Published in final edited form as:

J Dev Behav Pediatr. 2020 ; 41(8): 596–604. doi:10.1097/DBP.0000000000000827.

Screening for Autism Spectrum Disorder in High-Risk Younger Siblings

Kathryn Bradbury, PhD¹, Diana L. Robins, PhD², Marianne Barton, PhD¹, Lisa V. Ibañez, PhD³, Wendy L. Stone, PhD³, Zachary E. Warren, PhD⁴, Deborah Fein, PhD¹

¹Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut

²A.J. Drexel Autism Institute, Drexel University, Philadelphia, Pennsylvania

³Department of Psychology, University of Washington, Seattle, Washington

⁴Vanderbilt Kennedy Center Treatment and Research Institute for ASD, Department of Pediatrics, Vanderbilt University Medical Center, Nashville, Tennessee

Abstract

Objective—Most Autism Spectrum Disorder (ASD) screening measures have been developed for use with low-risk (LR) children; however, measures may perform differently in high-risk (HR) younger sibling populations. The current study sought to investigate the performance of an ASD screening measure, the Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M-CHAT-R/F), in a sample of HR younger siblings and directly compare its performance to that in a LR sample.

Methods—HR younger siblings ($n = 187$) and LR children ($n = 15,848$) were screened using the M-CHAT-R/F. Screen positive children completed comprehensive evaluations. M-CHAT-R/F psychometric properties and clinical characteristics were compared across samples.

Results—The M-CHAT-R/F demonstrated a significantly higher screen positive rate and ASD detection rate in the HR sample compared to the LR sample. Children with ASD in the HR sample had stronger verbal, nonverbal, and overall cognitive abilities compared to children with ASD in the LR sample in spite of comparable ASD severity and adaptive functioning. High positive predictive value of the M-CHAT-R at initial screen, with only incremental change after Follow-Up, suggests that Follow-Up is less critical in HR than LR samples. A significantly lower number of changed responses during Follow-Up further supports improved reporting accuracy of parents with ASD experience compared to parents less familiar with ASD.

Conclusion—The findings suggest that the M-CHAT-R/F can distinguish between ASD and Non-ASD at 18–24 months in a HR sibling sample, with performance comparable to or better than its performance in the general population.

Keywords

Autism Spectrum Disorder; High-risk younger siblings; Screening

INTRODUCTION

The American Academy of Pediatrics recommends universal screening for Autism Spectrum Disorders (ASD) at 18- and 24-month well-child visits in addition to ongoing developmental surveillance and broad developmental screening.^{1,2} Universal screening using standardized measures with timely referrals for at-risk children has been shown to lower the age of ASD diagnosis close to the second birthday;³ this is especially important in minority populations and those with lower socio-economic status, who may otherwise experience disproportionately greater delays.⁴ Despite these efforts, the median age of ASD diagnoses in the United States is still over the age of four years (52 months).⁵

Parenting experience, especially that with another child with ASD, impacts reporting of concerns for later-born children. Parents who have older children with ASD tend to report more concerns about later-born children than do parents who have older typically developing children or first-time parents.^{6,7} Concerns raised by parents who have older children with ASD appear to be valid, as they are correlated with scores on standardized assessment measures and are predictive of later ASD diagnosis.^{6,8,9} However, the predictive validity of parent concerns varies by domain (i.e., social, communication, etc.) and remains lower than that of standardized measures, further emphasizing the need for more formal and systematic screening.⁸

Despite the substantial body of literature indicating that parents of children at high risk (HR) for developing ASD endorse a greater number of more predictive and specific concerns,^{6–9} little research has been conducted to determine how differential parent concern translates to performance on standardized ASD screening measures. Screening in HR populations is critical given the high proportion of children who go on to develop significant developmental delays, including ASD (for review, see Arpino et al).¹⁰ However, most level 1 ASD screening measures have not been developed explicitly for use in this population and thus, may not display adequate predictive validity.¹¹ For instance, high false positive rates (22–43%)^{12,13} have been seen in very premature infants screened with the Modified-Checklist for Autism in Toddlers with Follow-up (M-CHAT/F),¹⁴ which are likely related to their high susceptibility to varied neurodevelopmental impairments. Similarly elevated screen positive rates (40%) on the M-CHAT/F have been observed in other high risk populations.¹⁵

Younger siblings of children with ASD of unknown etiology are at increased risk for developing ASD compared to children in the general population (10–20 % recurrence risk)^{1,16} as well as other developmental delays and subclinical ASD symptoms that may need treatment,¹⁷ which makes screening in this population essential. Parenting experience with ASD symptoms may lead to over-endorsement of such symptoms in younger siblings because of elevated parent concern or conversely, to more accurate endorsement due to increased familiarity with symptoms.

To date, there have been few studies investigating screening for ASD in a HR younger sibling population. Findings indicate that clinician-rated, interactive screening measures

(i.e., Screening Tool for Autism in Two-year-olds [STAT] and Autism Observation Scale for Infants [AOSI]) identify behavioral markers that differentiate HR younger siblings that go on to develop ASD (HR-ASD) from those that do not (HR-nonASD) early in the first two years of life (STAT PPV = .68;¹⁸ AOSI PPV = .78¹⁹). Similarly, studies of parent-report questionnaires (i.e., Parent Observation of Early Markers Scale [POEMS], First Year Inventory [FYI], Autism Parent Screen for Infants [APSI], M-CHAT-Revised with Follow-up [M-CHAT-R/F], Child Behavior Checklist [CBCL]/1.5–5) provide strong evidence that parents of HR-ASD children are able to identify and report at-risk behaviors predictive of ASD within the first two years of life with fairly good discriminability.^{20–24} The M-CHAT-R/F shows good discriminability at 18 months (PPV = .78),²³ whereas the Pervasive Developmental Problem Scale on the CBCL/1.5–5 does so at 24 months (PPV = .91).²⁴ Some measures show good discriminability between HR-ASD and HR-nonASD children even earlier, such as the APSI at 6 months (PPV = .47),²² POEMS at 9 months (PPV = .29),²⁰ and the FYI at 12 months (PPV = .71).²¹ Two studies have looked at screening across multiple time points,^{20,22} whereas others focused on a single time point^{21,23,24} with diagnostic outcome determined at 18 to 36 months of age across studies. However, the previous literature has not addressed how HR-ASD screening directly compares to that in a LR population by including a comparison group of LR children who are later diagnosed with ASD.

