4-13-2018

M-CHAT-R/F Performance in a High-Risk Infant Sibling Population

Kathryn Bradbury
University of Connecticut - Storrs, kathryn.bradbury@uconn.edu

M-CHAT-R/F Performance in a High-Risk Infant Sibling Population

Kathryn Rae Bradbury, PhD

University of Connecticut, 2018

Most Autism Spectrum Disorder (ASD) screening measures have been developed for level 1 screening in low-risk (LR) children. Measures may perform differently in high-risk (HR) populations, such as younger siblings of children with ASD, due to parent factors (i.e., experience with ASD) and child factors (i.e., increased prevalence of ASD symptoms). 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 (n = 187) and directly compare its performance to that in a LR comparison sample (n = 15,848). 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. High predictive power of the M-CHAT-R at initial screen, with limited incremental change after Follow-Up, suggests that the Follow-Up is less critical in a HR sample. A significantly reduced rate of changed responses during Follow-Up further supports improved reporting accuracy of parents with ASD experience. Response to name was the best discriminating item regardless of risk group, although other discriminating items differed. HR-ASD toddlers had higher cognitive functioning compared to LR-ASD toddlers in the face of comparable ASD severity. This pattern of results is indicative of differential parent responding as opposed to behavioral differences between risk groups. In sum, the findings of the current study suggest that the M-CHAT-R/F can accurately distinguish between ASD and Non-ASD within the first two years of life in a HR sibling sample.

M-CHAT-R/F Performance in a High-Risk Infant Sibling Population

Kathryn Rae Bradbury

B.A., Colgate University, 2008

M.A., University of Connecticut, 2014

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2018

APPROVAL PAGE

Doctor of Philosophy Dissertation

M-CHAT-R/F Performance in a High-Risk Infant Sibling Population

Presented by

Kathryn Rae Bradbury, B.A., M.A.

Major Advisor _____
Deborah Fein, Ph.D.

Associate Advisor _____
Marianne Barton, Ph.D.

Associate Advisor _____
Diana L. Robins, Ph.D.

University of Connecticut
2018

ACKNOWLEDGEMENTS

First and foremost, I would like to thank all of the participating children, families, and pediatricians without whom this work would not be possible. I would also like to thank all of the clinicians, graduate students, and undergraduate assistants of the Early Detection Lab past and present for their dedication to the M-CHAT(-R) and ASD screening research. Additional thanks to our collaborators – Diana Robins, Ph.D. now at the A.J. Drexel Autism Institute, Wendy Stone, Ph.D. at the University of Washington, and Zachary Warren, Ph.D. at Vanderbilt University - and their study teams who were instrumental in the recruitment and collection of data from our high-risk sample. I am proud to be part of this work to further ASD screening research in the hopes of earlier detection and better outcomes for children with ASD and their families. This work was funded in part by a grant through the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

To my advisors, Deborah Fein, Ph.D. and Marianne Barton, Ph.D., I cannot begin express how grateful I am for your support, guidance, and mentorship throughout my graduate career at UConn as well as on this project. To Diana Robins, Ph.D., my associate advisor and creator of the M-CHAT(-R), your expertise and guidance on this project has been invaluable. A special thank you to committee members, Letitia Naigles, Ph.D. and Rhiannon Smith, Ph.D., who were willing to serve on my dissertation committee and to provide expertise from related fields. I would also like to acknowledge Jeffrey Burke, Ph.D. for consultation on a subset of statistical analyses.

Finally, a huge thank you to my social support – my fellow lab mates, cohort members, family and friends, and last but certainly not least, my partner. I could not have come this far without you.

Table of Contents

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by deficits in social communication and the presence of restricted interests and repetitive behaviors (American Psychiatric Association [APA], 2013). The Centers for Disease Control and Prevention (CDC; 2016) estimates that 1 in every 68 children (8-year olds) have ASD, with higher rates in males than in females (4.5 to 1). In 4-year-old children, a lower prevalence estimate (1 in 75) and less unbalanced sex ratio (3.3 to 1) has been found (Christensen et al., 2016). Despite evidence indicating that reliable diagnoses can be made around age two (Eaves & Ho, 2004; Kleinman, Ventola, et al., 2008; Turner & Stone, 2007), most children with ASD are still not diagnosed until close to four years of age (CDC, 2016).

Parent Report of Concerns in ASD

Parents play a unique and critical role in alerting physicians to concerns that might not be detected through pediatrician surveillance alone during a brief well-child visit (Dulcan et al., 1990). In general, parent report tends to be very sensitive (.80), but not very specific for behavioral, emotional, or communication concerns (Glascoe & Dworkin, 1995). Regardless of later diagnosis (ASD or Non-ASD), communication-related or language concerns are the most common type of concern raised by parents (Chawarska et al., 2007; Coonrod & Stone, 2004; De Giacomo & Fombonne, 1998; Herlihy, Knoch, Vibert, & Fein, 2013; Hess & Landa, 2012; Siegel, Pliner, Eschler, & Elliott, 1988; Young, Brewer, & Pattison, 2003), while socio-emotional concerns are typically less commonly reported (De Giacomo & Fombonne, 1998; Hess & Landa, 2012). Simultaneous concerns across two or more developmental domains are more indicative of later ASD diagnosis (Siegel et al., 1988). Taken together, these findings suggest that parents may be more sensitive to overt behaviors, such as abnormal language development, than more subtle behaviors, such as social development. However, when social concerns are voiced, especially in high-risk (HR) samples, they demonstrate increased specificity when using standardized test results as the "gold standard" compared to other types of concerns (Hess & Landa, 2012). In most of these studies, parent concerns were elicited using broad, open-ended questions, such as "What first concerned you about your child's development," which were then categorized by type of concern. In contrast,

questionnaires which probe each area of development, such as the Parents' Evaluations of Developmental Status (PEDS; Glascoe, 1997), may help to elicit additional concerns to which parents are less sensitive. Moreover, specific questioning about social communication skills and abilities may be necessary to elicit social concerns, as opposed to broad, open-ended questions (Coonrod & Stone, 2004).

There are mixed findings as to whether or not the age of first concerns differs by diagnostic group (i.e., ASD vs. Non-ASD). Coonrod and Stone (2004) and Siegel et al. (1988) found, on average, parents report first concerns around 18 months, regardless of later diagnostic status (i.e., ASD vs. Non-ASD). In contrast, Zuckerman, Lindly

, and Sinche (2015) showed that parents of children with ASD report earlier concerns than parents of children with ID/DD, and parents of children with comorbid ASD and ID/DD reporting the earliest concerns. Increased ASD severity and social skill deficits may also lead to earlier age of first concern within children who are later diagnosed with ASD (Coonrod & Stone, 2004).

Despite early reporting of concerns, there is often a lag between parents' first report of concerns and when a diagnosis is obtained. Siegel et al. (1988) found a lag of approximately one year between the report of first concern and the age of first diagnosis, and lag of approximately two additional years beyond initial diagnosis before obtaining a definitive ASD diagnosis. Zuckerman et al. (2015) found a similar delay of approximately 2.7 years between report of first concern and age at ASD diagnosis, with over 44 percent of their sample experiencing greater than a three year delay. Parents of children later diagnosed with ASD were more likely to receive a reassuring provider response compared to those who were later diagnosed with ID/DD, and that was associated with increased delays between a discussion with a provider about concerns and an ASD diagnosis (Zuckerman et al., 2015).

Research findings have been more consistent with respect to the quantity of parent concerns. By 24 months, parents of children later diagnosed with ASD report more concerns compared to parents of children who are later diagnosed with another developmental diagnosis, such as developmental delay (DD), specific language impairment (SLI), or typical development (TD; Hess & Landa, 2012; Veness et al., 2011). However, comparable rates of concern are observed at eight and 12 months (Veness et al.,

2011). At 14 months, parent report of concerns differs in some samples (Chawarska et al., 2007), but not in others (Hess & Landa, 2012). Increased reporting of concerns in ASD is also associated with familial factors, such as higher levels of maternal depression (Talbott, Nelson, & Tager-Flusberg, 2015) and poorer psychological well-being (Karp, Ibañez, Warren, & Stone, 2017), as well as greater symptom severity in older affected siblings (Talbott et al., 2015).

Parenting experience, specifically the opportunity to make comparisons to other children, also impacts reporting of concerns for later-born children. This experience has long been known to be more impactful than formal knowledge of development on parents' willingness to raise concerns with health care providers and on the accuracy of their concerns (Glascoe, Altemeier, & MacLean, 1989). Experienced parents have been shown to report earlier concerns than first-time parents (De Giacomo & Fombonne, 1998; Siegel et al., 1988). The presence of older siblings with ASD, in particular, is related to earlier report of concerns about later-born children in some samples (Herlihy et al., 2013) but not in others (Chawarska et al., 2007).

Parents who have older children with ASD tend to report more concerns about later-born children than do parents who have older TD children (McMahon, Malesa, Yoder, & Stone, 2007; Ozonoff et al., 2009; Sacrey et al., 2015). Furthermore, these parents report more concerns for later-born children who later receive an ASD diagnosis compared to those who do not (Sacrey et al., 2015). In a sample of HR infants, rate of parent concerns at 12 months was predictive of later ASD diagnosis at age 3, while the type of parent concern indicative of later ASD diagnosis varied by time point (Sacrey et al., 2015). Concerns about sensory and motor skills at 6 months were predictive of later ASD, while other concerns were not predictive until later – social concerns at 12 months, communication concerns at 15 months, and RRBs at 18 months (Sacrey et al., 2015). This pattern of concern appears to follow the typical development of skills in these domains. Development of motor skills is especially prominent before the onset of language, while the presence of atypical behaviors, such as repetitive behaviors, may not be observed until later. Although, the literature suggests that repetitive behaviors, may arise as early as 12 months (Wolff et al., 2014), contrary to what was previously thought.

Parent concerns appear to be valid, as they are correlated with scores on standardized assessment measures and predictive of later ASD diagnosis (Hess & Landa, 2012; McMahon et al., 2007; Ozonoff et al., 2009). However, the predictive validity of parent concerns differs by domain and remains lower than that of standardized measures, further emphasizing the need for more formal and systematic screening (Hess & Landa, 2012). Several studies (Hess & Landa, 2012; Glascoe & Dworkin, 1995) promote the elicitation of parent concerns as a pre-screening that warrants follow-up with standardized measures.

Consistent with this view, the American Academy of Pediatrics recommends ASD-specific screening at 18- and 24-month well-child visits in addition to ongoing developmental surveillance and broad developmental screening (Johnson & Myers, 2007). Implementation of universal screening using standardized measures has been shown to lower the age of diagnosis to close to two years old (Robins et al., 2014); this is especially important in minority populations and those with lower socio-economic status, who may otherwise experience even greater delays (Herlihy et al., 2014). Familial factors, including higher maternal education and parenting experience with other children, especially those with ASD, also contribute to earlier diagnosis (Bickel, Bridgemohan, Sideridis, & Huntington, 2015).

Formal screening using evidenced-based screening measures improves early detection over clinical surveillance alone; however, a combined approach including both, in conjunction with informal elicitation of parental concerns, proves to be the most effective method (Miller et al., 2011). Universal screening also facilitates identification of ASD-specific signs where parents do not have any concern about ASD (Mathews, King, Kupzyk, & Lake, 2014). While universal screening identifies more cases, physicians may still override screening results, which may lead to continued delays in diagnosis (Thomas, Spragins, Mazloum, Cronkhite, & Maru, 2016).

ASD Screening in High-Risk Populations

Despite a substantial body of literature on differential parent concerns in children at HR for developing ASD, little research has been conducted to determine how differential parental concern translates to performance on formal ASD screening measures. Screening research is particularly important as screening provides a prospective approach to collecting parent concerns which is less likely

to be biased than the retrospective report of parent concerns that is frequently collected as part of ASD diagnostic interviews. Screening in HR populations is critical given the high proportion of children who go on to develop severe psychopathology, including ASD (For review, see Arpino et al., 2010). However, most ASD screening measures have not been developed explicitly for use in this population and thus, may not display adequate predictive validity. To this end, Oosterling et al. (2009) examined several ASD screening instruments (i.e., ESAT [Dietz, Swinkels, van Daalen, Van Engeland, & Buitelaar, 2006; Swinkels et al., 2006], SCQ [Rutter, Bailey, & Lord, 2003], ITC [Wetherby & Prizant, 2002], and key items of the CHAT (Baron-Cohen, Allen, & Gillberg, 1992) and found that no instrument demonstrated adequate sensitivity, specificity, PPV, or NPV at either the level of a total score or individual items in a HR population (due to physician referral or screen positive on ESAT) aged 8 to 44 months.

Most research on screening for ASD in HR populations conducted thus far has focused on children that have been deemed at-risk due to prematurity, positive screen on another measure, or pediatrician concern. In extremely premature infants, M-CHAT/F (Robins, Fein, & Barton, 1999) screen positive rates are significantly higher than compared to low-risk (LR) populations, ranging from 21 to 41% (Kuban et al., 2009; Limperopoulos et al., 2008; Luyster et al., 2011; Moore, Johnson, Hennessy, & Marlow, 2012). However, studies caution that screen positive results should be interpreted within the context of other neurodevelopmental impairments, as they may represent a high false positive rate given high correlations between sensory impairments and screen positive rates (Kuban et al., 2009; Luyster et al., 2011; Moore et al., 2012). High screen positive rates on the M-CHAT/F in these populations are consistent with elevated screen positive rates (20%; Stephens et al., 2012) on other ASD screening measures, such as the PDDST-II (Siegel, 2004), even though the latter instrument was developed for screening with more impaired populations. It is important to note that few studies of ASD screening in extremely premature infants have examined diagnostic accuracy of positive screens. In one study, Johnson et al. (2010) found that while 15.8% of children screened positive on the SCQ (Rutter et al., 2003), only 8% received an actual ASD diagnosis; thus, providing further support of a large proportion of false positives contributing to high screen positive rates in these HR extremely premature populations.

Similarly, elevated screen positive rates (40%) on the M-CHAT/F have been observed in a HR population referred from early intervention (EI) sites or due to pediatrician concerns (Kleinman, Robins, et al., 2008; Pandey et al., 2008). However, unlike Johnson et al.'s (2010) findings in premature infants, children referred from EI who screened positive on the M-CHAT/F were much more likely to receive an ASD diagnosis, as evidenced by a high positive predictive value (PPV) of .76. Furthermore, Kleinman, Robins, et al. (2008) found that the Follow-Up interview was not as critical in the HR population to reduce false positives, as it only raised the PPV from .60 to .76, compared to a large increase in PPV from .11 to .65 in the LR sample. Of note, children diagnosed with ASD from either sample (LR or HR) did not differ with respect to clinical characteristics. Pandey et al. (2008) found similarly high M-CHAT/F PPV rates (.74 to .79) in a larger HR sample also referred from EI providers, irrespective of age at screen (i.e., less than 24 months vs. greater than 24 months). It is likely that false positive rates are lower in EI-referred populations and in HR younger siblings than in extremely premature infants, who are susceptible to substantially more varied neurodevelopmental impairments that may or may not be directly associated with ASD (See Arpino et al., 2010 for review).

ASD Screening in High-Risk Younger Siblings

In an effort to better understand the early emergence of ASD symptoms and their development over time, research in the field has focused on infant (i.e., younger) siblings of children with ASD (for review, see Szatmari et al., 2016). The term younger siblings will be used instead of infant siblings in the current paper as the study participants are toddlers. This population is at increased risk for developing ASD, with recurrence risk ranging from 10-20% (Constantino, Zhang, Frazier, Abbacchi, & Law, 2010; Messinger et al., 2015; Ozonoff et al., 2011; Szatmari, 1999). Diagnosis is relatively stable in HR younger siblings, with 83% of children diagnosed at 18 months and 92% diagnosed at 24 months continuing to meet criteria for ASD at 36 months (Zwaigenbaum et al., 2016). Those HR younger siblings who do not ultimately develop ASD often have other developmental delays and subclinical ASD symptoms that are also frequently in need of treatment (Gamliel, Yirmiya, Jaffe, Manor, & Sigman, 2009;

Messinger et al., 2013; Toth, Dawson, Meltzoff, Greenson, & Fein, 2007), which makes screening in this population essential.

