Published in final edited form as:

J Autism Dev Disord. 2017 October ; 47(10): 3180–3188. doi:10.1007/s10803-017-3230-1.

Comparison of Autism Screening in Younger and Older Toddlers

Raymond Sturner, MD¹, Barbara Howard, MD¹, Paul Bergmann, MA², Lydia Stewart, MHS⁴, and Talin E. Afarian, BA³

¹Pediatrics, Johns Hopkins U. S. of Med., Balt., MD, United States

²Foresight Logic, Inc., St. Paul, MN

³Loyola U., Balt., MD., United States

⁴Johns Hopkins Bloomberg School of Public Health, Balt. MD, United States

Abstract

This study examined the effect of age at completion of an autism screening test on item failure rates contrasting older (>20 months) with younger (<20 months) toddlers in a community primary care sample of 73,564 children. Items related to social development were categorized into one of three age sets per criteria from Inada, Kamio & Koyama (2010). Younger toddlers produced higher rates of item failure than older toddlers and items in both of the later acquired item sets had higher probability rates for failure than the earliest acquired item set (prior to 8 months). Use of the same items and the same scoring throughout the target age range for autism screening may not be the best strategy for identifying the youngest toddlers at risk for autism.

Keywords

autism screening; M-CHAT; developmental screening; test development

Introduction

Autism Spectrum Disorder (ASD) is known to be impairing and prevalent (1 in 68 children in the U.S.) (Wingate et al., 2014). Since it has been established that early evidence-based interventions lead to improved long-term outcomes (National Research Council, 2001; Howlin, Magiati & Charman, 2009), early detection has become a national priority. The American Academy of Pediatrics (AAP) and the Center for Disease Control has recommended screening all children at 18 and 24 months (American Academy of Pediatrics,

Corresponding author: Raymond Sturner MD; rsturner@childhealthcare.org; phone: 410-353-3016; fax: 530-484-2347.

Conflict of Interest: "This study was conducted by the Center for Promotion of Child Development through Primary Care and its for-profit subsidiary, Total Child Health (TCH), Inc. CHADIS, the web-tool used in the study, was developed by Dr. Sturner and his spouse, Dr. Howard.

Dr Sturner is Director of the Center and Dr. Howard is President of TCH. Both are members of the Board of Directors of both entities and are paid consultants to both entities.

Compliance with Ethical Standards:

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

2006; CDC, 2015). The current most widely used ASD screen in primary care is the Modified Checklist for Autism in Toddlers (M-CHAT) (Robins, Fein, Barton & Green, 2001). The M-CHAT is among the tools recommended for ASD screening by the AAP (AAP, 2006) and major advocacy groups, such as Autism Speaks (Autism Speaks, 2015), and The National Autism Association (National Autism Association, 2016). However, a recent US Preventive Services Task Force report (2015) has suggested that universal screening at these ages should await more research on screening in asymptomatic toddlers in primary care pediatrics, in part, because of limitations in the demonstrated accuracy of the tools being recommended (USPSTF, 2015). The Task Force particularly noted that “there is insufficient evidence to determine if certain risk factors modify the performance characteristics of ASD screening tests, such as the age at which screening is performed” (USPSTF, 2015).

One limitation of current ASD screening research is the near absence of validation studies with representative samples of screen-negative children. A notable exception was a report of a large subsample of a Norwegian birth cohort recruited to complete the M-CHAT at 18 months and noting any ASD diagnosis identified with the help of both subsequent screening and family initiated concerns. This study revealed that the 18 month screen (without a follow-up interview) only identified about a third of the cases ultimately identified (Stenberg et al., 2014). In fact, a similar miss rate of 38% (Baird et al., 2000) was found for the original CHAT (Baron-Cohen et al., 1996), which was then modified as the M-CHAT for purposes of improving sensitivity at this age. Another cohort study of 18 month M-CHAT screening in Japan followed children through a three year diagnostic testing of those initially positive plus subsequent screen positives and community referrals. This study showed a higher sensitivity of 0.73 but used a lower than standard threshold for a positive M-CHAT screen requiring a failure rate of 17% of the population (Kamio et al., 2014). The low positive predictive value (PPV) of 0.12 of the M-CHAT was improved to 0.46 with a reduction in sensitivity to 0.48 through using the M-CHAT follow-up interview conducted 1 ½ to 2 months later by phone (Kamio et al., 2014). These studies together suggest that alternative scoring of the M-CHAT might be required at 18 months.

There is reason to believe that the M-CHAT may perform differently at 18 months of age than at 24 months. M-CHAT authors found lower PPV in younger low risk toddlers (0.28) compared to older ones (0.61) even using the recommended follow-up interview (M-CHAT/F) (Pandey et al., 2008) and recent studies have reported similar findings (Sturner et al., 2013). Another study by the M-CHAT authors suggested that younger toddlers may have a lower PPV, though that difference was not statistically significant (Robins et al., 2014). In that study, however, M-CHAT (and confirmatory interview, M-CHAT/F) was completed an average of two months after the standard 18-month visit. This two-month lag may be enough time to allow for improvement in potential issues of developmental immaturity that might be expected when the actual screening closely follows AAP recommendations for age of screening.

