

# UC Davis

## UC Davis Previously Published Works

### Title

Sex differences in early autism screening using the Modified Checklist for Autism in Toddlers, Revised, with Follow-Up (M-CHAT-R/F)

### Permalink

<https://escholarship.org/uc/item/9gg6n865>

### Journal

Autism, 27(7)

### ISSN

1362-3613

### Authors

Eldeeb, Sherief Y

Ludwig, Natasha N

Wieckowski, Andrea Trubanova

et al.

### Publication Date

2023-10-01

### DOI

10.1177/13623613231154728

Peer reviewed



Published in final edited form as:

Autism. 2023 October ; 27(7): 2112–2123. doi:10.1177/13623613231154728.

## Sex Differences in Early Autism Screening Using the Modified Checklist for Autism in Toddlers, Revised, with Follow-Up (M-CHAT-R/F)

Sherief Y. Eldeeb, BA<sup>1</sup>, Natasha N. Ludwig, PhD<sup>2</sup>, Andrea Trubanova Wieckowski, PhD<sup>1</sup>, Mary F. S. Dieckhaus, MS<sup>3</sup>, Yasemin Algur, MPH<sup>4</sup>, Victoria Ryan, MS<sup>4</sup>, Sarah Dufek, PhD<sup>5</sup>, Aubyn Stahmer, PhD<sup>5</sup>, Diana L. Robins, PhD<sup>1</sup>

<sup>1</sup>A.J. Drexel Autism Institute, Drexel University, Philadelphia, PA,

<sup>2</sup>Kennedy Krieger Institute, John Hopkins School of Medicine, Baltimore, MD

<sup>3</sup>Department of Psychological Sciences, University of Connecticut, Storrs, CT,

<sup>4</sup>Department of Epidemiology & Biostatistics, Drexel University, Philadelphia, PA,

<sup>5</sup>Department of Psychiatry and Behavioral Sciences, University of California, Davis MIND Institute, Sacramento, CA

### Abstract

Males are more likely to be diagnosed with autism than females, and at earlier ages, yet few studies examine sex differences in screening. This study explored sex differences in psychometric properties, recommended cut-off scores, and overall scores of the Modified Checklist for Autism in Toddlers, Revised, with Follow-Up (M-CHAT-R/F). Participants were 28,088 toddlers enrolled in four early detection of autism studies. Children ( $n = 731$ ) at high-likelihood for autism attended evaluations after screening and/or primary care clinician concern. Females were less likely to screen at high-likelihood for autism at each stage of screening and therefore less likely to be invited for evaluations. Positive Predictive Value (PPV) was significantly lower among females than males, but sensitivity was similar. False positive (FP) females were likely to have another developmental delay. Cut-off scores for males and females matched recommended guidelines. Final scores on the M-CHAT-R/F did not differ between males and females diagnosed with autism, but did for the overall sample identified at high-likelihood for autism. Our findings suggest that females are less likely to be referred for evaluations, but the M-CHAT-R/F accurately identifies both males and females with autism at established cut-offs. Future research should examine methods to reduce FP in females.

### Lay Abstract

---

Correspondence concerning this article should be addressed to Diana L. Robins, A.J. Drexel Autism Institute, 3020 Market Street, Suite 560, Philadelphia, PA 19104. dlr76@drexel.edu.

Dr. Robins is a co-owner of M-CHAT LLC, which receives royalties from parties that license use of the M-CHAT in electronic products. No royalties were received for any of the data presented in the current study. Dr. Robins sits on the advisory board of Quadrant Biosciences Inc, for which she receives an honorarium. The other authors have indicated they have no financial relationships or conflicts of interest relevant to this article to disclose.

This study examined a widely used autism screening tool, the Modified Checklist for Autism in Toddlers, Revised, with Follow-Up (M-CHAT-R/F) to identify differences in screening for autism between toddler males and females. Examining sex differences in screening for autism in toddlerhood is important as it determines who will be referred for evaluations and receive diagnoses, which is critical for access to autism-specific early intervention. This study found that females were less likely to screen positive and be invited for evaluations compared to males. Females at high-likelihood (HL) for autism were less likely to be diagnosed with autism, which decreases confidence in the screener's results. Importantly, the M-CHAT-R/F accurately identified both males and females with autism. Future research should examine ways to improve accuracy in screening results for females.

## Keywords

autism; sex differences; toddler; screening

---

Although autism can be reliably diagnosed in toddlers (Landa et al., 2013; Pierce et al., 2019), most children are first identified later in childhood (Maenner et al., 2021). Delayed diagnosis causes children to miss the opportunity for autism-specific early intervention, which contributes to poorer outcomes (Anderson, Liang, & Lord, 2014; Landa, 2018; Vivanti et al., 2016). Given the positive effects associated with early intervention (Fuller and Kaiser, 2020; Fuller et al., 2020; Zwaigenbaum et al., 2015), there is a pressing need to improve early detection practices to streamline diagnosis and access to services for young autistic children.

The Modified Checklist for Autism in Toddlers, Revised, with Follow-Up (M-CHAT-R/F; Robins et al., 2009) is one of the most widely used autism screening tools. Although estimates of M-CHAT sensitivity vary (range = 22% – 100%; Carbone et al., 2020; Chlebowski et al., 2013; Guthrie et al., 2019; Robins et al., 2014; Wieckowski et al., in press), there is consensus that not all children who screen at high-likelihood (HL) for autism are ultimately diagnosed with autism, although most are diagnosed with another developmental disorder (Robins et al., 2014). However, reducing the number of false positive (FP) screens is important to shorten waitlists and facilitate timely access to autism-specific services for both female and male autistic children.

Recent prevalence estimates indicate that 4.2 males are diagnosed for every female with autism (Maenner et al., 2021). Studies examining sex differences in core autism symptoms in young children display equivocal findings. Some toddler autism studies suggest that autistic males demonstrate more restricted and repetitive patterns of behavior (RRBs), and fewer social communication impairments compared to females (Hartley & Sikora, 2009; Lawson et al., 2018; Ros-Demarize et al., 2020; Sipes et al., 2011). In contrast, a study examining retrospective caregiver ratings of preschool behavior in children later diagnosed with autism indicated that girls engaged in more complex social imitation than boys and that females later diagnosed with autism used mimicking in social situations, whereas male counterparts were more likely to isolate themselves (Hiller et al., 2016). Further, other studies have found minimal to no sex differences in toddlers/preschoolers (Mussey et al.,

2017; Reinhardt et al., 2015; Van Wijngaarden-Cremers et al., 2014). The equivocal findings may impede optimizing early detection for both females and males.

Indeed, evidence indicates sex differences in the timing of an autism diagnosis. Specifically, females were diagnosed significantly later than males (i.e., 20.19 months) despite similar age of first caregiver concerns (McDonnell et al., 2020). Notably, this effect was moderated by verbal IQ such that stronger verbal skills were associated with later diagnosis more so for females than males, consistent with findings of a 1.8 year delay in childhood diagnosis of Asperger's disorder for females compared to males (Beeger et al., 2013). Evidence of sex-based disparity in the timing of an autism diagnosis illustrates the urgent need to improve early detection practices to facilitate access to early interventions for autistic females.

