35***	−.14	−.32**	−.35***	−.02	−.28**
Visual Reception	−.31**	−.14	−.29**	−.29**	−.03	−.24*
Fine Motor	−.22*	−.01	−.17	−.23*	.01	−.17
ADOS-2 calibrated severity score[a]						
Social Affect	.49***	.12	.42***	.48***	.12	.39***
Restricted/Repetitive Behavior	.28**	.08	.24*	.27**	.05	.21*
Total	.52***	.15	.46***	.51***	.12	.42***

Note. SC = Social Communication and Social Interaction; RRB = Restricted and Repetitive Behaviors; ADOS-2 = Autism Diagnostic Observation Schedule–Second Edition.
[a]Spearman correlations.
*$p < .05$, using false discovery rate–adjusted p values. **$p < .01$. ***$p < .001$.

across developmental domains and ASD severity. In considering community-viable screening tools and the imperative to provide intensified surveillance to at-risk infants, we found that the SORF was correlated with standard assessments of social communication and language development typically conducted by SLPs and psychologists within the context of diagnostic evaluations. The SORF was not associated with measures of cognitive, motor, or adaptive socialization skills, suggesting the utility of the SORF as a screener and not the sole measure of overall functioning. A comprehensive evaluation following the SORF would enhance the overall diagnostic conceptualization of the child and drive treatment recommendations. By employing the SORF during the administration of an early communication assessment measure, opportunity exists to implement ASD screening by incorporating SLPs. A majority of parents of children later diagnosed with ASD first begin to express concerns about their child's development because of early delays in the development of speech (Herlihy et al., 2015; Karp et al., 2017). Moreover, multidisciplinary approaches to diagnostic evaluations with an SLP on the team to help administer the diagnostic battery may be a cost-effective and efficient strategy for timely diagnosis (Williams-Arya et al., 2019). Thus, early social communication delays present in children with ASD suggest that SLPs are a critical professional group in supporting the screening and assessment of young children for ASD (American Speech-Language-Hearing Association, 2006). The SORF may therefore capitalize on the SLPs' involvement with at-risk toddlers and increase the likelihood of earlier and focused screening efforts for ASD. Finally, as expected, the 12-month SORF was strongly associated with 24-month autism severity and developmental skills.

Findings by Dow et al. (2017) support the utility of the SORF within a community sample, and results of the current study reflect the tool's utility in a sample of younger infants that includes HR siblings. Our results suggest a cutoff score slightly lower than the previous work (Dow et al., 2017), providing additional evidence that at-risk cohorts may require further screening efforts to capture more subtle vulnerabilities in early social development (Micheletti et al., 2019). Our cohort of HR siblings demonstrated higher nonverbal ability than Dow et al.'s (2017) community sample, which could reflect the unique developmental trajectory observed in HR sibling samples (Chawarska et al., 2014; Ozonoff et al., 2015). It is noteworthy that, in this study, the SORF accurately identified toddlers who both were very young and had high scores on measures of developmental ability. The lower cutoff score for this younger age underscores the importance of screening for subtle social communication delays if early cognitive development is in the expected range. By attempting to study an observational measure based on *DSM-5* diagnostic criteria for ASD in a very young HR sample, our results demonstrate utility of the SORF to detect core features to a significant degree in the first year. Given the sensitive nature of raising concerns about ASD among providers and parents as early as 12 months prior to full symptom expression (Pierce et al., 2011), these results provide further support for a reliable observational screening measure, particularly for younger siblings requiring further surveillance. By utilizing a tool that may "catch" these signs earlier, speeding up the referral process for evaluation and early intervention can potentially address the cascading developmental effects as ASD unfolds in early childhood (Wetherby et al., 2014).