The AAP recommends a second autism screen at 24 months because of the well-documented phenomenon of developmental regression shown to occur roughly between 18 and 24 months. A meta-analysis of studies of regression estimated that 32% of ASD cases have a

history of such regression (Barger, Campbell & McDonough, 2013). Additionally, prospective studies suggest that ASD symptoms emerge gradually (Ozonoff, Heung, Byrd, Hansen & Hertz-Picciotto, 2008), which may result in a lower number of endorsed autism-specific items in the younger children. Conversely, since some autism-related behaviors represent delays in development, there may be a tendency for some M-CHAT items to be failed by younger children who later are developmentally typical. The characteristic odd and repetitive behaviors associated with autism may not appear until the end of the second year (Ozonoff et al., 2008), which should favor ease of identification in older toddlers. It would, therefore, not be surprising if screening strategies need to vary by age to capture the dynamic natural history of the disorder. While screening tests cannot be expected to identify children at an age before the disorder is actually manifest, there are other children who have the disorder but signs are less obvious. For these reasons it might be best to postpone screening until signs are clearer, which seems to occur at about two years. On the other hand, since the earlier the intervention, the better the outcome, children who can be identified six months earlier have an obvious benefit and this fact should stimulate efforts to carefully consider strategies to achieve this end. To the extent that it is possible to achieve the goal of the AAP of adapting tests to have acceptable accuracy suitable for universal administration at 18 months, it will be useful to understand if “performance characteristics of ASD screening tests are affected by age” at the item level.

The first aim of this study is to compare rates of M-CHAT item failures in younger toddlers (<20 months) to those of older toddlers (20+ months). The second aim of this study is to attempt to explain any age-related item performance. We hypothesized that an excessive failure rate in the younger toddlers may be explained by higher failure rates of more recently acquired social milestones. To address this question we applied the criteria, set forth by Inada et al. (2010), establishing age milestones for the 16 social development items of the M-CHAT based on data collection in a typically developing community population. The milestones were assigned to one of three age sets: 1) behaviors that emerge before 8 months, 2) behaviors that emerge at around 12 months, and 3) behaviors that emerge between 15 and 17 months. Specifically the three sets of items identified by Inada et al. were described as follows: Group 1: “The first set of behaviors which are related to self/other dyadic interaction emerge before 8 months of age (Enjoying being swung, Interest in other children, Enjoying peek-a-boo, Eye contact, Response to smile, and Response to name); Group 2: “... includes triadic interactions of self/object/other [that] emerge at around 12 months (Pretend play, Imperative pointing, Declarative pointing, Imitation of acts, Point following, and Attracting parent’s attention); Group 3: “...contains items related to complex triadic interactions [that] emerge between 15 and 17 months (Functional play, Bringing object to show, Gaze following, and Social reference).” Our study is unique in utilizing a large enough database to explore item performance over the entire 18 to 24 month interval in question.

Method

Participants

Item-level data from 73,564 M-CHAT screenings in primary care practices across the US completed using an online delivery system (CHADIS, 2016) were collected for children aged 16–30 months. Data was collected between September 2009 and February 2014. The screenings were split into two age groups: younger (<20 months; $n = 33,908$) and older (20+ months; $n=39,656$) toddlers using an age cut-off of 20 months, which was found in a previous study to be the point at which age became an important predictor variable to autism diagnosis (Sturner et al., 2013). Figure 1 presents the sample selection procedure.

Data came from 280 pediatric primary care offices located in every region of the United States but with predominance in the Mid-Atlantic and Northeast (67% based on zip code), and 33% from the Southeast, Midwest, and West Coast regions of the United States.

The sample was 48.5% female. The overall mean age of children was 21.6 months ($s.d.=3.6$). Mean ages for the younger and older age groups were 18.2 ($s.d.=0.7$) and 24.6 ($s.d.=2.1$), respectively. Figure 2 presents the distribution of ages. Additional demographic data was not available for individual children but 49% of the practices completed a survey describing their practice at the group level. Sixty-six percent of these practices described their sites as suburban, while 20% were urban and 14% rural. The average proportion of Medicaid participants within the U.S. practices was 25.3% with a range between 0 and 95%. Racial distribution of the sample was 54.4% white, 18.2% black, 9.3% Asian with 1.8% other races. The sample population was reported as 17% Hispanic. Tables 1, 2, and 3 describe the subjects by age, gender, and age groupings both with adjustment for prematurity and without.

