

Accurate Autism Screening at the 18-Month Well-Child Visit Requires Different Strategies than at 24 Months

Raymond Sturner^{1,2,10} · Barbara Howard^{1,3} · Paul Bergmann^{4,5} · Tanya Morrel³ · Rebecca Landa^{1,6} · Kejuana Walton^{3,7} · Danielle Marks^{1,8,9}

Published online: 31 July 2017
© Springer Science+Business Media, LLC 2017

Abstract Accuracy of autism screening using M-CHAT plus the follow-up interview (M-CHAT/F) for children screened positive at 18-months was compared to screening at 24-months. Formal ASD testing was criterion for a community sample of M-CHAT positive children ($n=98$), positive predictive value (PPV) was 0.40 for the M-CHAT and 0.58 for the M-CHAT/F. MCHAT/F PPV was 0.69 among children 20+ months compared to 0.36 for <20 months. Multivariate analyses incorporating data from the Ages and Stages Questionnaire, MacArthur-Bates Communicative Development Inventory, M-CHAT and M-CHAT/F results, and M-CHAT items suggest language variables carry greatest relative importance in contributing to an age-based algorithm with potential to improve PPV for toddlers <20 months to the same level as observed in older toddlers.

Keyword Autism screening · Developmental screening · M-CHAT · Toddler development

Introduction

Current prevalence estimates suggest that autism spectrum disorder (ASD) now affects 1 in 68 children (Christensen et al. 2016). Early evidence-based intervention for children with ASD is associated with improved long-term outcomes, making early detection vital (National Research Council 2001; Howlin et al. 2009). The American Academy of Pediatrics (AAP) has issued guidelines for ASD screening at 18 and 24 months (American Academy of Pediatrics 2006). Although earlier detection is improving, average age at first diagnosis remains over age 3 years (Fountain et al. 2011). The U. S. Preventive Services Task Force concluded that the evidence for efficacy of screening in community settings is insufficient (Siu and U. S. Preventive Services Task Force 2016), and specifically 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” (Siu and U. S. Preventive Services Task Force 2016). However, the AAP continues its recommendation due to the prevalence and the importance of early detection. Therefore, much work remains to be done to improve early screening tools and practices to assure the adequacy at all recommended ages.

The Modified Checklist for Autism in Toddlers (M-CHAT) is an ASD screening instrument widely used in primary care (Robins et al. 2001). The M-CHAT is among the tools recommended for ASD screening by the AAP (American Academy of Pediatrics 2006), and major advocacy groups, such as Autism Speaks (2016). The M-CHAT

✉ Raymond Sturner
rsturner@childhealthcare.org

¹ Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA

² Center for Promotion of Child Development through Primary Care, Baltimore, MD 21210, USA

³ Total Child Health, Baltimore, MD 21210, USA

⁴ Foresight Logic, Inc., Saint Paul, MN, USA

⁵ PrairieCare Institute, Minneapolis, MN, USA

⁶ Kennedy Krieger Institute, Baltimore, MD 21287, USA

⁷ Baltimore Healthy Start, Baltimore, MD 21218, USA

⁸ Sinai Hospital, Baltimore, MD 21215, USA

⁹ Present Address: Maternal and Child Health Unit, Public Health Division, Wyoming Department of Health, Evanston, WY, USA

¹⁰ 6017 Altamont Place, Baltimore, MD 21210, USA

has replaced the original CHAT (Baron-Cohen et al. 1992) which was shown to have excellent specificity and positive predictive value (PPV) for early identification of autistic children but overlooked too many affected children (Baird et al. 2000). While studies of the M-CHAT suggested that the modifications of the test corrected the problem with sensitivity, these studies were conducted in samples heavily weighted with children already enrolled in clinical programs rather than only children presenting for screening in community primary care practices, which would tend to inflate estimates of sensitivity. While more recent M-CHAT studies did extend to large community pediatric samples (Chlebowski et al. 2013; Robins et al. 2014) clarifying estimates of predictive validity that can be expected in primary care screening, accurate assessment of sensitivity has been limited for all autism screening tests, including the M-CHAT, because costs and feasibility issues preclude evaluating community samples of children passing screening tests. However, when a large community sample of children who had M-CHAT screening at 18 months was followed with additional questionnaire screens at 3, 5, and 7 years with evaluations occurring 2 ½ to 8 years later, 2/3's of the children confirmed with ASD in the community had not been identified (Stenberg et al. 2014) indicating limited sensitivity. Another study of 18-month M-CHAT screening followed up at age 3 years showed a higher sensitivity (0.73) when the pass/fail cut point was decreased over the standard one but also causing a referral rate so high (17%) as to be of questionable feasibility (Kamio et al. 2014).

Since it is known that about a third of autism cases present with developmental regression (Barger et al. 2013) and these children may actually appear typically behaving at the 18-month visit, we can expect that a symptom report screening test may be less sensitive at that age than at 2 years, justifying recommendation of a second screen by the American Academy of Pediatrics. However, since earlier intervention has benefits, screening at the younger age is warranted if screening tools have an adequate proportion of positive screens confirmed by diagnostic testing, i.e., the PPV of the test.

There is reason to believe that the M-CHAT may be less accurate when administered at 18 months than at age 2. M-CHAT authors found lower PPV of 0.28 in younger toddlers from a community sample compared to older toddlers (0.61) even when the recommended follow-up interview (M-CHAT/F) (Pandey et al. 2008) was used. A more recent study in a community sample suggested that younger toddlers may have a lower PPV, though that difference was not statistically significant (Robins et al. 2014). However, it should be noted that in this study the required follow-up interview was completed an average of 2 months after the 18-month visit when more symptoms may have emerged or parent observation and concern may have been sharpened.

In the initial validation study of the M-CHAT, parents reviewed items they endorsed as positive with a research assistant over the telephone to determine if the parents would confirm their responses when probed with additional questions. Additional studies have used a formalized structured follow-up interview to improve PPV from 0.36 to 0.72 overall and from 0.11 to 0.68 in a community primary care sample (Kleinman et al. 2008). This follow-up interview (known now as M-CHAT/F) has been designated as a “highly recommended” critical part of a two-stage identification process (Chlebowski et al. 2013; Robins et al. 2014). Recently, it has been shown that the follow-up interview can be feasibly incorporated into routine check-up visits and is reliable when done with electronic decision support by primary care providers (PCPs) (Sturmer et al. 2016). Because use of a screen with low PPV increases unnecessary worries and expense for families with false positive results, M-CHAT screening without the follow-up interview for positives is presently not recommended in low risk populations such as routine primary care pediatrics (Robins et al. 2014).