Limitations and Future Directions

Results from this study suggest good sensitivity and specificity of the SORF at 12 months in a group of HR infants with ASD but are limited by a relatively small sample size and a lack of a comparison group of infants with non-ASD developmental delays. In our current sample, the number of toddlers with non-ASD developmental and/or language delays was small, limiting our ability to include them in the current analysis. This is a critically important issue in the development and evaluation of an early detection measure. Future research is needed to replicate these results in larger HR and LR samples, inclusive of developmental delay, language delays, broader autism phenotype features, and diagnostic confirmation in the early childhood years. Given that our sample was not a community-based sample of participants but rather an infant sibling group, this tool should also be further tested across a wider range of infants and toddlers recruited with different ascertainment methods. Our study shows promising results for detecting and measuring early ASD indicator behaviors in a cohort of HR infants when the SORF is utilized alongside an early communication measure. Furthermore, this observational measure presents opportunity to integrate additional naturalistic samples of behavior and communication, including the home setting as used by Dow et al. (2020). Future studies should examine whether SLPs without specialized experience evaluating very young children at risk for ASD, as well as multidisciplinary clinicians, can successfully rate social communication and behavior using this observational tool. Future studies are also needed to explore the use of the SORF within community practice as an option to maximize screening and speed up referral to diagnostic evaluation when ASD concerns are present.

Author Contributions

Moira L. Pileggi: Conceptualization (Equal), Writing – original draft (Lead), Writing – review & editing (Lead). **Natalie Brane:** Conceptualization (Equal), Writing – original draft (Lead), Writing – review & editing (Equal). **Jessica Bradshaw:** Formal analysis (Lead), Writing – original draft (Equal), Writing – review & editing (Equal). **Abigail Delehanty:** Writing – original draft (Lead), Writing – review & editing (Lead). **Taylor Day:** Conceptualization (Equal), Writing – original draft (Supporting). **Courtney McCracken:** Formal analysis (Supporting). **Jennifer Stapel-Wax:** Conceptualization (Equal), Investigation (Equal), Writing – original draft (Supporting), Writing – review & editing (Supporting).

Amy M. Wetherby: Investigation (Lead), Methodology (Lead), Writing – review & editing (Equal).

Acknowledgments

This research was supported in part by the National Institute of Mental Health Grant P50MH100029 (PI: Ami Klin) and Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant R01HD078410 (PI: Amy M. Wetherby). This work was supported in part by the National Institute of Mental Health Grant K23-MH120476 (PI: J. B.) and the National Institute of Deafness and Communication Disorders Grant R21DC071252 (PI: J. B.). The content is solely the responsibility of the authors and does not necessarily represent the official views of these federal agencies.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

American Speech-Language-Hearing Association. (2006). *Roles and responsibilities of speech-language pathologists in diagnosis, assessment, and treatment of autism spectrum disorders across the life span: Position statement*. http://www.asha.org/policy

Bai, D., Yip, B. H. K., Windham, G. C., Sourander, A., Francis, R., Yoffe, R., Glasson, E., Mahjani, B., Suominen, A., Leonard, H., Gissler, M., Buxbaum, J. D., Wong, K., Schendel, D., Kodesh, A., Breshnahan, M., Levine, S. Z., Parner, E. T., Hansen, S. N., . . . Sandin, S. (2019). Association of genetic and environmental factors with autism in a 5-country cohort. *JAMA Psychiatry, 76*(10), 1035–1043. https://doi.org/10.1001/jamapsychiatry.2019.1411

Barton, M. L., Dumont-Mathieu, T., & Fein, D. (2012). Screening young children for autism spectrum disorders in primary practice. *Journal of Autism and Developmental Disorders, 42*(6), 1165–1174. https://doi.org/10.1007/s10803-011-1343-5

Bradshaw, J., Steiner, A. M., Gengoux, G., & Koegel, L. (2015). Feasibility and effectiveness of very early intervention for infants at-risk for autism spectrum disorder: A systematic review. *Journal of Autism and Developmental Disorders, 45*(3), 778–794. https://doi.org/10.1007/s10803-014-2235-2

Bryson, S. E., & Zwaigenbaum, L. (2014). Autism Observation Scale for Infants. In V. B. Patel, V. R. Preedy, & C. R. Martin (Eds.), *Comprehensive guide to autism* (pp. 299–310). Springer. https://doi.org/10.1007/978-1-4614-4788-7_12

Bryson, S. E., Zwaigenbaum, L., McDermott, C., Rombough, V., & Brian, J. (2008). The Autism Observation Scale for Infants: Scale development and reliability data. *Journal of Autism and Developmental Disorders, 38*(4), 731–738. https://doi.org/10.1007/s10803-007-0440-y