ASD diagnostic data was unavailable for these children as the M-CHAT data was collected from routine clinical care without a protocol for referral and tracking.

The online delivery system includes an option for clinician-completed follow-up interviews (M-CHAT/F). However, those data are considered outside the scope of this investigation of online parent response to the M-CHAT/M-CHAT-R.

Measures

The Modified Checklist for Autism in Toddlers (M-CHAT): The 23-item Modified Checklist for Autism in Toddlers (M-CHAT; Robins et al., 2001) was used for this study. The M-CHAT is a parent-report checklist with items presented in a yes/no format. The M-CHAT was derived from the Checklist for Autism in Toddlers (CHAT) (Baron-Cohen, Allen, & Gliberg, 1992). The original validation sample included 1,293 children, 18–30 months of age, with 65 children receiving diagnostic evaluations. A discriminant function analysis of that sample assumed that non-risk children did not have autism and found that children could be best classified into the autistic or non-autistic groups based on the failure of any 2 items of the 6 items considered “critical”, or any 3 items overall.

A recently revised scoring algorithm follows an analysis of 16,071 screens. This new algorithm drops three items, removes designation of critical items, sets any 3 or more failed items as the threshold for requiring the follow-up interview conducted by a trained clinician (Robins et al., 2014), and includes rewording of some of the items. The new scoring was derived initially from cohorts using the 23-item version prior to the minor wording changes (Chlebowski, Robins, Barton & Fein, 2013; Robins et al., 2014). The 20-item version is now referred to as the M-CHAT-R. This study approximates the 20-item M-CHAT-R by extracting the 20 items from the M-CHAT dataset without implementing the rewording of any items.

Analyses

Overall differences in the expected number of failed items between younger and older age groups were investigated by fitting the data to negative binomial regression models controlling for M-CHAT screening result (on both the 20-item and 23-item algorithms) and gender. Negative binomial regression models were also used to investigate the effect of age on failed item counts among those who passed the M-CHAT screen (both 20-item and 23-item versions), controlling for gender. Truncated negative binomial models were used when modeling the subset of children who failed the M-CHAT screen in order to account for the absence of cases with zero failed items. Percent differences and marginal effects were then calculated.

In order to determine if gestational age of prematurely born children affected differences in failure rates, we repeated the analyses on a sample of children with adjusted age ≥ 16 months and < 31 months ($n=73,586$). Descriptive statistics (age, gender) and comparison of models fit to the adjusted and unadjusted samples showed no clinically meaningful nor statistically significant differences; therefore, adjusted ages were not considered further in the analyses.

Item level failure was modeled using logistic regression. Models were fit for the entire sample, for toddlers who failed the M-CHAT screen, and for toddlers who passed the M-CHAT screen, using both the 20-item and 23-item algorithms.

Results

Overall, the older 23-Item scoring algorithm generated 4,668 (6.3%) failed screens, which is significantly more than the 20-item scoring algorithm that generated 3,168 (4.3%) failed screens ($z = 17.15$, $p < 0.001$) on the same children.

Younger toddlers (< 20) produced higher rates of overall failure regardless of whether age was adjusted for prematurity or which M-CHAT scoring algorithm was used (see Table 4). Among those who passed the M-CHAT, regardless of the scoring algorithm, the average number of failed items for younger children was significantly higher than for older children. In contrast, among those who failed the M-CHAT, the average number of failed items for younger children was significantly lower (see Table 5).

In order to differentiate the effects of age, gender, and screening results on item performance, we evaluated percentage change in the expected number of failed items as the other factors were held constant. Holding gender and age group constant, having a positive MCHAT-23 screen increases the expected number of failed items by 1040% (Table 6).

Holding age group and MCHAT-23 screen result constant, being male increases the expected number of failed items by 6.8%. Holding gender and MCHAT-23 screen result constant, the expected number of failed items among younger children increases by 3.0%.

Among those who failed the MCHAT-23 screen, being male increases the expected number of failed items by 5.6% when controlling for age group. Interestingly, controlling for gender, the expected number of failed items among younger children who failed the MCHAT-23 screen *decreases* by 8.4%.

Among those who passed the MCHAT-23 screen, being male increases the expected number of failed items by 8.0% when holding age group constant. In contrast to those who failed the MCHAT-23, the expected number of failed items among younger children who passed the MCHAT-23 screen *increases* by 9.4% when controlling for gender.

The same pattern was found when the approximate emulation of the 20 item M-CHAT version was used (See Table 7).

The pattern of age differences lends support for the hypothesis that recently acquired social milestones are more likely to be associated with higher failure rates in younger toddlers for those items than in older toddlers. When items are sorted by probability of younger children failing the item, the earliest acquired items (group 1 items) as defined by Inada, et al (2010) all had lower probabilities for failure than any items in either of the other two sets acquired at a later age except one (item 17: “Does your child look at things you are looking at?”; see Table 8). However, there was no observed difference in probabilities of item failure between the two more recent item groups (groups 2 & 3) among the youngest children.

