See the Video Abstract at jdbp.org

Sensitivity and Specificity of 2 Autism Screeners Among Referred Children Between 16 and 48 Months of Age

Louisa A. Salisbury, MD,* Jonathan D. Nyce, MD,† Charles D. Hannum, MD,‡
R. Christopher Sheldrick, PhD,§ Ellen C. Perrin, MD‡

ABSTRACT: ***Objective:*** **Autism screening is recommended by the American Academy of Pediatrics and the Centers for Disease Control and Prevention at ages 18 and 24 months. Popular screening tests have been validated for the age range of 16 to 30 months. However, only a minority of children with autism spectrum disorder (ASD) are identified by age 3 years, and many are not identified until after they enter school. Thus, we aimed to measure the sensitivity and specificity of 2 available screening tests for ASDs in children older than 30 months.** ***Methods:*** **We assessed the sensitivity and specificity of 2 ASD screening tools administered to parents of children who were referred to a developmental clinic between the ages of 16 and 48 months: the Modified Checklist for Autism in Toddlers (M-CHAT) and the Parent's Observations of Social Interactions (POSI), which is a component of a comprehensive screening instrument called, the Survey of Well-being of Young Children.** ***Results:*** **Both the M-CHAT and the POSI had acceptable sensitivity (≥75%) among children across the age range studied. Their specificity was limited by the fact that the study was conducted in a developmental referral clinic.** ***Conclusion:*** **Two readily available screening tools, the POSI and the M-CHAT, have acceptable sensitivity in evaluating risk for autism in children at least to age 48 months. Further research should investigate their sensitivity and specificity when used in primary care settings.**

(*J Dev Behav Pediatr* 39:254–258, 2018)

Screening for autism spectrum disorder (ASD) is recommended by the Centers for Disease Control and Prevention (CDC) and the American Academy of Pediatrics (AAP) for children between 18 and 30 months old.[1–3] Pediatricians increasingly report systematic screening at recommended ages.[4] Nevertheless, the average age of diagnosis of ASD in the United States remains above 4 years, which is higher than the age recommended for optimal initiation of interventions.[5–7] Although reports of the average age of diagnosis are sensitive to survey methodology and vary across racial, socioeconomic, and geographic ranges of populations studied, recent evidence suggests that only a minority of children with ASD are identified before their third birthday (when eligibility ends for Part C Early Intervention services), and one-third to one-half are identified after school age.[8] In particular, systematic delays in the age of diagnosis are documented among children with minority ethnic/racial backgrounds, low socioeconomic status, and those with milder symptoms.[5,9]

Screening at or before a child's second birthday (as recommended by the AAP and the CDC) is clearly 1 effective approach to reducing the age of diagnosis for ASD.[3,4] However, early screening will not detect all children with ASD for at least 2 reasons. First, it is difficult at best for any program of screening to achieve 100% sensitivity. Not all children attend pediatric visits.[10] When they do, not all complete evidence-based screening instruments, often because of literacy or language barriers.[11] The sensitivity of ASD screeners themselves fall below 100%, and some proportion of children who screen positive typically do not follow through with recommended evaluations.[12] Together, these factors demonstrate the challenges to identifying all children with ASD.

Second, symptoms may not be fully detectable before age 30 months for all children with ASD. Longitudinal studies of children at high risk of ASD demonstrate that some children who do not meet diagnostic criteria at 18 months old go on to qualify for an ASD diagnosis at age 3 years.[13] Children with milder symptoms of ASD are typically identified later than those with more severe symptoms.[8] Children who demonstrate milder symptoms at early ages yet ultimately experience impairment attributable to ASD present a challenge to the effectiveness of early screening.[7]

From the *Department of Pediatrics, Mammoth Hospital, Mammoth Lakes, CA; †Kravis Children's Hospital at Mount Sinai, New York, NY; ‡Division of Developmental-Behavioral Pediatrics, Floating Hospital for Children, Tufts Medical Center, Boston, MA; §Boston University School of Public Health, Boston, MA.

Received June 2017; accepted November 2017.