Chawarska, K., Shic, F., Macari, S., Campbell, D. J., Brian, J., Landa, R., Hutman, T., Nelson, C. A., Ozonoff, S., Tager-Flusberg, H., Young, G. S., Zwaigenbaum, L., Cohen, I. L., Charman, T., Messinger, D. S., Klin, A., Johnson, S., & Bryson, S. (2014). 18-month predictors of later outcomes in younger siblings of children with autism spectrum disorder: A Baby Siblings Research Consortium study. *Journal of the American Academy of Child & Adolescent Psychiatry, 53*(12), 1317–1327.e1. https://doi.org/10.1016/j.jaac.2014.09.015

Constantino, J. N. (2013). *Social Responsiveness Scale*. Springer.

Constantino, J. N., Lajonchere, C., Lutz, M., Gray, T., Abbacchi, A., McKenna, K., Singh, D., & Richard, T. D. (2006). Autistic social impairment in the siblings of children with pervasive developmental disorders. *The American Journal of Psychiatry, 163*(2), 294–296. https://doi.org/10.1176/appi.ajp.163.2.294

Dow, D., Day, T. N., Kutta, T. J., Nottke, C., & Wetherby, A. M. (2020). Screening for autism spectrum disorder in a naturalistic home setting using the Systematic Observation of Red Flags (SORF) at 18–24months. *Autism Research, 13*(1), 122–133. https://doi.org/10.1002/aur.2226

Dow, D., Guthrie, W., Stronach, S. T., & Wetherby, A. M. (2017). Psychometric analysis of the Systematic Observation of Red Flags for autism spectrum disorder in toddlers. *Autism, 21*(3), 301–309. https://doi.org/10.1177/1362361316636760

Dumont-Mathieu, T., & Fein, D. (2005). Screening for autism in young children: The Modified Checklist for Autism in Toddlers (M-CHAT) and other measures. *Mental Retardation and Developmental Disabilities Research Reviews, 11*(3), 253–262. https://doi.org/10.1002/mrdd.20072

Esler, A. N., Bal, V. H., Guthrie, W., Wetherby, A., Ellis Weismer, S., & Lord, C. (2015). The Autism Diagnostic Observation Schedule, Toddler Module: Standardized severity scores. *Journal of Autism and Developmental Disorders, 45*(9), 2704–2720. https://doi.org/10.1007/s10803-015-2432-7

Herlihy, L., Knoch, K., Vibert, B., & Fein, D. (2015). Parents' first concerns about toddlers with autism spectrum disorder: Effect of sibling status. *Autism, 19*(1), 20–28. https://doi.org/10.1177/1362361313509731

Ingersoll, B., & Wainer, A. (2014). The broader autism phenotype. In F. R. Volkmar, S. J. Rogers, R. Paul, & K. A. Pelphrey (Eds.), *Handbook of autism and pervasive developmental disorders* (4th ed.). Wiley. https://doi.org/10.1002/9781118911389.hautc02

Karp, E. A., Ibañez, L. V., Warren, Z., & Stone, W. L. (2017). Brief report: What drives parental concerns about their 18-month-olds at familial risk for autism spectrum disorder? *Journal of Autism and Developmental Disorders, 47*(5), 1535–1541. https://doi.org/10.1007/s10803-017-3060-1

Landa, R., & Garrett-Mayer, E. (2006). Development in infants with autism spectrum disorders: A prospective study. *The Journal of Child Psychology and Psychiatry, 47*(6), 629–638. https://doi.org/10.1111/j.1469-7610.2006.01531.x

Maenner, M. J., Shaw, K. A., Baio, J., Washington, A., Patrick, M., DiRienzo, M., Christensen, D. L., Wiggins, L. D., Pettygrove, S., Andrews, J. G., Lopez, M., Hudson, A., Baroud, T., Schwenk, Y., White, T., Rosenberg, C. R., Lee, L.-C., Harrington, R. A., Huston, M., . . . Dietz, P. M. (2020). Prevalence of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2016. *MMWR Surveillance Summaries, 69*(4), 1–12. https://doi.org/10.15585/mmwr.ss6904a1

McCathren, R. B., Yoder, P. J., & Warren, S. F. (2000). Testing predictive validity of the Communication Composite of the Communication and Symbolic Behavior Scales. *Journal of Early Intervention, 23*(1), 36–46. https://doi.org/10.1177/105381510002300108 01