With these weaknesses in screening at the more valuable early age, it is fortunate that progress in developing screening tools is being made. There is evidence that some children can be identified as early as 1 year of age via a parent questionnaire surveying autism symptoms (Reznick et al. 2007) although numbers of children being overlooked, PPV, and clinical utility are not yet defined. A standard language measure [MacArthur-Bates Communicative Development Inventory (CDI)] has shown promise to further enhance those predictions among 1-year-olds (Veness et al. 2012), although this lengthy measure is not being advocated for use as an ASD screen. Further exploration of the value of standard language measures for enhancing prediction of autism outcomes in young children seems warranted. Recently, a widely used general developmental screen, the Ages and Stages Questionnaires—3rd Ed. (ASQ) (Squires and Bricker 2009) has been shown to be a possibly clinically valuable pre-screen for ASD with M-CHAT used as the second stage (Hardy et al. 2015).

The objective of this study was to compare the accuracy of autism screening between younger and older toddlers using the results of the M-CHAT completed by parents and the follow-up interview completed by PCPs and to explore strategies to improve screening accuracy. We studied the M-CHAT along with the ASQ completed by parents online prior to routine check-up visits at 18 and 24 months in conjunction with automated presentation of M-CHAT/F items to the PCP and instant rescoring at the time of the visit. We then conducted multivariate analyses of screening data to explore the potential for enhancing the PPV of an autism screening method by incorporating individual M-CHAT item responses with elements of a broad band

developmental screen and a standard language measure that was administered to parents of children screening positive on the M-CHAT who were either M-CHAT/F positive or negative.

Methods

Screening

PCPs screened children before routine 18- and 24-month visits with the M-CHAT and ASQ as per AAP recommendations. Both tools were completed by parents either at home or on a computer or touch pad in the waiting room via the online system called CHADIS (CHADIS.com 2017). Data collection occurred between September, 2009 and February, 2013. Forty-seven self-selected pediatricians from 22 offices used the M-CHAT/F. These Maryland practices were mostly suburban (73%), with some rural (18%) and urban locations (9%). Practice estimation of office-level demographics showed that 31% of children were Medicaid insured (range 5–65%) with 39% white; 33% African American; 16% Asian; and 8% Hispanic. Of the total of 5071 children screened with the M-CHAT, 341 (6.7%) were positive. The follow-up interview was completed by the PCP with parents of children with a positive M-CHAT (see Fig. 1 for sample flow). The M-CHAT was automatically scored, and for each positive screen, the PCP was prompted by the online system with the specific questions to follow up on failed items needed to complete the M-CHAT follow-up interview. The online system continually rescored the M-CHAT/F as items were asked and a positive screen was either confirmed or refuted by the scoring logic. The M-CHAT/F, administered online by a PCP has been shown to be as accurate as when administered by specially trained autism center staff (Sturner et al. 2016).

Diagnostic Testing and Assessment

All children with a positive M-CHAT, regardless of subsequent M-CHAT/F determination, were recruited for diagnostic testing at Kennedy Krieger Institute (KKI) Center for Autism and Related Disorders. Consent for participation was obtained using a protocol approved by The Johns Hopkins University School of Medicine Institutional Review Board (IRB). The IRB approved the initial phase of screening (M-CHAT) as a standard component of care requiring no study consent. There was no charge for the evaluation and the parents received a subject participation stipend. Parents were asked to complete a mailed copy of the age appropriate MacArthur-Bates Communicative Development Inventory (CDI) (Fenson et al. 2007) prior to the KKI evaluation. If the CDI was not completed by

the time of the evaluation, the parent was asked to do so during the evaluation session. In addition, the M-CHAT/F interview was repeated by a KKI staff member over the phone at the time the appointment was made. Diagnostic testing was completed using the Autism Diagnostic Observation Schedule, 2nd Ed. (ADOS-2) (Luyster et al. 2009) administered by certified research reliable speech and language pathologists. The same evaluator administered the Mullen Scales of Early Learning (MSEL) (Mullen 1995). KKI evaluators were aware of the study design with knowledge of M-CHAT positive results but were blind to PCP M-CHAT/F interview results and other test results. A total of 113 parents (18% of M-CHAT positive children) consented for this phase of the study and 98 (16% of M-CHAT positives) completed all parent-report measures.

Measures

Screening Instruments

The Modified Checklist for Autism in Toddlers (Robins et al. 2001) with follow-up interview (Kleinman et al. 2008) (M-CHAT/F): The 23-item 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 et al. 1992). The original validation sample included 1293 children 18–30 months old with 58 children receiving diagnostic evaluation and 38 diagnosed with ASD. In this initial study only cases where the M-CHAT result was confirmed by a “phone interview” were given a diagnostic evaluation. Only seven of these were “low risk” by being selected from primary care pediatric settings with the remaining being selected from those already in an early intervention program (Robins et al. 2001). 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 a score of any two of six items considered ‘critical’, or any three items overall. M-CHAT together with a confirmatory phone interview yielded a sensitivity of 0.87 and specificity of 0.99, with a PPV of 0.80. The critical items included those concerning joint attention (proto-declarative pointing, bringing to show, following a point), interest in other children, responding to name, and imitation.

The addition of a formalized post-questionnaire telephone interview (M-CHAT/F) increased sensitivity and specificity to 0.97 and 0.99 (Kleinman et al. 2008). The follow-up interview is described in a 26-page manual that was used to create the online version used in this study. A recently revised scoring (M-CHAT-R) follows an analysis in a larger sample (16,071) that eliminates three of the items, changed the item order, simplified the language

Fig. 1 Study sample flow



of some of the items and requires any three or more failed items for an overall fail. The study reported here was completed prior to this revision of the M-CHAT and therefore used the original scoring.

Ages and Stages Questionnaires 3rd Edition (ASQ) (Squires and Bricker 2009): The ASQ is a parent-report screen for developmental problems among infants and toddlers, including 21 separate questionnaires at 2-month age intervals between 2 and 60 months. Each questionnaire

includes six questions with a “yes, sometimes, and not yet” response set in each of five domains: communication, gross motor, fine motor, problem solving, and personal-social. Psychometrics, based on data from over 18,000 respondents, yielded agreement of 86% between ASQ and other measures of development (range 83–88%). The ASQ is among those endorsed for general developmental screening by the American Academy of Pediatrics (1994).

The MacArthur-Bates Communicative Development Inventory (CDI) (Fenson et al. 2007): The CDIs are parent-report measures designed to evaluate the communicative skills of young, typically developing children from their “early signs of comprehension, to their first nonverbal gestural signals, to the expansion of early vocabulary and the beginnings of grammar.” The CDI: words and gestures (infant form) is designed for use with 8- to 16-month old children. The CDI: words and sentences (Toddler form) is designed for use with 16- to 30-month old children. Both were normed on a large population of monolingual English-speaking children, representing a socioeconomic status that was higher than the national average based on 1990 U.S. census information.