The current study directly compares screening in a HR sample of younger siblings to screening in a large LR sample. For the latter, we used all of the English-speaking families from the M-CHAT-R/F validation study by Robins et al.³ Because the parents in the HR sample have prior experience in observing ASD signs and symptoms, and the prevalence in younger siblings is higher than in the general population, we hypothesize that (1) the M-CHAT-R/F will have a higher internal consistency in the HR sample, (2) the screen positive rate will be higher in the HR sample, (3) the M-CHAT-R/F will demonstrate higher positive predictive value (PPV) in the HR sample, and (4) parents of children in the HR sample will be more accurate in their reporting of at-risk behaviors, leading to fewer items changing between initial screen and Follow-Up interview.

METHODS

Participants

A total of 241 toddlers at high familial risk (HR) for ASD, defined as the presence of at least one confirmed affected older full or half sibling (proband), were screened (see Figure 1) at University of Connecticut (UConn), University of Washington (UW), and Vanderbilt University (VU) from 2009 to mid-2014. HR toddlers were recruited through ASD resource fairs, clinical research or treatment involving their older sibling with ASD, or through well-child screening at pediatric providers' offices. Fifty-four toddlers did not meet inclusion criteria due to being the twin sibling of a proband ($n = 1$), known neurological impairment ($n = 3$), gestational age < 37 weeks ($n = 3$), outside of screening age range ($n = 11$), refusal ($n = 8$), unconfirmed proband diagnosis ($n = 4$), lost to follow-up ($n = 16$), or for reasons unknown ($n = 8$). The final HR sample consisted of 187 screened toddlers. In some analyses, an additional 12 children were excluded due to incomplete follow-up ($n = 3$) or incomplete

evaluation ($n = 9$), for a total of 175 toddlers between 16 and 30 months old ($M = 21.2$, $SD = 4.1$) with complete data.

The LR comparison group consisted of all children from the M-CHAT-R/F validation study³ who were screened in English at well-child visits at a pediatric provider's office between 2009 and early 2013. All children were included except for those screened in Spanish ($n = 329$) and those with insufficient English proficiency ($n = 15$), previous ASD diagnosis ($n = 4$), medical diagnosis precluding evaluation ($n = 13$), withdrawal from study ($n = 2$), or screened outside of the eligible screening range ($n = 10$). The final LR sample consisted of 15,848 screened toddlers. In some analyses, an additional 448 children were excluded due to incomplete follow-up ($n = 304$) and incomplete evaluations ($n = 144$). Thus, the LR sample with complete data consisted of 15,400 toddlers screened at UConn ($n = 5,612$) and Georgia State University ($n = 9,788$) between 16 and 30 months old ($M = 20.96$, $SD = 3.30$). See Table 1 for demographics by risk group.

Procedures

All procedures were approved by the Institutional Review Board at each participating site. Caregivers of participating HR and LR children completed the M-CHAT-R (i.e., initial screen) in English. Structured Follow-Up interviews were completed with positive screens via telephone and those continuing to screen positive were evaluated. Evaluations were conducted by a clinician team consisting of a licensed clinical psychologist or developmental pediatrician and a psychology doctoral student or trained research assistant. Follow-Up interviews and evaluations were conducted with a select portion of screen negative cases when pediatricians raised specific concerns regarding ASD (e.g., red flag), children screened positive on another screening measures (i.e., STAT), or children failed the screener through an alternative scoring method. Detailed information about a subset of screen negative HR cases at VU was previously published in Weitlauf et al.,²³ and detailed information about ascertaining missed cases in the LR sample was published in Robins et al.³

Probands (older siblings) of screen positive cases completed similar evaluations as their younger siblings for diagnostic confirmation. Diagnosis of probands of screen negative cases was confirmed via review of a previous evaluation report and completion of the Social Communication Questionnaire (SCQ)²⁵ with a cutoff of 11 due to better specificity and sensitivity.²⁶ Different methods were utilized for confirmation of proband diagnosis as most screen negative cases were not evaluated and as such, neither were their older siblings.

Measures

The M-CHAT-R/F is a 20-item, parent-completed, ASD-specific screener validated for use in children between the ages of 16 and 30 months.³ Going forward, M-CHAT-R refers to initial screen, while M-CHAT-R/F refers to the two-stage screening process with Follow-Up administered for medium risk scores. A positive screen on the M-CHAT-R consists of three or more at-risk responses out of 20 total items, whereas a positive screen on the Follow-Up (M-CHAT-R/F) consists of two or more at-risk responses. The Follow-Up interview consists of additional structured questions to confirm at-risk responses. Of the children who received

“high fail” scores on the M-CHAT-R (i.e., 8), eight bypassed Follow-Up and were immediately eligible for evaluation, as described in Robins et al.³

Evaluations were the same for both HR and LR groups and included direct assessment of cognitive and social-communication skills as well as parent report of development, ASD symptomatology, and adaptive skills collected via interview. ASD symptomatology was assessed using observational information from the Autism Diagnostic Observation Schedule (Second Edition; ADOS[–2]),²⁷ along with parental report using the Toddler Autism Symptom Inventory (TASI)²⁸ or the Autism Diagnostic Interview, Revised (ADI-R),²⁹ and the Childhood Autism Rating Scale (Second Edition; CARS[–2]).³⁰ The sample was further characterized using the Mullen Scales of Early Learning (MSEL),³¹ Vineland Adaptive Behavior Scales – Second Edition, Survey Interview Form (VABS-II),³² and a developmental history form. Due to floor effects on the MSEL, developmental quotients (DQ) were derived using age-equivalents as mental age ($DQ = MA/CA * 100$), as has been done elsewhere in the literature.²¹ Verbal DQ scores were derived by averaging the age equivalents on the receptive and expressive language domains, while nonverbal DQ scores were derived by averaging the age equivalents on the visual reception and fine motor domains.