Previous studies have not found significant sex differences in the *age* of first caregiver concerns, but have documented differences in the *types* of early caregiver concerns that emerge (Dillon et al., 2021; Hiller et al., 2016; McDonnell, 2020; Ramsey et al., 2018). Caregivers of males reported more concerns about RRBs than caregivers of females among toddlers (Ramsey et al., 2018), preschoolers (Hiller et al., 2016), and children and adolescents (Dillon et al., 2021). Caregivers also report greater concern about externalizing behavior in females and internalizing behavior in males during the preschool years (Hiller et al., 2016), and greater concern about social interactions for males diagnosed with autism, compared to females diagnosed with autism and children diagnosed with another developmental disability (Little et al., 2017). These findings of possible sex differences in caregiver concerns may inform improvements to caregiver-report screeners to close the gap between age of first concerns and autism diagnosis for females.

Males tend to score higher (indicative of more autism characteristics) than females on autism screening tools in the general population (ASQ; Auyeung et al., 2008; Baron-Cohen et al., 2001; Baron Cohen et al., 2006; CAST; Williams et al., 2008; SRS; Constantino & Todd, 2003). A similar pattern emerges for *toddler*-specific screeners in the general population, including the Qualitative-Checklist for Autism in Toddlers (Allison et al., 2008) and the M-CHAT-R/F (Øien et al., 2017). In contrast, among autistic children, no sex differences have emerged in the M-CHAT-R/F total score (Øien et al., 2017; Ros-Demarize et al., 2020) or likelihood of endorsement of most items on the M-CHAT among autistic children, although caregivers endorsed “difficulty with imitation” more often in females compared to males and “difficulty following a point” more often in males than in females (Øien et al., 2017).

Considering findings suggesting possible sex differences in the early clinical presentation of autism, combined with evidence of later autism diagnosis for females, there is an urgent need to systematically explore sex differences in early detection tools for autism. The present study explored sex differences in the performance of the M-CHAT-R/F in a large sample of toddlers screened during well-child primary care visits. We predicted that males would be more likely than females to screen at HL of autism on initial and follow-up screening, and that sensitivity and positive predictive value of the M-CHAT-R/F would be stronger in males compared to females. Exploratory analyses examined whether sex-specific cut-off scores would improve the utility of the M-CHAT-R/F for females.

## Methods

### Participants

Participants included 28,088 toddlers (14,331 males, 13,757 females), aged 14.07–30.98 months, who participated in one of four studies examining early detection of autism between 2009 and 2020, across four universities located in Philadelphia, Pennsylvania; Storrs, Connecticut; Atlanta, Georgia; and Sacramento, California. Toddlers classified as HL of autism on the M-CHAT-R/F, another study-specific screener, or whose primary care clinician (PCC) indicated concern for autism were invited for a no-cost evaluation at their university site ( $n = 1,112$ ; 728 male, 384 female); 731 toddlers (65.7%; 488 male, 243 female) attended a diagnostic evaluation (see Table 1). Of those evaluated, 677 (92.6%; 451 male, 226 female) were HL based on initial screen and/or surveillance. Eight children (4 male, 4 female) were missing information about reason for evaluation.

**Inclusion and Exclusion Criteria**—For inclusion in parent studies, participants were screened using the English or Spanish M-CHAT-R/F (Robins et al., 2009) during at least one primary care well-child visit prior to 31 months of age. When children had more than one M-CHAT-R/F, only initial screen data were included. Exclusion criteria for this study included (a) missing information on child's sex ( $n = 398$ ), (b) three or more items missing on the M-CHAT-R/F, which makes classification of score indeterminate ( $n = 99$ ), and (c) completed diagnostic evaluation prior to initial M-CHAT-R/F screen ( $n = 63$ ). Missing data on the M-CHAT-R/F were not recoded. In addition, evaluation data were excluded (but screening data were kept) for 786 participants whose final outcome was undetermined: (a) screen positive on the M-CHAT-R/F but family did not attend evaluation, or (b) evaluation was invalid, due to severe motor or sensory impairment, or low compliance. Thirteen children classified as medium-likelihood (ML) who did not receive Follow-Up but attended evaluation were excluded from analyses specifically examining Follow-Up scores, but were included in all other analyses with their initial score counting as their final score.

**Description of Studies**—Participants data were aggregated from four early detection of autism studies that screened toddlers with M-CHAT-R/F during well-child visits, and invited children at HL of autism for no-cost evaluation: 1) Early Detection of Pervasive Developmental Disorders (R01HD039961, 2009–2014), screened toddler during 18- and/or 24-month well-child visits, validated the M-CHAT-R/F, and demonstrated improved utility compared to the original M-CHAT (Robins et al., 2014). 2) Validation of Web-Based Administration of the M-CHAT-R/F (Autism Speaks #8368, 2012–2016) screened using electronic delivery at 18- and/or 24-month visits; Follow-Up was administered in the same electronic session (54%) or by phone with a trained member of the research staff, validating electronic screening (46%; Attar et al., in press). 3) Early Detection of Autism (R01HD039961, 2014–2020) randomized pediatric practices to one of three screening schedules, beginning at 12-, 15-, or 18-months and encouraged rescreening at 18, 24, and 36 months. Caregivers completed the M-CHAT-R/F electronically or on paper during all visits except 12-months, demonstrating that earlier and repeated screening detects autism (Wieckowski et al., 2021). 4) Connecting the Dots (R01MH115715, 2017–2022), screened electronically during well-child visits at 18 and 24 months to relate primary care detection

to outcomes at age 5 (McClure et al., 2021). See supplemental tables 1 and 2 for study/site subgroups.

## Measures

### Screening

**Demographic Information.** Caregivers of screened toddlers reported demographics after enrolling: child's age, sex, race, ethnicity, and maternal education (as a measure of socioeconomic status).

**Modified Checklist for Autism in Toddlers, Revised, with Follow-up (M-CHAT-R/F; Robins et al., 2009).** The M-CHAT-R/F is a two-part autism screener validated in 16- to-30-month old children (Robins et al., 2014). The initial screen is a 20-item caregiver questionnaire assessing communication, joint attention, and pretend play. Toddlers who initially scored a 0–2 were classified as low-likelihood (LL). Caregivers of toddlers who scored 3–7 (medium-likelihood, ML) completed structured Follow-Up, and had a final score of 2 or higher was classified as high-likelihood (HL) of autism and 0–1 as LL. Toddlers who scored an 8 or above on initial M-CHAT-R were classified as HL of autism

**Clinician Concern.** PCCs were asked to indicate concern for autism at each screening visit.

**Diagnostic Evaluation Measures**—Across all studies, the clinical best estimate diagnosis incorporated data from the Toddler Autism Symptom Interview (TASI; Barton et al., 2012; Coulter et al., 2021) or Autism Diagnostic Interview, Revised (ADI-R; Rutter et al., 2003), Autism Diagnostic Observation Schedule (ADOS), original or 2<sup>nd</sup> Edition (Lord et al., 1999; Lord et al., 2012), Mullen Scales of Early Learning (Mullen, 1995), Vineland Adaptive Behavior Scales, 2<sup>nd</sup> or 3<sup>rd</sup> Edition (Sparrow et al., 2005; Sparrow et al., 2016) and a Medical, Developmental, and Family History form.