Diagnostic Assessments

Autism Diagnostic Observation Schedule, 2nd Ed. (ADOS-2) (Luyster et al. 2009): The ADOS-2, a semi-structured behavior observation assessment of social and communication skills, is the recognized standard test for diagnosis of ASD across age, developmental level, and language skills. The assessment relies on a series of ‘planned social occasions’ designed to elicit specific social and communicative situations in a standardized way. The ADOS includes four separate modules. Each module takes approximately 40–60 min to administer. Modules 1 through 4 provide cutoff scores for autism and autism spectrum classifications. Modules 1 through 3 also provide a comparison score indicating level of autism spectrum-related symptoms compared to children with ASD who are the same age and have similar language skills. Data collection for this study was completed before availability of the Autism Diagnostic Observation Schedule, Second Edition, Toddler Module and before the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, redefinition of autism and social communication disorder. The dependent variable was the KKI clinical impression of ASD after performing the ADOS-2 and MSEL.

Mullen Scales of Early Learning (MSEL) (Mullen 1995): The MSEL is a standardized developmental test for children between 3 and 69 months of age, with five subscales: gross motor, fine motor, visual reception, receptive language and expressive language. An early learning composite (ELC) is generated using scores from the fine motor,

receptive language, expressive language, and visual reception scales. The ELC provides a developmental quotient, which is used to determine the expected levels of communicative and social functioning which help to inform the clinical impression of autism along with the Autism Diagnostic Observation Schedule-2 (ADOS-2) results.

Analyses

All analyses were conducted using the statistical computing platform, R (Core Team 2015), and Stata/IC 14.1 (StataCorp 2015). Descriptive statistics were tabulated on demographic characteristics of the sample. Sensitivity was calculated based on the sample of children for which diagnostic data were available and therefore does not account for children in the larger sample who screened negative on the M-CHAT.

Applicable receiver operating characteristics (ROC) and 95% Confidence Intervals were estimated for the M-CHAT screen alone, and the M-CHAT with follow-up interview (M-CHAT/F). In both cases, the criterion variable was the KKI clinical ASD determination made after consideration of the results of the ADOS-2 and MSEL.

Comparisons of estimated ROCs were conducted between age groups and between M-CHAT and M-CHAT/F using z-tests and two one-sided tests of equivalence (TOST). Twenty months was determined to be the cutoff between age groups that maximized age related screening differences.

Since children with false positive screens for ASD may nevertheless have developmental difficulties benefitting from early intervention, we also examined screening performance against a combined criterion of scores >1.5 SD below the mean on two or more Mullen subscales or >2 SD on any one Mullen subscale (common state eligibility criteria e.g., per nectac.org 2015), and/or receiving an ASD diagnosis.

Using the R package, rpart: Recursive Partitioning and Regression Trees (Therneau et al. 2015) based on the concepts of classification and regression tree (CART) analysis (Breiman et al. 1984), we explored alternative algorithms for determining positive ASD screens by including individual M-CHAT items, M-CHAT items corrected by the M-CHAT/F, ASQ scale scores, and CDI scale scores as independent variables. The dependent variable in the model was the same KKI ASD determination used as the criterion variable for the Screening ROCs. CART is a partitioning algorithm with potential utility in clinical practice due to the decision tree nature of the results. While the current sample is too small and narrow (i.e., no M-CHAT negatives are included in this study) to produce a CART model robust enough to form a basis for recommending changes to current screening protocols, its use as an exploratory tool

may provide direction for further research as we seek better ways to identify younger toddlers at risk for autism.

The recursive partitioning algorithm used in this study is a non-parametric, nonlinear approach to modeling effects and interactions in the data without making assumptions about an underlying stochastic model of the data. In other words, there is no attempt to fit a function or estimate parameters. The algorithm utilizes the non-probabilistic Gini splitting criteria to determine which candidate predictor variables split the cases into the most homogeneous groups with regard to the dependent variable (in our case, KKI clinical ASD determination). It is an approach that assumes the relationships within the data are complex and unknown, and follows a stepwise process of creating locally optimal partitions within the context of parent partitions. As such, CART (as well as most other recursive partitioning algorithms) is well suited to exploring a small, highly dimensional dataset for insights into potential interactions among, and relative importance of, predictor variables.

The CART model was derived from a sample of all children from both age groups with non-missing data for criterion variable ($n=98$). Using resubstitution on the overfitted model, we estimated three sets of ROCs based on the entire sample ($n=98$), younger toddlers (<20 months; $n=48$), and older toddlers (20+ months; $n=50$). We then compared the estimated ROCs for the CART model between age groups and against the M-CHAT and M-CHAT/F. While the CART ROCs are overinflated and not useful for generalizing the model to a larger population, they do provide a means to understand potential differences between younger and older toddlers in what might be important to consider when screening for ASD.

A beneficial feature of CART is the use of surrogate predictor variables when a case is missing data for the primary splitter at any particular node. As such, a variable may appear in the tree many times either as a primary or surrogate splitter given its overall relative importance (Therneau et al. 2015). This provides an opportunity to consider the potential relative importance of the independent variables as applied to predicting a KKI clinical ASD determination. We calculated an overall measure of importance for each independent variable using goodness of split and adjusted agreement measures generated for each primary and surrogate splitter at each node.

Results

Demographic characteristics of the subjects completing diagnostic testing with complete data are presented in Table 1. Mean age is 22.9 months (s.d.=6.1) with a range of 14.7–40.8 months. The M-CHAT was assigned for their “18 month” or “24 month” visit, but parents may

complete the tool at a slightly older or younger age than the M-CHAT standardization age range. As typical for children who screen positive for ASD (Halladay et al. 2015), 74.5% are male. The demographics show a range of backgrounds roughly similar to U.S. national data except that Hispanics are underrepresented. The majority of the respondents were biological mothers (90%), and had a college degree (67.4%). With the exception of respondent education ($\chi^2=12.7$, $p=.048$), there were no significant differences in demographic characteristics between the two age groups.

The PPV of the MCHAT/F on the entire sample ($n=98$) was significantly higher than the PPV of the M-CHAT alone (0.58 vs. 0.40, $z=-2.48$, $p=.013$; see Table 2). While the difference between M-CHAT and M-CHAT/F remained significant among the older toddlers (0.69 vs. 0.48, $z=-2.16$, $p=.031$), it was greatly diminished among younger toddlers (0.36 vs. 0.31, $z=-0.46$, $p=.643$).