Disclosure: E. C. Perrin and R. C. Sheldrick are among the coauthors of the Parent's Observations of Social Interactions (POSI), which is a component of the Survey of Well-being of Young Children (SWYC). The remaining authors declare no conflict of interest.

Address for reprints: Ellen C. Perrin, MD, 800 Washington Street, #854, Tufts Medical Center, Boston, MA 02111; e-mail: eperrin@tuftsmedicalcenter.org.

Logically, reducing the age of diagnosis requires identifying children before their risk would otherwise have been identified. Thus, screening can be helpful at any age if it is effective in detecting cases that were missed at earlier ages. Based on recent evidence regarding the wide distribution of ages at which children with ASD are first identified,[14] some have called for later ASD screening to supplement earlier efforts.[15] Unfortunately, there is a significant evidence gap with respect to the accuracy of ASD screening instruments for children after age 30 months.[16]

The objective of this study was to evaluate the accuracy of 2 currently available screening tools for the detection of ASD—the Parent's Observations of Social Interactions (POSI) and the Modified Checklist for Autism in Toddlers (M-CHAT)—among children who were referred for evaluation between 31 and 48 months of age as well as children who were referred in the recommended age range of 16 to 30 months.

The M-CHAT is a freely available widely used screener for ASD.[16–19] Although an evidence review conducted by the United States Preventive Health Services Task Force (USPSTF) was unable to assess its sensitivity due to variability in methodology and populations across studies, it did conclude that the M-CHAT was effective in detecting ASD with a positive predictive value approaching 50% among children 16 to 30 months old, when combined with the M-CHAT follow-up interview.[20] Also freely available, the POSI is a brief screener for ASD that was initially evaluated in both a referral center and a primary care sample and compared with the M-CHAT.[21] Findings from this study suggested that among children aged 16 to 30 months, the POSI is more sensitive than the M-CHAT (estimates of sensitivity = 83% in primary care and 89% in a referral clinic compared with 50% and 71%, respectively, for the M-CHAT) but less specific (estimates of specificity = 74% in primary care and 54% in a referral clinic compared with 84% and 62%, respectively, for the M-CHAT).[21]

At the time of this study, neither checklist had been systematically evaluated beyond the age of 30 months, and we hypothesized that both would demonstrate adequate sensitivity among children at least up to 48 months of age, thus including most children referred for language and general developmental delays, and all children who are 3 years old. Although evidence of screening accuracy from referred populations should be interpreted with caution because biases are likely to inflate estimates of accuracy (as noted by the USPSTF[20]), results may be useful as an indicator of the "upper bound" of a screener's sensitivity in the general population.

METHODS

Participants and Procedures

We conducted a retrospective chart review of children referred to a developmental-behavioral pediatric clinic between July 1, 2010, and June 30, 2013, for evaluations to clarify diagnoses of developmental delays, autism spectrum disorder, and other neurobehavioral disorders. The study was approved by the institutional review board of the medical center. Before their evaluation, parents were routinely asked to complete an intake packet that included several checklists including both the Parent's Observations of Social Interactions (POSI) and the Modified Checklist for Autism in Toddlers (M-CHAT) screening tools. Children then received comprehensive evaluations that included a thorough history, physical examination, observation of play and parental interactions, and direct testing. Testing was adapted to the specific referral question and often included the Autism Diagnostic Observation Schedule, Preschool Language Scale, and Mullen Scales of Early Learning. In light of testing results and in consultation with other professionals (e.g., speech/language pathologist, clinical psychologist, social worker, as indicated), board-certified developmental-behavioral pediatricians assigned diagnoses based on clinical judgment.

The charts of all children aged 16 to 48 months who were seen for evaluation at the center over the 3-year period were reviewed. Adequate literacy in English was required. Demographic information including insurance type, primary language, prematurity, and maternal education was collected from the patient's chart. Additional data included scores of POSI and M-CHAT screening tools, and the final diagnoses based on the full evaluation, as recorded by the clinician in the clinic's electronic medical record system.