## Procedure

Screening for all studies occurred in pediatric practices during well-child visits using electronic and/or paper M-CHAT-R/F. For paper forms, caregivers of children with ML scores were contacted by research staff to complete Follow-Up over the phone. For electronic screeners, Follow-Up questions displayed immediately after completion of the initial screener. Electronic screens were automatically scored, and PCCs were able to view the results. Caregivers were not given screening results directly for any of the studies. In study 1, all screen positive participants were eligible for Follow-Up and were reclassified as LL or HL based on the results; the ML threshold was established based on this sample. In studies 2, 3, and 4, only children in ML range received the Follow-Up. For the purpose of analyses in this study, Follow-Up data were only included for children who should have received Follow-Up based on the current protocol. Children were re-screened at later ages depending on original study procedures; although only data from first screens were included in the current study, final diagnosis of children identified with autism after rescreening was used to identify false negative cases. Screeners were offered in both English and Spanish at the Connecticut, Sacramento, and Philadelphia sites due to capacity for clinical evaluations in Spanish; Atlanta was an English-only site. Caregivers waived documentation

of consent at screening and provided written, informed consent at evaluations. Toddlers classified as HL of autism based on M-CHAT-R/F, other study-specific screeners, and/or whose PCC reported an autism concern were invited for no-cost diagnostic evaluations. Average time between HL of autism classification and diagnostic evaluation was 4.40 months ( $SD = 4.43$  months). Evaluations occurred at university clinics or pediatric offices; teams included a licensed psychologist, certified school psychologist, or developmental pediatrician and a trainee. Each member of assessment teams was research-reliable on all measures they administered. Clinical best-estimate diagnosis was based on *International Classification of Diseases, 10th edition* (ICD-10; World Health Organization, 2004), or *Diagnostic and Statistical Manual of Mental Disorders, 4th edition, text revision* (DSM-IV-TR; APA, 2000) or *5th edition* (DSM-5; APA, 2013) criteria. Diagnoses of Autistic Disorder, Pervasive Developmental Disorder, Not Otherwise Specified, Atypical Autism, Childhood Autism, and Asperger's Syndrome were grouped into an autism classification. When autism was ruled out, other developmental disorders (DD) were considered, including Global Developmental Delay or Language Delay, or the child was determined to have no diagnosis (ND). Caregivers received oral and written feedback about diagnoses including recommendations and information about local intervention resources. When caregivers declined to complete the M-CHAT-R Follow-Up or evaluation, the PCC was informed. Institutional review boards approved each of the four studies; three studies had approval at each university and Connecting the Dots used a single IRB on which other universities relied. PCCs and caregiver advocates contributed to study design, implementation, and dissemination of findings.

### Statistical Analyses

To evaluate difference in the distribution of autism likelihood classifications between males and females,  $2$  (sex)  $\times$   $3$  (likelihood level) chi-square analyses were run, followed by pairwise comparisons of autism likelihood level through  $2 \times 2$  chi-square analyses, providing the appropriate effect size ( $V$  or  $\phi$ ). Similarly, two additional chi-square analyses were run to explore the relationship between sex and likelihood classification of Follow-Up scores and evaluation attendance among those who screened positive. To evaluate the association of autism diagnosis and sex with final scores on the M-CHAT-R/F among those who screened as HL of autism, a two-way Analysis of variance (ANOVA) was run, with sex and diagnosis (ASD, non-ASD) as independent variables. Exploratory  $2$  (sex)  $\times$   $2$  (item endorsement) chi-square analyses were run to examine potential item-level sex differences among diagnosed children. For the M-CHAT-R/F psychometric properties, sensitivity (i.e., detecting autism when truly present) was calculated by dividing the number of true positive cases (TP; i.e., positive screen and received autism diagnosis) by the total number of children diagnosed with autism. In the study samples pooled for this analysis, not all children who screened positive received a diagnostic evaluation. The sensitivity value reported in this paper is computed based on the subset of screen positive children who attended the evaluation, and as such, it should be considered an estimate and interpreted with caution. Positive predictive value for autism ( $PPV_{\text{autism}}$ ; likelihood that positive result is a true autism case) was calculated by dividing TP for autism by all screen positives, whereas PPV for any developmental disability ( $PPV_{\text{DD}}$ ; likelihood that positive result indicates autism or another DD) was calculated by dividing TP for autism or DD diagnosis by all screen positives.

TP and false positive (FP) cases were determined by M-CHAT-R/F results, whereas false negative (FN) cases were detected by PCC concern, positive M-CHAT-R/F rescreen, or other positive screener result. Due to the lack of confirmatory evaluations among the whole sample to confirm true negatives (TN), specificity and negative predictive value were not included. Chi-square analyses were run to compare components of sensitivity (TP to FN), initial and final PPV<sub>autism</sub> (TP for autism to FP) and final PPV<sub>DD</sub> (TP for autism or DD to FP) by sex. Due to the strong influence of the Follow-Up interview, PPV was examined both for initial and final scores. For these calculations, children classified as LL (not evaluated) were presumed not to have autism. The two proportion Z-test (Lowry, n.d.) was examined to compare the improvement of PPV from initial to final scores in males and females. Exploratory ROC analyses were conducted to investigate different cutoff scores at initial and final M-CHAT-R/F screening for males and females.

## Results

### Autism Likelihood Classification from M-CHAT-R/F

**Initial Score Classification and Item-Level**—The proportion of males and females classified in the low (LL), medium (ML), and high (HL) range on the initial M-CHAT-R significantly differed ( $X^2(2, 28,088) = 87.28, p < .001, V = .056$ ; see Table 2). Pairwise comparisons indicated that on initial M-CHAT-R, females were less likely than males to be classified as HL vs. LL (OR = .48, 95% CI [.37, .61];  $X^2(1, 26212) = 37.98, p < .001, \phi = -.038$ ), HL vs. ML (OR = .67, 95% CI [.52, .87];  $X^2(1, 2182) = 9.14, p = .003, \phi = -.065$ ), and ML vs. LL (OR = .71, 95% CI [.64, .78];  $X^2(1, 27782) = 51.69, p < .001, \phi = -.043$ ). Notably, in the subsample diagnosed with autism, there was no significant difference between males and females in the proportion of HL vs. ML classification on the initial M-CHAT-R ( $X^2(1, 272) = .20, p = .658, \phi = -.027$ ) or on any of the exploratory item-level analyses ( $ps > .05$ ). Effect sizes of analyses were small.