We did not observe a significant difference nor equivalence of M-CHAT PPV (without follow-up interview) between younger and older toddlers (see Table 3). However, M-CHAT/F PPV (0.69 vs. 0.36, $z=-3.32$, $p<.001$) and sensitivity (0.75 vs. 0.33, $z=-4.14$, $p<.001$) were markedly higher among older toddlers as compared to younger toddlers, and the false discovery rate was half that of the younger age group (0.31 vs. 0.64, $z=3.32$, $p<.001$). The difference in specificity was clinically non-meaningful and statistically indeterminate (0.73 vs. 0.69).

Use of the combined criterion (ASD/developmental delay) increased the M-CHAT and M-CHAT/F PPVs to 0.68 and 0.88, respectively. The M-CHAT/F PPV for the younger age group increased to 0.71, and for the older age group, it increased to 0.96 (see Table 4).

CART analyses yielded a model with much higher but artificially inflated performance estimates due to the relatively small sample size, over fitting of the model and utilization of re-substitution (see Table 5). Direct comparisons of performance between the resulting model and the M-CHAT or M-CHAT/F are therefore neither meaningful nor valid. However, the estimates do provide insight into relative differences of model performance between age groups. Differences between age groups on PPV (0.72 vs. 0.88, younger vs. older, respectively), sensitivity (0.87 vs. 0.92), and Specificity (0.85 vs. 0.88) were all substantially smaller for the CART model as compared to M-CHAT/F, an improvement for the younger children. The difference in PPV between age groups was statistically significant ($z=-1.96$, $p=.50$). For sensitivity and specificity, the differences were not significant and statistical equivalence was indeterminate.

This study provided interesting preliminary information on the relative importance of those predictor variables that contributed to generation of the CART model (Table 6). Language variables carried largest relative importance in

Table 1 Demographics

	Overall (n=98)	Under 20 (n=48)	20 Plus (n=50)	χ^2	p value
Gender	–	–	–	0.333	.564
Female	25.5%	44.0%	56.0%	–	–
Male	74.5%	50.7%	49.3%	–	–
Race	–	–	–	2.718	.257
Non-White	34.7%	39.6%	28.0%	–	–
White	65.3%	58.3%	72.0%	–	–
Missing	1.0%	2.1%	0.0%	–	–
Hispanic	–	–	–	6.005	.050
Yes	5.1%	0.0%	10.0%	–	–
No	93.9%	97.9%	90.0%	–	–
Missing	1.0%	2.1%	0.0%	–	–
Insurance	–	–	–	7.216	.065
Private	75.5%	79.2%	72.0%	–	–
MA/CHIPS	20.4%	12.5%	28.0%	–	–
None	1.0%	2.1%	0.0%	–	–
Missing	3.1%	6.2%	0.0%	–	–
Respondent relationship to child	–	–	–	3.784	.151
Biological mother	89.8%	93.8%	86.0%	–	–
Other	9.2%	4.2%	14.0%	–	–
Missing	1.0%	2.1%	0.0%	–	–
Respondent education	–	–	–	12.702	.048*
≤12th grade education	5.1%	4.2%	6.0%	–	–
HS diploma or GED	8.2%	2.1%	14.0%	–	–
Some college	16.3%	10.4%	22.0%	–	–
Associate degree	3.1%	2.1%	4.0%	–	–
Bachelor degree	30.6%	31.2%	30.0%	–	–
Graduate degree	33.7%	43.8%	24.0%	–	–
Missing	3.1%	6.2%	0.0%	–	–

Table 2 Comparison of positive predictive values between M-CHAT and M-CHAT/F (criterion: ASD diagnosis)

	PPV	SE	95% confidence interval	Difference		Equivalence				
				Z	p	Z ₁	p ₁	Z ₂	p ₂	
All ages										
M-CHAT	0.40	0.05	0.30	0.49	–2.48	.013*	3.88	<.001	–1.08	.860
M-CHAT/F	0.58	0.05	0.48	0.67						
<20 months										
M-CHAT	0.31	0.07	0.18	0.44	–0.46	.643^	1.50	.067	0.57	.283
M-CHAT/F	0.36	0.07	0.22	0.49						
20+ months										
M-CHAT	0.48	0.07	0.34	0.62	–2.16	.031*	3.17	<.001	–1.14	.873
M-CHAT/F	0.69	0.07	0.56	0.82						
Difference*										
Equivalence~	<20 months n=48									
Indeterminate^	20+ months n=50									

contributing to the age-based algorithm described above for predicting autism diagnoses. Five of the top ten items in relative importance were from the CDI and two were

language items from the M-CHAT (Responds to name; Understands what people say) with the words produced scale included.

Table 3 Comparison of positive predictive value, sensitivity and specificity between younger and older age groups (criterion: ASD diagnosis)

	PPV	SE	95% confidence interval		Difference		Equivalence		p_1	Z_2	p_2
			Z	p	Z	p					
M-CHAT PPV											
<20 months	0.31	0.07	0.18	0.44	-1.69	.090^	2.70	.003	-6.8	.752	
20+ months	0.48	0.07	0.34	0.62							
M-CHAT/F PPV											
<20 months	0.36	0.07	0.22	0.49	-3.32	<.001*	4.31	<.001	-2.33	.990	
20+ months	0.69	0.07	0.56	0.82							
M-CHAT/F sensitivity											
<20 months	0.33	0.07	0.20	0.47	-4.14	<.001*	5.14	<.001	-3.15	.999	
20+ months	0.75	0.06	0.63	0.87							
M-CHAT/F specificity											
<20 months	0.73	0.06	0.60	0.85	0.38	.703^	0.71	.239	1.47	.071	
20+ months	0.69	0.07	0.56	0.82							
M-CHAT/F false discovery rate											
<20 months	0.64	0.07	0.51	0.78	3.32	<.001*	-2.33	.990	4.31	<.001	
20+ months	0.31	0.07	0.18	0.44							
Difference*											
Equivalence~	<20 months	n=48									
Indetermi- nate^	20+ months	n=50									

Table 4 Comparison of positive predictive value, sensitivity and specificity between younger and older age groups (criterion: ASD/developmental delay diagnosis)

	PPV	SE	95% confidence interval		Difference		Equivalence				
			Z	p	Z ₁	p ₁	Z ₂	p ₂			
M-CHAT PPV											
<20 months	0.65	0.07	0.51	0.78	-0.79	.430^	1.85	.032	0.27	.392	
20+ months	0.72	0.06	0.60	0.84							
M-CHAT/F PPV											
<20 months	0.71	0.07	0.59	0.84	-3.34	<.001*	4.69	<.001	-1.99	.977	
20+ months	0.96	0.03	0.91	1.01							
M-CHAT/F sensitivity											
<20 months	0.32	0.07	0.19	0.45	-3.68	<.001*	4.67	<.001	-2.69	.996	
20+ months	0.69	0.07	0.57	0.82							
M-CHAT/F specificity											
<20 months	0.76	0.06	0.64	0.88	-2.26	.024*	3.64	<.001	-0.88	.811	
20+ months	0.93	0.04	0.86	1.00							
M-CHAT/F false discovery rate											
<20 months	0.29	0.07	0.16	0.41	3.34	<.001*	-1.99	.977	4.69	<.001	
20+ months	0.04	0.03	0.00	0.09							
Difference*											
Equiva- lence~	<20 months n = 48										
Indetermi- nate^	20+ months n = 50										