Diagnoses were made based upon expert clinical judgment using all available information according to DSM-IV-TR³³ criteria for Autistic Disorder and Pervasive Developmental Disorder, Not Otherwise Specified (PDD-NOS). A diagnosis of PDD-NOS refers to children with impairment in reciprocal social communication and either deficits in nonverbal communication or the presence of restricted, repetitive behaviors that did not meet full DSM-IV-TR criteria for a specific ASD, such as Autistic Disorder.³³ Of note, DSM-IV-TR criteria was used for the entire sample for consistency’s sake despite revised criteria published during data collection (i.e., in 2013). Children not meeting ASD criteria were diagnosed with Global Developmental Delay, Developmental Language Disorder, or another appropriate DSM-IV-TR diagnosis. Those with subclinical symptoms were labeled as “At-Risk for ASD” (i.e., subclinical ASD symptoms) or “No Diagnosis” (i.e., other notable subclinical symptoms) and were included in the non-ASD group. The label “developmental delay or concern” is used going forward to describe all children with a non-ASD diagnosis, except those with typical development. For the purposes of this study, typical development was operationalized as not meeting ASD cutoffs on ASD measures, having no delays of greater than one and half standard deviations in any cognitive or adaptive domain, and not meeting criteria for any other DSM-IV-TR diagnosis.

Statistical Analyses

Demographic characteristics across HR sites and across HR and LR samples were compared using chi-squared test, one-way analysis of variance, or student’s t-test, as appropriate. Clinical characteristics of children who screened positive and received an ASD diagnosis were compared across risk status using a series of student’s t-tests. Bonferroni correction was used to correct for multiple comparisons and resulted in an adjusted α value of .004 (.05/12 comparisons). Internal consistency for the M-CHAT-R was calculated for each sample individually and then compared across samples using Feldt Test.³⁴ Screen positive

rates were calculated as a percentage of children in each sample who screened positive at specific points in the screening process. These percentages were used to calculate the ASD detection rate and PPV in each sample, which were then compared across samples using chi-squared test or Fischer's Exact Test (FET), as appropriate. FET was used when expected cell counts were less than 5. M-CHAT-R/(F) total scores, change scores, and delays between initial screening and follow-up were compared across samples using student's t-tests. Data were analyzed using SPSS Version 25.

RESULTS

Demographic Characteristics of HR and LR Samples

Demographic characteristics of HR toddlers were similar across study sites with respect to age at screen or evaluation, sex, race, or maternal education (p 's = .366 - .868). The HR and LR samples with complete data were comparable with respect to demographic variables, except the LR sample had more children of color ($p < .001$, $\phi = .06$) and a lower proportion of males ($p = .044$, $\phi = .02$) compared to the HR sample, with very small effect sizes (see Table 1).

Characteristics of ASD Groups

Clinical characteristics of all children who screened positive and received an ASD diagnosis can be found in Table 2. Autism severity is measured by the ADOS(-2) and the CARS(-2), cognitive functioning by the MSEL, and adaptive skills by the VABS-II. There were no significant differences between HR and LR children diagnosed with ASD in autism severity on any of the ADOS(-2) Calibrated Severity Scores or the CARS(-2) total score. HR-ASD children showed higher Verbal DQ, Non-Verbal DQ, and Overall DQ on the MSEL. After correction for multiple comparisons, there were no differences in parent-reported adaptive skills on any of the VABS-II domains.

Performance of the M-CHAT-R/F

As evaluation of screen negative children was limited in the current study, sensitivity and specificity calculations, as well as ROC analyses, were not conducted. Instead, data analyses focused on internal consistency, screen positive rates, PPV for ASD and for other developmental diagnoses.

Reliability—Internal consistency for the M-CHAT-R was significantly greater for the HR sample (Cronbach's $\alpha = .88$) than for the LR sample (Cronbach's $\alpha = .64$), as determined by Feldt Test³⁴ ($W = .34$, $p < .0001$).

Screen Positive Rates—Of the 187 HR children who were screened, 76 (40.6%) screened positive on initial screening (see Figure 1). Follow-Up interviews were conducted for 65 (85.5%) of the eligible children; eight children with high scores bypassed Follow-Up per the two-stage screening procedures described in Robins et al.³ and three children had incomplete Follow-Up. Of those who completed Follow-Up, most children (87.7%) in the HR sample continued to screen positive. Compared to the LR sample, the HR sample had a significantly higher screen positive rate at both initial screen (HR: 40.6%; LR 7.1%; p

$< .001$, $\phi = .14$) and retained the screen positive status at Follow-Up (HR: 87.7%; LR: 37.1%; $p < .001$, $\phi = .26$) with small to medium effect sizes, respectively.