**Final Score Among Those Who Completed Follow-Up**—Among children in the ML category who completed Follow-Up, a significant relationship between sex and Follow-Up outcome was found (OR = .72, 95% CI [.59, .88];  $X^2(1, 1610) = 10.23, p = .001, \phi = -.08$ ); females were less likely than males to be classified as HL of autism. Two-way ANOVA examined the association of autism diagnosis and sex among those who screened at HL (final) of autism. Final scores significantly differed ( $F(1, 727) = 30.56, p < .001, \eta_p^2 = .040$ ) between children evaluated who received an autism diagnosis ( $M = 5.30, SD = 4.15$ ) and those who did not ( $M = 3.70, SD = 3.30$ ), but there were no significant differences between males and females on final score ( $p = .922$ ), or interaction between sex and autism diagnosis ( $p = .538$ ). Effect sizes of analyses were small.

### Psychometric Properties, Autism Rates, and Evaluation Attendance

Overall estimated sensitivity of M-CHAT-R/F was strong, but there was no relationship between sensitivity and sex ( $X^2(1, 319) = .11, p = .743, \phi = -.018$ ); see Table 3. On both the initial M-CHAT-R (TP<sub>male</sub> = 22.5%, TP<sub>female</sub> = 12.8%; FP<sub>male</sub> = 77.5%, FP<sub>female</sub> = 87.2%; OR = .50, 95% CI [.38, .67];  $X^2(1, 1468) = 22.19, p < .001, \phi = -.123$ ) and final M-CHAT-R/F (TP<sub>male</sub> = 51.2%, TP<sub>female</sub> = 37.4%; FP<sub>male</sub> = 48.8%, FP<sub>female</sub> = 62.6%; OR =



.57, 95% CI [.40, .82],  $X^2(1, 555) = 9.42, p = .002, \phi = -.13$ ) scores, females were less likely than males to have a TP versus a FP score. Notably, PPV significantly improved by a factor of 2.92 times for females ( $z = 7.46, p < .0001$ ), 2.28 times for males ( $z = 10.01, p < .0001$ ), and 2.51 times overall ( $z = 12.80, p < .0001$ ) after Follow-Up compared to after initial score. When the final M-CHAT-R/F was examined in a combined autism and DD diagnosis group, females were less likely than males to be classified as a TP versus a FP ( $TP_{\text{male}} = 89.5\%$ ,  $TP_{\text{female}} = 79.7\%$ ;  $FP_{\text{male}} = 10.5\%$ ,  $FP_{\text{female}} = 20.3\%$ ;  $OR = .46$ , 95% CI [.28, .75];  $X^2(1, 555) = 12.13, p < .001, \phi = -.135$ ).

Autism rates were examined by combining TP and FN cases, compared to the total sample. The rate of autism was 1.63% for males and 0.62% for females; with approximately 2.8 males receiving an autism diagnosis for every female. Among children classified as HL, there was no significant association between males and females on evaluation attendance ( $X^2(1, 1112) = 1.57, p = .21, \phi = .038$ ). Across all analyses, effect sizes were small.

### Exploratory ROC Analysis

Due to the high rates of FP for females in our sample, an exploratory receiver operating curve (ROC) analysis was conducted to investigate whether optimal cut-off scores at initial and final M-CHAT-R/F screening differ for males and females. Area under the curve (AUC) metrics showed strong performance across cut-off scores at both initial and final screening (see Table 4). The sample showed optimal performance for both sensitivity and specificity at a score of  $\geq 2.5$  at initial screen and at a score of  $\geq 1.5$  at final screen in both the overall sample as well as separately for both males and females (see Table 5). Our findings are consistent with the M-CHAT-R/F authors' recommended cut-off scores of 3 at initial screen and 2 at final screen for optimal utility both for males and females. Although the ROC analyses conducted above included the original M-CHAT-R/F validation sample, excluding the sample resulted in consistent findings, so the larger sample was maintained to maximize power.

### Discussion

The goal of this study was to examine sex differences in toddler screening for autism using the M-CHAT-R/F in unselected community samples. Overall, the M-CHAT-R/F demonstrates good psychometric properties across sex; however, sex differences emerged.

#### Psychometric Properties of the M-CHAT-R/F by Sex

As expected, females were less likely than males to be classified as HL of autism at all stages of the screening process using the M-CHAT-R/F; however, the accuracy of a HL screen result in predicting an autism diagnosis differed based on sex. This is reflected in lower PPVs (i.e., a higher proportion of FP to TP cases) for females compared to males. Positively, the addition of the Follow-Up significantly improved the PPV for both sexes (i.e., over two-fold for males and approximately three-fold for females). As such, findings support the utility of the 2-stage M-CHAT-R/F, both for females and males.

Despite a higher number of FP screens in females compared to males, it is encouraging that there were no sex differences in the estimated sensitivity of the tool. This result

provides support that the M-CHAT-R/F readily identifies both males and females who have autism who are detectable during toddlerhood, with the caveat that not all children received confirmatory evaluations; children with LL of autism based on multiple screens and clinician surveillance were not evaluated and were presumed to be TN. However, we also interpret estimates of sensitivity among toddlers with caution, understanding that not all children with autism will be detected at this young age (Robins, 2020). The lower rate of HL classification among females may be due to lower prevalence of autism in females. This finding was strengthened by the fact that sex differences did not emerge across HL- or ML-levels on initial total scores among those who were ultimately diagnosed with autism. Additionally, exploratory ROC analysis supported recommended cut-off scores for the M-CHAT-R/F for both males and females. Similar to previous research examining sex differences in other autism-detecting measures, our exploratory ROC analysis supports the current recommended risk cut-off scores for the M-CHAT-R/F both for males and females (Kaat et al., 2021; Kalb et al., 2022). In developing neurodevelopmental screening measures, it is important to optimize sensitivity considering the negative consequences of not identifying a child with the disorder. Delayed diagnosis for an autistic child can impede access to autism-specific interventions at a critical time during brain development. Thus, despite a higher number of false positive screens for females compared to males, strong sensitivity across sex should be recognized as evidence that the M-CHAT-R/F readily identifies both autistic males and females.

In this sample, sex differences remained when considering risk for all developmental disabilities detected by M-CHAT-R/F. PPV of any DD was significantly higher for males compared to females. Nonetheless, clinicians can be assured that most (i.e., 79.7%) females, and males (89.5%) classified as HL of autism on the M-CHAT-R/F are diagnosed with autism or another developmental disorder.

It is unclear why there may be a higher number of false positive screens on the M-CHAT-R/F for females compared to males. In general, PPV of a test increases as base rate increases, and vice versa. As such, lower PPV for females compared to males may be due to the higher prevalence of autism (more TP cases proportional to FP cases) in males compared to females, although we found fewer positive results overall (both TP and FP) for females compared to males. As there was no difference in estimated sensitivity of the M-CHAT-R/F between males and females, it is possible that there are fewer autistic females but that the M-CHAT-R/F is still successfully catching those at high-likelihood. Similarly, there is a higher prevalence of other neurodevelopmental disorders for males compared to females. For example, the male to female ratio is 2:1 for intellectual disability (Roepers, 2008), 1.5 to 1 for language delay (Shriberg et al., 1999), and 3 to 1 for attention deficit hyperactivity disorder (Willcutt, 2012). As such, differential sex prevalence may also explain lower PPV for females compared to males when considering risk for neurodevelopmental conditions more broadly. Furthermore, it is crucial to keep in mind that the M-CHAT-R/F, as well as all other autism screening and diagnostic tools, were developed based primarily on the male autism phenotype given higher prevalence of autism in males compared to females. As a result, there is currently debate within the field about whether current diagnostic tools are less sensitive to the female phenotype compared to the male phenotype (D’Mello et al., 2022). It is therefore possible that the greater FP rate may be an artefact of missing

females who truly are autistic. Reliable and valid biomarkers demonstrating equivalence of underlying autism independent of behavioral diagnostic systems will allow examination of whether M-CHAT-R/F and other tools must be adjusted to increase accuracy in females.