Table 5 Comparison of positive predictive value, sensitivity and specificity of CART model between younger and older toddler age groups (criterion: ASD diagnosis)—words produced included

Measure	Value	SE	95% confidence interval		Difference		Equivalence			
					Z	p	Z ₁	p ₁	Z ₂	p ₂
PPV										
<20 months	0.72	0.06	0.60	0.85	-1.96	.050*	3.206	<.001	-0.718	.076
20+ months	0.88	0.05	0.79	0.97						
Sensitivity										
<20 months	0.87	0.05	0.77	0.96	-0.80	.425^	2.393	.008	0.798	.215
20+ months	0.92	0.04	0.84	0.99						
Specificity										
<20 months	0.85	0.05	0.75	0.95	-0.53	.599^	1.983	.024	0.93	.176
20+ months	0.88	0.07	0.80	0.97						
Difference*										
Equivalence~	N = 48 for <20 month group									
Indeterminate^	N = 50 for 20 months group									

Table 6 Relative variable importance—CART model between younger and older toddler age groups (criterion: ASD diagnosis)—words produced included

Variable	Description	Importance
mcDiWFNTile	CDI word forms percentile	1.00
mchatAdjAge	Age at time of M-CHAT, adjusted for prematurity	0.71
mcDiWENTile	CDI word endings percentile	0.38
mfuFinal14PF	MCHAT/F item 14: responds to name	0.38
mcDiWPNTile	CDI words produced percentile	0.33
mfuFinal02PF	MCHAT/F item 2: interested in others	0.33
mcDiCXNTile	CDI sentence complexity percentile	0.29
mcDiM3LNTile	CDI mean 3 longest sentences percentile	0.24
mfuFinal21PF	MCHAT/F item 21: understands what people say	0.19
mfuFinal22PF	MCHAT/F item 22: stares at nothing/wanders	0.19
mfuFinal18PF	MCHAT/F item 18: unusual finger movements	0.14
mfuFinal08PF	MCHAT/F item 8: plays properly	0.10
mfuFinal20PF	MCHAT/F item 20: wonder if deaf	0.10
mfuFinal12PF	MCHAT/F item 12: Smiles in response	0.05
mchatFUResult	M-CHAT/F screening result	0.05
mfuFinal03PF	MCHAT/F item 3: likes climbing	0.05
mfuFinal15PF	MCHAT/F item 15: looks when you point	0.05
mfuFinal17PF	MCHAT/F item 17: looks at same thing	0.05
mfuFinal05PF	MCHAT/F Item 5: pretends	0.05
Gender	Gender	0.05
mcDiTooOld50	Child was too old for CDI	0.05

The CDI words produced subscale is likely to be too long to be included in a practical screening regimen. Therefore, the CART analyses were repeated excluding CDI words produced Score. While the relative importance of variables changes, the ROCs remained similar to the original model. Four of the top ten items were from the CDI when the words produced Scale was excluded (Table 7).

CART analyses using the combined criterion of ASD/Developmental Delay diagnosis again resulted in changes to relative variable importance and, as expected, higher ROCs. Using the combined criterion, the ASQ

developmental screen rose in relative importance (Tables 8, 9).

Discussion

Consistent with prior M-CHAT research this study supports use of the M-CHAT/F to identify M-CHAT false positives among toddlers older than 20 months of age, thereby substantially reducing the number of unnecessary referrals for comprehensive autism evaluation (Kleinman

Table 7 Relative variable importance—CART model between younger and older toddler age groups (criterion: ASD diagnosis)—no. words produced

Variable	Description	Importance
mcdiWFNTile	CDI word forms percentile	1.00
mchatAdjAge	Age at time of M-CHAT, adjusted for prematurity	0.74
mcdiWENTile	CDI word endings percentile	0.43
mfuFinal14PF	MCHAT/F item 14: responds to name	0.39
mfuFinal02PF	MCHAT/F item 2: interested in others	0.35
mcdiCXNTile	CDI sentence complexity percentile	0.30
mfuFinal21PF	MCHAT/F item 21: understands what people say	0.22
mfuFinal22PF	MCHAT/F item 22: stares at nothing/wanders	0.17
mcdiM3LNtile	CDI mean 3 longest sentences percentile	0.17
mfuFinal18PF	MCHAT/F item 18: unusual finger movements	0.17
mfuFinal08PF	MCHAT/F item 8: plays properly	0.09
mfuFinal20PF	MCHAT/F item 20: wonder if deaf	0.09
mfuFinal12PF	MCHAT/F item 12: smiles in response	0.09
mfuFinal15PF	MCHAT/F item 15: looks when you point	0.04
mfuFinal17PF	MCHAT/F item 17: looks at same thing	0.04
mfuFinal05PF	MCHAT/F item 5: pretends	0.04
mchatFUResult	M-CHAT/F screening result	0.04
mcdiTooOld50	Child was too old for CDI	0.04
mfuFinal10PF	MCHAT/F item 10: looks you in the eye	0.04

Table 8 Relative variable importance—CART model between younger and older toddler age groups (criterion: ASD/developmental delay diagnosis)—words produced Included

Variable	Description	Importance
mcdiM3LNtile	CDI mean 3 longest sentences percentile	1.00
mcdiWPNTile	CDI words produced percentile	0.72
mcdiWFNTile	CDI word forms percentile	0.68
mchatAdjAge	Age at time of M-CHAT, adjusted for prematurity	0.40
mcdiCXNTile	CDI sentence complexity percentile	0.36
mcdiTooOld50	Child was too old for CDI	0.20
mcdiWENTile	CDI word endings percentile	0.20
asq3Communication	ASQ3 communication scale	0.16
asq3FineMotor	ASQ3 fine motor scale	0.08
asq3GrossMotor	ASQ3 gross motor scale	0.08
mfuFinal01PF	MCHAT/F item 1: enjoys being swung	0.04
mfuFinal20PF	MCHAT/F item 20: wonder if deaf	0.04

Table 9 Relative variable importance—CART model between younger and older toddler age groups (criterion: ASD/developmental delay diagnosis)—no. words produced