While screening in HR younger siblings is vital, it is also more complicated than screening in the general, LR population, especially since many of measures were not explicitly developed for use in a HR population. Macari et al. (2012) determined that HR younger siblings may display a unique constellation of symptoms early on that lead to a later ASD diagnosis which may differ from symptoms in children in the LR population who go on to receive an ASD diagnosis. In addition, the higher prevalence of sub-clinical ASD symptoms in this population (Constantino et al., 2006; Pickles et al., 2000; Schwichtenberg, Young, Sigman, Hutman, & Ozonoff, 2010) may make it more difficult to distinguish those who will go on to have ASD from those who will not. Parents of HR younger siblings may also be influenced by other factors including parenting experience (McMahon et al., 2007; Ozonoff et al., 2009; Sacrey et al., 2015) and psychological well-being (Karp et al., 2017; Talbott et al., 2015), adding a further layer of complication. These factors are all likely to interact and to contribute to differential screener performance in HR populations that is not yet well understood.

To date, there have been few studies published investigating screening for ASD in a HR younger sibling population. See Table 1 for summary of studies. Two studies (Stone, McMahon, & Henderson, 2008; Zwaigenbaum et al., 2005) have been conducted using clinician-rated observations and four studies (Feldman et al., 2012; Rowberry et al., 2015; Sacrey et al., 2016; Weitlauf, Vehorn, Stone, Fein, & Warren, 2015) have been conducted using parent-report questionnaires. Sample sizes of HR younger siblings range from 59 (Stone et al., 2008) to 204 (Sacrey et al., 2016). Most studies consisted of screening at one or two time points, while two studies (Feldman et al., 2012; Sacrey et al., 2016) repeatedly screened HR children at six time points. While it may be important to conduct ongoing screening, especially in HR younger siblings where patterns of onset may show increased variability (Bryson et al., 2007; Macari et al., 2012), familiarity of the items on a repeated measure must be considered when interpreting their psychometric properties. Moreover, Sacrey et al. (2016) noted that parents received feedback on their ratings, which in turn may have impacted subsequent ratings.

From this point on, all HR younger siblings who later receive an ASD diagnosis will be referred to as HR-ASD, all HR younger siblings who later receive a Non-ASD diagnosis will be referred to as HR-NonASD, and all LR typically developing controls will be referred to as LR-TDC. Clinician-rated measures of observation have demonstrated the ability to differentiate between HR-ASD from HR-NonASD and LR-TDC as early as 12 months on the Autism Observation Scale for Infants (AOSI; Bryson, McDermott, Rombough, Brian, & Zwaigenbaum, 2008; Zwaigenbaum et al., 2005) and 14 months on the Screening Tool for Autism in Two-year-olds (STAT; Stone, Coonrod, Turner, & Pozdol, 2004; Stone et al., 2008). Zwaigenbaum et al. (2005) identified the following behavioral risk markers at 12 months to be predictive of ASD diagnosis at 24 months: atypical eye contact and visual tracking, disengagement of visual attention, reduced orienting to name, imitation, social smiling, and social interest, as well as atypical reactivity and sensory-oriented behaviors. Stone et al. (2008) examined an extension of the STAT to younger age ranges and found adequate predictive power (PPV = .68) for a modified cut-score of 2.75 in children 14 months and older. However, specific behavioral markers that differentiated HR-ASD from HR-NonASD were not identified. The sample also included children who were referred due to physician concern (n = 12), which make comparisons to the other studies with only HR younger siblings more complicated. Taken together, these findings indicate that clinician-rated screening measures are able to identify behavioral markers that differentiate HR-ASD from other HR younger siblings early in the first two years of life.

More recently, several studies have examined the ability of parent-rated screening measures to differentiate between these groups. Feldman et al. (2012) was the only study to examine this in a measure specifically designed for screening in a HR sibling population, the Parent Observation of Early Markers Scale (POEMS). The POEMS (Feldman et al., 2012) was developed to monitor HR younger siblings over the course of their development based on items from the ADI-R (Lord, Rutter, & Le Couteur, 1994), ADOS (Lord et al., 2000), and the Childhood Autism Rating Scale (CARS; Schopler, Reichler, & Renner, 1995). It includes items specific to ASD, as well as items that tap into behavioral and emotional concerns. Parents rate the items on a 4-point scale, similar to the CARS. Feldman et al. (2012) found that POEMS

scores were significantly higher for HR-ASD children compared to HR-NonASD children beginning at 9 months and all following screening time points through 24 months. Items that differentiated between diagnostic groups at 9 months included: interest in faces, shifts attention to person, mood, response to name, and waiting. Despite the ability to differentiate HR-ASD children from other HR children by 9 months, the overall PPV for the POEMS was quite low (.21), with the highest predictive power seen at 9 months (PPV = .29; Feldman et al., 2012). One major limitation of the study was the very small sample of children who were ultimately diagnosed with ASD (n = 9) and the reliance on community diagnosis for diagnostic classifications.

Sacrey et al. (2016) took a similar approach to Feldman et al. (2012) and examined performance of a parent-completed screening measure, the Autism Parent Screen for Infants (APSI), at six time points over the course of development beginning at 6 months and continuing through 24 months. The APSI is a 26-item questionnaire based on the AOSI (Bryson et al., 2008), where for each behavior, parents rate whether the child "definitely" or "possibly" exhibits the behavior or "no" he/she does not exhibit the behavior. Sacrey et al. (2016) found that total scores on the APSI across all screening time points beginning at 6 months differentiated HR-ASD from HR-NonASD and LR-TDC at 36 months. Individual items that differentiated between diagnostic groups were visual tracking and vocalization with parents. By 18 months, the HR-ASD group showed differential performance on 21 of the 26 items on the APSI (Sacrey et al., 2016). PPV ranged from .47 (6 months) to .68 (18 months), while the optimal sensitivity and specificity balance was achieved at the youngest age range (6 months). Recently, when Sacrey et al. (2017) compared parent-report on the APSI to clinician ratings on the AOSI, they found relatively poor agreement between ratings. However, this was likely due to the fact that parents identified an additional 9 items beyond those observed by clinicians that differentiated between HR-ASD and HR-NonASD children, emphasizing the importance of parent report when screening for ASD. Results indicate that parents can identify at-risk behaviors that may be less likely to appear during a brief observational screening measure, such as repetitive motor behaviors, unusual sensory responses, and reactivity.

In another study, Rowberry et al. (2015) examined whether the First Year Inventory (FYI; Baranek, Watson, Crais, & Reznik, 2003), an ASD screening instrument developed for 12-month-old children, could distinguish HR-ASD from HR-NonASD and HR children with typical development (HR-TD). They found that the FYI *Imitation* construct (social imitation skills, early motor, and vocal skills) was the best at discriminating between these outcomes, with large effect sizes (Cohen's *d*s greater than 1). Parent ratings on the FYI for social behaviors, but not sensory or repetitive behaviors, were correlated with clinician ratings on the Autism Diagnostic Observation Schedule – Toddler Module (ADOS-T; Lord, Luyster, Gotham, & Guthrie, 2012). These results suggest that parents are noticing and reporting differences in social behavior more than sensory and repetitive behaviors (Rowberry et al., 2015). Although, this discrepancy may indicate that sensory and repetitive behaviors may be less likely to be observed in an evaluation as in Sacrey et al. (2017).

Last, Weitlauf et al. (2015) examined the performance of the M-CHAT-R/F (Robins, Fein, & Barton, 2009) at 18 months in a subset of HR younger siblings (n = 74) from a single site within the current larger scale, multi-site study. PPV estimates for the HR sibling population (PPV = .78) were significantly higher than that of a LR population (PPV = .475; Robins et al., 2014), which is expected given high recurrence risk of ASD in younger siblings (10-20%; Constantino et al., 2010; Messinger et al., 2015; Ozonoff et al., 2011; Szatmari, 1999). Consistent with findings from LR populations (Robins et al., 2014), PPV (.96) was higher for identifying children with any developmental concerns opposed to solely ASD (PPV = .78; Weitlauf et al., 2015). Of note, children at-risk for ASD, defined by the presence of ASD symptoms not definitive enough to warrant diagnosis (typically due to age or complexity) were included in the HR-ASD group. The inclusion of these children may have contributed to higher PPVs seen in this study. Again, differential performance on individual items for the HR-ASD group was not examined.

Taken together, these studies provide strong evidence that parents of HR-ASD children are able to identify and report at-risk behaviors within the first two years of life that predict later ASD diagnosis with fairly good discriminability. However, the previous literature has not thoroughly addressed how

ASD screening in this population compares directly to ASD screening in a LR population. Several studies only utilized a HR sample (Feldman et al., 2012; Stone et al., 2008; Weitlauf et al., 2015), while others included a LR control group of typically developing children without a family history of ASD (Rowberry et al., 2015; Sacrey et al., 2016; Zwiagenbaum et al., 2005). None of the studies conducted thus far has included LR children who are later diagnosed with ASD. Moreover, while previous work has discussed screening in the context of ASD symptomatology, few studies have used a large, well-characterized sample and interpreted their findings within this context.

The Present Study

The current study seeks to extend research on ASD screening in HR siblings to determine how parenting experience (i.e., having an older child on the spectrum) and child characteristics (i.e., increased ASD symptomatology) impacts responses about subsequent children on a standardized ASD screening measure, the M-CHAT-R/F. In contrast to previous studies, the current study directly compares screening in a HR sample of younger siblings to screening in a large LR sample to better understand how these measures perform differently in these two populations. To this aim, psychometric properties of the M-CHAT-R were examined, including screen positive rate, ASD detection rate, PPV, and internal consistency. The values were then compared to the psychometrics of the M-CHAT-R with a LR sample, a subset of the sample from the M-CHAT-R/F validation study by Robins et al. (2014). Item-level analyses were conducted to identify potential differences in item performance between the HR and LR samples. Discriminant function analyses (DFAs) were conducted to determine which items were best at discriminating between ASD and Non-ASD within each of the samples. A series of two-way between-subjects ANOVAs were run to examine the impact of risk status and diagnostic group upon ASD severity, cognitive ability, and adaptive functioning. Exploratory simple main effects analyses were run to compare the HR-ASD and LR-ASD groups.

Given previous research, we hypothesized the following – (1) the HR sample will have a higher screen positive rate than the LR sample, (2) the M-CHAT-R/F will demonstrate higher predictive power (i.e., higher PPV) in the HR sample compared to the LR sample, (3) 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, and (4) items that best discriminate between ASD and Non-ASD cases on the M-CHAT-R will be comparable in HR and LR samples. Given the very limited research on behavioral differences between LR and HR samples with ASD, exploratory analyses were conducted examining clinical characteristics and no specific hypotheses were made.

Methods

Participants

Toddlers at high familial risk (HR) for ASD were recruited directly through ASD resource fairs or indirectly through clinical research or treatment involving an older sibling with ASD. High familial risk was defined by the presence of at least one confirmed affected older full or half sibling (proband). A total of 241 HR toddlers were screened with the Modified Checklist for Autism in Toddlers, Revised with Follow-up (M-CHAT-R/F; Robins et al., 2009) at one of three sites – University of Connecticut (UConn) in Storrs, CT, University of Washington (UW) in Seattle, WA, and Vanderbilt University (VU) in Nashville, TN. Fifty-four toddlers were excluded as they did not meet study inclusion criteria or for other reasons. HR toddlers were excluded if they were a twin of the proband (n = 1), had a severe sensory or motor impairment to preclude testing, a pre-existing diagnosis of ASD or other known neurological condition prior to screening (n = 3), a birth weight < 2000g, or a gestational age < 37 weeks (n = 3). Children were also excluded if they were screened outside of the eligible screening range (16 to 30 months; n = 11), if parents refused participation (n = 8), or ASD diagnosis of proband was not confirmed (n = 4). Eight children were excluded for unknown reasons and 16 children were lost to follow-up during the enrollment process.

Children with incomplete data (i.e., incomplete Follow-Up or incomplete evaluation) were also excluded from analyses (n = 12) – three did not complete Follow-Up and 9 did not complete evaluation. Completers were more likely to identify their race as White compared to non-completers, χ^2 (1, n = 179) = 10.19, p = .001, ϕ = .24. No other systematic differences were observed, such as age at screen, sex, or

maternal education. The final HR sample consisted of 175 toddlers between 16 and 30 months (M = 21.2, SD = 4.1) who were screened at UConn (n = 23), UW (n = 49), or VU (n = 103). See Table 2 for demographics by study site; there were no significant differences among the sites.

The current study utilized a subsample of children from the M-CHAT-R/F validation study (Robins et al., 2014) as a LR comparison group screened during well-child visits at the pediatrician office. Only children who completed screening in English were included in the LR comparison sample (329 children screened in Spanish were excluded). An additional 44 children were excluded as they did not meet inclusion criteria or for other reasons. Specifics of recruitment and inclusion/exclusion criteria for the LR sample are described elsewhere (Robins et al., 2014). The final LR sample consisted of 15,400 toddlers screened at UConn (n = 5,612) and Georgia State University (GSU; n = 9,788) between 16 and 30 months (M = 20.96, SD = 3.30). See Table 3 for demographics by risk group. The LR sample was 50.7% male, and 59.9% identified as White. Mothers of children in the LR sample were generally highly educated (51.5% college educated). The HR and LR samples were comparable with respect to demographic variables, except the LR sample had more children of color (χ^2 [1, n = 14,516] = 45.59, p < .001, ϕ = .06) and a lower proportion of males (χ^2 [1, n = 15,321] = 4.04, p = .044, ϕ = .02). However, these differences had very small effect sizes. There was no group difference in maternal education.

Procedure

All procedures were approved by the Institutional Review Board at each participating site. Caregivers of participating children completed the M-CHAT-R in English. The caregivers of children who screened positive were contacted and asked the structured Follow-Up questions to confirm at-risk responses. Children who continued to screen positive were invited to attend a free developmental and diagnostic evaluation at their local university conducted by a clinician team consisting of a licensed clinical psychologist or developmental pediatrician and a clinical psychology doctoral student or trained research assistant. Evaluations included direct assessment of cognitive and social-communication skills as well as parent report of development, ASD symptomatology, and adaptive skills collected via interview.

At the end of the evaluation, parents were provided with verbal feedback, including any appropriate diagnoses and recommendations, and they were sent a comprehensive report via mail.

ASD diagnoses were made based upon expert clinical judgment using all available information according to DSM-IV-TR (APA, 2000) criteria. Children who did not meet criteria for ASD were diagnosed with other developmental diagnoses, such as developmental delay or developmental language disorder, as appropriate. Children who did not meet criteria for any DSM-IV-TR criteria, but who had significant subclinical ASD symptoms of concern were labeled as "At-Risk for ASD" (i.e., broader autism phenotype) and those with other notable subclinical symptoms were labeled as "No diagnosis". 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 defined as not meeting ASD criteria on ASD measures, having no delays of greater than one and half standard deviations on any cognitive or adaptive domain, and not meeting criteria for any other clinical diagnosis in DSM-IV-TR.

For screen positive cases in the HR sample, a similar evaluation was also conducted with each child's older sibling with ASD (proband) for diagnostic confirmation. Diagnosis of probands for children who screened negative on the M-CHAT-R/F (and were not evaluated) was confirmed via review of a previous evaluation report and parent completion of the SCQ (Rutter et al., 2003).

Efforts were made to detect missed cases (children with ASD who screened negative on the M-CHAT-R/F) in both groups. Pediatricians were asked to "flag" children about whom they had ASD-specific concerns. One site (VU) also followed up with a subset of screen negative cases, results of which were previously published in Weitlauf et al. (2015). For more details about ascertaining missed cases in the LR sample, see Robins et al. (2014). All supplemental methods for detecting false negative cases triggered automatic evaluation.

Measures

Modified Checklist for Autism in Toddlers, Revised with Follow-up (M-CHAT-R/F). 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 (Robins et al., 2014). A positive screen on the M-CHAT-R consists of three or more at-risk responses out of 20 total items, while 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 questions to confirm at-risk responses. As this project was conducted concurrently with the M-CHAT-R/F validation study, a two-stage screening process as described in Robins et al. (2014) was utilized for a portion of the sample. Eight children who received "high fail" scores on the M-CHAT-R (i.e., $\geq$ 8) bypassed Follow-Up and were immediately eligible for evaluation.