The male-to-female autism prevalence ratio was 2.8 in this sample, which is somewhat lower than the current estimate of 4.2 in 8-year-old children but is more similar to the 3.4 ratio in 4-year-old children derived by CDC autism surveillance (Maenner et al., 2021; Shaw et al., 2021). It is important to note that children identified with autism early in life may be more cognitively impacted than those identified later (Christensen et al., 2019) and autistic females are more likely to be diagnosed with co-occurring intellectual disability (Fombonne, 2003; Loomes et al., 2017). As such, it is likely that our sample of children includes more children who will eventually be diagnosed with intellectual disability than the general population of autistic individuals across the lifespan.

### **M-CHAT-R/F Total Score by Sex**

Differences in final scores were also examined among children classified as HL of autism in order to explore sex differences in continuous M-CHAT-R score. It is important to note that the M-CHAT-R was not initially developed to be used in this manner; however, total score has been used across many autism screening and diagnostic tools as a continuous measure of autism characteristics, and there was a significant difference between scores of children with and without autism highlighting the M-CHAT-R/F's construct validity as an autism screener. Consistent with previous research, there were no sex differences in total score among autistic children; however, contrary to Øien and colleagues (2017), there were also no sex differences in total score among those without autism. Methodological differences between these studies may have contributed to differences. Specifically, Øien and colleagues used the original M-CHAT and did not conduct the Follow-Up. In the present study, the M-CHAT-R/F was used and final scores after Follow-Up were examined; adding this critical second stage of screening may effectively reduce potential sex differences. Additionally, Øien and colleagues examined scores for all children, not only those deemed at HL for autism. In the present study, we examined total scores only for children deemed at HL of autism because the variance in scores for children not at HL of autism was much lower than for children classified as HL of autism due to the vast majority of children scoring 0–1, which violates statistical assumptions of ANOVA. The lack of sex differences among children classified as HL of autism based on the M-CHAT-R/F in this sample is consistent with recent literature examining total scores in a group of children referred for autism evaluation due to clinical concerns about autism (Ros-Demarize et al., 2020). Taken together, these data support that total score does not differ across males and females among children at HL of autism.

### **Limitations and Future Directions**

The findings from the current study must be interpreted within the context of several limitations. Research has demonstrated more consistent autism sex differences later in development, and this cross-sectional design does not follow children as they age. In addition, it is likely that some children who do not meet clinically significant criteria in toddlerhood will do so later in development (Ozonoff, 2018); evaluation later in childhood

is expected to lead to identification of more autistic children, consistent with recent record review studies after children are four years or older (Carbone et al., 2020; Guthrie et al., 2019). It is also possible that if presentation of autism differs in females versus males, the M-CHAT-R/F, a tool primarily validated by detecting autism in toddler boys, may be mistuned and missing girls who would otherwise be referred for evaluations.

The sensitivity of our sample should be interpreted as an estimation rather than a true value, as FNs were identified as missed cases through concurrent clinician surveillance, later positive toddler or preschool M-CHAT-R/F rescreen, or other positive toddler screen, rather than confirming TN classification in every LL child. While we acknowledge this limitation, we also note that CDC estimates of autism prevalence show a decrease from 4-year-olds compared to 8-year-olds (Shaw et al., 2021). Therefore, it is possible the decreased rate of autism detected in this study is due to the even younger sample of primarily 18–24-month-olds, who may show lower prevalence estimates than are reported in older children. Similarly, it is important to note that not all children who were identified as HL for autism were evaluated. Evaluation non-attendance did not significantly differ by sex and was not due to lack of evaluation availability, but to parents declining to attend.

While exploratory examination of M-CHAT-R/F item-level sex differences based on chi-square analyses did not emerge, future research should use strategies such as differential item functioning or sex-specific algorithms or item weighting that may maximize psychometric properties based on sex, to consider M-CHAT-R/F equivalence or lack thereof across males and females. Future research also should explore clinician and caregiver perceptions of the M-CHAT-R/F for boys and girls and whether the higher FP rate in females impacts referral practices.

In order to obtain a large sample of females with autism, four studies, including the sample used to validate the M-CHAT-R/F, were aggregated for this study, and study-specific differences in FN detection may impact results. Finally, small effect sizes should be interpreted with caution when contextualizing the results given the relatively large sample size.

## Conclusion

Our study found that autistic males and females screened positive on M-CHAT-R/F at similar rates, highlighting similar sensitivity across males and females as a strength of the tool. We found no indication that the detection of autism in females would benefit from different risk score thresholds than the current guidelines recommended for the M-CHAT-R/F. Even in early childhood when there is less expected differentiation of autism symptomology, however, females are still less likely to be classified as HL of autism and referred for evaluations compared to males. Our results highlight a need to improve the screening process in females, as even when females are referred they are more likely to be FP for autism than TP, although they were still likely to have other DDs and benefit from detection. The higher FP rate in females compared to males may influence clinician confidence in referring families at this crucial earlier age. Future research should focus on examining methods to decrease FPs in the screening process for females.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank the healthcare clinicians, toddlers, and their families for participating in this study; as well as the many individuals involved in data collection. In addition, we would like to thank the late Lauren Adamson, Ph.D. for her contributions to this work.

The study was supported by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development grant R01HD039961, Autism Speaks grant 8368, NICHD grant P50HD103526, and the National Institute of Mental Health, R01MH115715.