McWilliam, R. A., & Ware, W. B. (1994). The reliability of observations of young children's engagement: An application of generalizability theory. *Journal of Early Intervention, 18*(1), 34–47. https://doi.org/10.1177/105381519401800104

Micheletti, M., McCracken, C., Constantino, J. N., Mandell, D., Jones, W., & Klin, A. (2019). Research review: Outcomes of 24- to 36-month-old children with autism spectrum disorder vary by ascertainment strategy: A systematic review and meta-analysis. *The Journal of Child Psychology and Psychiatry, 61*(1), 4–17. https://doi.org/10.1111/jcpp.13057

Miller, M., Iosif, A.-M., Hill, M., Young, G. S., Schwichtenberg, A. J., & Ozonoff, S. (2017). Response to name in infants

developing autism spectrum disorder: A prospective study. *The Journal of Pediatrics, 183,* 141–146.e1. https://doi.org/10.1016/j.jpeds.2016.12.071

Morgan, L., Delehanty, A., Dillon, J. C., Schatschneider, C., & Wetherby, A. M. (2020). Measures of early social communication and vocabulary production to predict language outcomes at two and three years in late-talking toddlers. *Early Childhood Research Quarterly, 51,* 366–378. https://doi.org/10.1016/j.ecresq.2019.12.005

Mullen, E. (1995). *Mullen Scales of Early Learning*. AGS.

Ozonoff, S., Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., Rogers, S. J., Rozga, A., Sangha, S., Sigman, M., Steinfeld, M. B., & Young, G. S. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(3), 256–266.e2. https://doi.org/10.1016/j.jaac.2009.11.009

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., Bryson, S., Carver, L. J., Constantino, J. N., Dobkins, K., Hutman, T., Iverson, J. M., Landa, R., Rogers, S. J., Sigman, M., & Stone, W. L. (2011). Recurrence risk for autism spectrum disorders: A Baby Siblings Research Consortium study. *Pediatrics, 128*(3), e488–e495. https://doi.org/10.1542/peds.2010-2825

Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., Chawarska, K., Macari, S. L., Messinger, D., Stone, W. L., Zwaigenbaum, L., & Iosif, A.-M. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: A Baby Siblings Research Consortium study. *The Journal of Child Psychology and Psychiatry, 56*(9), 988–998. https://doi.org/10.1111/jcpp.12421

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. *The Journal of Pediatrics, 159*(3), 458–465.e6. https://doi.org/10.1016/j.jpeds.2011.02.036

Pierce, K., Gazestani, V. H., Bacon, E., Barnes, C. C., Cha, D., Nalabolu, S., Lopez, L., Moore, A., Pence-Stophaeros, S., & Courchesne, E. (2019). Evaluation of the diagnostic stability of the early autism spectrum disorder phenotype in the general population starting at 12 months. *JAMA Pediatrics, 173*(6), 578. https://doi.org/10.1001/jamapediatrics.2019.0624

Rowberry, J., Macari, S., Chen, G., Campbell, D., Leventhal, J. M., Weitzman, C., & Chawarska, K. (2015). Screening for autism spectrum disorders in 12-month-old high-risk siblings by parental report. *Journal of Autism and Developmental Disorders, 45*(1), 221–229. https://doi.org/10.1007/s10803-014-2211-x

Rutter, M., Bailey, A., & Lord, C. (2003). *The Social Communication Questionnaire*. WPS.

Sandin, S., Reichenberg, A., Lichtenstein, P., Kuja-Halkola, R., Hultman, C., & Larsson, H. (2017). The heritability of autism spectrum disorder. *JAMA, 318*(12), 1182–1184. https://doi.org/10.1001/jama.2017.12141

Siu, A. L. (2016). Screening for autism spectrum disorder in young children: US Preventive Services Task Force recommendation statement. *JAMA, 315*(7), 691–696. https://doi.org/10.1001/jama.2016.0018

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales* (2nd ed.). AGS.