Variable	Description	Importance
mcdiM3LNtile	CDI mean 3 longest sentences percentile	1.00
mcdiWFNTile	CDI word forms percentile	0.63
mchatAdjAge	Age at time of M-CHAT, adjusted for prematurity	0.40
mcdiCXNTile	CDI sentence complexity percentile	0.37
mcdiWENTile	CDI word endings percentile	0.30
mcdiTooOld50	Child was too old for CDI	0.20
asq3Communication	ASQ3 communication scale	0.17
asq3FineMotor	ASQ3 fine motor scale	0.10
asq3GrossMotor	ASQ3 gross motor scale	0.10
mfuFinal01PF	MCHAT/F item 1: enjoys being swung	0.03
mfuFinal20PF	MCHAT/F item 20: wonder if deaf	0.03

et al. 2008; Robins et al. 2001). However, this study also shows that use of the M-CHAT/F does not substantially improve the relatively low PPV of the M-CHAT in low risk toddlers younger than 20 months, confirming the published M-CHAT study (Pandey et al. 2008) which compared screening results between younger and older toddlers, and included the use of the M-CHAT/F. Importantly, our results better represent the true performance of the M-CHAT/F for younger toddlers because, in contrast to these other studies, the M-CHAT/F was completed at nearly the same time as the 18-month check-up rather than via telephone interviews an average of 3 months later when tests for ASD all have higher PPV (Robins 2008). The implication is that recommendations of methods for autism screening at the time of the standard 18-month visit may need to differ from those for the 24-month check-up. These issues are likely to be less relevant to populations of known at risk children, such as those already being followed by early intervention programs. Further exploration of these data suggest that including additional language and developmental data might provide a broader range of variables needed to differentiate these age groups.

Another study of screening 18-month-olds in Japan attempted to improve prediction by lowering the threshold for a positive M-CHAT (using one critical item instead of two) and completing the follow-up interview by phone at 19–20 months (Kamio et al. 2014). The lower threshold and somewhat different scoring of the M-CHAT resulted in a high referral rate (17%) but the PPV was still only 0.46. Strategies other than simply lowering the threshold are needed to make ASD screening as accurate for younger toddlers as it is for older ones, especially since referral rates are already higher for the younger toddlers as seen in this study and that of Pandey et al. (2008). The false discovery rate was twice as high in the younger group than the older one in this study making referral of all positives impractical.

It should be noted that while use of a follow-up interview may improve PPV and conserve scarce evaluative resources, it will not do anything to improve sensitivity and may even overlook children who ultimately have an ASD diagnosis identified by the wider filter of the initial parent response items (Stenberg et al. 2014). These results underline the need to consider early identification as an iterative process over time rather than counting on a single screen. When the limitations of current screening tools are recognized, strategies may be considered for including the screening tools as part of a promising system of “developmental surveillance” using discrete observations by specially trained PCPs (Barbaro and Dissanayake 2010). While there is already a general recommendation by the AAP for ongoing surveillance in addition to screening, it may be time to consider bolstering the surveillance component

with this more targeted and enhanced approach, at least for the younger toddlers. In addition, clinicians need to be aware that prediction improves after 20 months. There may be some cases when re-screening the children who fail the M-CHAT/F at the 18-month visit 2 months later might be deemed appropriate. This would, of course, not include cases where there is reason for concern based on issues raised through a surveillance process or current parent or clinician concern.

The results of this study and that of Pandey and colleagues are quite similar, with PPV for predicting a clinical diagnosis of ASD in this study of 0.69 for the older toddlers and 0.36 for the younger ones, compared to 0.61 and 0.28 for the corresponding age groups of Pandey. These studies also point to differences in PPV when considering children with either developmental delays or ASD outcomes (ASD/DD) using a common criteria for offering developmental services in U. S. government sponsored early intervention services to define developmental difficulties. For the combined ASD/DD criterion PPV was higher for the older age group in both studies, 0.96 in this study and 0.95 in the Pandey study, compared to 0.71 and 0.72 in the younger age group. This suggests that, while most of the children being referred from M-CHAT screening at the 18 month visit may be found to have a developmental difficulty, the difficulty is less likely to be ASD than something else. This does not support the claim that the M-CHAT is an autism specific screen at this age. While the M-CHAT is one of the tools considered to be “autism specific” rather than a “broad band” developmental screen, our study confirms that children failing the M-CHAT and M-CHAT/F screen who are not determined to have an ASD diagnosis are still likely to have developmental difficulties severe enough to qualify for early intervention services. We may, therefore, want to think of the MCHAT as a broader clinical developmental screening tool that has the added benefit of picking up on autism—rather than counting on the tool to be the complete answer for autism identification.

There are a number of possible explanations regarding the difficulty in effectively detecting ASD in younger children. For example, the age of onset of symptoms varies, with 22–41% of cases regressing from apparently typical development or only mild delays to signs of ASD after 16 months (Barger et al. 2013). In those cases, most show clear signs by 24 months, warranting the second recommended screening age (Academy of Pediatrics 2006). In addition, atypical odd and repetitive behaviors may not appear until the end of the second year (e.g., Veness et al. 2012). Prospective studies indicate that ASD symptoms may emerge gradually in the toddler age group (Landa et al. 2007, 2013; Ozonoff et al. 2008). Specifically children misclassified at 18 months in a prospective studies of siblings of a proband with ASD were higher functioning, and their

autism symptoms increased between 18 and 36 months (Chawarska et al. 2014). Children who present for asymptomatic screening are more likely to be higher functioning than those who are identified because of specific parental concern simply because they have more reassuring developmental features as well as less autistic features. The yes/no format of the M-CHAT may also not capture the “sometimes” status of the behavior at a point in time when it is not fully emerged through development and may be better captured by a more dimensional screen such as the Quantitative or Q-CHAT (Allison et al. 2008). However, in addition to the children who are actually typical before regression, it also may not be reasonable to expect some bright children with mild ASD to be detectable as young toddlers.

Between 18 and 24 months is a time of great expressive language emergence. Evidence of adequate language development appeared to be an important differentiator between children who did not show evidence for ASD from those who did. This study provides evidence that beginning use of more advanced language forms as measured by a MacArthur-Bates Communicative Development Inventory has potential to be one of the major factors differentiating risk from non-risk for ASD in both age groups. For children who are positive on the M-CHAT screen for autism phenotype, additional items indicating that the child is beginning to exhibit language forms that are just emerging at that developmental age contribute to more accurate decision rules for autism evaluation referrals. However, use of language measures standardized by age have not been a part of autism specific screening tests such as M-CHAT, which uses the same items and same scoring across this entire age range. Some of the additional language items which improve prediction are relatively brief ones, but the CDI vocabulary inventory (words produced subscale) is probably too lengthy (680 words) to be considered practical for inclusion in a screening battery, even as a follow-up. However, existing short forms should be explored and, recently, an even more abbreviated computer adaptive version has been developed which is highly correlated with the longer version (Makransky et al. 2016). Issues of efficiency should be weighed in the development of screening tests suitable for younger toddlers. However, even without the use of this vocabulary subscale, other briefer CDI items still enhance prediction for the youngest group of toddlers almost as well. In addition, if the salience of language items in multivariate models predicting autism diagnoses in toddlers can be replicated, this may have implications for rethinking the centrality of language in the way autism is defined, such as when comparing the prominence of language functioning in DSM-4 criteria vs. its lesser role in DSM-5.