## References

- Allison C, Baron-Cohen S, Wheelwright S, Charman T, Richler J, Pasco G, & Brayne C (2008). The Q-CHAT (Quantitative Checklist for Autism in Toddlers): A normally distributed quantitative measure of autistic traits at 18–24 months of age: preliminary report. *Journal of Autism and Developmental Disorders*, 38(8), 1414–1425. 10.1007/s10803-007-0509-7 [PubMed: 18240013]
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. Washington, DC: American Psychiatric Association.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. Washington, D.C: American Psychiatric Association.
- Anderson DK, Liang JW, & Lord C (2014). Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, 55(5), 485–494. 10.1111/jcpp.12178 [PubMed: 24313878]
- Attar S, Bradstreet LE, Ramsey RL, Kelly K, & Robins DL (in press). Validation of the electronic Modified Checklist for Autism in Toddlers, Revised, with Follow-Up: A nonrandomized controlled trial. *The Journal of Pediatrics*
- Auyeung B, Baron-Cohen S, Wheelwright S, & Allison C (2008). The Autism Spectrum Quotient: Children's Version (AQ-Child). *Journal of Autism and Developmental Disorders*, 38(7), 1230–1240. 10.1007/s10803-007-0504-z [PubMed: 18064550]
- Baron-Cohen S, Hoekstra RA, Knickmeyer R, & Wheelwright S (2006). The Autism-Spectrum Quotient (AQ)—Adolescent Version. *Journal of Autism and Developmental Disorders*, 36(3), 343. 10.1007/s10803-006-0073-6 [PubMed: 16552625]
- Baron-Cohen S, Wheelwright S, Skinner R, Martin J, & Clubley E (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5–17. 10.1023/A:1005653411471 [PubMed: 11439754]
- Barton ML, Boorstein H, Dumont-Mathieu T, Herlihy LE, & Fein D (2012). Toddler autism Symptom Inventory (TASI).
- Begeer S, Mandell D, Wijnker-Holmes B, Venderbosch S, Rem D, Stekelenburg F, & Koot HM (2013). Sex differences in the timing of identification among children and adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 43(5), 1151–1156. 10.1007/s10803-012-1656-z [PubMed: 23001766]
- Carbone PS, Campbell K, Wilkes J, Stoddard GJ, Huynh K, Young PC, & Gabrielsen TP (2020). Primary care autism screening and later autism diagnosis. *Pediatrics*, 146(2), E20192314. 10.1542/peds.2019-2314 [PubMed: 32632024]
- Chlebowski C, Robins DL, Barton M, Fein DA (2012, May 17-19). The Modified Checklist for Autism in Toddlers: A follow up study investigating the early detection of autism spectrum disorders in a low risk sample [Poster presentation]. International Meeting for Autism Research, Toronto, Ontario, Canada. <https://imfar.confex.com/imfar/2012/webprogram/Paper9999.html>.
- Christensen DL, Maenner MJ, Bilder D, Constantino JN, Daniels J, Durkin MS, ... & Dietz P (2019). Prevalence and characteristics of autism spectrum disorder among children aged 4 years—early

- autism and developmental disabilities monitoring network, seven sites, United States, 2010, 2012, and 2014. *MMWR Surveillance Summaries*, 68(2), 1.
- Constantino JN, & Todd RD (2003). Autistic traits in the general population: A twin study. *Archives of General Psychiatry*, 60(5), 524–530. 10.1001/archpsyc.60.5.524 [PubMed: 12742874]
- Coulter KL, Barton ML, Boorstein H, Cordeaux C, Dumont-Mathieu T, Haisley L, Herlihy L, Jashar DT, Robins DL, Stone WL, Fein DA (2021). The Toddler Autism Symptom Inventory: Use in diagnostic evaluations of toddlers. *Autism*, 1–14.
- D’Mello AM, Frosch IR, Li CE, Cardinaux AL, & Gabrieli JDE (2022). Exclusion of females in autism research: Empirical evidence for a “leaky” recruitment-to-research pipeline. *Autism Research: Official Journal of the International Society for Autism Research*, 15(10), 1929–1940. 10.1002/aur.2795 [PubMed: 36054081]
- Dillon EF, Kanne S, Landa RJ, Annett R, Bernier R, Bradley C, Carpenter L, Kim SH, Parish-Morris J, Schultz R, Wodka EL, Wodka EL, & the SPARK Consortium. (2021). Sex differences in autism: Examining intrinsic and extrinsic factors in children and adolescents enrolled in a national ASD cohort. *Journal of Autism and Developmental Disorders*. Advance online publication. 10.1007/s10803-021-05385-y
- Fombonne E (2003). Epidemiological surveys of autism and other pervasive developmental disorders: an update. *Journal of Autism and Developmental Disorders*, 33(4), 365–382. [PubMed: 12959416]
- Fuller EA, Oliver K, Vejnosa SF, & Rogers SJ (2020). The effects of the Early Start Denver Model for children with autism spectrum disorder: a meta-analysis. *Brain Sciences*, 10(6), 368. [PubMed: 32545615]
- Fuller EA, & Kaiser AP (2020). The effects of early intervention on social communication outcomes for children with autism spectrum disorder: A meta-analysis. *Journal of Autism and Developmental Disorders*, 50(5), 1683–1700. [PubMed: 30805766]
- Giarelli E, Wiggins LD, Rice CE, Levy SE, Kirby RS, Pinto-Martin J, & Mandell D (2010). Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. *Disability and Health Journal*, 3(2), 107–116. 10.1016/j.dhjo.2009.07.001 [PubMed: 21122776]
- Guthrie W, Wallis K, Bennett A, Brooks E, Dudley J, Gerdes M, Pandey J, Levy SE, Schultz RT, Miller JS (2019). Accuracy of autism screening in a large pediatric network. *Pediatrics*, 144(4), e20183963. 10.1542/peds.2018-3963. [PubMed: 31562252]
- Hartley SL, & Sikora DM (2009). Sex differences in autism spectrum disorder: an examination of developmental functioning, autistic symptoms, and coexisting behavior problems in toddlers. *Journal of Autism and Developmental Disorders*, 39(12), 1715. 10.1007/s10803-009-0810-8 [PubMed: 19582563]
- Hiller RM, Young RL, & Weber N (2016). Sex differences in pre-diagnosis concerns for children later diagnosed with autism spectrum disorder. *Autism*, 20(1), 75–84. 10.1177/1362361314568899 [PubMed: 25717130]
- Kaat AJ, Shui AM, Ghods SS, Farmer CA, Esler AN, Thurm A, Georgiades S, Kanne SM, Lord C, Kim YS, & Bishop SL (2021). Sex differences in scores on standardized measures of autism symptoms: a multisite integrative data analysis. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 62(1), 97–106. 10.1111/jcpp.13242 [PubMed: 32314393]
- Kalb LG, Singh V, Hong JS, Hologue C, Ludwig NN, Pfeiffer D, Reetzke R, Gross AL, & Landa R (2022). Analysis of race and sex bias in the autism diagnostic observation schedule (ADOS-2). *JAMA Network Open*, 5(4), e229498. 10.1001/jamanetworkopen.2022.9498 [PubMed: 35471566]
- Landa RJ (2018). Efficacy of early interventions for infants and young children with, and at risk for, autism spectrum disorders. *International Review of Psychiatry*, 30(1), 25–39. 10.1080/09540261.2018.1432574 [PubMed: 29537331]
- Landa RJ, Gross AL, Stuart EA, & Faherty A (2013). Developmental trajectories in children with and without autism spectrum disorders: the first 3 years. *Child Development*, 84(2), 429–442. 10.1111/j.1467-8624.2012.01870.x [PubMed: 23110514]
- Lawson LP, Joshi R, Barbaro J, & Dissanayake C (2018). Gender differences during toddlerhood in autism spectrum disorder: A prospective community-based longitudinal follow-up study. *Journal of Autism and Developmental Disorders*, 48(8), 2619–2628. 10.1007/s10803-018-3516-y [PubMed: 29497988]