Screeners Performance—The following analyses were conducted with only the children who screened positive on the M-CHAT-R and completed Follow-Up (HR: $n = 65$; LR: $n = 910$). Due to the nature of Follow-Up (confirmation of at-risk responses), item responses could only change from “at-risk” to “not-at-risk” during Follow-Up questioning. As anticipated, the HR sample had significantly higher initial M-CHAT-R total scores (out of a possible 20 items; HR: $M = 6.5$, $SD = 2.9$; LR: $M = 4.7$, $SD = 2.5$; $t(71.1) = -4.93$, $p < .001$, $d = .67$) and M-CHAT-R/F [Follow-Up] total scores (HR: $M = 5.0$, $SD = 3.1$; LR: $M = 1.7$, $SD = 2.4$; $t(69.3) = -8.27$, $p < .001$, $d = 1.18$) than the LR sample. Item responses were substantially less likely to change on Follow-Up in the HR sample ($M = 1.60$ items, $SD = 1.58$) than in the LR sample ($M = 3.07$ items, $SD = 1.80$), $t(973) = 6.44$, $p < .001$, $d = .87$. Of note, Follow-Up was completed after a shorter delay after initial screening in the HR sample ($M = .75$ months, $SD = 1.12$) than in the LR sample ($M = 2.77$ months, $SD = 2.34$), $t(68.8) = 11.17$, $p < .001$, $d = 1.10$.

Diagnostic Rates—Sixty-five children (57 screen positives and eight high-fails) were determined eligible for evaluation in the HR sample. Of the 65 eligible children, 56 completed evaluations and 34 (61%) of those children received an ASD diagnosis. As expected, the ASD detection rate in HR children (19.4%) was significantly higher than that in the LR sample (0.65%), $p < .001$, $\phi = .20$. All children who screened positive on the initial M-CHAT-R questionnaire and who had complete data were included in the Positive Predictive Value (PPV) analyses for initial screen, and all children who screened positive at Follow-Up with complete data were included in the Follow-Up PPV analyses (see Table 3). PPV for ASD in the HR sample was significantly higher than that of the LR sample at initial screen (HR = 0.531; LR = 0.138; $p < .001$, $\phi = .28$), but comparable at Follow-Up (HR = 0.607; LR = 0.488; $p = .113$, $\phi = .10$). Furthermore, PPV for the HR sample did not differ between initial screen and Follow-Up either for ASD ($p = .403$) or for any developmental delay or concern (DD; FET $p = .12$), whereas PPV for the LR sample was higher after Follow-Up both for ASD ($p < .001$, $\phi = .35$) and any DD ($p < .001$, $\phi = .56$). The M-CHAT-R/F continued to demonstrate very high PPVs for any DD in both risk groups (HR = .982; LR = .941; FET $p = .310$).

DISCUSSION

The current study sought to build upon existing literature on ASD screening in HR younger siblings by examining the performance on the M-CHAT-R/F in a relatively large, well-characterized sample of HR and LR toddlers. Clinical comparisons indicated that within the ASD sample, HR children showed stronger cognitive abilities despite comparable adaptive skills and ASD severity. This pattern is consistent with recent research comparing large multiplex (HR) and simplex (LR) samples with ASD from the Autism Genetic Research Exchange (AGRE) database.³⁵ It is possible that parents of HR children may have increased knowledge and awareness of these social behaviors, regardless of compensatory behaviors used by children with stronger cognitive abilities. In contrast, parents of LR children may be less likely to notice the subtle social deficits if their child has stronger cognitive skills, which

may mask the areas of challenge. This idea is further supported by research indicating that children with ASD and comorbid developmental delay are more likely to be detected earlier.³⁶ However, Berends et al.³⁵ found that while children with ASD in multiplex families had higher cognitive abilities than the affected child in simplex families, they did not differ from one another, suggesting that birth order (i.e., parenting experience) is not likely to be the driving factor behind the observed differences in cognitive abilities between risk groups.

Several predictions about the performance of the M-CHAT-R/F were based the assumption that the prevalence rate of ASD in HR younger sibs would be higher than in the LR sample. This assumption was supported: the ASD detection rate for HR younger siblings in the current study (19.4%) was significantly greater than in the LR comparison sample (.65%). The HR detection rate was similar to some recent recurrence rate estimates (19.5%),¹⁶ but higher than others (10%).¹ Of note, the LR sample detection rate was lower than the current ASD prevalence rates of 1 in 59 (1.69%) in 8-year-olds.⁵ These prevalence rates are based upon educational and medical records at age 8 and we would not expect to identify all of these cases at our mean age of 21 months.

We predicted that the M-CHAT-R would have better internal consistency in the HR sample. This prediction was supported: greater internal consistency was seen in the HR sample compared to the LR sample (.88 vs .64). This finding may reflect the increased prevalence of ASD and subclinical ASD symptoms in HR samples.¹⁷ Additionally, of the children who screened positive at initial screen, parents of HR children endorsed more at-risk items than parents of LR children (6.7 vs 4.1) at initial screen and at follow-up (5.1 vs 1.3). Thus, higher overall screening scores in the HR sample may have also contributed to higher internal consistency in this sample.

The second prediction was that children in the HR sample would have a higher screen positive rate than those in the LR sample. As expected, the screen positive rate was higher in HR toddlers (40.6%) compared to the LR sample (7%) on initial screen, and after follow-up (HR 88% and LR 37% of those screening positively at initial screen). We also predicted a higher PPV for the M-CHAT-R/F in the HR sample. As PPV is affected by base rate, a higher PPV in the HR sample would be expected given the higher base rate of ASD in this population compared to LR samples and given the HR parents' prior experience. As hypothesized, the PPV of the initial M-CHAT-R/F in the HR sample was significantly higher than that of the LR comparison sample (.531 versus .138), with small to medium effect sizes. After Follow-Up, the PPV increased to .607 in the HR sample and .488 in the LR group. PPV's of the two groups differed at initial screen but not on Follow-Up. Furthermore, the PPV for the LR group improved significantly with the Follow-Up, but the increase in the HR group was not significant. These findings suggest that Follow-Up has reduced clinical utility in this HR sample.