Social Communication Questionnaire (SCQ). The SCQ (Rutter et al., 2003) is a 40-item parent-report questionnaire used to assess ASD symptomatology in children aged four and older. The SCQ demonstrates good internal consistency (α = .90) and strong convergent validity with the ADI (r = .71; (Berument, Rutter, Lord, Pickles, & Bailey, 1999). The initial standardization sample suggested a cutoff of greater than or equal to 15, which yielded a relatively good sensitivity (.85) and specificity (.75; (Berument et al., 1999). More recent research recommends lower cutoffs, such as 11 (Wiggins, Bakeman, Adamson, & Robins, 2007) or 13 (Snow & Lecavalier, 2008), in younger children as the measure is not as sensitive in these populations. The SCQ was used in the current study to confirm diagnosis of probands with a cutoff of 11. The Current version was used for probands younger than six years old, while the Lifetime version was used for probands six years and older.

Autism Diagnostic Observation Schedule (ADOS). The ADOS (Lord et al., 2000) is a play-based diagnostic instrument used to elicit social communication and restricted and repetitive behaviors. Different modules are used based on a child's age and language ability. Inter-rater reliability was high for Module 1 (Mk_w .78) and Module 2 (Mk_w of .70; (Lord et al., 2000). Most children received the ADOS as the ADOS-2 (Lord et al., 2012) was still in development at study start. Modules 1 and 2 were used in the current study. To facilitate comparison across modules, ADOS calibrated severity scores were calculated for Overall Severity, Social Communication Severity, and RRB Severity, using algorithms developed by Gotham, Pickles, and Lord (2009) and Hus, Gotham, and Lord (2014) respectively. ADOS Calibrated

Severity Scores are comparable to ADOS-2 Comparison Scores and were used to characterize the sample. ADOS scores were also considered during diagnostic decision making in the current study.

Developmental History Form. Demographic characteristics, developmental history, and family history were collected through an investigator-developed Developmental History Form. Information collected on this form was used in characterizing the current sample. Maternal education was used as a proxy for socio-economic status.

Parent Interview of ASD Symptoms. Parent report of ASD symptomatology was collected through two separate semi-structured interviews dependent upon a child's age – the Toddler Autism Symptom Inventory (TASI; Barton, Boorstein, Dumont-Mathieu, Herlihy, & Fein, 2012) for children under three years old and the Autism Diagnostic Interview, Revised (ADI-R; Lord et al., 1994) for older children. Both interviews were developed based upon DSM-IV-TR criteria of Autistic Disorder and are used to assess the presence of symptoms. The TASI is investigator-developed and is currently undergoing reliability testing and validation. Results from the interviews were considered for clinical best estimate of diagnosis.

Childhood Autism Rating Scale (CARS). The CARS (Schopler et al., 1995) is a 15-item clinician rating scale of ASD severity based upon observation and parent report. While a cut-off of 30 is typically used for Autistic Disorder, a cutoff of 25.5 has been proposed for ASD more broadly (Chlebowski, Green, Barton, & Fein, 2010). Each item is rated on a 7-point scale ranging from 1 ("within normal limits for that age") to 4 ("severely abnormal for that age"), with half point scores possible between the anchors. Internal consistency of the CARS is high ($\alpha = .94$) and inter-rater reliability is good (.71; Schopler et al., 1995). A recent factor analysis of the CARS provided support for a three-factor solution (Social Communication, Stereotyped Behaviors and Sensory Sensitivities, and Emotional Reactivity) in a sample of two-year-olds (Moulton, Bradbury, Barton, & Fein, 2016). Factor scores are calculated by the average score across each of items that load on the factor, resulting in a score ranging from 1 to 4, which is interpreted similarly to individual CARS items. CARS total scores and factor scores

were used to characterize the sample. CARS ratings were also considered during clinical best estimate of diagnosis.

Mullen Scales of Early Learning (MSEL). The MSEL (Mullen, 1995) is a standardized test of cognitive development for use in children between one to 68 months old and consists of five domains – gross motor, visual reception, fine motor, receptive language, and expressive language. Each domain yields a t-score and combines to yield an overall score called the Early Learning Composite. Inter-scorer reliability is high, ranging from .91 to .99 across all age ranges (Mullen, 1995). The gross motor scale was not administered in the current study. Due to floor effects on the MSEL, developmental quotients (DQ=MA/CA*100) were derived using age equivalents as mental age, as has been done elsewhere in the literature (e.g., Munson et al., 2008; Rowberry et al., 2015). 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. MSEL scores were used to characterize the sample.

Vineland Adaptive Behavior Scales – Second Edition, Survey Interview Form (VABS-II). The VABS-II (Sparrow, Cicchetti, & Balla, 2005), is a semi-structured parent interview used to assess a child's adaptive behavior skills across four domains – communication, socialization, daily living, and motor skills. Each domain yields a standard score and combine to yield an overall Adaptive Behavior Composite score. Inter-rater reliability is high, ranging from .73 to .87 across domains in the zero to six years of age cohort, and test-retest reliability was also high, ranging from .82 to .86 in the zero to two years of age cohort (Sparrow et al., 2005). VABS-II scores were used to characterize the sample.

Results

Performance of the M-CHAT-R/F

No significant differences were observed in demographic variables across HR study sites, so the samples were combined for all subsequent analyses. ***Internal consistency*** for the M-CHAT-R was significantly greater for the HR sample (Cronbach's α = .88) than for the LR sample (Cronbach's α =

.64), as determined by Feldt Test (W = .34, p < .0001; Feldt, 1969). This suggests that M-CHAT-R items perform better as aspects of a single construct in the HR sample than the LR sample.

Of the 187 HR children who were screened, 76 (40.6%) screened positive on initial screening. See Figure 1 for flow chart of screening results for HR sample. Follow-Up interviews were conducted for 65 (85.5%) of the eligible children to confirm at-risk responses. Eight children who had high fail scores on initial screening ($\geq$ 8) bypassed Follow-Up per the two-stage screening procedures described in Robins et al. (2014). In most cases, Follow-Up was completed shortly after initial screening. Parents were contacted several times before they were considered lost to follow-up. The delay between M-CHAT-R completion and Follow-Up was significantly shorter in the HR sample (M = .63 months, SD = .87) than in the LR sample (M = 3.73 months, SD = 4.57), t (181.2) = 15.46, p < .001, d = .69). After Follow-Up, most children (n = 65; 85.5%) in the HR sample continued to screen positive (including the eight high fail children) and were eligible for evaluation. Fifty-six evaluations were completed and 34 (19.4%) of HR children received an ASD diagnosis.

In an effort to identify false negatives (ASD cases that screened negative), 44 evaluations were conducted of screen negative cases due to reported physician concern (n = 1) or as part of VU's follow up on screen negative cases (n = 43). Description of these procedures and characterization of these children is provided elsewhere (Weitlauf et al., 2015). Only two additional ASD cases were detected using these procedures, which represents 4.5% of evaluated screen negative cases. Both of these cases were identified through VU's follow up on screen negatives. Extrapolated to the full set of children who screened negative, it suggests that approximately 5 of the 111 screen negative children are likely false negatives. As evaluation of screen negative children was limited in the current study, reliable sensitivity and specificity calculations, as well as ROC analyses, could not be conducted.

A similar flow chart showing screening results for the LR comparison sample can be found in Figure 2. Compared to the LR sample, the HR sample had a significantly higher screen positive rate at both initial screen (LR = 7.1%; HR = 40.6%; χ^2 [1, n = 16,035] = 300.1, p < .001, ϕ = .14) and at Follow-Up (LR = 37.1%; HR = 87.7%; χ^2 [1, n = 975] = 64.32, p < .001, ϕ = .26) with small to medium effect

sizes, respectively. In the HR sample, 34 ASD cases were detected (19.4% detection rate), which was significantly higher than in the LR sample (.65% detection rate; χ^2 [1, n = 15,575] = 715.42, p < .001, ϕ = .20).

Positive predictive value (PPV) is the ratio of children who screened positive and received an ASD diagnosis (i.e., true positives) to all children who screened positive (i.e., true positives and false positives). See table 4 for PPVs by risk group and screening stage. For these analyses, any DD refers to children who received an ASD diagnosis, another developmental diagnosis, or raised significant clinical concern at evaluation. PPV for ASD in the HR sample was significantly higher than that of the LR sample at initial screen (χ^2 [1, n = 841] = 65.62, p < .001, ϕ = .28), but not at Follow-Up (χ^2 [1, n = 261] = 2.51, p = .113, ϕ = .10). Furthermore, PPV for the HR sample did not differ between initial screen and Follow-Up for either ASD (χ^2 [1, n = 120] = 0.70, p = .403, ϕ = .08) or for any DD (FET = .12, ϕ = .16), while PPV for the LR sample was higher after Follow-Up for both ASD (χ^2 [1, n = 982] = 119.51, p < .001, ϕ = .35) and any DD (χ^2 [1, n = 982] = 313.31, p < .001, ϕ = .56). Taken together, these results suggest that Follow-Up was not as critical in the HR sample as it was in the LR sample to reduce false positives. For the HR group, the M-CHAT-R/F continued to demonstrate a very high PPV (.982) for any DD, as in the LR group (.941).

Item-Level Performance of the M-CHAT-R/F

The following analyses were completed with the subset of children who screened positive on the M-CHAT-R and completed Follow-Up (HR: n = 59; LR: n = 956). As anticipated, the HR sample had significantly higher M-CHAT-R total scores (HR: M = 6.7, SD = 2.9; LR: M = 4.1, SD = 2.5; t (63.3) = -6.80, p < .001, d = .06) and M-CHAT-R/F total scores (HR: M = 5.1, SD = 3.2; LR: M = 1.3, SD = 2.2; t (61.4) = -9.18, p < .001, d = 1.72) than the LR sample. 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. Total change scores were calculated for each subject based on the number of items whose responses changed from M-CHAT-R to Follow-Up. Item responses were substantially less likely to

change in the HR sample (M = 1.64, SD = 1.64) than in the LR sample (M = 2.90, SD = 1.71), t (1013) = 5.51, $p < .001$, d = .74, even though the average number of items failed was higher in the HR group.

Exploratory item-level analyses were conducted to better understand the response pattern within the HR sample compared to the LR sample by examining the frequency of which each item changed from an "at-risk" response to "not-at-risk" when asked during Follow-Up (see Table 5). To account for multiple comparisons, a Bonferroni correction was utilized (.05/20 comparisons) leading to an adjusted alpha value of .0025.

Parents of LR children were more likely to change their response during Follow-Up than parents of HR children on 12 of the 20 items, with small to medium effect sizes (ϕs = .24 - .38). On average, parents of LR children changed their response 69.0% of the time (range: 48 - 89.1%) compared to 30.2% of the time (range: 7.4 - 100%) for parents of HR children. Differential response change was most substantive on Item 15 (imitation; ϕ = .38), Item 3 (pretend play; ϕ = .35), and Item 1 (follow point; ϕ = .34), which represent relatively subtle social behaviors. The two items that changed equally across risk groups inquired about overt skills such as walking (Item 13) or enjoying climbing (Item 4).

Discriminability of the M-CHAT-R/F

An exploratory discriminant function analysis (DFA) was conducted in an effort to further characterize item performance in the HR sample and to determine the ability of the 20 items on the M-CHAT-R to predict a best estimate ASD diagnosis. For the purposes of the DFA, HR children were classified into two groups – HR children who received an ASD diagnosis (n = 34) and all other HR children (M-CHAT-R(/F) screen negatives and those who received Non-ASD diagnoses; n = 130). Eleven HR cases were excluded from the DFA analysis due to missing item-level data. Standardized canonical discriminant function coefficients for each M-CHAT-R item are presented in Table 6. In the HR sample, 8 items were identified as differentiating well between ASD and Non-ASD. The top three items most predictive of ASD on the M-CHAT-R were Item 10 (*response to name*) followed by Item 2 (*wonder if deaf*) and then Item 5 (*unusual finger movements*) in descending order. The DFA correctly classified 28

out of the 34 ASD cases (82.4%) and only misclassified 3 of the 134 non-ASD cases (i.e., false positives; 2.3%); overall, the complete M-CHAT-R/F discriminated between ASD and non-ASD cases in a HR sample with 94.5% accuracy. Of the six ASD cases that were misclassified as non-ASD, two received diagnoses of PDD-NOS and four received diagnoses of Autistic Disorder. Two cases screened negative on initial screen and did not receive Follow-Up, and the remaining three cases scored relatively low on the M-CHAT-R/F, suggesting parent underreporting. For the non-ASD cases misclassified as ASD, two displayed significant subclinical ASD symptoms and were classified as At-Risk for ASD and the other case had developmental concerns and did not receive any diagnosis.

Differential item failure on the M-CHAT-R between HR-ASD and HR-NonASD groups was analyzed to better understand response patterns. With respect to the most predictive items determined by the DFA, HR-ASD cases were significantly more likely to fail Items 10 (*response to name*) and 2 (*wonder if deaf*) than were the HR-NonASD cases. While Item 5 (*unusual finger movements*) was highly predictive of ASD, it was rarely endorsed by parents of either HR-ASD or HR-NonASD cases. See Table 6 for differential failure rates for all M-CHAT-R items.

DFA was also run in the LR sample to determine whether the M-CHAT-R/F best discriminating items were comparable to that of the HR sample. As done in the HR sample, the LR sample was divided into two groups – LR children who received an ASD diagnosis (n = 108) and all other LR children (n = 14,245). Cases with missing data (n = 1047) were excluded from the LR DFA analysis. See table 7 for the standardized canonical discriminant function coefficients. In the LR sample, 98.5% of cases were correctly classified, which was higher than in the HR sample, likely due to the very high number of Non-ASD cases in the LR sample. The DFA correctly classified 78 of the 108 ASD cases (72.2%) and only misclassified 185 of the 14,245 non-ASD cases (1.2%). Seven items were identified as best discriminating between LR-ASD and LR-NonASD. As in the HR sample, Item 10 (*response to name*) was the best at discriminating between ASD and non-ASD. In the LR sample, the next best discriminating items were Item 7 (*point to show*) and Item 9 (*show to share*). While Items 2 (*wonder if deaf*) and 15 (*imitation*) discriminated well in the HR sample, they discriminated poorly in the LR sample. Conversely,

Items 7 (*point to show*) and 9 (*show to share*) worked well to discriminate between ASD and non-ASD in the LR sample but performed poorly in the HR sample. Item 16 (*follow gaze*) was the only other item that served as a good discriminating item in both samples.

Clinical Characteristics of the Samples

Clinical data from all evaluated children who screened positive (LR: $n = 205$; HR: $n = 56$), broken down by diagnostic group (ASD vs. Non-ASD) and risk status (LR vs. HR) can be found in Table 8. The distribution of ASD and Non-ASD diagnoses was similar regardless of risk status, χ^2 (1, $n = 261$) = 2.51, $p = .113$, $\phi = .10$. Almost half of the screen positive children received an ASD diagnosis within each risk group (LR: 48.8%; HR: 60.7%). The Non-ASD groups were composed of children determined to be At-Risk for ASD (LR: $n = 4$; HR: $n = 6$), to have a global developmental delay (LR: $n = 50$; HR: $n = 4$), developmental language disorder (LR: $n = 23$; HR: $n = 1$), other DSM-IV diagnosis (LR: $n = 1$; HR: $n = 1$), no diagnosis (LR: $n = 15$; HR: $n = 9$), or typical development (LR: $n = 12$; HR: $n = 1$).

A series of two-way between subjects analysis of variances (ANOVAs) were conducted to explore the impact of risk status (LR vs. HR) and diagnostic group (ASD vs. Non-ASD) upon measures of screening (M-CHAT-R Total Score, M-CHAT-R/F Total Score, Change Score), ASD severity (ADOS SA CSS, ADOS RRB CSS, ADOS Overall CSS, CARS Total, CARS SC Factor, CARS ER Factor, CARS SSBS Factor) , cognitive ability (MSEL Verbal Developmental Quotient, MSEL Non-Verbal Developmental Quotient, MSEL Developmental Quotient), and adaptive functioning (VABS-II Communication, Daily Living, Motor, Socialization, Adaptive Behavior Composite). See Table 8 for descriptives and Table 9 for main effects and interactions.

Given the aims of the current study, the main effect of risk status and the interaction between risk status and diagnostic group were of most interest. Additionally, exploratory simple effects analyses were run (adjusted for multiple comparisons using Fisher's LSD) even when an interaction was not present, to gain a better understanding of potential impact of risk status within the ASD group (see Table 10). Scores on several measures violated Levene's test of homogeneity of variances and use of more stringent *p*

values (.01) was recommended to reduce Type I error rates (Weinberg & Abramowitz, 2008). Bonferroni corrections were also considered to correct for multiple comparisons. Ultimately, an adjusted α value of .003 (.05/18) was used as the Bonferroni correction was even more stringent than the recommended value (.01). Effect sizes were measured using partial eta squared (η^2_p), and strength of effect sizes were interpreted using Cohen's (1988) criteria of .01 for small effects, .06 for medium effects, and .14 for large effects.