Discussion

M-CHAT overall failure rates are higher in younger toddlers compared to older ones. This is despite known higher rates of confirmed autism in the older group (Barger et al., 2013). While we can confirm that the revised scoring used in the 20-item version M-CHAT-R results in lower overall failure rates, failure rates are still higher in the younger age group.

Developmental tests for toddlers typically use age and prematurity-adjusted scoring. Autism screening tests presume that the phenotype is atypical of normal developmental progression for the entire validation age range of the test, but this assumption may not be appropriate for the younger age range of the test. Adjusting for the gestational age of prematurely born children may assure that the lower developmental age of the test is not violated. For these reasons we explored the effect of adjusting for prematurity on excessive item failure rates in the younger group and found that it did not affect the differences between age groups. This suggests that autism screening issues in the younger aged toddlers are not simply due to an excess of developmentally inappropriate children in the younger-aged subsample. While

premature children have been shown to have excessive rates of diagnosed autism (Mahoney, Minter, Burch & Stapel-Wax, 2013) these data tend to come from the most extremely low weight and premature children who may have neurological damage (Guy et al., 2014), not simply delays that are artifacts due to gestational age adjustments. There is evidence that the M-CHAT can be a useful screen for autism in these populations (Guy et al., 2014; Dudova et al., 2014) although these studies were completed with children at two years of age, obviating the issues of developmental age potentially making them ineligible for a test at the lower age limit (16 months chronological age) or confounding results.

The items previously shown to be acquired earliest (Inada et al., 2010) are not the ones showing higher failure rates in the younger toddlers. However, there was no difference between items in sets acquired between 12 and 14 months from those acquired between 15 and 17 months – both were associated with higher rates of young child failures. The item with the largest younger/older failure difference in our study was not one of the social milestones but the item about walking that is a parent report milestone attained by 75% at 14 months (Størvold, Aarethun & Bratberg, 2013). Using an age of acquisition for a milestone outside of two or even three standard deviations of the test age may result in a more desirable failure rate than more arbitrary cuts near the lowest 25% as for this M-CHAT/M-CHAT-R item and possibly others. It is not clear if the higher positive rates for more recently acquired milestones were due to parental misinterpretation of behaviors that are not as well established or simply due to wide distribution of milestone onset among typically developing children.

It is of interest that the item with the highest probability of passing in the younger age compared to the older toddler age group was item 11, the hyperacusis response to noise (“Does your child ever seem oversensitive to noise? (e.g. plugging ears).” It is not clear why this is true. It is possible that the increase in the endorsement of the hyperacusis item to a rather frequent response (15.7%) in the over 20 month group may mirror known maturational changes of the auditory system in normal children through these ages (Araceli, Pablo & Penélope, 2016). The possibility that hyperacusis as an ASD symptom may tend to appear after 18 months in children with ASD requires investigation. The three items that were found to have the highest failure rates in younger toddlers in this study are the same items which tend to have the lowest specificities for predicting autism outcomes in other studies including large samples of M-CHAT negative children providing item data (Kim, Joseph & Faisler, 2016; Srisinghasongkrami, et. al., 2016). For example, in a ten year follow-up of former premature infants screened at two years, these three items were the only ones with specificities under 0.95 for predicting an ASD diagnosis: with specificities of 0.77 for item 11 (sensitivity to noise), 0.86 for item 22 (stares at nothing, wanders), 0.88 for item 18 (unusual finger movements) (Kim, et. al., 2016). These items also had the highest failure rates in older toddlers not ultimately having an ASD diagnosis (Kim et. al., 2016). We speculate that in typical younger toddlers who have even higher rates of odd behaviors, such as “stares at nothing; wanders” in our study, specificity might become even more unfavorable and may therefore contribute to the lower rates of positive predictive value seen in the younger toddlers. Overall, it seems important to better understand the developmental trajectories of screening test items if we are to develop tests that distinguish emergence of a deviant direction of development from the range of normal developmental progress. The

large community sample of M-CHAT results reported here provides insights into issues related to specificity not evident in the test validation studies completed without data on toddlers screening negative.

While M-CHAT item and overall failure rates are generally higher in younger toddlers compared to older ones, among those who failed M-CHATs older toddlers failed more items than younger toddlers. This finding adds to the impression that higher item failures in younger toddlers are: more likely to be due to developmental immaturity than to ASD symptoms than for older toddlers; potential reasons that the M-CHAT is a more definitive screen for older toddlers than younger ones, and that screening accuracy is lower in the younger group in some studies (Pandey et al., 2008; Sturner et al., 2013). The two-stage approach of Kamio et al., of using a lower threshold for a positive for the first stage parent screen may have promise, except that the requirement for scheduling such a large number of interviews after the visit for those failing is likely to be impractical. While a follow-up interview at the moment of care has been shown to be feasible and reliable by a primary care provider (Sturner et al., 2016), we do not know if the follow-up interview completed at the moment of care, rather than a month or two later, would have similar results if scoring is truly age sensitive. The pattern of age and gender sensitivity by item identified here might be a guide for further age and gender adjusted or weighted scoring for both initial and follow-up stages of the M-CHAT.