Furthermore, we expected that parents of children in the HR sample would be more accurate in their endorsement of at-risk behaviors due to their experience with ASD and likely increased understanding of items, leading to fewer items changing from "at-risk" to "not-at-risk" between initial screen and Follow-Up interview. Consistent with this prediction, items were less likely to change at Follow-Up in the HR sample compared to the LR sample (1.6

vs 2.9 items). Parents of older children with ASD may be more accurate in their reporting of ASD at-risk behaviors for later-born children as they may be more likely to recognize ASD-related behaviors whereas less familiar parents may be less certain about their presence or absence. Parents' endorsement of at-risk behaviors on the screener is also reflective of parents' willingness to report these behaviors. Parents of HR children may be more willing to endorse at-risk behaviors for later-born children as they may be more familiar with the process of evaluation and intervention and may recognize the importance of earlier detection and intervention in addressing developmental concerns.

Limitations and Strengths

There are several limitations to the current study, such as only screening at one time point, differences in ascertainment and attrition across samples, and limited follow-up on screen-negative cases. Screening in the current study took place during a toddler well-child visit (e.g., 18 to 24 months) that was later than some of previous HR screening studies, which may not be ideal given potential differences in onset of symptoms in HR younger siblings compared to LR children.³⁷ However, it is still valuable to screen at these later ages as cognitive ability may impact the emergence and detection of ASD with stronger cognitive abilities, as seen in our HR sample, associated with later age of onset or detection.³⁶

Ascertainment differences between samples likely contributed to differences in race and ethnicity, sex, and the amount of time elapsed between initial screening and Follow-Up. Specifically, the LR sample was collected through broad-based screening at pediatric well-child visits, whereas the HR sample was primarily self-referred and was informed about the study through research or clinical contact with an ASD-affected older sibling. In the LR sample, explicit efforts were made to include pediatric sites that served a more racially and economically diverse population. Additionally, a significant portion of the HR sample was lost to follow-up. Of the HR children completing the study, more children identified as Non-Hispanic White compared to another race/ethnicity, potentially contributing to the racial differences seen between the LR and HR samples. Of note, the effect size of differences between groups in race and ethnicity was quite small ($\phi = .06$). With respect to differences in sex distribution, the HR sample was less evenly distributed (i.e., more males) than the LR sample, although again the effect size was very small ($\phi = .02$). While this may represent a potential bias towards screening male younger siblings over female younger siblings, it may be warranted as ASD recurrence risk is approximately three times greater in male younger siblings than female younger siblings.¹⁶

Differences in time elapsed between screen and Follow-Up across groups is another limitation. The HR sample was more likely to complete Follow-Up and to be evaluated quickly compared to the LR sample. Thus, the LR group was significantly older at Follow-Up (~24 months) than the HR sample (~22 months). It is important to consider the delay between screen and Follow-Up, as longer delays, three months on average in the LR sample, may allow a child to develop some of the skills inquired about on the screener. Differing study procedures across samples, such as postal delays in receiving screeners from pediatrician offices in the LR sample, likely contributed to differences in time elapsed. However, it is possible that differential levels of parent concern (i.e., more concerned parents

returned calls more quickly) may have also contributed. Furthermore, there may be increased awareness about ASD in the more recent sibling sample, compared to the somewhat earlier LR sample collected, contributing to higher study completion rates.

Finally, minimal follow-up on screen negative cases precluded our ability to definitively calculate sensitivity and specificity and run an ROC analysis. In the current sample, 4.5% of screen negatives that were evaluated were identified as false negative cases; if extrapolated to the full sample of HR toddlers screened, this suggests that 5 children were potentially missed. Although following up on all screen negative cases would provide more complete data, it is critical to consider the cost (research time and effort) of such an endeavor, especially if it may only lead to the detection of a small number of ASD cases. Furthermore, as discussed above, the ASD detection rate in the current HR sample was 19.4%, which was comparable to prevalence rates in published literature, suggesting that most ASD cases were likely detected.

Despite these limitations, the current study has a number of strengths. First, the study adds to the limited body of literature examining screening in a HR sibling sample using a well-known and validated ASD screening measure. The study improves upon past screening studies in HR siblings with a larger sample size and provides a direct comparison of psychometric properties of a screening tool in a HR sample compared to a very large LR comparison sample. Additionally, unlike many screening studies in HR samples,^{12,13} the current study examined diagnostic accuracy and predictive power. Last, the sample is well-characterized using standardized measures of ASD severity, cognitive ability, and adaptive functioning and the findings are interpreted within this context. Furthermore, case confirmation utilized rigorous research reliable methods as opposed to community or record review of ASD diagnosis as in other studies.

Clinical Implications

The current study provides strong evidence that parent-completed screening measures can effectively detect ASD in HR siblings under the age of two years. Despite potentially subtle and inconsistent behavioral markers of ASD in HR younger siblings during the first few years of life,³⁶ parent detected, at-risk behaviors successfully discriminated between ASD and Non-ASD in a HR sample in the current study. Screening using parent-completed measures is significantly more cost-effective than using other more invasive or time-intensive measures. Importantly for physicians or other healthcare professionals who work with high-risk siblings, it demonstrates that the structured Follow-Up is not necessary to reduce false positives in HR siblings, reducing the amount of time needed for screening in this population. Screening for HR younger siblings is especially critical if it leads to the detection of higher functioning children with ASD, as past research highlights the importance of stronger cognitive skills and better motor skills at age 2 in achieving a positive outcome.³⁸ With these skills intact, these higher-functioning children may be more responsive to intervention. Screen positive HR children who do not go on to receive an ASD diagnosis are highly likely to receive other developmental diagnoses needing intervention, which is consistent with past research.¹⁷ Only one screen positive child in the current HR sample was determined to be typically developing. Based on these findings, we recommend

that physicians refer HR children who screen positive on the M-CHAT-R directly for developmental evaluation.