As anticipated, the main effect for diagnostic group was significant at the adjusted significance levels on all screening and clinical characteristics, except for change score and VABS-II Motor SS, with effect sizes ranging from small to very large based upon Cohen's (1988) criteria (η^2_p = .04 -.50). When risk status was collapsed, there was no main effect of diagnostic group on change score, F (1, 252) = .001, p = .976, η^2_p < .001, indicating that the average number of items that changed when asked at Follow-Up did not differ between toddlers diagnosed with ASD and those diagnosed with other disorders (Non-ASD). Additionally, the ASD and Non-ASD groups showed a similar level of motor skills as measured by the VABS-II, F (1, 244) = .63, p = .429, η^2_p = .003.

With Bonferroni correction, no significant interactions between risk status and diagnostic group were observed on any screening or clinical measure. This indicates that there is no significant difference in the effect of risk status for ASD or Non-ASD toddlers. Thus, all significant main effects were interpreted.

On screening measures, significant main effects for risk status were seen on the M-CHAT-R/F Total Score and the Change Score with small to moderate effect sizes. Collapsed across diagnostic group, HR toddlers (M = 5.8, SD = 2.8) scored significantly higher at Follow-Up than the LR toddlers (M = 4.5, SD = 2.6), F (1, 252) = 9.16, p = .003, η^2_p = .04. Interestingly, the average change score (the average number of items that change between initial screener and Follow-Up) was lower for the HR toddlers (M = 1.4, SD = 1.4) compared to the LR toddlers (M = 2.3, SD = 1.7), F (1, 252) = 13.74, p < .001, η^2_p = .05.

No significant interactions were seen between risk status and diagnostic group on either of these measures.

After correcting for multiple comparisons, a main effect for risk status was not observed on any measures of ASD severity indicating that LR and HR toddlers showed comparable levels of ASD severity. This held true within the ASD group, with the exception of the CARS ER Factor Score. Simple main effects analysis showed that within the ASD group, the HR-ASD toddlers were more emotionally reactive than the LR-ASD toddlers, F (1, 232) = 12.47, $p < .001$, $\eta^2 = .05$.

Significant main effects of risk status were present on verbal, nonverbal, and overall cognitive ability as measured by MSEL developmental quotients with moderate to large effect sizes. Collapsed across diagnostic groups, HR toddlers (M = 68.6, SD = 26.4) had stronger verbal abilities compared to LR toddlers (M = 56.3, SD = 23.7) with a medium effect size, F (1, 246) = 19.91, $p < .001$, $\eta^2_p = .08$. HR toddlers scored in the mild intellectual disability range, while the scores of the LR toddlers were consistent with profound intellectual disability. As would be expected, nonverbal scores were higher in both risk groups. However, there was still a substantial main effect of risk status on nonverbal ability where HR toddlers scored in the low average range (M = 88.0, SD = 18.8) while the LR toddlers' scores (M = 73.7, SD = 20.0) fell in the borderline range of functioning, F (1, 247) = 27.53, $p < .001$, $\eta^2_p = .10$. In terms of overall cognitive functioning, the HR toddlers scored in the borderline range (M = 78.3, SD = 20.4) while the scores of LR toddlers (M = 65.1, SD = 20.2) were suggestive of comorbid moderate intellectual disability, F (1, 246) = 28.08, $p < .001$, $\eta^2_p = .10$. Simple main effects analyses indicate that these findings hold true within the ASD group as well. HR-ASD toddlers had significantly higher verbal ($\eta^2 = .05$), non-verbal ($\eta^2 = .10$), and overall cognitive ability ($\eta^2 = .08$) compared to LR-ASD toddlers, with medium effect sizes. Of note, the standard deviations of both risk groups were quite large on all measures of cognitive ability suggesting that while the means fell within the descriptive ranges mentioned above, there was significant variability of cognitive functioning within both groups.

Last, with respect to adaptive functioning, no significant main effects of risk status were observed on any VABS-II domain when more stringent *p* values were used. Average scores on Communication, Socialization, Daily Living, Motor, and overall adaptive functioning (Adaptive Behavior Composite) were comparable across risk status, with diagnostic group collapsed and within only the ASD group.

In sum, the ASD group performed worse than the Non-ASD on almost all measures, as would be expected. The HR group was higher functioning than the LR group on all measures of cognitive functioning in the face of comparable adaptive functioning and ASD severity across groups. Exploratory analyses suggested that these patterns held true within the ASD group as well, such that HR-ASD toddlers demonstrated higher cognitive functioning than the LR-ASD group, while they were comparable with respect to ASD severity.

Discussion

The current study sought to build upon existing literature on ASD screening in HR infant siblings by examining the performance on the M-CHAT-R/F in a relatively large sample of HR toddlers. Unlike past studies, the current study directly compared screening results in HR younger siblings to screening in a large-scale, low-risk sample drawn from the M-CHAT-R/F validation study (Robins et al., 2014). Response patterns and clinical characteristics were also compared across risk type.

Results indicate that the M-CHAT-R/F can successfully discriminate between ASD and Non-ASD in a HR population between 18 to 24 months with good predictive power. Similar to M-CHAT/F findings in other HR populations (Kleinman, Robins, et al., 2008), the Follow-Up interview was not necessary to reduce false positives in the current HR sample. Low change rates during Follow-Up also suggest parents with ASD experience are more accurate in their reporting of concerns on later-born children. Exploratory item-level analyses suggest differential discrimination patterns exist in LR and HR samples, likely due to the increased prevalence of the subclinical ASD symptoms in HR siblings or parents' familiarity with these symptoms. Within the ASD group, HR-ASD toddlers showed higher cognitive functioning compared to LR-ASD toddlers. Other adaptive skills were comparable. Similar

ASD symptom severity was observed across groups further supporting the idea that these findings represent differential reporting based upon parent experience with ASD as opposed to behavioral differences among the samples.

Performance of the M-CHAT-R/F

Increased internal consistency was seen in the HR sample compared to the LR sample (.88 vs .64), suggesting that ASD at-risk behaviors may hold together better as a single construct in HR toddlers. This is not surprising given evidence of the broader autism phenotype and increased prevalence of subclinical ASD symptoms (Constantino et al., 2006; Pickles et al., 2000; Schwichtenberg et al., 2010), which is not seen in low-risk populations. Possible differences in how HR parents answer the questions, given their experience with ASD, may also contribute to these findings.

The screen positive rate was higher in HR toddlers compared to the LR sample. The screen positive rate at the initial screen on the M-CHAT-R for HR toddlers in the current study was 40.6%, which is consistent with a previous screening study conducted with a subset of 18-month-old HR toddlers from the current sample (40%; Weitlauf et al., 2015). It is also comparable to screen positive rates (45%) in HR sibling samples using other screening measures, such as the STAT (Stone et al., 2008). Elevated screen positive rates in HR toddlers in this sample and others are similar to the elevated M-CHAT screen positive rates seen in other HR groups, such as extremely premature infants, which range from 21-41% (Kuban et al., 2009; Limperopoulous et al., 2008; Luyster et al., 2011; Moore et al., 2012) and in children referred from EI or other providers (40%; Kleinman, Robins et al., 2008). This suggests that high M-CHAT-R/F screen positive rates are quite consistent, regardless of the reason for their risk.

In order to determine predictive validity of M-CHAT-R screen positives, Follow-Up interviews and diagnostic evaluations were conducted. The current study found an ASD detection rate of 19.4% in the HR toddlers, which is highly consistent with recent recurrence rates estimates of 18.7% (Ozonoff et al., 2011) and 19.5% (Messinger et al., 2015) for HR infant siblings. This consistency suggests that most of the ASD cases were likely detected through parent completion of the M-CHAT-R/F in the HR sample and few cases were missed. As expected, a 19.4% detection rate in the HR sample was significantly

greater than the .65% detection rate in the LR comparison sample, whose rate was about 40% of the current ASD prevalence rates in 8-year-olds (1 in 68 or 1.47%) reported by the CDC (2016) and about 50% of the current ASD prevalence rates in 4-year-olds (1.34%; Christensen et al., 2016). It is important to note that these prevalence rates are based upon educational and medical records at ages 8 and 4 respectively, and we would not expect to identify all of these cases around the age of two.

As hypothesized, the PPV for the M-CHAT-R in the HR sample was significantly higher than that of the LR comparison sample (.531 vs. .138), with small to medium effect sizes. After Follow-Up, the PPV raised to .607 in the HR sample, however, this was not significantly higher than the PPV at initial screen (.531). It is important to note that the PPVs found in the current sample are somewhat lower than the PPVs for HR samples reported elsewhere. Lower PPVs compared to Kleinman, Robins et al. (2008; PPV = .76) and Pandey et al. (2008; PPVs = .74 - .79) using the M-CHAT/F in an EI-referred population are expected and may speak to the increased difficulty in distinguishing between ASD and the broader autism phenotype in the HR sibling population. Weitlauf and colleagues (2015) also found notably higher PPVs (.783 after Follow-Up), in a subset of the current study sample of HR toddlers, although this may be due to the inclusion of children deemed "at-risk" for ASD in their PPV calculations. In the current study, children deemed "at-risk" for ASD were not included in the true positive count, as one of the main objectives of the study was to determine how parent report on the M-CHAT-R/F could differentiate between HR younger siblings who were later diagnosed with ASD and those with subclinical ASD symptoms or other developmental concerns.

Relatively high PPVs were also found in HR children screened with the FYI at 12 months (.71; Rowberry et al., 2015). The PPV in the current study after Follow-Up (.607) was comparable to the PPV seen at the optimal cutoff of the Autism Parent Screen for Infants (APSI; Sacrey et al., 2016) at 6 months (.47), but lower than PPVs when screening in the age range comparable to the current study (.68 -.79; Sacrey et al., 2016). Higher PPV rates in the Sacrey et al. (2016) may be due to more nuanced items on the APSI compared to the M-CHAT-R/F. For instance, for RRBs the M-CHAT-R/F only asks about unusual finger movements, while the APSI asks about hand use/holding objects, unusual sensory

behaviors, insistence on a particular object, resist play/fixed play routines, all of which distinguish between HR-ASD and HR-NonASD in the Sacrey et al. study.

In contrast, the PPV in the current sample (.607) was substantially higher than the PPV seen on the POEMS overall (.21), at the optimal screening age of 9 months (.29) or at ages similar to the current study (.20 - .29; Feldman et al., 2012). The POEMS' PPVs seem low across the board, which may be due to their reliance upon evaluations conducted in the community. As a delay in diagnosis is common among community samples (CDC, 2016), it is likely that ASD cases were missed. Additionally, community evaluations may be less rigorous, less reliant on standardized measures, and more variable in their interpretation of ASD symptoms. These factors likely contributed to the lowered the PPV seen in the Feldman et al. (2012) study. Current PPVs were also expectedly higher than that of ASD screening in extremely premature infants despite similar screen positive rates, as screening in extremely premature infants led to high false positive rates (Johnson et al., 2010).

Furthermore, PPV for the samples did not differ after Follow-Up in the current study. Consistent with findings from Kleinman, Robins et al. (2008) with an EI-referred HR sample, the Follow-Up interview was not as necessary in the current sample of HR toddlers to reduce false positives, as it did not significantly increase the PPV of the screening instrument (.531 to .607). This suggests that parents of older children with ASD may be more accurate in their reporting of ASD at-risk behaviors for later-born children. Parents' endorsement of at-risk behaviors on the screener is reflective of both the detection of these behaviors as well as the willingness to report these behaviors. Parents' of HR children may be more willing to report concerns 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.

Item-Level Performance of the M-CHAT-R/F

The accuracy of parents of HR toddlers was further supported by exploratory item-level analyses examining the percentage of time each at-risk item changed when asked during Follow-Up. Item responses changed on average 69% of the time in the LR sample and only 30% of the time in the HR

sample. Response change rates between the HR and LR samples differed on most items on the M-CHAT-R/F (12 of 20), even following Bonferroni corrections, with small to medium effect sizes. Items that more frequently changed in the LR population with the largest effect sizes (ϕs = .34 - .38) tended to be subtler social behaviors, with which parents of LR children may not be as familiar. This is consistent with previous research suggesting that experience with older children, especially older children with ASD, may sensitize parents to more subtle delays (Herlihy et al., 2013; McMahon et al., 2007; Ozonoff et al., 2009; Sacrey et al., 2015).

Despite a high frequency of changed responses, the average number of items that change at Follow-Up in the LR and HR samples is low (1.4 items in the HR group and 2.3 items on average in the LR group). However, when considering the cutoff on the M-CHAT-R/F is 2 items, the small average change score becomes more clinically meaningful. When diagnostic groups are collapsed, the LR groups had significantly higher change scores than the HR group indicating that the HR was less in need of the scaffolding that the Follow-Up provides. Of note, there was no main effect of diagnostic status, such that parents of ASD toddlers and Non-ASD toddlers (collapsed across risk status), were equally likely to change their responses at Follow-Up. No interaction between risk status and diagnostic group was seen on average change scores. In sum, these results suggest that parents of an older affected child were more accurate in their initial reporting, regardless of whether or not the younger sibling had ASD.

Discriminability of the M-CHAT-R/F

Discrimination ability of individual items was also explored and compared across samples. In both samples, response to name was the most discriminating item between ASD and Non-ASD groups. This item has previously been shown to be a good discriminator between ASD and Non-ASD in a LR population on the M-CHAT(-R)/F (Robins, Fein, Barton, & Green, 2001). In a HR sample, *response to name* has been shown to differentiate HR-ASD toddlers from HR-NonASD toddlers on the POEMS at 9 months (Feldman et al., 2012) and on the APSI and AOSI at 12 months (Sacrey et al., 2016; Zwaigenbaum et al., 2005), while on the FYI, *response to name* differentiated HR-ASD groups from LR-TD groups, but not from HR-NonASD groups at 12 months (Rowberry et al., 2015). *Response to name*

alone at 12 months has been shown to have a specificity of .89 and a sensitivity of .50 for ASD diagnosis at 24 months in a combined sample of HR and LR children, although it appears to differentiate better in a HR sample (Nadig et al., 2007).

Imitation is another discriminating behavior that consistently distinguishes HR-ASD from HR-NonASD (Feldman et al., 2012; Rowberry et al., 2015; Sacrey et al., 2016; Zwaigenbaum et al., 2005), as seen in the current study. The FYI Imitation construct, which includes items about motor, vocal, and social imitation, was best at discriminating HR-ASD from other HR-NonASD and LR-TD groups (Rowberry et al., 2015).

Finally, *eye contact* was shown to discriminate well between HR-ASD from HR-NonASD in the current sample but not in the LR sample, which is supported by the literature. *Eye contact* has been shown to differentiate HR-ASD from HR-NonASD on the AOSI and APSI beginning at 12 months (Sacrey et al., 2016; Zwaigenbaum et al., 2005). Moreover, Brian et al. (2008) found *eye contact* to be one of the top three discriminating items on the AOSI based upon a DFA using only the items that differed between the HR groups.

There were several items that discriminated well between ASD and NonASD in the current HR sample that have not been previously identified in the literature. One of which was *wondering if deaf*, the second best discriminating item in the HR sample. This item tries to ascertain if a child responds to surrounding sounds and people, and thus can be considered a social orienting behavior. *Walking* also differentiated HR-ASD from HR-NonASD in the current sample. Delayed fine and gross motor skills on standardized assessment have been shown to differentiate between HR-ASD and HR-NonASD beginning at 14 months (Landa & Garrett Mayer, 2006) and parent-reported motor concerns differentiate HR-ASD, HR-NonASD, and LR-TDC as early as 6 months (Sacrey et al., 2015). Last, there were some behaviors that have been shown to differentiate in some previous samples but did not do so in the current sample, such as social interest (Feldman et al., 2012; Sacrey et al., 2016; Zwaigenbaum et al., 2005). This may be due to nuanced differences in items, such as *interest in faces* (Feldman et al., 2015), *sustained interest*

and pleasure in interaction (Sacrey et al., 2016; Zwaigenbaum et al., 2005), and *interest in other children* on the M-CHAT-R/F (Robins et al., 2014).