The M-CHAT/M-CHAT-R is not the only parent report autism screening tool to use the same set of items across the recommended target screening ages with the same scoring algorithm and without a gestational age adjustment. Accuracy in screening children with prematurity is especially important because both gestational age and medical complications associated with prematurity have been found to be associated with increased incidence of autism diagnoses (Kuzniewicz et al., 2014). We reviewed parent-report based primary ASD screening tools recommended by clinical practice guidelines in the U.S. and U.K. covering ages 18 and 24 months. Of the thirteen screening instruments identified for use in this age range, only one instrument – the stage one Pervasive Developmental Disability Screening Test-2 (PDDST-II) —utilized different items at ages 18 and 24 months (Siegel, 2004). Additionally, only the Communication and Symbolic Behavior Scales Developmental Profile (CSBS-DP) provided scoring adjustment based on age norms, with different cut-offs for 18 and 24-month composite scores (Wetherby, Brosnan-Maddox, Peace & Newton, 2008). This age adjustment in scoring may be a factor in the CSBS-DP showing better sensitivity to autism diagnoses than the M-CHAT (Dudova et al., 2014). However, none of the screening tools reviewed contained consideration or adjustment for prematurity, per se or gender. While the desire to keep the screening process simple is understandable, the cost may include less accuracy at the youngest ages.

Limitations

The 20 item version of the M-CHAT used in this study does not include the minor wording changes in some of the items as published in the current M-CHAT-R version now recommended by the authors (Robins et al., 2014) for clinical application. Another limitation is that the data presented here did not include follow-up interviews. It is possible

that clinicians using the structure of the follow-up interview may be able to correct some parental misperceptions of behaviors of the younger toddlers. In addition, this data does not include diagnostic evaluations. On the other hand, there may be absolute limitations to the potential of any parent report measure at a given point in time. These limitations might be compensated for or even replaced by a promising system of “developmental surveillance” using discrete observations by specially trained primary care clinicians from late infancy through the toddler years (Barbaro & Dissanayake, 2010). It was hoped, however, that this large dataset might help to focus on item level data to better understand how tests might be modified for the younger toddler ages rather than concluding that the test is inadequate or screening should not be done at the earlier age.

Conclusion

Higher failures rates for more recently acquired milestones together with the known relatively lower prevalence of true positive cases at earlier ages may be contributing factors to the lower accuracy of autism screening in younger versus older toddlers found by some investigators. Further exploration into age and gender sensitive scoring and alternative screening strategies for younger toddlers at the standard 18-month pediatric check-up visit, who are generally under 20 months, should be a priority to improve early detection.

Acknowledgments

We are grateful to all of the participating practices using CHADIS for clinical care. This project was funded in part by a grant from the NIMH grant no. R44MH085399

References

- American Academy of Pediatrics. Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics*. 2006; 118(1):405–420. [PubMed: 16818591]
- Araceli RC, Pablo MMJ, Penélope GH. Maturation changes of the auditory system in normal preschool children assessed by brainstem auditory evoked potentials (BAEPs). *Clinical Neurophysiology*. 2016; 127(9):e320–e321. DOI: 10.1016/j.clinph.2016.05.351
- Autism Speaks. Modified checklist for autism in toddlers, revised (M-CHAT-R). 2016. Retrieved August 3, 2016, from <http://www.autismspeaks.org/what-autism/diagnosis/screen-your-child>
- Baird G, Charman T, Baron-Cohen S, Cox A, Swettenham J, Wheelwright S, Drew A. A screening instrument for autism at 18 months of age: a 6-year follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2000; 39(6):694–702. DOI: 10.1097/00004583-200006000-00007 [PubMed: 10846303]
- Barbaro J, Dissanayake C. Prospective identification of autism spectrum disorders in infancy and toddlerhood using developmental surveillance: the social attention and communication study. *Journal of Developmental & Behavioral Pediatrics*. 2010; 31(5):376–385. DOI: 10.1097/DBP.0b013e3181df7f3c [PubMed: 20495475]
- Barger BD, Campbell JM, McDonough JD. Prevalence and onset of regression within autism spectrum disorders: A meta-analytic review. *Journal of Autism and Developmental Disorders*. 2013; 43(4): 817–828. DOI: 10.1007/s10803-012-1621-x [PubMed: 22855372]
- Baron-Cohen S, Cox A, Baird G, Swettenham J, Nightingale NATASHA, Morgan KATE, ... Charman T. Psychological markers in the detection of autism in infancy in a large population. *The British Journal of Psychiatry*. 1996; 168(2):158–163. [PubMed: 8837904]
- Baron-Cohen S, Allen J, Giliberg C. Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *British Journal of Psychiatry*. 1992; 161:839–843. [PubMed: 1483172]