Future Directions

Future studies should more systematically follow screen negative cases to identify HR younger siblings who may be missed during screening. While the current study demonstrates that the M-CHAT-R/F performs differently in a HR younger sibling sample, the data cannot tease apart the impact of parent experience on endorsing at-risk ASD behaviors versus the impact of child factors (i.e., increased prevalence of ASD and subclinical symptoms) upon the screening process. The impact of other parent factors (e.g., maternal depression) and family factors (e.g., proband severity) could also be examined to extend the literature on their impact upon reporting of parent concerns to screening.

Conclusions

The current study investigated the performance of the M-CHAT-R/F in a HR sibling sample compared to performance in a large LR sample. Results indicate that the M-CHAT-R/F can successfully discriminate between ASD and Non-ASD in HR toddlers. Low response change rates and limited incremental improvement in PPV at Follow-Up provides evidence that parents with ASD experience are more accurate in their reporting of concerns on later-born children compared to parents who do not have an older child with ASD. Future research is needed to better understand other contributing factors to differential ASD screening performance in HR siblings.

Acknowledgements

The authors are grateful for the participation of children and families in this research study and the support of the research teams at University of Connecticut, Georgia State University, University of Washington, and Vanderbilt University. The authors would also like to acknowledge funding from: *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (R01HD039961), Maternal and Child Health Bureau Grant R40MC00270, the University of Connecticut's Research Foundation Faculty Grant, and University of Connecticut Doctoral Dissertation Fellowship. Dr. Kathryn Bradbury is now in the Child Development and Rehabilitation Center at Oregon Health & Science University. Dr. Diana Robins was in the Department of Psychology and Neuroscience Research Institute at Georgia State University during data collection for this study.

Author Disclosures

Drs. Robins, Fein, and Barton are co-owners of M-CHAT, LLC. M-CHAT, LLC licenses use of our intellectual property, the Modified Checklist for Autism in Toddlers (M-CHAT), for use in commercial products. None of the work in the current study generates any royalties or other income. Data presented in the current paper is from the freely available paper version of the M-CHAT-R. Dr. Robins is on the advisory board for Quadrant Biosciences, Inc and receives funding for her research from the National Institutes of Health. Dr. Warren has served as a consultant for Adaptive Technology Consulting, LLC and Hoffman La Roche. Dr. Warren reports that these relationships have no conflict of interest for the current work. The remaining authors declare that they have no conflict of interest.

REFERENCES

1. Hyman SL, Levy SE, Myers SM, Council on Children with Disabilities, Section on Developmental and Behavioral Pediatrics. Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics*. 2020;145(1):e20193447. doi:10.1542/peds.2019-3447. [PubMed: 31843864]
2. Lipkin PH, Macias MM, Council on Children with Disabilities, Section on Developmental and Behavioral Pediatrics. Promoting optimal development: identifying infants and young children with

developmental disorders through developmental surveillance and screening. *Pediatrics*. 2020;145(1):e20193449. doi:10.1542/peds.2019-3449. [PubMed: 31843861]

3. Robins DL, Casagrande K, Barton M, Chen C-MA, Dumont-Mathieu T, Fein D. Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics*. 2014;133(1):37–45. doi:10.1542/peds.2013-1813. [PubMed: 24366990]
4. Herlihy LE, Brooks B, Dumont-Mathieu T, et al. Standardized screening facilitates timely diagnosis of autism spectrum disorders in a diverse sample of low-risk toddlers. *J Dev Behav Pediatr*. 2014;35(2):85–92. doi:10.1097/DBP.0000000000000014. [PubMed: 24509053]
5. Baio J, Wiggins L, Christensen DL, et al. Prevalence of autism spectrum disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2014. *MMWR Surveill Summ*. 2018;67(6):1–23. doi:10.15585/mmwr.ss6706a1.
6. Ozonoff S, Young GS, Steinfeld MB, et al. How early do parent concerns predict later autism diagnosis? *J Dev Behav Pediatr*. 2009;30(5):367–375. [PubMed: 19827218]
7. Herlihy L, Knoch K, Vibert B, et al. Parents' first concerns about toddlers with autism spectrum disorder: effect of sibling status. *Autism*. 2015;19(1):20–28. doi:10.1177/1362361313509731. [PubMed: 24216070]
8. Hess CR, Landa RJ. Predictive and concurrent validity of parent concern about young children at risk for autism. *J Autism Dev Disord*. 2012;42(4):575–584. doi:10.1007/s10803-011-1282-1. [PubMed: 21584850]
9. Richards M, Mossey J, Robins DL. Parents' concerns as they relate to their child's development and later diagnosis of autism spectrum disorder. *J Dev Behav Pediatr*. 2016;37(7):532–540. doi:10.1097/DBP.0000000000000339. [PubMed: 27541581]
10. Arpino C, Compagnone E, Montanaro ML, et al. Preterm birth and neurodevelopmental outcome: a review. *Childs Nerv Syst*. 2010;26(9):1139–1149. doi:10.1007/s00381-010-1125-y. [PubMed: 20349187]
11. Oosterling IJ, Swinkels SH, Van der Gaag RJ, et al. Comparative analysis of three screening instruments for autism spectrum disorder in toddlers at high risk. *J Autism Dev Disord*. 2009;39(6):897–909. doi:10.1007/s10803-009-0692-9. [PubMed: 19205862]
12. Luyster RJ, Kuban KCK, O'Shea TM, et al. The modified checklist for autism in toddlers in extremely low gestational age newborns: individual items associated with motor, cognitive, vision and hearing limitations. *Paediatr Perinat Epidemiol*. 2011;25(4):366–376. doi:10.1111/j.1365-3016.2010.01187.x. [PubMed: 21649679]
13. Moore T, Johnson S, Hennessey E, et al. Screening for autism in extremely preterm infants: problems in interpretation. *Dev Med Child Neurol*. 2012;54(6):514–520. doi:10.1111/j.1469-8749.2012.04265.x. [PubMed: 22458327]
14. Robins DL, Fein D, Barton ML. Modified Checklist for Autism in Toddlers (M-CHAT). Storrs. 1999.
15. Pandey J, Verbalis A, Robins DL, et al. Screening for autism in older and younger toddlers with the Modified Checklist for Autism in Toddlers. *Autism*. 2008;12(5):513–535. doi:10.1177/1362361308094503. [PubMed: 18805945]
16. Messinger DS, Young GS, Webb SJ, et al. Early sex differences are not autism-specific: a Baby Siblings Research Consortium (BSRC) study. *Mol Autism* 2015;6(1):1. doi:10.1186/s13229-015-0027-y. [PubMed: 25705365]
17. Messinger D, Young GS, Ozonoff S, et al. Beyond autism: a baby siblings research consortium study of high-risk children at three years of age. *J Am Acad Child Adolesc Psychiatry*. 2013;52(3):300–308.e301. doi:10.1016/j.jaac.2012.12.011. [PubMed: 23452686]
18. Stone WL, McMahon CR, Henderson LM. Use of the Screening Tool for Autism in Two-Year-Olds (STAT) for children under 24 months. *Autism*. 2008;12(5):557–573. doi:10.1177/1362361308096403. [PubMed: 18805947]
19. Zwaigenbaum L, Bryson S, Rogers T, et al. Behavioral manifestations of autism in the first year of life. *Int J Dev Neurosci*. 2005;23(2–3):143–152. doi:10.1016/j.ijdevneu.2004.05.001. [PubMed: 15749241]