In a LR population, Veness et al. (2011) found items regarding *showing*, *pointing*, *sharing interest*, *gaze/point following*, and *requesting attention* all differentiated ASD from TD through parent report on the CSBS ITC at 24 months. Similarly, items regarding *showing*, *pointing*, and *point following*, among others, were also shown to discriminate well between ASD and Non-ASD on the M-CHAT/F in a LR population (Robins et al., 2001). The other items that show good discriminability on the M-CHAT-R/F based on a larger LR sample from Chlebowski, Robins, Barton, & Fein (2013) include: *response to name*, *social interest*, *social responsivity*, and *pretend play.* Four of the previously determined M-CHAT-R/F items continued to show good discriminability in the LR sample in the current study. However, only two of the items, those regarding social orienting (Item 10 – *response to name* and Item 2 – *wonder if deaf*), show good discriminability in the HR sample. *Following gaze* was another item that showed good discriminability in both HR and LR samples. The limited overlap between the best discriminating items on the M-CHAT-R/F in the LR sample compared to previous studies is not surprising given the difficulties in reliably replicating the "best" subset of items in new screening samples. In LR samples, these "best" discriminating items have been identified as an alternative score cutoff for the M-CHAT(-R)/F ($\geq$ 2 of Critical6 in the original screener or of Best7 in the revised version). More recent research by Robins et al. (2014) found that Best7 demonstrated less sensitivity and did not detect any additional cases compared to using a total score cutoff of 3, leading them to conclude that overall total score is the most appropriate score cutoff. Hence, the DFA conducted in the current sample is interpreted in an exploratory manner and not in an effort to identify a best subset of items or alternative scoring cutoff for use with HR children.

Taken together, these data suggest that key at-risk behaviors for ASD in the LR population may not be the same behaviors that distinguish between ASD and Non-ASD in a HR population, either because they are part of the broader autism phenotype (expected to be higher in the sibling group than the general population) or because parents experienced with ASD are more accurate in reporting them. Given

the relatively small sample size of HR toddlers in the current study, another DFA should be run with a larger sample to replicate the findings from the current study before more definitive conclusions can be drawn about differential patterns of discrimination in HR and LR samples.

Clinical Characterization of Samples

It is critical to interpret the results of the current study within the context of clinical characteristics of the samples. As anticipated, the main effect of diagnostic group (ASD vs. Non-ASD) was observed on almost all clinical measures. Because of how the children were assigned to diagnostic group, this would be inevitable on measures of ASD symptomatology or severity. The ASD group in the current study also showed weaker verbal and nonverbal cognitive abilities than the non-ASD group, with medium to large effect sizes. This finding is consistent with literature that suggests by 24 months, children with ASD show deficits in cognitive abilities compared to typically developing children and children with delayed language (Landa & Garrett-Mayer, 2006). The current study replicated previous findings of uneven cognitive abilities in some young children with ASD, with stronger non-verbal abilities and weaker verbal abilities (Joseph & Tager-Flusberg, 2002; Landa & Garrett Mayer, 2006; Munson et al., 2008). The ASD group also showed weaker adaptive skills on all VABS-II domains (communication, daily living, motor, and socialization) compared to the non-ASD group, which is also consistent with previous literature. Toddlers with ASD have been shown to exhibit a unique profile of adaptive skills that are discrepant from children with other developmental disorders, such as language impairment and developmental delay (Stone, Ousley, Hepburn, Hogan, & Brown, 2009). Socialization and communication skills have shown to be substantially lower for children diagnosed with ASD than those with other disorders, even when controlled for language abilities (Stone et al., 2009).

Due to the limited research comparing HR and LR samples that include a LR-ASD group, exploratory analyses were conducted and no specific hypotheses were made. In the current study, main effects of risk status were less common than main effects of diagnostic group. Clinical characteristics analyses were conducted with a subset of toddlers who screened positive and were evaluated, to reduce the impact of screen negative cases who were evaluated (to detect possible false negatives). To answer

our main question of interest – Do HR-ASD and LR-ASD cases differ? – we focused on simple main effects within the ASD group.

The HR-ASD group demonstrated higher verbal, non-verbal, and overall cognitive abilities than the LR-ASD group, with small to medium effect sizes. These findings are in contradiction to past research indicating that HR younger siblings (ASD and Non-ASD) have significantly lower verbal and non-verbal abilities (Rowberry et al., 2015) and lower overall cognitive ability (Sacrey et al., 2016) compared to LR-TD at 24 months; however, little is known how HR-ASD samples compare to LR children with ASD or other developmental delays. Amongst typically developing children of varying risk status, Rowberry et al. (2015) found no significant differences in verbal and non-verbal abilities between the HR-TD and LR-TD groups. Results of the current study are consistent with past studies linking dysmorphology and/or microcephaly, which is common in LR-ASD samples, with lower IQ (Miles et al., 2005). HR-ASD and LR-ASD groups had comparable adaptive skills as measured on the VABS-II.

Differences in ascertainment between the samples also cannot be ruled out as an explanation for the observed discrepancy in cognitive scores. However, these findings provide growing evidence that higher functioning children are detectable between 18-24 months through parent endorsement of at-risk behaviors on the M-CHAT-R. While screening in a HR sibling population is imperative due to increased prevalence of other developmental disorders, it is especially critical if it leads to the detection of higher functioning children with ASD. Past research highlights the importance of stronger cognitive skills and better motor skills at age 2 in achieving an optimal outcome (Sutera et al., 2007; Turner & Stone, 2007). With these skills intact, these higher-functioning children may be more responsive to intervention.

Typically, children with ASD and comorbid developmental delay are more likely to be detected earlier (Szatmari et al., 2016; Zuckerman et al., 2015). Increased evidence of developmental delay in the LR sample could suggest that children are failing items that are not representative of ASD-specific at-risk behaviors but represent more general delay. Wetherby et al. (2004) found that several previously identified red flags, including reduced responsivity to contextual clues, reduced pointing, reduced

vocalizations with consonants, and reduced playing with toys appropriately, do not adequately distinguish between ASD and DD in LR populations.

Finally, HR toddlers showed comparable levels of ASD severity to LR toddlers on most measures. Given the presence of increased subclinical ASD symptoms in the HR population (Constantino et al., 2006; Pickles et al., 2000; Schwichtenberg et al., 2010), one might expect significantly higher ASD severity in the HR toddlers. Within the ASD group, the HR-ASD toddlers showed increased emotional reactivity compared to LR-ASD toddlers, which is consistent with increased irritability and responsivity (typically to sensory stimuli) on parent-rated temperament measures in HR-ASD toddlers (Zwaigenbaum et al., 2005).

Taken together, the findings from our exploratory analyses comparing HR-ASD and LR-ASD groups revealed comparable ASD symptom severity and adaptive functioning with notable differences in cognitive functioning. The HR-ASD group showed higher verbal and non-verbal cognitive abilities compared to those of the LR-ASD group. Further research is needed to replicate and better understand these differences. These results provide preliminary evidence that caution may be warranted when generalizing findings from HR sibling studies to the development of ASD in LR children.

Limitations

There are several limitations that must be considered when interpreting the results of the current study. It examines the results of the M-CHAT-R/F in HR toddlers at only one screening time point, which may not be ideal for this population given the potential differences in onset of symptoms (Bryson et al., 2007; Macari et al., 2012). Moreover, the average screening age for the current study was around 20 months. This is substantially later than the age at which some of the other screening studies have demonstrated adequate differentiation between HR-ASD and HR-NonASD (Feldman et al., 2012; Rowberry et al., 2015; Sacrey et al., 2016). However, only about 40% of HR-ASD children are symptomatic by 18 months, highlighting the importance of screening at later time points, as in the current study, so as not to miss children with more subtle symptoms (Szatmari et al., 2016). While the current

study used a large LR comparison sample, the LR sample was ascertained differently than the HR sample and thus, family information on much of the LR sample could not be collected.

One major limitation is the minimal follow-up on screen negative cases, which ultimately precluded our ability to conduct reliable sensitivity and specificity calculations and run a ROC analysis. In a subset of the current sample, Weitlauf et al. (2015) systematically followed up on screen negative cases and found that approximately one fifth of children who screened negative on the M-CHAT-R/F at 18 months ultimately went on to receive an ASD/R (ASD or at-risk for ASD) diagnosis by 36 months. However, it is unclear how many of these cases represented true ASD diagnoses since at-risk cases were also included. In the current sample, 4.5% of the screen negative cases were detected using alternate methods to catch false negatives; if extrapolated to the full sample of HR toddlers screened, this suggests that 5 children were potentially missed. While 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% and comparable to prevalence rates in published literature, suggesting that most ASD cases were likely detected. Looking forward, PPV estimates may be more informative in assessing clinical use of screening tools, such as the M-CHAT-R/F, in HR sibling populations, than measures of sensitivity and specificity (Szatmari et al., 2016). Sensitivity and specificity are more reliant upon inherent characteristics of the measure and thus remain more consistent across risk status. Thus, PPV, which is easily influenced by prevalence rates of the disorder, provides more information about the effectiveness of screening measures across populations. Hence, there is a significant need to examine screening performance of ASD screening measures in a HR sibling population, as in the current study.

Racial and ethnic differences, as well as sex differences, were found between the LR sample and the HR sample, which were likely due to differences in ascertainment between the samples. Specifically, the LR sample was collected through implementation of broad-based screening at pediatric well-child visits and explicit efforts were made to include pediatric sites that served a more racially and

economically diverse population. In contrast, the HR sample was primarily a self-referred population who was informed about the study through research or clinical contact with an ASD-affected older sibling. Despite the statistical significance of the difference, the effect size ($\phi = .06$) was quite small according to Cohen's (1988) criteria (.10 = small effect). Furthermore, rates of ASD and detection rate of ASD using the M-CHAT-R/F have been shown to be similar across racial groups (Khowaja, Hazzard, & Robins, 2015), which suggests that differing racial composition across the samples is likely not driving the results seen in the current study. The sex distribution in the HR sample was less even (i.e., more males), which is expected given the increased prevalence rates for ASD in boys (CDC, 2016; Christensen et al., 2016). Moreover, the effect size was very small ($\phi = .02$).

Additionally, despite efforts to recruit and retain a diverse sample, a significant portion of the HR sample was lost to follow-up either during the enrollment process or afterward. Differential attrition during enrollment is unknown. However, HR children who completed the study were more likely to identify as White compared to those who did not, potentially contributing to the racial differences seen between the LR and HR samples. Refusal rates within the HR sample were high. While this may be initially surprising, it is consistent with previous research suggesting that parents of younger siblings may be less likely to enroll in research studies than LR community samples (Rogers et al., 2014).

Finally, there were differences in delay between screen and Follow-Up as well as between screen and evaluation, such that the HR sample was more likely to complete Follow-Up and be evaluated more quickly compared to the LR sample. The delay may suggest differential parent concern, specifically that parents of HR toddlers were more aware of the impact of an ASD diagnosis upon the family and thus, more likely to respond to phone calls from the researchers leading to reduced lag times between initial screen and Follow-Up. Alternatively, logistical reasons due to differing study procedures across samples, such as postal delays in receiving screeners from pediatrician offices in the LR sample, may have contributed to the results, independent of parent concern. It is likely these factors combined to result in these delays; the distinct contribution of parent concerns is unknown.

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. Thus, the LR group was significantly older at Follow-Up (~24 months) than the HR sample (~22 months). Results from a study examining M-CHAT/F performance in younger and older toddlers suggest that the M-CHAT/F had better predictive power in older toddlers, which they suggested could be due to more stable delays at that age (~ 24 months; Pandey et al., 2008). However, no differences in PPV were seen based upon age in the revised version of the screener (Robins et al., 2014), M-CHAT-R/F, which was used in the current study.

Despite these limitations, the current study has a number of strengths that should be highlighted. First and foremost, 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 screening tools in a HR sample compared to screening in a very large LR comparison sample. It also provides exploratory item-level analyses to better understand response patterns in HR toddlers compared those in a LR sample. Additionally, unlike many screening studies in HR populations more broadly, the current study examined diagnostic accuracy and predictive power. Lastly, the sample is well-characterized using standardized measures of ASD severity, cognitive ability, and adaptive functioning and the findings are interpreted within this context.

Clinical Implications

The current study provides strong evidence that parent-completed screening measures can effectively detect ASD in HR siblings around the age of two. Despite potentially subtle and inconsistent behavioral markers of ASD in HR younger siblings during the first year of life (for review, see Szatmari et al., 2016), parent detected, at-risk behaviors can successfully discriminate 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 and time-intensive measures, such as EEG or eye tracking methods. Importantly for physicians, it demonstrates that Follow-Up is not necessary to reduce false

positives in HR siblings, reducing the amount of time needed for screening in this population. 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 (Messinger et al., 2013; Toth et al., 2007). Only one child (1.8%) in the current HR sample that screened positive on the M-CHAT-R/F was determined to be typically developing upon further evaluation. 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 research should be conducted to address some of the aforementioned limitations of the current study. Ascertainment differences should be minimized in future studies by utilizing a prospective cohort approach. Should resources allow, future studies should more systematically follow up screen negative cases to identify HR younger siblings who may be missed during screening and to determine estimates of NPV. 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 the broader autism phenotype) upon the screening process. The impact of other parent factors (i.e., maternal depression) and family factors (i.e., 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 with good predictive power. 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, even when they are less cognitively delayed. Comparable ASD symptom severity was observed across groups further supporting the idea that these findings represent differential reporting based upon parent experience with

ASD as opposed to behavioral differences among the samples. Future research is needed to better understand the other factors contributing to differential ASD screening performance in HR siblings.

Tables

Table 1
Previous Screening Studies Conducted in High-Risk Younger Siblings

Study	Measure	*N*	Screening Time Points in Months	ASD Outcomes
Clinician-Rated Observation				
Stone et al., 2008	STAT	59	Once between 12-23 (*M* = 16.4)	For children 14 months and older, a higher cutoff of 2.75 led to a sensitivity of 0.93, specificity of 0.83, PPV of 0.68, and NPV of 0.97 for ASD diagnosis at 24 months.
Zwaigenbaum et al., 2015	AOSI	65	6, **12**	At 12 months, a 7 point cutoff led to a sensitivity of 0.84, specificity of 0.98, PPV of .78 and NPV of .91 for ASD diagnosis at 24 months.
Parent-Report Questionnaire				
Feldman et al., 2012	POEMS	108	3, 6, **9**, **12**, **18**, **24**	Overall, a total score cutoff of 70 led to a sensitivity of .74, specificity of .73, and PPV of .21. The best PPV (.29) was seen at 9 months, with a sensitivity of .57 and specificity of .84 for ASD diagnosis at 36 months.
Rowberry et al., 2015	FYI	71	**12**	Using a CART analysis, two items were identified (*Do you get the feeling that your baby plays or communicates with you less now than in the past?* and *Does your baby copy or imitate you when you make sounds or noises with your mouth?)* with a sensitivity of .63, specificity of .93, PPV of .71, and NPV of .89 for diagnosis at 24 months.
Sacrey et al., 2016	APSI	204	**6**, **9**, **12**, **15**, **18**, **24**	Optimal cutoff scores differed by screening time point with the highest cutoff scores at the youngest ages (15 at 6 months, 10 at 12 months, and 9 at 18 months). ROC analyses indicated best performance of the APSI at 6 months with a sensitivity of .67, specificity of 0.86, PPV of 0.47, and NPV of 0.83 for ASD diagnosis at 36 months.
Weitlauf et al., 2015	M-CHAT-R/F	74	**18**	At 18 months, a cutoff of 2 on the M-CHAT-R/F led to a sensitivity of 0.78, specificity of 0.72, PPV of .78 and a NPV of .72 for an ASD diagnosis at 18 months. A sensitivity of .78, specificity of .81, PPV of .75 and NPV of .83 was seen for an ASD diagnosis at 36 months.

Note. Sample sizes are provided for HR children only. Bold text refers to ages at which HR-ASD can be differentiated from HR-NonASD in each study. STAT = Screening Test for Autism in Two-Year-Olds. PPV = Positive predictive value. NPV = Negative predictive value. AOSI = Autism Observation Scale for Infants. POEMS = Parent Observation of Early Markers Scale. FYI = First Year Inventory. CART = Classification and Regression Tree. APSI = Autism Parent Screen for Infants. ROC = Receiver Operator Curve. M-CHAT-R/F = Modified Checklist for Autism in Toddlers, Revised with Follow-Up.