- Little LM, Wallisch A, Salley B, & Jamison R (2017). Do early caregiver concerns differ for girls with autism spectrum disorders?. *Autism*, 21(6), 728–732. 10.1177/1362361316664188 [PubMed: 27542396]
- Loomes R, Hull L, & Mandy WPL (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(6), 466–474. [PubMed: 28545751]
- Lord C, Rutter M, DiLavore P, & Risi S (1999). *Autism Diagnostic Observation Schedule (ADOS)*. Los Angeles: Western Psychological Services.
- Lord C, Rutter M, DiLavore P, Risi S, Gotham K, & Bishop SL (2012). *Autism Diagnostic Observation Schedule: ADOS-2*. Los Angeles: Western Psychological Services.
- Lowry R (n.d.). The Significance of the Difference Between Two Independent Proportions. VassarStats. Retrieved November 30, 2022 from [http://vassarstats.net/propdiff\\_ind.html](http://vassarstats.net/propdiff_ind.html).
- Maenner MJ, Shaw KA, Bakian AV, Bilder DA, Durkin MS, Esler A, Furnier SM, Hallas L, Hall-Lande J, Hudson A, Hughes MM, Patrick M, Pierce K, Poynter JN, Salinas A, Shenouda J, Vehorn A, Warren Z, Constantino JN, DiRienzo M, ... Cogswell ME (2021) Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2018. *MMWR Surveillance Summaries*, 70(11), 1–16. 10.15585/mmwr.ss7011a1
- McClure LA, Lee NL, Sand K, Vivanti G, Fein D, Stahmer A, & Robins DL (2021). Connecting the Dots: An RCT integrating standardized autism spectrum disorders screening, high-quality treatment, and long-term outcomes. *Trials*, 22, 319–331. 10.1186/s13063-021-05286-6 [PubMed: 33934719]
- McDonnell CG, DeLucia EA, Hayden EP, Penner M, Curcin K, Anagnostou E, Nicolson R, Kelley E, Georgiades S, Liu X, & Stevenson RA (2020). Sex differences in age of diagnosis and first concern among children with autism spectrum disorder. *Journal of Clinical Child & Adolescent Psychology*, 1–11. 10.1080/15374416.2020.1823850
- Mullen EM (1995). *Mullen Scales of Early Learning* (pp. 58–64). Circle Pines, MN: AGS.
- Mussey JL, Ginn NC, & Klinger LG (2017). Are males and females with autism spectrum disorder more similar than we thought?. *Autism*, 21(6), 733–737. 10.1177/1362361316682621 [PubMed: 28749236]
- Øien RA, Hart L, Schjølberg S, Wall CA, Kim ES, Nordahl-Hansen A, Eisemann MR, Chawarska K, Volkmar FR, & Shic F (2017). Parent-endorsed sex differences in toddlers with and without autism: Utilizing the M-CHAT. *Journal of Autism and Developmental Disorders*, 47(1), 126–134. 10.1007/s10803-016-2945-8 [PubMed: 27757737]
- Ozonoff S, Young GS, Brian J, Charman T, Shephard E, Solish A, & Zwaigenbaum L (2018). Diagnosis of autism spectrum disorder after age 5 in children evaluated longitudinally since infancy. *Journal of the American Academy of Child & Adolescent Psychiatry*, 57(11), 849–857. [PubMed: 30392626]
- Pierce K, Gazestani VH, Bacon E, Barnes CC, Cha D, Nalabolu S, Lopez L, Moore A, Pence-Stophaeros S, & Courchesne E (2019). Evaluation of the diagnostic stability of the early autism spectrum disorder phenotype in the general population starting at 12 months. *JAMA Pediatrics*, 173(6), 578–587. 10.1001/jamapediatrics.2019.0624 [PubMed: 31034004]
- Ramsey RK, Nichols L, Ludwig NN, Fein D, Adamson LB, & Robins DL (2018). Brief report: Sex differences in parental concerns for toddlers with autism risk. *Journal of Autism and Developmental Disorders*, 48(12), 4063–4069. 10.1007/s10803-018-3583-0 [PubMed: 29700708]
- Reinhardt VP, Wetherby AM, Schatschneider C, & Lord C (2015). Examination of sex differences in a large sample of young children with autism spectrum disorder and typical development. *Journal of Autism and Developmental Disorders*, 45(3), 697–706. 10.1007/s10803-014-2223-6 [PubMed: 25189824]
- Robins DL (2020). How do we determine the utility of screening tools? *Autism*, 24(2), 271–273. 10.1177/1362361319894170 [PubMed: 31856579]
- Robins DL, Casagrande K, Barton M, Chen CMA, 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. 10.1542/peds.2013-1813 [PubMed: 24366990]

- Robins DL, Fein D, & Barton M (2009). The Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M-CHAT-R/F). *Pediatrics*, 133, 37–45.
- Ropers HH (2008). Genetics of intellectual disability. *Current Opinion in Genetics & Development*, 18(3), 241–250. 10.1016/j.gde.2008.07.008 [PubMed: 18694825]
- Ros-Demarize R, Bradley C, Kanne SM, Warren Z, Boan A, Lajonchere C, Parkm J, & Carpenter LA (2020). autism symptoms in toddlers and preschoolers: An examination of sex differences. *Autism Research*, 13(1), 157–166. 10.1002/aur.2241 [PubMed: 31747131]
- Rutter M, Le Couteur A, & Lord C (2003). *The Autism Diagnostic Interview - Revised (ADI-R)*. Los Angeles, CA: Western Psychological Services.
- Shaw KA, Maenner MJ, Bakian AV, Bilder DA, Durkin MS, Furnier SM, Hughes MM, Patrick M, Pierce K, Salinas A, Shenouda J, Vehorn A, Warren Z, Zahorodny W, Constantino JN, DiRienzo M, Esler A, Fitzgerald RT, Grzybowski A, Hudson A, ... Cogswell ME (2021). Early identification of autism spectrum disorder among children aged 4 years - autism and developmental disabilities monitoring network, 11 Sites, United States, 2018. *MMWR Surveillance Summaries*, 70(10), 1–14. 10.15585/mmwr.ss7010a1
- Shriberg LD, Tomblin JB, & McSweeney JL (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language, and Hearing Research*, 42(6), 1461–1481. 10.1044/jslhr.4206.1461
- Sipes M, Matson JL, Worley JA, & Kozlowski AM (2011). Gender differences in symptoms of autism spectrum disorders in toddlers. *Research in Autism Spectrum Disorders*, 5(4), 1465–1470. 10.1016/j.rasd.2011.02.007
- Sparrow SS, Balla DA, & Cicchetti DV (2005). *Vineland II: Vineland Adaptive Behavior Scales*. American Guidance Service.
- Sparrow SS, Cicchetti DV, & Saulnier CA (2016). *Vineland Adaptive Behavior Scales-III*. Bloomington. MN: Pearson.
- Van Wijngaarden-Cremers PJ, van Eeten E, Groen WB, Van Deurzen PA, Oosterling IJ, & Van der Gaag RJ (2014). Gender and age differences in the core triad of impairments in autism spectrum disorders: A systematic review and meta-analysis. *Journal of Autism and Developmental Disorders*, 44(3), 627–635. 10.1007/s10803-013-1913-9 [PubMed: 23989936]
- Vivanti G, Dissanayake C, & The Victorian ASELCC Team (2016). Outcome for Children Receiving the Early Start Denver Model Before and After 48 Months. *Journal of Autism and Developmental Disorders*, 46(7), 2441–2449. 10.1007/s10803-016-2777-6 [PubMed: 27020055]
- Willcutt EG (2012). The prevalence of DSM-IV attention-deficit/hyperactivity disorder: A meta-analytic review. *Neurotherapeutics*, 9(3), 490–499. 10.1007/s13311-012-0135-8 [PubMed: 22976615]
- Wieckowski AT, Hamner T, Nanovic S, Porto K, Coulter K, Eldeeb S, Chen CA, Fein DA, Barton ML, Adamson LB, & Robins DL (2021). Early and repeated screening detects autism spectrum disorder. *Journal of Pediatrics*, 234, 227–235. NIHMS1683036. 10.1016/j.jpeds.2021.03.009 [PubMed: 33711288]
- Wieckowski AT, Williams LN, Rando J, Lyall K, & Robins DL (in press). Sensitivity and specificity of the Modified Checklist for Autism in Toddlers: A systematic review and meta-analysis. *JAMA Pediatrics*
- Williams JG, Allison C, Scott FJ, Bolton PF, Baron-Cohen S, Matthews FE, & Brayne C (2008). The childhood autism spectrum test (CAST): Sex differences. *Journal of Autism and Developmental Disorders*, 38(9), 1731–1739. 10.1007/s10803-008-0558-6 [PubMed: 18408991]
- World Health Organization. (2004). *The International Statistical Classification of Diseases and Health Related Problems ICD-10: Tenth Revision*. World Health Organization.
- Zwaigenbaum L, Bauman ML, Choueiri R, Kasari C, Carter A, Granpeesheh D, Mailloux Z, Smith Roley S, Wagner S, Fein D, Pierce K, Buie T, Davis PA, Newschaffer C, Robins D, Wetherby A, Stone WL, Yirmiya N, Estes A, Hansen RL, ... Natowicz MR (2015). Early intervention for children with Autism Spectrum Disorder under 3 years of age: Recommendations for practice and research. *Pediatrics*, 136 (Suppl 1), S60–S81. 10.1542/peds.2014-3667E [PubMed: 26430170]