The M-CHAT is a parent-reported checklist of a child's behavior that includes 23 yes/no questions with 6 "critical" questions (at the time of this study, the revised M-CHAT was not yet in wide use). The M-CHAT defines children who are "at risk" as those who fail 2 or more of the critical questions or 3 or more of the total 23 items. To increase the specificity of the test, the M-CHAT includes a follow-up interview to clarify survey responses and reduce false positive responses. No follow-up interview was performed because this was a retrospective chart review of cases that were all followed by a full evaluation.

The POSI is a parent-reported screening tool developed by a multi-institutional expert panel of developmental-behavioral pediatricians and psychologists with the aim of creating a short screening test for autism that would be both accurate and practical for use by primary care pediatricians. The development and design of the POSI screening tool are described in detail in an earlier publication.[21] The POSI is 1 component of a comprehensive developmental screening instrument called the Survey of Well-being of Young Children, which is available at www.TheSWYC.org. The POSI includes 7 questions, each with 5 possible responses. Following standard procedures, children are scored as "positive" if 3 or more answers are in the last 3 columns, thus requiring further evaluation.

Analytic Strategy

All data were double entered to ensure accuracy. To maximize sensitivity for both the M-CHAT and the POSI, responses in which both "yes" and "no" were marked, or in which written comments clearly described the

indicated behavior, were coded as positive. Screeners were scored if responses to no more than 2 items were missing. Demographic information and scored checklists were entered separately by coders who were unaware of the child's diagnoses. Rare ambiguities in the diagnoses that were recorded in the patient charts were resolved in discussion with the senior author, who was unaware of the screening test scores.

For both the M-CHAT and the POSI, sensitivity (the proportion of positive screens among children with ASD) and specificity (the proportion of negative screens among children without ASD) were calculated separately for children ages 16 to 30 months and children ages 31 to 48 months, and 95% confidence intervals were calculated based on a Clopper-Pearson interval of the binomial distribution. Because the study was conducted in a sample of clinically referred children, the prevalence of developmental disabilities—including but not limited to ASD—was expected to be high. Thus, we expected low specificity compared with studies of the use of screeners in general populations. Chi-square tests of independent proportions were used to compare sensitivities and specificities across age groups, and within each age range; the performance of the POSI and the M-CHAT were compared using McNemar tests of dependent proportions. In addition, Cronbach's alpha was used to provide a general estimate of the lower bound of reliability for each screener. All analyses were conducted in Stata version 12.

RESULTS

Complete data were obtained for 524 children between the ages of 16 and 48 months at the time of diagnostic evaluation; 39 (7.5%) were excluded based on insufficient data in their charts. Demographic characteristics of the 485 cases with complete data are presented in Table 1. Seventy-seven percent of the sample were boys, 37% reported nonwhite race, 18% reported Hispanic ethnicity, and 20% spoke a primary language other than English. As proxies for socioeconomic status, 46% had Medicaid and 7% had maternal education less than high school graduation. In addition, 19% had a history of prematurity. With the exception of age, there were no statistically significant differences in demographics between the older and younger age groups.

In our study sample, 61.2% (n = 297) of children were given a diagnosis of autism spectrum disorder (ASD). Among these children, 91% had a positive Autism Diagnostic Observation Schedule (ADOS) result—similar to the co-occurrence of positive test scores and clinical diagnoses of ASD in studies of ADOS validity.[22] Among children who did not receive a diagnosis of ASD, 56.5% were diagnosed with global developmental delay. In the younger age group (16–30 months), both the Parent's Observations of Social Interactions (POSI) and the Modified Checklist for Autism in Toddlers (M-CHAT) displayed sensitivities that were reliably above 70%, and, as expected, both displayed comparatively low specificity in our clinically referred sample (Table 2). In this age group, the POSI displayed higher sensitivity than the M-CHAT ($p < 0.01$), whereas the M-CHAT displayed higher specificity than the POSI ($p < 0.05$). Among children aged 31 to 48 months, the POSI displayed a sensitivity of 75.0%, which was lower than in the younger age group ($p < 0.001$) but still reliably above 70%. In this older age group, the M-CHAT had a sensitivity of 69.4%, and as expected, both screeners displayed comparatively low specificity. These estimates were not statistically different from estimates in the younger age group (Table 2). Although comparisons between the POSI and the M-CHAT in the older age group were not statistically significant, the POSI displayed higher sensitivity ($p = 0.34$), whereas the M-CHAT displayed higher specificity ($p = 0.13$).