Table 2
Demographic Characteristics of High-Risk Sample by Study Site

	UConn (n = 23)	UW (n = 49)	Vanderbilt (n = 103)	*Statistic*
Age at Screen in Months (*M*, *SD)*	21.6 (4.2)	21.1 (3.5)	21.1 (4.4)	F (2, 172) = .14, p = .868, η^2 = .001
Age at Evaluation in Months (*M*, *SD*)	23.9 (3.2)	23.7 (3.7)	22.5 (4.1)	F (2, 96) = 1.02, p = .366, η^2 = .02
Sex (*N*, %)				
Male	11 (47.8)	26 (55.3)	64 (62.1)	χ^2(2, n = 173) = 1.83, p = .400, V = .10
Female	12 (52.2)	21 (44.7)	39 (37.9)	
Race (*N*, %)				
White	19 (86.4)	35 (79.5)	89 (88.1)	χ^2 (2, n = 167) = 1.84, p = .398, V = .11
Of Color	3 (13.6)	9 (20.5)	12 (11.9)	
Maternal Education (*N*, %)				
No College Degree	6 (30.0)	20 (43.5)	42 (46.2)	χ^2 (2, n = 157) = 1.74, p = .418, V = .11
College Degree or Higher	14 (70.0)	26 (56.5)	49 (53.8)	

Note. UConn = University of Connecticut. UW = University of Washington. Vanderbilt = Vanderbilt University. Due to small sample size of each racial group, children of color were all combined into a single group.

Table 3
Demographic Characteristics by Risk Status

Demographic Characteristic	High-Risk (*n* = 175)	Low-Risk (*n* = 15,400)	*Statistic*
Age at Screen in Months (*M, SD)*	21.15 (4.08)	20.96 (3.30)	t (176.6) = -0.61, p = .542, d = .06
Age at Evaluation in Months (*M, SD)*	22.81 (3.98)	26.05 (5.48)	t (246.4) = 6.14, p < .001, d = .68
Sex (*N*, %)			χ^2 (1, n = 15,321) = 4.04, p = .044, ϕ = .02
Male	101 (58.4)	7680 (50.7)	
Female	72 (41.6)	7468 (49.3)	
Race (*N*, %)			χ^2 (1, n = 14,516) = 45.59, p < .001, ϕ = .06
White	143 (85.6)	8596 (59.9)	
Of Color	24 (14.4)	5753 (40.1)	
Maternal Education (*N*, %)			χ^2 (1, n = 13,888) = 1.68, p = .194, ϕ = .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.

Table 4
Psychometric Properties of the M-CHAT-R/F by Screen Stage and Risk Status

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

Table 5
Changed Responses on M-CHAT-R/F by Item and Risk Status

M-CHAT-R/F Item		High-Risk		Low-Risk		*Statistic*[a]
		n	*N* (%)	*n*	*N* (%)	
1	Follow point	27	2 (7.4)	102	49 (48.0)	χ^2 (1, *n* = 129) = 14.74, *p* < .001, ϕ = .34
2	Wonder if deaf	11	2 (18.2)	111	76 (68.5)	FET, *p* =.002, ϕ = .30
3	Pretend play	21	7 (33.3)	364	320 (88.2)	FET, *p* < .001, ϕ = .35
4	Enjoy climbing	3	2 (66.7)	41	35 (85.4)	FET, *p* = .413, ϕ = .13
5	Unusual finger movements	4	1 (25.0)	257	229 (89.1)	FET, *p* = .006, ϕ = .24
6	Point to ask	36	6 (16.7)	287	158 (55.1)	χ^2 (1, *n* = 323) = 18.86, *p* < .001, ϕ = .24
7	Point to show	45	8 (17.8)	376	213 (56.6)	χ^2 (1, *n* = 421) = 24.35, *p* < .001, ϕ = .24
8	Interest in other children	12	2 (16.7)	108	60 (55.6)	χ^2 (1, *n* = 120) = 6.54, *p* = .011, ϕ =.23
9	Show to share	26	6 (23.1)	195	121 (62.1)	χ^2 (1, *n* = 221) = 14.26, *p* < .001, ϕ = .25
10	Response to name	13	2 (15.4)	74	46 (62.2)	χ^2 (1, *n* = 87) = 9.78, *p* =.002, ϕ = .34
11	Social smile	4	1 (25.0)	53	36 (67.9)	FET, *p* = .119, ϕ = .23
12	Upset by noises	11	1 (9.1)	289	217 (75.1)	FET, *p* < .001, ϕ = .28
13	Walk	1	1 (100.0)	54	27 (50.0)	FET, *p* > .999, ϕ = .13
14	Eye contact	14	7 (50.0)	124	99 (79.8)	FET, *p* = .020, ϕ = .21
15	Imitation	14	5 (35.7)	95	79 (83.2)	FET, *p* < .001, ϕ = .38
16	Follow gaze	43	7 (16.3)	375	209 (55.7)	χ^2 (1, *n* = 418) = 24.05, *p* < .001, ϕ = .24
17	Seek attention	39	16 (41.0)	468	411 (87.8)	χ^2 (1, *n* = 507) = 59.32, *p* < .001, ϕ = .34
18	Understand	29	8 (27.6)	218	124 (56.9)	χ^2 (1, *n* = 247) = 8.23, *p* = .003, ϕ = .19
19	Social reference	37	11 (29.7)	307	236 (76.9)	χ^2 (1, *n* = 344) = 36.25, *p* < .001, ϕ = .33
20	Enjoy movement activities	7	2 (28.6)	38	29 (76.3)	FET, *p* = .023, ϕ = .37

Note. FET = Fisher's Exact Test. FET used when expected cell counts were less than 5. Sample size for each item will vary, given that items are only administered when a child screened positive and that specific item showed risk.

[a]Bonferroni-adjusted α = .0025 used to determine significance.

Table 6

Standardized Canonical Discriminant Function Coefficients and Item Failure Rates by Diagnostic Status in High-Risk Sample

M-CHAT-R Item		High-Risk		
		Coefficient	Non-ASD *N* (%)	ASD *N* (%)
10	Response to name	0.484	2 (1.4)	16 (45.7)
2	Wonder if deaf	0.393	4 (2.9)	10 (27.8)
5	Unusual finger movements	0.319	5 (3.6)	4 (11.1)
15	Imitation	0.317	5 (3.6)	11 (30.6)
14	Eye contact	0.295	3 (2.2)	16 (44.4)
16	Follow gaze	0.284	25 (18.1)	29 (80.6)
17	Seek attention	0.262	24 (17.4)	28 (77.8)
13	Walk	0.261	0 (0.0)	1 (2.8)
4	Enjoy climbing	-0.226	2 (1.4)	3 (8.3)
6	Point to ask	0.221	23 (16.5)	25 (69.4)
20	Enjoy movement activities	-0.212	5 (3.6)	2 (5.7)
11	Social smile	-0.179	2 (1.4)	5 (13.9)
19	Social reference	-0.168	21 (15.1)	23 (63.9)
1	Follow point	0.124	10 (7.2)	23 (63.9)
8	Interest in other children	0.099	4 (2.9)	12 (33.3)
9	Show to share	-0.081	12 (8.6)	19 (52.8)
3	Pretend play	0.065	11 (8)	16 (44.4)
18	Understand	0.035	13 (9.4)	24 (66.7)
7	Point to show	-0.007	26 (18.7)	29 (80.6)
12	Upset by noises	0	19 (13.7)	4 (11.1)

Table 7

Standardized Canonical Discriminant Function Coefficients and Item Failure Rates by Diagnostic Group in Low-Risk Sample

		Low-Risk		
M-CHAT-R Item		Coefficient	Non-ASD *N* (%)	ASD *N* (%)
10	Response to name	0.291	59 (0.4)	29 (25.4)
7	Point to show	0.264	406 (2.7)	87 (74.4)
9	Show to share	0.241	164 (1.1)	56 (47.9)
6	Point to ask	0.235	303 (2.0)	72 (61.5)
1	Follow point	0.222	90 (0.6)	41 (35.3)
18	Understand	0.208	239 (1.6)	59 (50.9)
16	Follow gaze	0.188	516 (3.4)	78 (67.2)
20	Enjoy movement activities	-0.135	71 (0.5)	5 (4.3)
8	Interest in other children	0.133	135 (0.9)	29 (24.8)
15	Imitation	0.099	93 (0.6)	25 (21.7)
19	Social reference	0.087	552 (3.6)	59 (50.9)
2	Wonder if deaf	0.054	197 (1.3)	18 (15.5)
17	Seek attention	-0.038	1060 (7.0)	63 (53.8)
3	Pretend play	0.032	840 (5.5)	57 (48.7)
12	Upset by noises	0.010	2192 (14.5)	28 (23.9)
5	Unusual finger movements	0.009	1907 (12.6)	26 (22.2)
14	Eye contact	-0.008	163 (1.1)	24 (20.5)
4	Enjoy climbing	-0.008	81 (0.5)	7 (6.0)
13	Walk	0.001	90 (0.6)	11 (9.4)
11	Social smile	0.000	64 (0.4)	13 (11.2)

Table 8
Descriptives for Two-way ANOVAs by Risk Status and Diagnostic Group

	Non-ASD				ASD			
	Low-Risk		High-Risk		Low-Risk		High-Risk	
Measure	*n*	Mean (*SD*)	*n*	Mean (*SD*)	*n*	Mean (*SD*)	*n*	Mean (*SD*)
M-CHAT-R Total Score	105	6.1 (2.4)	22	5.8 (2.0)	100	7.3 (3.2)	34	8.8 (3.3)
M-CHAT-R/F Total Score	105	3.7 (1.7)	22	4.7 (2.1)	100	5.3 (3.1)	29	6.7 (3.0)
Change Score	105	2.5 (1.8)	22	1.1 (1.2)	100	2.3 (1.7)	29	1.5 (1.5)
ADOS SA CSS	104	2.4 (1.4)	21	2.8 (1.7)	94	6.4 (1.7)	31	6.6 (1.9)
ADOS RRB CSS	104	3.0 (2.4)	21	4.0 (2.6)	94	7.0 (2.4)	31	7.7 (2.4)
ADOS Overall CSS	104	2.0 (1.4)	21	2.6 (1.7)	94	6.5 (2.1)	31	6.9 (2.0)
CARS SC Factor Score[a]	101	1.5 (0.3)	14	1.7 (0.5)	95	2.4 (0.4)	26	2.5 (0.5)
CARS ER Factor Score[a]	101	1.3 (0.3)	14	1.4 (0.2)	95	1.7 (0.5)	26	2.1 (0.5)
CARS SBSS Factor Score[a]	101	1.4 (0.3)	14	1.5 (0.3)	95	2.0 (0.4)	26	2.1 (0.5)
CARS Total Score	101	21.4 (3.6)	15	23.5 (5.1)	95	32.7 (5.1)	27	33.4 (6.2)
MSEL Verbal DQ	103	66.7 (21.5)	21	81.6 (24.7)	93	44.7 (20.6)	33	60.3 (24.4)
MSEL Non-Verbal DQ	103	80.7 (18.5)	21	91.9 (18.0)	94	66.1 (18.7)	33	85.6 (19.1)
MSEL DQ	103	73.7 (18.7)	21	86.8 (19.4)	93	55.5 (17.3)	33	73 (19.4)
VABS-II Communication SS	104	85.1 (13.3)	22	88.4 (12.5)	94	71.9 (12.2)	33	74.7 (16.1)
VABS-II Daily Living SS	104	90.2 (15.5)	17	87.3 (10.1)	94	79.7 (14.6)	32	82.4 (12.6)
VABS-II Motor SS	104	88.7 (13.6)	18	91.6 (7.5)	94	84 (11.6)	32	90 (11.7)
VABS-II Socialization SS	104	86.7 (11.5)	21	86.5 (9.9)	94	77.4 (9.9)	33	76.6 (11.5)
VABS-II ABC SS	103	85.3 (13.1)	17	84.0 (7.2)	94	75.2 (9.9)	32	77.8 (11.5)

Note. Change Score = Number of items that changed from At-Risk to Not-at-risk during Follow-Up. ADOS CSS = Autism Diagnostic Observation Schedule Calibrated Severity Score (Range, 1-10; ASD Cutoff, 4). SA = Social Affect. RRB = Restricted Repetitive Behavior. CARS = Childhood Autism Rating Scale (Range, 15-60; ASD Cutoff, 25.5). SC = Social Communication. ER = Emotional Reactivity. SBSS = Stereotyped Behaviors and Sensory Sensitivities. MSEL DQ = Mullen Scales of Early Learning Developmental Quotient (M = 100, SD = 15). VABS-II = Vineland Adaptive Behavior Scales, Second Edition. SS = Standard Score (M = 100, SD = 15). ABC = Adaptive Behavior Composite.
[a]CARS Factor Scores range from 1(Within Normal Limits) to 4 (Severely Abnormal).

Table 9

Two-Way Between-Subjects ANOVAs Examining Effect of Risk Status and Diagnostic Group on Clinical Measures

	Main Effect						Interaction		
	Risk Status			Diagnostic Group			Risk Status x Diagnostic Group		
Measure	*F*	*p*[a]	η^2_p	*F*	*p*[a]	η^2_p	*F*	*p*[a]	η^2_p
M-CHAT-R Total Score	1.91	.168	.01	23.10	<.001	.08	4.33	.038	.02
M-CHAT-R/F Total Score	9.16	.003	.04	20.48	<.001	.08	0.26	.613	.001
Change Score	13.74	<.001	.05	0.001	.976	<.001	2.29	.131	.01
ADOS SA CSS	1.56	.214	.01	230.60	<.001	.48	0.19	.662	.001
ADOS RRB CSS	5.10	.025	.02	99.33	<.001	.29	0.17	.679	.001
ADOS Overall CSS	3.66	.057	.02	248.48	<.001	.50	0.03	.869	<.001
CARS SC Factor Score	1.35	.246	.01	150.79	<.001	.39	0.75	.387	.003
CARS ER Factor Score	5.20	.024	.02	54.84	<.001	.19	4.20	.041	.02
CARS SBSS Factor Score	0.85	.358	.004	86.48	<.001	.27	0.12	.725	.001
CARS Total Score	2.84	.093	.01	166.78	<.001	.42	0.81	.369	.003
MSEL Verbal DQ	19.91	<.001	.08	40.17	<.001	.14	0.01	.918	<.001
MSEL Non-Verbal DQ	27.53	<.001	.10	12.76	<.001	.05	2.00	.159	.01
MSEL DQ	28.08	<.001	.10	30.96	<.001	.11	0.59	.444	.002
VABS II Communication SS	2.19	.140	.01	43.35	<.001	.15	0.02	.903	<.001
VABS II Daily Living SS	0.002	.967	<.001	10.28	.002	.04	1.36	.224	.01
VABS II Motor SS	4.87	.028	.02	2.51	.114	.01	0.63	.429	.003
VABS II Socialization SS	0.10	.749	<.001	32.35	<.001	.12	0.03	.865	<.001
VABS II ABC SS	0.11	.743	<.001	18.50	<.001	.07	1.00	.318	.004

Note. η^2_p = Partial Eta Squared. Change Score = Number of items that changed from At-Risk to Not-at-risk during Follow-Up. ADOS CSS = Autism Diagnostic Observation Schedule Calibrated Severity Score. SA = Social Affect. RRB = Restricted Repetitive Behavior. CARS = Childhood Autism Rating Scale. SC = Social Communication. ER = Emotional Reactivity. SBSS = Stereotyped Behaviors and Sensory Sensitivities. MSEL DQ = Mullen Scales of Early Learning Developmental Quotient. VABS-II = Vineland Adaptive Behavior Scales, Second Edition. SS = Standard Score. ABC = Adaptive Behavior Composite.

[a]Bonferroni-adjusted α = .003 was used to determine significance.