While the ASQ developmental screening items did not significantly contribute beyond the CDI items to improve prediction of autism for young toddlers, this routine screen

does make a significant contribution to a model predicting developmental delay as well as ASD for children who were positive on the M-CHAT screening test. It should be noted that the screening power of the language items found in this study may result from language being a better proxy for overall developmental level or IQ than the ASQ. This study also suggests that a different pattern of behaviors may characterize emerging ASD in the 18-month visit age group than seen as toddlers approach age 2 years. In addition, demonstration of more advanced language abilities at the younger toddler age may reassure that there is no ASD diagnosis obviating the need for additional assessments.

The decision trees developed through CART analyses models are too exploratory at this point for clinical application. If replicated, computer systems presenting screening items to parents could vary item presentation depending on the pattern of prior item responses to make a more accurate screening process. Some parents will have a short screening session and others somewhat longer with the variable pattern of item presentation generated through an algorithmic item selection process.

Conclusions

The known heterogeneous etiology of ASD (e.g., Scherer and Dawson 2011) and its dynamic natural history present challenges that require different screening and assessment strategies at different ages. The two age groups described in this study are of practical relevance to child health care delivery because they correspond to two of the standard well-child visits that are ones at which autism screening is recommended by the American Academy of Pediatrics (18 and 24 months) (American Academy of Pediatrics 2006).

There is an understandable desire of screening test developers to employ simple scoring such as a total score for all ages of children. However, such ease of use comes at the cost of precision. While confirmation of the strategies suggested in this study is essential before clinical application, optimal scoring is likely to be complex, differ by age and gender and be difficult to remember and score unless aided by a computer. Computer-aided parent report screening technologies now allow for more sophisticated algorithms without placing undue burden on parents or clinicians.

Limitations

There are limitations to this study that must be considered when interpreting results. First, while a community pediatric practice population was sampled, there was no attempt to assess children passing the screening. Also, the subsample consenting and completing diagnostic

testing was skewed by not being fully representative since many parents declined participation. While there was an attempt to identify missed cases by recruiting all children who screened positive before the follow-up interview, the estimate of sensitivity for the entire community sample remains artificially high because all the screened negative children were not evaluated. Fewer than half of the PCPs in participating offices joined the study and administered a M-CHAT/F when indicated, and only a third of M-CHAT/F positives were evaluated at the KKI Autism center. Additionally, the sample size is relatively small. However, the sampling procedures at the 18-month visit were the same as at 24 months with similar demographic characteristics suggesting that these sampling issues are unlikely to confound the issue of age differences, the central issue of this report. However, by using the office assignment to a “18 month” or “24 month visit” rather than limiting subject enrollment by actual age, the sample included a few children who were a bit younger and a bit older than the standardization of the test. It should also be noted that the version of the M-CHAT used here was the 23-item earlier version, but we do not have a reason to believe that the slight item and wording modifications would affect the younger toddler age group differently than the older group.

Acknowledgments This project was funded by a grant from the NIMH Grant No. R44MH085399. The participating primary care pediatricians and their office staffs in the state of Maryland are acknowledged for their cooperation during the project and continuing implementation of the screenings described after the conclusion of the project.

Author Contributions RS and BH conceived of the study, participated in its design and coordination and drafted the manuscript; RL participated in the design and interpretation of the data; TM participated in the design and coordination of the study. KW and DM participated in the coordination of the study and assisted with the measurement; PB participated in the design of the study and performed the statistical analysis; All authors read and approved the final manuscript.

Compliance with Ethical Standards

Disclosure 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. The terms of the arrangement were managed by The Johns Hopkins University in accordance with its conflict of interest policies.

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.

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.
- American Academy of Pediatrics, Committee on Children With Disabilities (1994). Developmental surveillance and screening of infants and young children. *Pediatrics*, *93*, 863.
- American Academy of Pediatrics (2006). Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics*, *118*(1), 405–420.
- Autism Speaks (2016). *Modified checklist for autism in toddlers, revised (M-CHAT-R)*. 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. (2000). 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*, *39*(6), 694–702. doi:10.1097/00004583-200006000-00007.
- Barbaro, J., & Dissanayake, C. (2010). Prospective identification of autism spectrum disorders in infancy and toddlerhood using developmental surveillance: The social attention and communication study. *Journal of Developmental & Behavioral Pediatrics*, *31*(5), 376–385. doi:10.1097/DBP.0b013e3181df7f3c.
- Barger, B. D., Campbell, J. M., & McDonough, J. D. (2013). Prevalence and onset of regression within autism spectrum disorders: A meta-analytic review. *Journal of Autism and Developmental Disorders*, *43*(4), 817–828. doi:10.1007/s10803-012-1621-x.
- Baron-Cohen, S., Allen, J., & Gillingham, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *British Journal of Psychiatry*, *161*, 839–843.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth Statistical Press.
- CHADIS (2017). CHADIS (Child Health and Development Interactive System). <http://www.chadis.com/site/>. Accessed 15 July 2017.
- Chawarska, K., Shic, F., Macari, S., Campbell, D. J., Brian, J., Landa, R., Hutman, T., Nelson, C. A., Ozonoff, S., Tager-Flusberg, H., Young, G. S., Zwaigenbaum, L., Cohen, I. L., Charman, T., Messinger, D. S., Klin, A., Johnson, S., Bryson, S. (2014). 18-Month predictors of later outcomes in younger siblings of children with autism spectrum disorder: A baby siblings research consortium study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(12), 1317–1327. doi:10.1016/j.jaac.2014.09.015.
- Chlebowski, C., Robins, D. L., Barton, M. L., & Fein, D. (2013). Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics*, *131*(4), e1121–e1127. doi:10.1542/peds.2012-1525.
- Christensen, D. L., Baio, J., Van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J. N., Daniels, J., Durkin, M. S., Fitzgerald, R. T., Kurzius-Spencer, M., Lee, L. C., Pettygrove, S., Robinson, C., Schulz, E., Wells, C., Wingate, M. S., Zahorodny, W., & Yeargin-Allsopp, M. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *MMWR Surveillance Summaries*, *65*(3), 1–23. doi:10.15585/mmwr.ss6503a1.