20. Feldman MA, Ward RA, Savona D, et al. Development and initial validation of a parent report measure of the behavioral development of infants at risk for autism spectrum disorders. *J Autism Dev Disord.* 2012;42(1):13–22. doi:10.1007/s10803-011-1208-y. [PubMed: 21387111]
21. Rowberry J, Macari S, Chen G, et al. Screening for autism spectrum disorders in 12-month-old high-risk siblings by parental report. *J Autism Dev Disord.* 2015;45(1):221–229. doi:10.1007/s10803-014-2211-x. [PubMed: 25149178]
22. Sacrey LAR, Bryson S, Zwaigenbaum L, et al. The Autism Parent Screen for Infants: predicting risk of autism spectrum disorder based on parent-reported behavior observed at 6–24 months of age. *Autism.* 2018;22(3):322–334. doi:10.1177/1362361316675120. [PubMed: 29671640]
23. Weitlauf AS, Vehorn AC, Stone WL, et al. Using the M-CHAT-R/F to identify developmental concerns in a high-risk 18-month-old sibling sample. *J Dev Behav Pediatr.* 2015;36(7):497–502. doi:10.1097/DBP.0000000000000194. [PubMed: 26171765]
24. Rescorla LA, Winder-Patel BM, Paterson SJ, et al. Autism spectrum disorder screening with the CBCL/1½–5: findings for young children at high risk for autism spectrum disorder. *Autism.* 2017;23(1):29–38. doi:10.1177/1362361317718482. [PubMed: 28931307]
25. Rutter M, Bailey A, Lord C. *The Social Communication Questionnaire: Manual.* Los Angeles, CA: Western Psychological Services; 2003.
26. Wiggins LD, Bakeman R, Adamson LB, et al. The utility of the Social Communication Questionnaire in screening for autism in children referred for early intervention. *Focus Autism Other Dev Disabil.* 2007;22(1):33–38. doi:10.1177/10883576070220010401.
27. Lord C, Rutter M, DiLavore P, et al. *Autism Diagnostic Observation Schedule–2nd Edition (ADOS-2).* Los Angeles, CA: Western Psychological Services; 2012.
28. Barton ML, Boorstein HC, Dumont-Mathieu TM, et al. *Toddler Autism Symptom Inventory.* Storrs, CT: Self-Published; 2012.
29. Lord C, Rutter M, Le Couteur A. *Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders.* *J Autism Dev Disord.* 1994;24(5):659–685. [PubMed: 7814313]
30. Schopler E, Reichler RJ, Renner BR. *Childhood Autism Rating Scale - Second Edition.* Los Angeles, CA: Western Psychological Services; 2010.
31. Mullen EM. *Mullen Scales of Early Learning: Manual.* Minneapolis: NCS Pearson, Inc.; 1995.
32. Sparrow SS, Cicchetti DV, Balla DA. *Vineland Adaptive Behavior Scales, Second Edition.* Circle Pines: AGS Publishing; 2005.
33. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* American Psychiatric Association; 2000.
34. Feldt LS. A test of the hypothesis that Cronbach's alpha or Kuder-Richardson Coefficient twenty is the same for two tests. *Psychometrika.* 1969;34(3):363–373. doi:10.1007/BF02289364.
35. Berends D, Dissanayake C, Lawson LP. Differences in cognition and behaviour in multiplex and simplex autism: Does prior experience raising a child with autism matter? *J Autism Dev Disord.* 2019;49(8):3401–3411. doi:10.1007/s10803-019-04052-7. [PubMed: 31102196]
36. Szatmari P, Chawarska K, Dawson G, et al. Prospective longitudinal studies of infant siblings of children with autism: lessons learned and future directions. *J Am Acad Child Adolesc Psychiatry.* 2016;55(3):179–187. doi:10.1016/j.jaac.2015.12.014. [PubMed: 26903251]
37. Macari SL, Campbell D, Gengoux GW, et al. Predicting developmental status from 12 to 24 months in infants at risk for Autism Spectrum Disorder: a preliminary report. *J Autism Dev Disord.* 2012;42(12):2636–2647. doi:10.1007/s10803-012-1521-0. [PubMed: 22484794]
38. Turner LM, Stone WL. Variability in outcome for children with an ASD diagnosis at age 2. *J Child Psychol Psychiatry.* 2007;48(8):793–802. doi:10.1111/j.1469-7610.2007.01744.x. [PubMed: 17683451]

Figure 1.