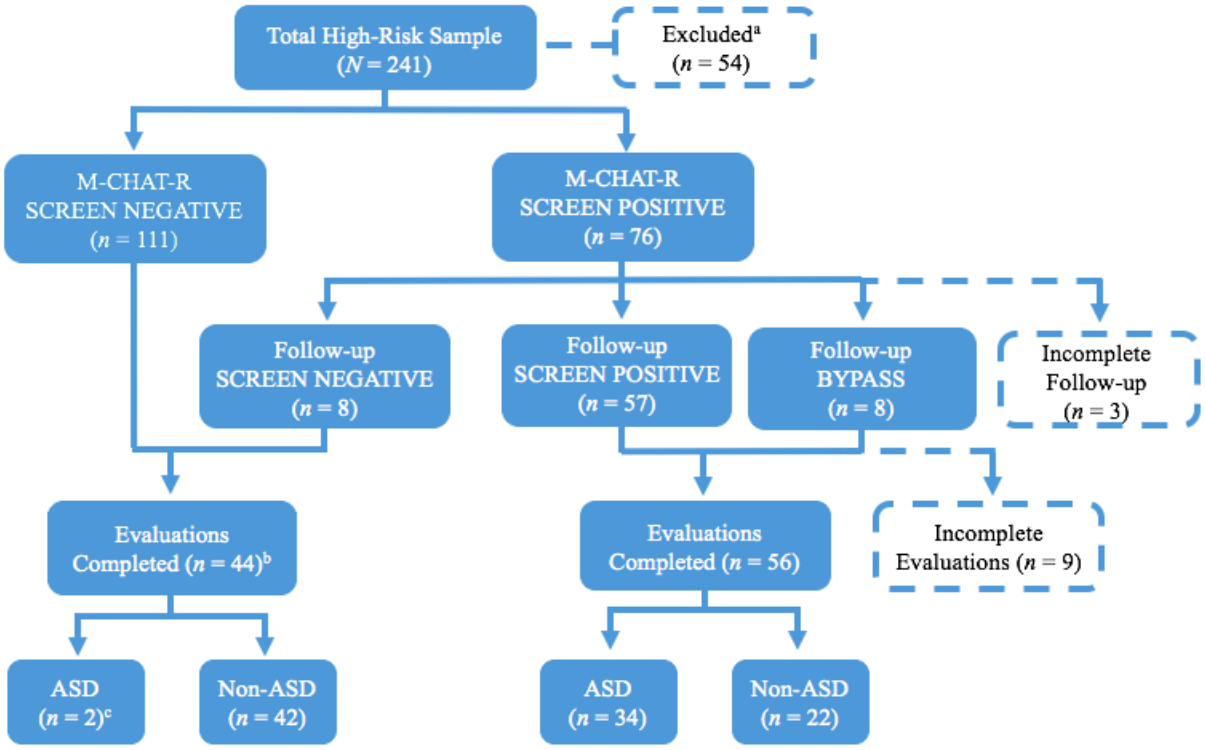
Table 10
Simple Main Effects of Risk Status on Clinical Measures Within ASD Group

Measure	Estimated Mean Difference	SEM	*F*	*p*	η^2
ADOS SA CSS	0.21	0.34	0.38	.537	.002
ADOS RRB CSS	0.71	0.50	1.98	.161	.01
ADOS Overall CSS	0.49	0.37	1.78	.183	.01
CARS SC Factor Score	0.02	0.09	0.06	.809	<.001
CARS ER Factor Score	0.32	0.09	12.47	<.001	.05
CARS SBSS Factor Score	0.04	0.09	0.22	.643	.001
CARS Total Score	0.64	1.02	0.40	.527	.002
MSEL Verbal DQ	15.62	4.42	12.48	<.001	.05
MSEL Non-Verbal DQ	19.46	3.77	26.62	<.001	.10
MSEL DQ	17.47	3.72	22.04	<.001	.08
VABS II Communication SS	2.78	2.68	1.08	.299	.004
VABS II Daily Living SS	2.72	2.97	0.83	.362	.003
VABS II Motor SS	6.01	2.51	5.75	.017	.02
VABS II Socialization SS	-0.83	2.18	0.14	.705	.001
VABS II ABC SS	2.52	2.33	1.16	.282	.01

Note. Fisher's LSD was used to correct for multiple comparisons. Estimated Mean Differences presented as mean of the LR sample subtracted from the mean of the HR sample. SEM = Standard Error of the Mean. ADOS CSS = Autism Diagnostic Observation Schedule Calibrated Severity Score. SA = Social Affect. RRB = Restricted Repetitive Behavior. CARS = Childhood Autism Rating Scale. SC = Social Communication. ER = Emotional Reactivity. SBSS = Stereotyped Behaviors and Sensory Sensitivities. MSEL DQ = Mullen Scales of Early Learning Developmental Quotient. VABS-II = Vineland Adaptive Behavior Scales, Second Edition. SS = Standard Score. ABC = Adaptive Behavior Composite.

Figures

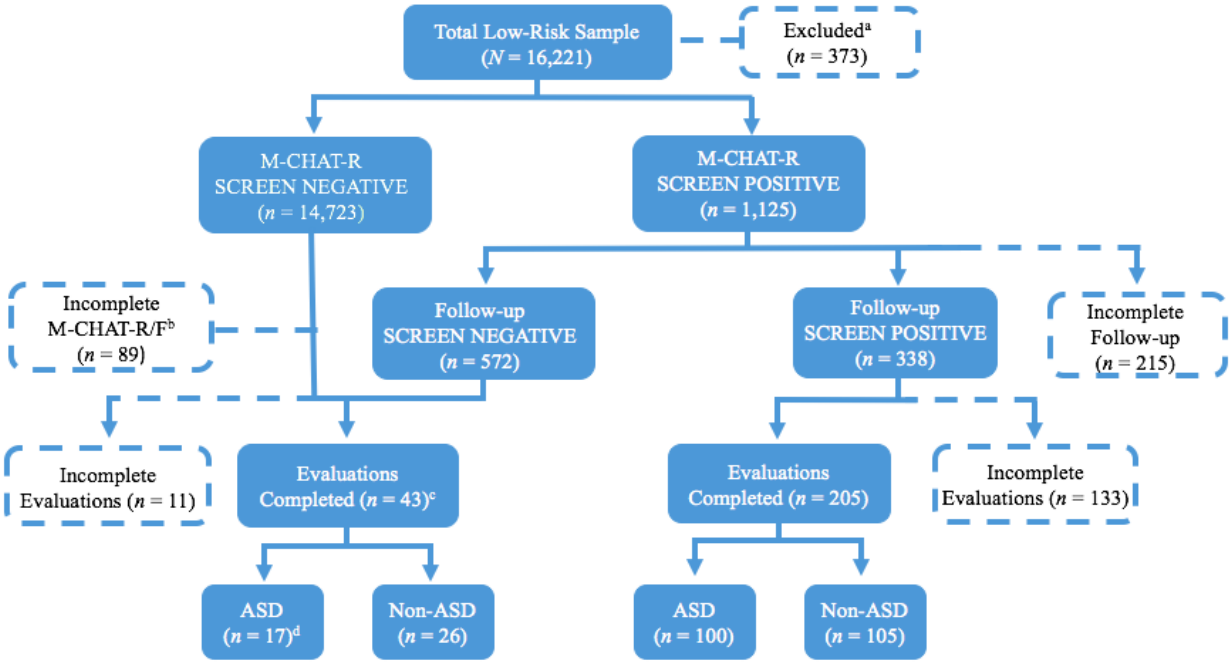
Figure 1
Flow Chart of Screening Results for High-Risk Sample



[a]Children were excluded if they were a twin (*n* = 1), known neurological impairment (*n* = 3), gestational age < 37 weeks (*n* = 3), outside of screening age range (*n* = 11), refused (*n* = 8), diagnosis of older sibling could not be confirmed (*n* = 4), lost to follow-up (*n* = 16), and unknown (*n* = 8).
[b]Screen negative cases were detected through pediatrician red flag (*n* = 1) and VU follow-up on screen negatives (*n* = 43).
[c]Both cases were detected through VU follow-up on screen negatives cases.

Figure 2

Flow Chart of Screening Results for Low-Risk Sample



[a]Cases were excluded if the M-CHAT-R was not completed in English (*n* = 329) and for not meeting inclusion criteria (*n* = 44).
[b]Children who had incomplete M-CHAT-R/F based on alternate scoring criteria. See Robins et al. (2014) for details.
[c]Screen negative cases were detected through pediatrician red flag (*n* = 20), STAT screen positive (*n* = 20), or alternate scoring (*n* = 3).
[d]ASD cases were detected through pediatrician red flag (*n* = 9), STAT screen positive (*n* = 6), and alternate scoring (*n* = 2).

References

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition, Text Revision. Washington, DC: American Psychiatric Association

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders,* Fifth Edition. Arlington, VA: American Psychiatric Association.

Arpino, C., Compagnone, E., Montanaro, M. L., Cacciatore, D., De Luca, A., Cerulli, A., … Curatolo, P. (2010). Preterm birth and neurodevelopmental outcome: A review. *Child's Nervous System*, *26*(9), 1139–1149. doi:10.1007/s00381-010-1125-y

Baranek, G. T., Watson, L. R., Crais, E., & Reznick, S. (2003). *First Year Inventory (FYI) 2.0*. Chapel Hill, NC: University of North Carolina at Chapel Hill.

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *The British Journal of Psychiatry, 161*, 839–843.

Barton, M.L., Boorstein, H., Dumont-Mathieu, T., Herlihy, L.E., & Fein, D. (2012). *Toddler ASD Symptom Interview (TASI)*. Self-published.

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism Screening Questionnaire: Diagnostic validity. *The British Journal of Psychiatry, 175*, 444–451.

Bickel, J., Bridgemohan, C., Sideridis, G., & Huntington, N. (2015). Child and family characteristics associated with age of diagnosis of an Autism Spectrum Disorder in a tertiary care setting. *Journal of Developmental and Behavioral Pediatrics, 36*(1), 1–7. doi:10.1097/DBP.0000000000000117

Brian, J., Bryson, S. E., Garon, N., Roberts, W., Smith, I. M., Szatmari, P., & Zwaigenbaum, L. (2008). Clinical assessment of autism in high-risk 18-month-olds. *Autism, 12*(5), 433-456. doi:10.1177/1362361308094500

Bryson, S.E., McDermott, C., Rombough, V., Brian, J., & Zwaigenbaum, L. (2008). The Autism Observation Scale for Infants (AOSI): Scale development and reliability data. *Journal of Autism and Developmental Disorders, 38*, 731-738.

Bryson, S. E., Zwaigenbaum, L., Brian, J., Roberts, W., Szatmari, P., Rombough, V., & McDermott, C. (2007). A prospective case series of high-risk infants who developed autism. *Journal of Autism and Developmental Disorders, 37*(1), 12–24. doi: 10.1007/s10803-006-0328-2

Center for Disease Control and Prevention (CDC). (2016). Prevalence and characteristics of Autism Spectrum Disorder among children aged 8 years— Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *Morbidity and Mortality Weekly Reports (MMWR)*, 65(SS-3), 1–23.

Chawarska, K., Paul, R., Klin, A., Hannigen, S., Dichtel, L. E., & Volkmar, F. (2007). Parental recognition of developmental problems in toddlers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 37*(1), 62–72. doi:10.1007/s10803-006-0330-8

Chlebowski, C., Green, J. A., Barton, M. L., & Fein, D. (2010). Using the Childhood Autism Rating Scale to diagnose Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 40*(7),

787–799. doi:10.1007/s10803-009-0926-x

Chlebowski, C., Robins, D.L., Barton, M.L., & Fein, D. (2013). Large-scale use of the Modified Checklist for Autism in Low-Risk Toddlers. *Pediatrics, 121*(4), e1121-e1127. doi:10.1542/peds.2012-1525

Christensen, D. L., Bilder, D. A., Zahorodny, W., Pettygrove, S., Durkin, M. S., Fitzgerald, R. T., … Yeargin-Allsopp, M. (2016). Prevalence and characteristics of Autism Spectrum Disorder among 4-year-old children in the Autism and Developmental Disabilities Monitoring Network. *Journal of Developmental and Behavioral Pediatrics, 37*(1), 1-8.

Cohen, J.W. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Constantino, J. N., Zhang, Y., Frazier, T., Abbacchi, A. M., & Law, P. (2010). Sibling recurrence and the genetic epidemiology of autism. *American Journal of Psychiatry*, *167*(11), 1349–1356. doi:10.1176/appi.ajp.2010.09101470

Coonrod, E. E., & Stone, W. L. (2004). Early concerns of parents of children with autistic and nonautistic disorders. *Infants & Young Children*, *17*(3), 258.

De Giacomo, A., & Fombonne, E. (1998). Parental recognition of developmental abnormalities in autism. *European Child & Adolescent Psychiatry, 7*(3), 131–136.

Dietz, C., Swinkels, S., van Daalen, E., van Engeland, H., & Buitelaar, J. K. (2006). Screening for autistic spectrum disorder in children aged 14–15 months II: Population screening with the Early Screening of Autistic Traits Questionnaire (ESAT). *Journal of Autism and Developmental Disorders, 36*(6), 713– 722. doi:10.1007/s10803-006-0114-1

Dulcan, M. K., Costello, E. J., Costello, A. J., Edelbrock, C., Brent, D., & Janiszewski, S. (1990). The pediatrician as gatekeeper to mental health care for children: Do parents' concerns open the gate? *Journal of the American Academy of Child & Adolescent Psychiatry*, *29*(3), 453–458. doi:10.1097/00004583-199005000-00018

Eaves, L. C., & Ho, H. H. (2004). The very early identification of Autism: Outcome to age 4 1/2–5. *Journal of Autism and Developmental Disorders*, *34*(4), 367–378. doi:10.1023/B:JADD.0000037414.33270.a8.

Feldman, M. A., Ward, R. A., Savona, D., Regehr, K., Parker, K., Hudson, M., … Holden, J. J. A. (2012). Development and initial validation of a parent report measure of the behavioral development of infants at risk for Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 42*(1), 13–22. doi:10.1007/s10803-011-1208-y

Feldt, L. S. (1969). A test of the hypothesis that cronbach's alpha or kuder-richardson coeffient twenty is the same for two tests. *Psychometrika, 34*(3), 363-373. doi:10.1007/BF02289364

Gamliel, I., Yirmiya, N., Jaffe, D. H., Manor, O., & Sigman, M. (2009). Developmental trajectories in

siblings of children with autism: Cognition and language from 4 months to 7 years. *Journal of Autism and Developmental Disorders, 39*(8), 1131–1144. doi:10.1007/s10803-009-0727-2

Glascoe, F. P. (1997). *Parents' Evaluations of Developmental Status*. Nashville, TN: Ellsworth and Vandermeer Press.

Glascoe, F. P., & Dworkin, P. H. (1995). The role of parents in the detection of developmental and behavioral problems. *Pediatrics*, *95*(6), 829–836.

Glascoe, F. P., Altemeier, W. A., & MacLean, W. E. (1989). The importance of parents' concerns about their child's development. *Archives of Pediatrics & Adolescent Medicine, 143*(8), 955.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 39,* 693-705. doi:10.1007/s10803-008-0674-3

Herlihy, L. E., Brooks, B., Dumont-Mathieu, T., Barton, M. L., Fein, D., Chen, C. -M., & Robins, D. L. (2014). Standardized screening facilitates timely diagnosis of Autism Spectrum Disorders in a diverse sample of low-risk toddlers. *Journal of Developmental and Behavioral Pediatrics, 35,* 85-92. doi:10.1097/DBP.0000000000000014

Herlihy, L., Knoch, K., Vibert, B., & Fein, D. (2013). Parents' first concerns about toddlers with Autism Spectrum Disorder: Effect of sibling status. *Autism*. doi:10.1177/1362361313509731

Hess, C. R., & Landa, R. J. (2012). Predictive and concurrent validity of parent concern about young children at risk for autism. *Journal of Autism and Developmental Disorders, 42*(4), 575–584. doi:10.1007/s10803-011-1282-1

Hus, V., Gotham, K., & Lord, C. (2014). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders, 44*, 2400-2412. doi:10.1007/s10803-012-1719-1

Johnson, C. P., & Myers, S. M. (2007). Identification and evaluation of children with Autism Spectrum Disorders. *Pediatrics*, *120*(5), 1183–1215. doi:10.1542/peds.2007-2361

Johnson, S., Hollis, C., Kochhar, P., Hennessy, E., Wolke, D., & Marlow, N. (2010). Autism Spectrum Disorders in extremely preterm children. *The Journal of Pediatrics, 156* (4), 525-531. doi:10.1016/j.jpeds.2009.10.041

Joseph, R. M., Tager-Flusberg, H., & Lord, C. (2002). Cognitive profiles and social-communicative functioning in children with Autism Spectrum Disorder. *Journal of Child Psychology and Psychiatry, 43*(6), 807-821. doi:10.1111/1469-7610.00092

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. doi:10.1007/s10803-017-3060-1

Khowaja, M. K., Hazzard, A. P., & Robins, D. L. (2015). Sociodemographic barriers to early detection of autism: Screening and evaluation using the M-CHAT, M-CHAT-R, and Follow-Up. *Journal of Autism and Developmental Disorders*, *45*(6), 1797–1808. doi:10.1007/s10803-014-2339-8

Kleinman, J. M., Robins, D. L., Ventola, P. E., Pandey, J., Boorstein, H. C., Esser, E. L., … Fein, D. (2008). The Modified Checklist for Autism in Toddlers: A follow-up study investigating the early detection of Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *38*(5), 827–839. doi:10.1007/s10803-007-0450-9

Kleinman, J. M., Ventola, P. E., Pandey, J., Verbalis, A. D., Barton, M., Hodgson, S., … Fein, D. F. (2008). Diagnostic stability in very young children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *38*(4), 606–615. doi:10.1007/s10803-007-0427-8

Kuban, K. C. K., O'Shea, T. M., Allred, E. N., Tager-Flusberg, H., Goldstein, D. J., & Leviton, A. (2009). Positive screening on the Modified Checklist for Autism in Toddlers (M-CHAT) in extremely low gestational age newborns. *The Journal of Pediatrics*, *154*(4), 535–540.e1. doi:10.1016/j.jpeds.2008.10.011

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

Limperopoulos, C., Bassan, H., Sullivan, N. R., Soul, J. S., Robertson, R. L., Moore, M., … du Plessis, A. J. (2008). Positive screening for autism in ex-preterm infants: Prevalence and risk factors. *Pediatrics*, *121*(4), 758–765. doi:10.1542/peds.2007-2158

Lord, C., Luyster, R., Gotham, K., & Guthrie, W. (2012). *Autism Diagnostic Observation Schedule, 2nd edition.* Torrance, CA: Western Psychological Services.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223. doi:10.1023/A:1005592401947

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685.