Stone, W. L., McMahon, C. R., & Henderson, L. M. (2008). Use of the Screening Tool for Autism in Two-Year-Olds (STAT) for children under 24 months: An exploratory study. *Autism, 12*(5), 557–573. https://doi.org/10.1177/1362361308096403

Watson, L. R., Crais, E. R., Baranek, G. T., Dykstra, J. R., & Wilson, K. P. (2013). Communicative gesture use in infants with and without autism: A retrospective home video study. *American Journal of Speech-Language Pathology, 22*(1), 25–39. https://doi.org/10.1044/1058-0360(2012/11-0145)

Wetherby, A. M., Guthrie, W., Woods, J. J., Schatschneider, C., Holland, R. D., Morgan, L., & Lord, C. (2014). Parent-implemented social intervention for toddlers with autism: An RCT. *Pediatrics, 134,* 1084–1093. https://doi.org/10.1542/peds.2014-0757

Wetherby, A. M., & Prizant, B. M. (2002). *Communication and Symbolic Behavior Scales–Developmental Profile (CSBS-DP)*. Brookes.

Wetherby, A. M., Woods, J. J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life. *Journal of Autism and Developmental Disorders, 34*(5), 473–493. https://doi.org/10.1007/s10803-004-2544-y

Wetherby, A. M., Woods, J. J., Nottke, C., Stronach, S. T., Dow, D., & McCoy, D. (2016). *Systematic Observation of Red Flags of Autism Spectrum Disorder* [Unpublished manual]. Florida State University.

Williams-Arya, P., Anixt, J., Ehrhardt, J., & Manning-Courtney, P. (2019). Improving access to diagnostic assessments for autism spectrum disorder using an Arena model. *Journal of Developmental & Behavioral Pediatrics, 40*(5), 161–169. https://doi.org/10.1097/DBP.0000000000000694

Wolff, J. J., Jacob, S., & Elison, J. T. (2018). The journey to autism: Insights from neuroimaging studies of infants and toddlers. *Development and Psychopathology, 30*(2), 479–495. https://doi.org/10.1017/S0954579417000980

Yoder, P., Stone, W. L., Walden, T., & Malesa, E. (2009). Predicting social impairment and ASD diagnosis in younger siblings of children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 39*(10), 1381–1391. https://doi.org/10.1007/s10803-009-0753-0

Zwaigenbaum, L., Bauman, M. L., Fein, D., Pierce, K., Buie, T., Davis, P. A., Newschaffer, C., Robins, D. L., Wetherby, A., Choueiri, R., Kasari, C., Stone, W. L., Yirmiya, N., Estes, A., Hansen, R. L., McPartland, J. C., Natowicz, M. R., Carter, A., Granpeesheh, D., . . . Wagner, S. (2015). Early screening of autism spectrum disorder: Recommendations for practice and research. *Pediatrics, 136*(Suppl. 1), S41–S59. https://doi.org/10.1542/peds.2014-3667D

Zwaigenbaum, L., Bryson, S., & Garon, N. (2013). Early identification of autism spectrum disorders. *Behavioural Brain Research, 251,* 133–146. https://doi.org/10.1016/j.bbr.2013.04.004

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. https://doi.org/10.1016/j.ijdevneu.2004.05.001

Appendix

Systematic Observation of Red Flags of Autism Spectrum Disorder (SORF; Wetherby et al., 2016)

Impairment in Social Communication and Social Interaction

1. Limited sharing warm, joyful expressions
2. Flat affect or reduced facial expressions
3. Limited sharing interests
4. Lack of response to name or social bids
5. Poor eye gaze directed to faces
6. Limited use of conventional gestures—showing and pointing
7. Uses person's hand/body as a tool without gaze
8. Limited use of consonant sounds in vocal communication
9. Limited coordination of nonverbal communication
10. Less interest in people than objects
11. Limited sharing of reciprocal social play

Restricted and Repetitive Patterns of Behavior, Interests, or Activities

12. Repetitive movements with objects
13. Repetitive movements or posturing of body
14. Repetitive speech or intonation
15. Ritualized patterns of behavior
16. Marked distress over change
17. Excessive interest in particular objects, actions, or activities
18. Clutches particular objects
19. Sticky attention to objects
20. Fixated interests on parts of objects
21. Lack of or adverse response to specific sounds, textures, or other sensory stimuli
22. Unusual sensory exploration/excessive interest in sensory aspects of environment