Flow Charts of Screening Results for High-Risk and Low-Risk Samples

^aScreen negative cases were detected through pediatrician red flag ($n = 1$) and Vanderbilt University follow-up on screen negatives ($n = 43$).

^bBoth cases were detected through Vanderbilt University follow-up on screen negatives cases.

^cScreen negative cases were detected through pediatrician red flag ($n = 20$), STAT screen positive ($n = 20$), or alternate scoring ($n = 3$).

^dASD cases were detected through pediatrician red flag ($n = 9$), STAT screen positive ($n = 6$), and alternate scoring ($n = 2$).

Table 1

Demographic Characteristics of Subjects with Complete Data by Risk Status

Demographic Characteristic	High-Risk (<i>n</i> = 175)	Low-Risk (<i>n</i> = 15,400)	Statistic
Age at Screen in Months (<i>M</i> , <i>SD</i>)	21.15 (4.08)	20.96 (3.30)	$t(176.6) = -0.61, p = .542, d = .06$
Age at Evaluation in Months (<i>M</i> , <i>SD</i>)	22.81 (3.98)	26.05 (5.48)	$t(246.4) = 6.14, p < .001, d = .68$
Sex (<i>N</i> , %)			$\chi^2(1, n = 15,321) = 4.04, p = .044, \phi = .02$
Male	101 (58.4)	7680 (50.7)	
Female	72 (41.6)	7468 (49.3)	
Race (<i>N</i> , %)			$\chi^2(1, n = 14,516) = 45.59, p < .001, \phi = .06$
White	143 (85.6)	8596 (59.9)	
Of Color	24 (14.4)	5753 (40.1)	
Maternal Education (<i>N</i> , %)			$\chi^2(1, n = 13,888) = 1.68, p = .194, \phi = .01$
No College Degree	68 (43.3)	6662 (48.5)	
College Degree or Higher	89 (56.7)	7069 (51.5)	

Note. Due to small sample size of each racial group, children of color were all combined into a single group for analyses. For the LR sample of color, 58% identified as Black, 16% identified as bi- or multiracial, 13% identified as Asian or Pacific Islander, 13% identified as Hispanic/Latino, and less than 1% identified as American Indian or Alaskan Native.

Table 2

Clinical Characteristics of Children with ASD by Risk Status

	High-Risk		Low-Risk				
Measure	<i>n</i>	Mean (<i>SD</i>)	<i>n</i>	Mean (<i>SD</i>)	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
Autism Diagnostic Observation Schedule (ADOS; Second Edition)							
Social Affect CSS	31	6.6 (1.9)	94	6.4 (1.7)	−0.57	.57	0.11
RRB CSS	31	7.7 (2.4)	94	7.0 (2.4)	−1.41	.16	0.29
Overall CSS	31	6.9 (2.0)	94	6.5 (2.1)	−1.16	.25	0.20
Childhood Autism Rating Scale (CARS; Second Edition)							
Total Score	27	33.4 (6.2)	95	32.7 (5.1)	−0.55	.58	0.12
Mullen Scales of Early Learning (MSEL)							
Verbal DQ	33	60.3 (24.4)	93	44.7 (20.6)	−3.57	.001	0.69
Non-Verbal DQ	33	85.6 (19.1)	94	66.1 (18.7)	−5.10	< .001	1.03
DQ	33	73 (19.4)	93	55.5 (17.3)	−4.82	< .001	0.95
Vineland Adaptive Behavior Scales – Second Edition (VABS-II)							
Communication SS	33	74.7 (16.1)	94	71.9 (12.2)	−0.91	.37	0.20
Daily Living SS	32	82.4 (12.6)	94	79.7 (14.6)	−0.94	.35	0.20
Motor SS	32	90 (11.7)	94	84 (11.6)	−2.53	.01	0.52
Socialization SS	33	76.6 (11.5)	94	77.4 (9.9)	0.40	.69	.07
Adaptive Behavior Composite	32	77.8 (11.5)	94	75.2 (9.9)	−1.19	.24	0.24

Note. ADOS CSS = Calibrated Severity Score (Range, 1–10; ASD Cutoff, 4). RRB = Restricted Repetitive Behavior. CARS(–2) Total Score range 15–60; ASD Cutoff, 25.5. MSEL DQ = Developmental Quotient ($M = 100$, $SD = 15$). VABS-2 SS = Standard Score ($M = 100$, $SD = 15$). Bonferroni correction was used to correct for multiple comparisons and resulted in an adjusted α value of .004 (.05/12 comparisons).

Table 3

Psychometric Properties of the M-CHAT-R/F by Screen Stage and Risk Status

	High-Risk (<i>n</i> = 64)			Low-Risk (<i>n</i> = 777)			
	TP	FP	PPV	TP	FP	PPV	<i>Statistic</i>
After M-CHAT-R Screen Positive							
ASD	34	30	0.531	107	670	0.138	$\chi^2(1, n = 841) = 65.62, p < .001, \phi = .28$
Any DD	58	6	0.906	202	575	0.260	$\chi^2(1, n = 841) = 115.63, p < .001, \phi = .37$
After M-CHAT-R/F Screen Positive							
ASD	34	22	0.607	100	105	0.488	$\chi^2(1, n = 261) = 2.51, p = .113, \phi = .10$
Any DD	55	1	0.982	193	12	0.941	FET, <i>p</i> = .310, $\phi = .08$

Note. TP = True Positive. FP = False Positive. PPV = Positive Predictive Value. ASD = Autism Spectrum Disorder. Any DD = Any developmental delay or concern. FET = Fisher's Exact Test. FET was used when expected cell counts were less than 5.