Luyster, R. J., Kuban, K. C. K., O'Shea, T. M., Paneth, N., Allred, E. N., & Leviton, A. (for ELGAN Study investigators). (2011). The Modified Checklist for Autism in Toddlers in extremely low gestational age newborns: Individual items associated with motor, cognitive, vision and hearing limitations. *Paediatric and Perinatal Epidemiology*, *25*(4), 366–376. doi:10.1111/j.1365-3016.2010.01187.x

Macari, S. L., Campbell, D., Gengoux, G. W., Saulnier, C. A., Klin, A. J., & Chawarska, K. (2012). Predicting developmental status from 12 to 24 months in infants at risk for Autism Spectrum Disorder: A preliminary report. *Journal of Autism and Developmental Disorders, 42,* 2636-2647. doi:10.1007/s10803-012-1521-0

Mathews, T. L., King, M. L., Kupzyk, K. A., & Lake, C. M. (2014). Findings and implications of developmental screening for high-risk children referred to a tertiary developmental disability center. *Journal of Pediatric Health Care*, *28*(6), 507–515. doi:10.1016/j.pedhc.2014.03.002

McMahon, C. R., Malesa, E. E., Yoder, P. J., & Stone, W. L. (2007). Parents of children with Autism Spectrum Disorders have merited concerns about their later-born infants. *Research and Practice for*

Persons with Severe Disabilities, 32(2), 154–160.

Messinger, D. S., Young, G. S., Webb, S. J., Ozonoff, S., Bryson, S. E., Carter, A., … Zwaigenbaum, L. (2015). Early sex differences are not autism-specific: A Baby Siblings Research Consortium (BSRC) study. *Molecular Autism, 6*(1), 1. doi:10.1186/s13229-015-0027-y

Messinger, D., Young, G. S., Ozonoff, S., Dobkins, K., Carter, A., Zwaigenbaum, L., … Sigman, L. (2013). Beyond autism: A baby siblings research consortium study of high-risk children at three years of age. *Journal of the American Academy of Child & Adolescent Psychiatry, 52*(3), 300–308.e1. doi:10.1016/j.jaac.2012.12.011

Miles, J. H., Takahashi, T. N., Bagby, S., Sahota, P. K., Vaslow, D. F., Wang, R. E., … Farmer, J. E. (2005). Essential versus complex autism: Definition of fundamental prognostic subtypes. *American Journal of Medical Genetics, 135A,* 171-180. doi:10.1002/ajmg.a.30590

Miller, J. S., Gabrielsen, T., Villalobos, M., Alleman, R., Wahmhoff, N., … Segura, B. (2011). The each child study: Systematic screening for Autism Spectrum Disorders in a pediatric setting. *Pediatrics, 127*(5), 866–871. doi:10.1542/peds.2010-0136

Moore, T., Johnson, S., Hennessy, E., & Marlow, N. (2012). Screening for autism in extremely preterm infants: Problems in interpretation. *Developmental Medicine & Child Neurology, 54*(6), 514–520. doi:10.1111/j.1469-8749.2012.04265.x

Moulton, E., Bradbury, K., Barton, M., & Fein, D. (2016). Factor analysis of the childhood autism rating scale in a sample of two year olds with an Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders.* Advance online publication. doi:10.1007/s10803-016-2936-9

Mullen, E. M. (1995). *Mullen Scales of Early Learning: Manual*. Minneapolis, MN: NCS Pearson, Inc.

Munson, J., Dawson, G., Sterling, L., Beauchaine, T., Zhou, A., Koehler, E., … Abbott, R. (2008). Evidence for latent classes of IQ in young children with Autism Spectrum Disorder. *American Journal on Mental Retardation, 113*(6), 439-452. doi:10.1352/2008.113:439-452

Nadig, A. S., Ozonoff, S., Young, G. S., Rozga, A., Sigman, M., & Rogers, S. J. (2007). A prospective study of response to name in infants at risk for autism. *Archives of Pediatrics & Adolescent Medicine, 161*(4), 378–383. doi:10.1001/archpedi.161.4.378

Oosterling, I. J., Swinkels, S. H., Jan van der Gaag, R., Visser, J. C., Dietz, C., & Buitelaar, J. K. (2009). Comparative analysis of three screening instruments for Autism Spectrum Disorder in toddlers at high risk. *Journal of Autism and Developmental Disorders, 39,* 897-909. doi:10.1007/s10803-009-0692-9

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., … Stone, W. L. (2011). Recurrence risk for Autism Spectrum Disorders: A Baby Siblings Research Consortium study. *Pediatrics*. doi:10.1542/peds.2010-2825

Ozonoff, S., Young, G. S., Steinfeld, M. B., Hill, M. M., Cook, I., Hutman, T., … Sigman, M. (2009). How early do parent concerns predict later autism diagnosis? *Journal of Developmental and Behavioral Pediatrics, 30*(5), 367–375.

Pandey, J., Verbalis, A., Robins, D. L., Boorstein, H., Klin, A. M. I., Babitz, T., … Fein, D. (2008).

Screening for autism in older and younger toddlers with the Modified Checklist for Autism in Toddlers. *Autism*, *12*(5), 513–535. doi:10.1177/1362361308094503

Pickles, A., Starr, E., Kazak, S., Bolton, P., Papanikolaou, K., Bailey, A., … Rutter, M. (2000). Variable expression of the autism broader phenotype: Findings from extended pedigrees. *Journal of Child Psychology and Psychiatry*, *41*(4), 491–502.

Robins, D. L., Casagrande, K., Barton, M., Chen, C.-M. A., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the Modified Checklist for Autism in Toddlers, Revised with Follow-up (M-CHAT-R/F). *Pediatrics*, *133*(1), 37–45. doi:10.1542/peds.2013-1813

Robins, D. L., Fein, D., & Barton, M. L. (1999). *Modified Checklist for Autism in Toddlers (M-CHAT)*. Self-published.

Robins, D. L., Fein, D., & Barton, M. L. (2009). *The Modified Checklist for Autism in Toddlers, Revised with Follow-up (M-CHAT-R/F).* Self-published.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *31*(2), 131–144.

Rogers, S. J., Vismara, L., Wagner, A. L., McCormick, C., Young, G., & Ozonoff, S. (2014). Autism treatment in the first year of life: A pilot study of infant start, a parent-implemented intervention for symptomatic infants. *Journal of Autism and Developmental Disorders*, *44*(12), 2981–2995. doi:10.1007/s10803-014-2202-y

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. doi:10.1007/s10803-014-2211-x

Rutter, M., Bailey, A., & Lord, C. (2003). *Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services.

Sacrey, L. A. R., Bryson, S., Zwaigenbaum, L., Brian, J., Smith, I. M., Roberts, W., … Garon, N. (2016). 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.* Advance online publication. doi:10.1177/1362361316675120

Sacrey, L. A. R., Zwaigenbaum, L., Bryson, S., Brian, J., Smith, I. M., Roberts, W., … Armstrong, V. (2015). Can parents' concerns predict Autism Spectrum Disorder? A prospective study of high-risk siblings from 6 to 36 months of age. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(6), 470–478. doi:10.1016/j.jaac.2015.03.014

Sacrey, L. A., Zwaigenbaum, L., Bryson, S. E., Brian, J. A., Smith, I. M., Roberts, W., … Garon, N. (2017, May). *Parent and clinician agreement in early behavioral signs in 12-month-old infants at-risk for Autism Spectrum Disorder: A high-risk sibling cohort.* Paper presented at the International Meeting for Autism Research, San Francisco, CA. Abstract retrieved from www.autism-insar.org/IMFARarchives.

Schopler, E., Reichler, R. J., & Renner, B. R. (1995). *Childhood Autism Rating Scale: Manual.* Los

Angeles, CA: Western Psychological Services.

Schwichtenberg, A. J., Young, G. S., Sigman, M., Hutman, T., & Ozonoff, S. (2010). Can family affectedness inform infant sibling outcomes of Autism Spectrum Disorders? *Journal of Child Psychology and Psychiatry, 51*(9), 1021–1030. doi:10.1111/j.1469-7610.2010.02267.x

Siegel, B. (2004). *Pervasive Developmental Disorders Screening Test - II*. San Antonio, TX: PsychCorp.

Siegel, B., Pliner, C., Eschler, J., & Elliott, G. R. (1988). How children with autism are diagnosed: Difficulties in identification of children with multiple developmental delays. *Journal of Developmental and Behavioral Pediatrics*, *9*(4), 199–204.

Snow, A. V., & Lecavalier, L. (2008). Sensitivity and specificity of the Modified Checklist for Autism in Toddlers and the Social Communication Questionnaire in preschoolers suspected of having pervasive developmental disorders. *Autism*, *12*(6), 627–644. doi:10.1177/1362361308097116

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales, Second Edition*. Circle Pines, MN: AGS Publishing.

Stephens, B. E., Bann, C. M., Watson, V. E., Sheinkopf, S. J., Peralta-Carcelen, M., Bodnar, A., … Vohr, B. R. (2012). Screening for Autism Spectrum Disorders in extremely preterm infants. *Journal of Developmental and Behavioral Pediatrics*, *33*(7), 535–541. doi:10.1097/DBP.0b013e31825fd0af

Stone, W. L., Coonrod, E. E., Turner, L. M., & Pozdol, S. L. (2004). Psychometric properties of the STAT for early autism screening. *Journal of Autism and Developmental Disorders*, *34*(6), 691–701. doi:10.1007/s10803-004-5289-8

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. *Autism*, *12*(5), 557-573. doi:10.1177/1362361308096403

Stone, W. L., Ousley, O. Y., Hepburn, S. L., Hogan, K. L., & Brown, C. S. (2009). Patterns of adaptive behavior in very young children with autism. *American Journal on Mental Retardation, 104*(2), 187-199. doi:10.1352/0895-8017(1999)104<0187:POABIV>2.0.CO;2

Sutera, S., Pandey, J., Esser, E. L., Rosenthal, M. A., Wilson, L. B., Barton, M., … Fein, D. (2007). Predictors of optimal outcome in toddlers diagnosed with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *37*(1), 98–107. doi:10.1007/s10803-006-0340-6

Swinkels, S. H., Dietz, C., van Daalen, E., Kerkhof, I. H., van Engeland, H., & Buitelaar, J. K. (2006). Screening for autistic spectrum disorder in children aged 14 to 15 months I: The development of the Early Screening of Autistic Traits questionnaire (ESAT). *Journal of Autism and Developmental Disorders, 36*(6), 723–732. doi:10.1007/s10803-006-0115-0

Szatmari, P. (1999). Heterogeneity and the genetics of autism. *Journal of Psychiatry & Neuroscience*, *24*(2), 159–165.

Szatmari, P., Chawarska, K., Dawson, G., Georgiades, S., Landa, R., Lord, C., … Halladay, A. (2016). Prospective longitudinal studies of infant siblings of children with autism: Lessons learned and future directions. *Journal of the American Academy of Child and Adolescent Psychiatry, 55*(3), 179-187. doi:10.1016/j.jaac.2015.12.014

Talbott, M. R., Nelson, C. A., & Tager-Flusberg, H. (2015). Diary reports of concerns in mothers of infant siblings of children with autism across the first year of life. *Journal of Autism and Developmental Disorders, 45*(7), 2187–2199. doi:10.1007/s10803-015-2383-z

Thomas, R. E., Spragins, W., Mazloum, G., Cronkhite, M., & Maru, G. (2016). Rates of detection of developmental problems at the 18-month well-baby visit by family physicians' using four evidence-based screening tools compared to usual care: A randomized controlled trial. *Child: Care, Health and Development,* 1–12. doi:10.1111/cch.12333

Toth, K., Dawson, G., Meltzoff, A. N., Greenson, J., & Fein, D. (2007). Early social, imitation, play, and language abilities of young non-autistic siblings of children with autism. *Journal of Autism and Developmental Disorders, 37*(1), 145–157. doi:10.1007/s10803-006-0336-2

Turner, L. M., & Stone, W. L. (2007). Variability in outcome for children with an ASD diagnosis at age 2. *Journal of Child Psychology and Psychiatry, 48*(8), 793–802. doi:10.1111/j.1469-7610.2007.01744.x

Veness, C., Prior, M., Bavin, E., Eadie, P., Cini, E., & Reilly, S. (2011). Early indicators of Autism Spectrum Disorders at 12 and 24 months of age: A prospective, longitudinal comparative study. *Autism, 16*(2), 163-177. doi:10.1177/1362361311399936

Weitlauf, A. S., Vehorn, A. C., Stone, W. L., Fein, D., & Warren, Z. E. (2015). Using the M-CHAT-R/F to identify developmental concerns in a high-risk 18-month-old sibling sample. *Journal of Developmental and Behavioral Pediatrics, 36*(7), 497. doi:10.1097/DBP.0000000000000194

Weinberg, S. L. & Abramowitz, S. K. (2008). *Statistics using SPSS: An integrative approach.* New York, NY: Cambridge University Press.

Wetherby, A. M., Woods, 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.

Wetherby, A., & Prizant, B. (2002). *Communication and Symbolic Behavior Scales Developmental Profile First normed edition.* Baltimore, MD: Paul H Brookes Publishing.

Wiggins, L. D., Bakeman, R., Adamson, L. B., & Robins, D. L. (2007). The utility of the Social Communication Questionnaire in screening for autism in children referred for early intervention. *Focus on Autism and Other Developmental Disabilities, 22*(1), 33–38. doi:10.1177/10883576070220010401

Wolff, J. J., Botteron, K. N., Dager, S. R., Elison, J. T., Estes, A. M., Gu, H., … The IBIS Network. (2014). Longitudinal patterns of repetitive behavior in toddlers with autism. *Journal of Child Psychology and Psychiatry, 55*(8), 945-953. doi:10.1111/jcpp.12207

Young, R. L., Brewer, N., & Pattison, C. (2003). Parental identification of early behavioural abnormalities in children with autistic disorder. *Autism, 7*(2), 125–143.

Zuckerman, K. E., Lindly, O. J., & Sinche, B. K. (2015). Parental concerns, provider response, and timeliness of Autism Spectrum Disorder diagnosis. *The Journal of Pediatrics, 166*(6), 1431–9.e1. doi:10.1016/j.jpeds.2015.03.007

Zwaigenbaum, L., Bryson, S. E., Brian, J., Smith, I. M., Roberts, W., Szatmari, P., … Vaillancourt, T. (2016). Stability of diagnostic assessment for Autism Spectrum Disorder between 18 and 36 months in a high-risk cohort. *Autism Research, 9*(7), 790-800. doi:10.1002/aur.1585

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. doi:10.1016/j.ijdevneu.2004.05.001