- Centers for Disease Control and Prevention (CDC). Autism spectrum disorder: Screening and diagnosis for health care providers. 2015. Retrieved August 8, 2016, from <http://www.cdc.gov/ncbddd/autism/hcp-screening.html>
- Chlebowski C, Robins DL, Barton ML, Fein D. Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics*. 2013; 131(4):e1121–e1127. DOI: 10.1542/peds.2012-1525 [PubMed: 23530174]
- Dudova I, Markova D, Kasparova M, Zemankova J, Beranova S, Urbanek T, Hrdlicka M. Comparison of three screening tests for autism in preterm children with birth weights less than 1,500 grams. *Neuropsychiatric disease and treatment*. 2013; 10:2201–2208. DOI: 10.2147/NDT.S72921
- Guy A, Seaton SE, Boyle EM, Draper ES, Field DJ, Manktelow BN, Johnson S. Infants born late/moderately preterm are at increased risk for a positive autism screen at 2 years of age. *The Journal of pediatrics*. 2015; 166(2):269–275. DOI: 10.1016/j.jpeds.2014.10.053 [PubMed: 25477165]
- Howlin P, Magiati I, Charman T. Systematic review of early intensive behavioral interventions for children with autism. *American journal on intellectual and developmental disabilities*. 2009; 114(1):23–41. DOI: 10.1352/2009.114:23;nd41 [PubMed: 19143460]
- Inada N, Kamio Y, Koyama T. Developmental chronology of preverbal social behaviors in infancy using the M-CHAT: Baseline for early detection of atypical social development. *Research in Autism Spectrum Disorders*. 2010; 4(4):605–611. DOI: 10.1016/j.rasd.2009.12.003
- Kamio Y, Inada N, Koyama T, Inokuchi E, Tsuchiya K, Kuroda M. Effectiveness of using the Modified Checklist for Autism in Toddlers in two-stage screening of autism spectrum disorder at the 18-month health check-up in Japan. *Journal of autism and developmental disorders*. 2014; 44(1):194–203. DOI: 10.1007/s10803-013-1864-1 [PubMed: 23740200]
- Kim SH, Joseph RM, Frazier JA, O'Shea TM, Chawarska K, Alfred EN, Leviton A, Kuban KK. Predictive Validity of the Modified Checklist for Autism in Toddlers (M-CHAT) Born Very Preterm. *J Pediatr*. 2016; 178:101–7. [PubMed: 27592094]
- Kleinman JM, Robins DL, Ventola PE, Pandey J, Boorstein HC, Esser EL, Barton M. 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*. 2008; 38(5):827–839. DOI: 10.1007/s10803-007-0450-9 [PubMed: 17882539]
- Kuzniewicz MW, Wi S, Qian Y, Walsh EM, Armstrong MA, Croen LA. Prevalence and neonatal factors associated with autism spectrum disorders in preterm infants. *The Journal of pediatrics*. 2014; 164(1):20–25. DOI: 10.1016/j.jpeds.2013.09.021 [PubMed: 24161222]
- Mahoney AD, Minter B, Burch K, Stapel-Wax J. Autism spectrum disorders and prematurity: A review across gestational age subgroups. *Advances in Neonatal Care*. 2013; 13(4):247–251. DOI: 10.1097/ANC.0b013e31828d02a1 [PubMed: 23912016]
- Mawle E, Griffiths P. Screening for autism in pre-school children in primary care: Systematic review of English language tools. *International Journal of Nursing Studies*. 2006; 43(5):623–636. DOI: 10.1016/j.ijnurstu.2005.11.011 [PubMed: 16387303]
- National Autism Association. Signs and symptoms of autism. 2015. Retrieved August 8, 2016, from: <http://nationalautismassociation.org/resources/signs-of-autism/>
- National Research Council. Educating children with autism. National Academies Press; 2001.
- Ozonoff S, Heung K, Byrd R, Hansen R, Hertz-Picciotto I. The onset of autism: Patterns of symptom emergence in the first years of life. *Autism Research*. 2008; 1(6):320–328. DOI: 10.1002/aur.53 [PubMed: 19360687]
- Pandey J, Verbalis A, Robins DL, Boorstein H, Klin A, Babitz T, Fein D. Screening for autism in older and younger toddlers with the Modified Checklist for Autism in Toddlers. *Autism*. 2008; 12(5): 513–535. DOI: 10.1177/1362361308094503 [PubMed: 18805945]
- Robins DL, Casagrande K, Barton M, Chen CM, Dumont-Mathieu T, Fein D. Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics*. 2014; 133(1):37–45. DOI: 10.1542/peds.2013-1813 [PubMed: 24366990]
- Robins DL, Fein D, Barton ML, Green JA. 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*. 2001; 31(2):131–144. [PubMed: 11450812]