**Table 1**

Demographic Characteristics for all Screened Toddlers and for Toddlers who Completed an Autism Evaluation

	Toddlers screened <i>N</i> = 28,088		Toddlers evaluated <i>N</i> = 731	
	Males screened <i>n</i> = 14,331	Females screened <i>n</i> = 13,757	Males evaluated <i>n</i> = 488	Females evaluated <i>n</i> = 243
Age at screening/evaluation ( <i>M</i> , <i>SD</i> )	20.18 (3.19)	20.18 (3.18)	24.06 (5.31)	23.49 (5.56)
Race ( <i>n</i> , %)				
White/Caucasian	8,027 (56.0)	7,612 (55.3)	218 (44.7)	102 (42.0)
Black/African American	3,182 (22.2)	3,157 (22.9)	122 (25.0)	84 (34.6)
Asian/Native Hawaiian/Pacific Islander	711 (5.0)	648 (4.7)	32 (6.6)	15 (6.2)
American Indian/Alaska Native	32 (.2)	38 (.3)	4 (.8)	3 (1.2)
Bi-/Multi-racial	965 (6.7)	950 (6.9)	46 (9.4)	19 (7.8)
Missing/Other	1,414 (9.9)	1,352 (9.8)	66 (13.5)	20 (8.2)
Ethnicity ( <i>n</i> , %)				
Hispanic/Latine	1,580 (11.0)	1,452 (10.6)	95 (19.2)	32 (13.2)
Non-Hispanic/Latine	9,240 (64.5)	8,998 (65.4)	302 (61.9)	169 (69.5)
Unknown	3,511 (24.5)	3,307 (24.0)	91 (18.6)	42 (17.3)
Maternal Education ( <i>n</i> , %)				
High school or below	5,185 (36.2)	4,824 (35.1)	217 (44.5)	109 (44.9)
Post-secondary	8,113 (56.6)	7,822 (56.9)	257 (52.7)	115 (47.3)
Missing/Other	1,033 (7.2)	1,111 (8.1)	14 (2.9)	19 (7.8)
Diagnosis				
Autism	-	-	234 (48.0)	85 (35.0)
DD <sup>a</sup>	-	-	173 (35.5)	92 (37.9)
No Diagnosis	-	-	81 (16.6)	66 (27.2)

<sup>a</sup>Other Developmental Disorders

**Table 2**

## Sex Distribution in Initial and Follow-Up Screening Likelihood Levels

Likelihood classification	Sex		Total <i>n</i> (%)
	Male (%/within)	Female (%/within)	
Low (0–2)	13,019 (90.8)	12,887 (93.7)	25,906 (92.2)
Medium (3–7)	1,104 (7.7)	772 (5.6)	1,876 (6.7)
Low Follow-Up	517 (54.5)	414 (62.5)	931 (57.8)
High Follow-Up	431 (45.5)	248 (37.5)	679 (42.2)
High (8–20)	208 (1.5)	98 (0.7)	306 (1.1)

Note. 266 cases were excluded for missing data in Follow-Up from the Medium-Likelihood group.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3**

Frequency of Classification and M-CHAT-R/F Psychometric Properties for Males and Females

Sex	Final Classification				Psychometric Properties			
	TP	FP	FN	TN	Sensitivity	Positive Predictive Value (PPV)		
						Initial Scores (autism only)	Final scores (autism only)	Final scores (autism+DD)
Male	191	182	43	13,433	.816	.225	.512	.895
Female	68	114	17	13,254	.800	.128	.374	.797
Overall	259	296	60	26,687	.812	.186	.467	.863

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4**

Psychometric Properties of Different Cutoff Scores for Males and Females

Cut-off Score	Psychometrics by Sex			
	Male		Female	
	Sensitivity	Specificity	Sensitivity	Specificity
Initial				
<b>3+ (optimal)</b>	<b>.842</b>	<b>.950</b>	<b>.882</b>	<b>.961</b>
4+	.748	.976	.753	.984
5+	.637	.986	.624	.991
Final				
<b>2+ (optimal)</b>	<b>.859</b>	<b>.892</b>	<b>.859</b>	<b>.904</b>
3+	.776	.900	.682	.994
4+	.602	.994	.518	.996

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 5**

Receiver Operating Characteristic Curve Analyses of Optimal Cutoff Score by Sex

Sex	ROC Properties					
	AUC	SE	95% CI	Sensitivity	Specificity	<i>p</i>
Initial score (optimal cutoff of 3+)						
Male	.935	.011	[.91, .96]	.842	.950	<.001
Female	.968	.010	[.95, .99]	.882	.961	<.001
Overall	.945	.008	[.93, .96]	.853	.956	<.001
Final score (optimal cutoff of 2+)						
Male	.928	.011	[.91, .95]	.859	.892	<.001
Female	.930	.019	[.89, .97]	.859	.904	<.001
Overall	.930	.009	[.91, .95]	.859	.898	<.001

Note. AUC = Area under the curve, SE = Standard error, CI = Confidence interval