- Core Team (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Fenson, L., Marchman, V. A., Dale, P. S., Reznick, S., Thal, D., & Bates, E. (2007). *The MacArthur Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore: Paul H. Brookes Publishing Co.
- Fountain, C., King, M. D., & Bearman, P. S. (2011). Age of diagnosis for autism: individual and community factors across 10 birth cohorts. *Journal of Epidemiology and Community Health*, 65(6), 503–510. doi:10.1136/jech.2009.104588.
- Halladay, A. K., Bishop, S., Constantino, J. N., Daniels, A. M., Koenig, K., Palmer, K., Messinger, D., Pelphrey, K., Sanders, S. J., Singer, A. T., Taylor, J. L., Szatmari, P. (2015). Sex and gender differences in autism spectrum disorder: summarizing evidence gaps and identifying emerging areas of priority. *Molecular Autism*, 6, 36. doi:10.1186/s13229-015-0019-y.
- Hardy, S., Haisley, L., Manning, C., & Fein, D. (2015). Can screening with the ages and stages questionnaire detect autism? *Journal of Developmental and Behavioral Pediatrics: JDBP*, 36(7), 536–543. doi:10.1097/DBP.0000000000000201.
- Howlin, P., Magiati, I., & Charman, T. (2009). Systematic review of early intensive behavioral interventions for children with autism. *American Journal on Intellectual and Developmental Disabilities*, 114(1), 23–41. doi:10.1352/2009.114:23-41.
- Kamio, Y., Inada, N., Koyama, T., Inokuchi, E., Tsuchiya, K., & Kuroda, M. (2014). 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*, 44(1), 194–203. doi:10.1007/s10803-013-1864-1.
- Kleinman, J. M., Robins, D. L., Ventola, P. E., Pandey, J., Boorstein, H. C., Esser, E. L., & Barton, M. (2008). The modified checklist for autism in toddlers: a follow-up study investigating the early detection of autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 38(5), 827–839. doi:10.1007/s10803-007-0450-9.
- Landa, R. J., Gross, A. L., Stuart, E. A., & Faherty, A. (2013). Developmental trajectories in children with and without autism spectrum disorders: the first 3 years. *Child Development*, 84(2), 429–442. doi:10.1111/j.1467-8624.2012.01870.x.
- Landa, R. J., Holman, K. C., & Garrett-Mayer, E. (2007). Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. *Archives of General Psychiatry*, 64, 853–864. doi:10.1001/archpsyc.64.7.853.
- Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., Bishop, S., Esler, A., Hus, V., Oti, R., Richler, J., Risi, S., & Lord, C. (2009). The Autism Diagnostic Observation Schedule—Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(9), 1305–1320.
- Makransky, G., Dale, P. S., Havmose, P., Bleses, D. (2016). An IRT-based, computerized adaptive testing version of the MacArthur-Bates Communicative Development Inventory: Words & sentences (CDI:WS). *Journal of Speech, Language, and Hearing Research*, 59(2), 281–289.
- Mullen, E. M. (1995). *The Mullen Scales of Early Learning: AGS edition*. Circle Pines, MN: American Guidance Service, Inc.
- National Research Council (2001). *Educating children with autism*. Washington, DC: National Academies Press.
- Nectac (2016). *States and territories' definitions of/criteria for IDEA Part C eligibility*. Updated March 4, 2015; Accessed September 29, 2016, from http://nectac.org/pdfs/topics/earlyid/partc_elig_table.pdf. Nectac.org.
- Ozonoff, S., Heung, K., Byrd, R., Hansen, R., & Hertz-Picciotto, I. (2008). The onset of autism: Patterns of symptom emergence in the first years of life. *Autism Research*, 1(6), 320–328. doi:10.1002/aur.53.
- Pandey, J., Verbalis, A., Robins, D. L., Boorstein, H., Klin, A., Babitz, T., & Fein, D. (2008). Screening for autism in older and younger toddlers with the Modified Checklist for Autism in Toddlers. *Autism: The International Journal of Research and Practice*, 12(5), 513–535. doi:10.1177/1362361308094503.
- Reznick, J. S., Baranek, G. T., Reavis, S., Watson, L. R., & Crais, E. R. (2007). A parent-report instrument for identifying one-year-olds at risk for an eventual diagnosis of autism: the first year inventory. *Journal of autism and developmental disorders*, 37(9), 1691–1710.
- Robins, D. L. (2008). Screening for autism spectrum disorders in primary care settings. *Autism: The International Journal of Research and Practice*, 12, 537.
- Robins, D. L., Casagrande, K., Barton, M., Chen, C-MA., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics*, 133(1), 37–45. doi:10.1542/peds.2013-1813.
- Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 31(2), 131–144. DOI:10.1177/1362361308094502.
- Scherer, S. W., & Dawson, G. (2011). Risk factors for autism: translating genomic discoveries into diagnostics. *Human Genetics*, 130(1), 123–148. doi:10.1007/s00439-011-1037-2.
- Siu, A. L., & The US Preventive Services Task Force (USPSTF). (2016). Screening for autism spectrum disorder in young children. US preventive services task force recommendation statement. *JAMA: The Journal of the American Medical Association*, 315(7), 691–696. doi:10.1001/jama.2016.0018.
- Squires, J., & Bricker, D. (2009). *Ages & Stages Questionnaires, Third Edition (ASQ-3)*. Baltimore: Brookes Publishing. <http://www.agesandstages.com>.
- StataCorp (2015). *Stata statistical software: Release 14*. College Station, TX: StataCorp LP.
- Stenberg, N., Bresnahan, M., Gunnes, N., Hirtz, D., Hornig, M., Lie, K. K., & Schjølberg, S. (2014). Identifying children with autism spectrum disorder at 18 months in a general population sample. *Paediatric and Perinatal Epidemiology*, 28(3), 255–262. doi:10.1111/ppe.12114.
- Sturner, R., Howard, B., Bergmann, P., Morrel, T., Andon, A., Marks, D., Rao, P., & Landa, R. (2016). Autism screening with online decision support by primary care pediatricians aided by M-CHAT/F. *Pediatrics*, 138(3), e20153036. doi:10.1542/peds.2015-3036.
- Therneau, T., Atkinson, B., & Ripley, B. (2015). *rpart: Recursive partitioning and regression trees. R package version 4.1–9*. <https://cran.r-project.org/web/packages/rpart/rpart.pdf>.
- Veness, C., Prior, M., Veness, C., Eadie, P., Cini, E., Reilly, S. (2012). Early indicators of autism spectrum disorders at 12 and 24 months of age: a prospective, longitudinal comparative study. *Autism*, 16(2), 163–177.