- Siegel, B. Pervasive developmental disorders screening test-ii (pddst-ii). Harcourt; San Antonio (TX): 2004.
- Srisinghasongkram P, Pruksananonda C, Chonchaiya W. Two-Step Screening of the Modified Checklist for Autism in Toddlers in Thai Children with Language Delay and Typically Developing Children. *J Autism Dev Disord*. 2016; 46:3317–3329. [PubMed: 27460003]
- Stenberg N, Bresnahan M, Gunnes N, Hirtz D, Hornig M, Lie KK, Schjølberg S. Identifying children with autism spectrum disorder at 18 months in a general population sample. *Paediatric and perinatal epidemiology*. 2014; 28(3):255–262. DOI: 10.1111/ppe.12114 [PubMed: 24547686]
- Størvold GV, Aarethun K, Bratberg GH. Age for onset of walking and prewalking strategies. *Early human development*. 2013; 89(9):655–659. DOI: 10.1016/j.earlhumdev.2013.04.010 [PubMed: 23701748]
- Sturner, RA., Howard, B., Morrel, T., Bergmann, P., Walton, K., Marks, D., Landa, R. Accurate autism screening at the 18 month visit requires different strategies than at 24 months. Presented at the Pediatric Academic Societies annual meeting; Washington, D. C. 2013 May.
- U.S. Preventive Services Task Force. [accessed 3/14/17] Final Recommendation Statement: Autism spectrum disorder in young children: Screening. 2016. <http://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/autism-spectrum-disorder-in-young-children-screening>
- Wetherby AM, Brosnan-Maddox S, Peace V, Newton L. Validation of the infant-toddler checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. *Autism: The International Journal of Research and Practice*. 2008; 12(5):487–511. DOI: 10.1177/1362361308094501 [PubMed: 18805944]
- Wingate M, Kirby RS, Pettygrove S, Cunniff C, Schulz E, Ghosh T, ... Fitzgerald R. Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR Surveillance Summaries*. 2014; 63(2)

Figure 1.
Sample Selection

Figure 2.
Age distribution of sample (not adjusted for prematurity)

Table 1

Age by Gender by Adjustment for Prematurity

		Overall	Females	Males
Unadjusted				
	Age	Mean=21.6 (SD=3.6)	Mean=21.6 (SD=3.6)	Mean=21.6 (SD=3.6)
	N	N=73,564	N=35,674 (48.5%)	N=37,890 (51.5%)
Adjusted				
	Age	Mean=21.6 (SD=3.6)	Mean=21.6 (SD=3.6)	Mean=21.6 (SD=3.6)
	N	N=73,586	N=35,659 (48.5%)	N=37,865 (51.5%)

Table 2

Age Group by Gender by Adjustment for Prematurity

	Overall	Females	Males
Unadjusted			
Under 20	33,908 (46.1%)	16,504 (46.3%)	17,404 (45.9%)
20 Plus	39,656 (53.9%)	19,170 (53.7%)	20,486 (54.1%)
Adjusted			
Under 20	34,142 (46.4%)	16,625 (48.7%)	17,517 (51.3%)
20 Plus	39,444 (53.6%)	19,058 (48.3%)	20,386 (51.7%)

Table 3

Age by Age Group by Adjustment for Prematurity

	Overall	Under 20	20 Plus
Unadjusted			
Age	Mean=21.6 (SD=3.6)	Mean=18.2 (SD=0.7)	Mean=24.6 (SD=2.1)
N	N=73,564	N=33,908 (46.1%)	N=39,656 (53.9%)
Adjusted			
Age	Mean=21.6 (SD=3.6)	Mean=18.1 (0.7)	Mean=24.5 (SD=2.1)
N	N=73,586	N=34,142 (46.4%)	N=39,444 (53.6%)

Table 4

M-CHAT/M-CHAT-R Failure Rate by Age Group

Version	Prenaturity	Overall	< 20 Months	20+ Months	Difference	χ ²	p
M-CHAT	Unadjusted	6.35%	6.84%	5.93%	15.435 %	25.48	< .001
M-CHAT-R	Unadjusted	4.31%	4.68%	3.98%	17.59%	21.67	< .001
M-CHAT	Adjusted	6.34%	6.90%	5.85%	17.94%	33.93	< .001
M-CHAT-R	Adjusted	4.30%	4.73%	3.92%	20.92%	29.08	< .001

Unadjusted <20 (n=33,908)

Unadjusted 20+ (n=39,656)

Adjusted > 20 (n=34,142) Adjusted 20+ (n=39,444) Difference = (<20)-[20+]

Table 5
 Number of Failed M-CHAT/M-CHAT-R Items by Age Group and Screen Result (N's *)

Version	Screen Result	Overall	< 20 Months	20+ Months	Difference	t	p
M-CHAT	Either	0.687	0.714	0.664	7.53%	4.70	< .001
M-CHAT-R	Either	0.566	0.587	0.549	7.90%	4.09	< .001
M-CHAT	Negative	0.413	0.433	0.396	9.34%	7.50	< .001
M-CHAT-R	Negative	0.371	0.386	0.359	7.52%	5.66	< .001
M-CHAT	Positive	4.737	4.547	4.923	-7.64%	-4.31	< .001
M-CHAT-R	Positive	4.901	4.682	5.120	-8.55%	-4.28	< .001

* Numbers by age group and screen result:

<20 (n=33,908)	20+ (n=39,656)
M-CHAT <20 Negatives (n=31,590)	M-CHAT 20+ Negatives (n=37,306)
M-CHAT <20 Positives (n=2,318)	M-CHAT 20+ Positives (n=2,350)
M-CHAT-R <20 Negatives (n=32,320)	M-CHAT-R 20+ Negatives (n=38,076)
M-CHAT-R <20 Positives (n=1,588)	M-CHAT-R 20+ Positives (n=1,580)

Table 6

Factor and Percentage Changes in Expected Number of Failed MCHAT-23 Items

	Factor Change	% Change	<i>z</i>	<i>p</i>
<i>MCHAT-23 Overall</i>				
MCHAT-23 Fail vs Pass	11.401	1040.1	237.84	< .001
Male vs Female	1.068	6.8	6.73	< .001
Younger vs Older	1.030	3.0	3.01	.003
<i>MCHAT-23 Fails</i>				
Male vs Female	1.056	5.6	2.96	.003
Younger vs Older	0.916	−8.4	−4.89	<.001
<i>MCHAT-23 Passes</i>				
Male vs Female	1.080	8.0	6.41	<.001
Younger vs Older	1.094	9.4	7.52	<.001

Table 7

Factor and Percentage Changes in Expected Number of Failed MCHAT-20 Items

	Factor Change	% Change	<i>z</i>	<i>p</i>
<i>MCHAT-20 Overall (n=73,564)</i>				
MCHAT-20 Fail vs Pass	13.084	1208.4	222.79	< .001
Male vs Female	1.077	7.7	6.947	< .001
Younger vs Older	1.023	2.3	2.108	.035
<i>MCHAT-20 Fails (n=3,168)</i>				
Male vs Female	1.045	4.5	2.067	.039
Younger vs Older	0.907	−9.3	−4.679	< .001
<i>MCHAT-20 Passes (n=70,396)</i>				
Male vs Female	1.093	9.3	7.100	< .001
Younger vs Older	1.074	7.4	5.693	< .001

Table 8

MCHAT-23 Overall - Sorted by Under 20 Probability of Failing the Item

		Age Group					
Item	Set	Overall	Under 20	Over 20	Difference	z-Value	p-Value
11		0.142	0.123	0.158	-0.035	-13.757	<.001
22		0.097	0.101	0.094	0.007	3.058	.002
18		0.099	0.101	0.098	0.003	1.232	.218
19	2	0.049	0.058	0.040	0.018	11.649	<.001
23	3	0.042	0.043	0.041	0.002	1.596	.111
13	2	0.029	0.035	0.025	0.010	8.739	<.001
6	2	0.024	0.028	0.020	0.008	8.494	<.001
7	2	0.023	0.026	0.020	0.006	6.773	<.001
5	2	0.019	0.023	0.015	0.008	9.187	<.001
8	3	0.018	0.020	0.015	0.005	5.428	<.001
20		0.019	0.015	0.022	-0.007	-7.170	<.001
21		0.014	0.015	0.014	0.001	1.657	.097
15	2	0.014	0.015	0.014	0.001	1.005	.315
9	3	0.011	0.013	0.010	0.003	3.904	<.001
16		0.009	0.013	0.006	0.007	10.315	<.001
2	1	0.013	0.011	0.015	-0.004	-5.309	<.001
1	1	0.013	0.011	0.015	-0.004	-5.196	<.001
10	1	0.014	0.010	0.017	-0.006	-7.763	<.001

Item	Set	Age Group					p-Value
		Overall	Under 20	Over 20	Difference	z-Value	
17	3	0.011	0.010	0.012	−0.002	−2.949	.003
14	1	0.010	0.009	0.010	−0.001	−1.925	.054
3		0.007	0.006	0.007	−0.002	−2.659	.008
4	1	0.006	0.004	0.007	−0.003	−6.310	< .001
12	1	0.004	0.004	0.005	−0.001	−3.034	.002