Table 1. Demographic Information

	Ages 16–30 mo	Ages 31–48 mo	*p*
Complete data	271	214	
Sex, n (%)			
Female	68 (25)	44 (21)	
Male	203 (75)	170 (79)	n.s.
Age at survey, mo, mean (SD)	24.8 (4.0)	36.8 (4.3)	*
Hispanic ethnicity, n (%)	51 (19)	35 (16)	n.s.
Race			
African American	32 (12)	21 (10)	
Asian	15 (6)	18 (8)	
White	164 (61)	127 (59)	
Other	11 (4)	10 (5)	
Not reported	59 (22)	38 (18)	n.s.
English as first language, n (%)	219 (81)	167 (78)	n.s.
Insurance, n (%)			
Private	129 (48)	93 (44)	
Medicaid	123 (46)	97 (46)	
Unknown	19 (8)	24 (10)	n.s.
Birth history, n (%)			
Premature (<37 wk)	61 (23)	31 (15)	
Term	186 (69)	161 (76)	
Unknown	22 (8)	20 (9)	n.s.
Mother's education, n (%)			
Less than high school	24 (10)	12 (7)	
High school graduate	64 (26)	65 (36)	
Some college education	53 (21)	31 (17)	
College graduate	71 (29)	43 (24)	
Graduate education	35 (14)	29 (16)	
Unknown	24 (9)	34 (16)	n.s.

*$p < 0.001$.
n.s., not significant.

Table 2. Sensitivity and Specificity of M-CHAT and POSI in 2 Age Groups

Sensitivity		POSI	M-CHAT	*p*, McNemar Test (POSI vs M-CHAT)
Ages 16–30 mo	173 diagnosed with ASD	93.6% (95% CI: 88.9%–96.8%)	77.5% (95% CI: 70.5%–83.5%)	<0.001
Ages 31–48 mo	124 diagnosed with ASD	75.0% (95% CI: 66.4%–82.3%)	69.4% (95% CI: 60.4%–77.3%)	0.34
p value, χ^2 test (older vs younger)		<0.001	0.12	

Specificity		POSI	M-CHAT	*p*, McNemar Test (POSI vs M-CHAT)
Ages 16–30 mo	98 without ASD diagnosis	40.8% (95% CI: 31%–51.2%)	54.1% (95% CI: 43.7%–64.2%)	0.02
Ages 31–48 mo	90 without ASD diagnosis	47.8% (95% CI: 37.1%–58.6%)	58.9% (95% CI: 48%–69.2%)	0.13
p value, χ^2 test (older vs younger)		0.34	0.51	

ASD, autism spectrum disorder; CI, confidence interval; M-CHAT, Modified Checklist for Autism in Toddlers; POSI, Parent's Observations of Social Interactions.

DISCUSSION

Our primary aim in this study was to evaluate whether 2 currently available autism spectrum disorder screening tools, the POSI and the Modified Checklist for Autism in Toddlers (M-CHAT), are sufficiently accurate to be useful among children aged 31 to 48 months. We also examined their accuracy in the recommended screening age range of 16 to 30 months.

Findings are consistent with the hypothesis that both the POSI and the M-CHAT have acceptable sensitivity in both of these age groups, thus warranting further study in primary care pediatric populations. Comparable with previous estimates, the M-CHAT displayed higher specificity than the POSI among children aged 16 to 30 months, whereas the POSI displayed higher sensitivity than the M-CHAT. Results were similar in the older age group—the M-CHAT displayed higher specificity, but the POSI displayed higher sensitivity. Although both screeners may be suitable for screening among either younger or older children, decisions need to be made regarding whether to prioritize higher sensitivity or higher specificity. Prioritizing sensitivity will result in fewer missed cases (false negatives), whereas prioritizing specificity will result in fewer false positive cases. Thus, such decisions may depend on the availability of follow-up resources.

The main limitation of our study is that the sample consisted of children who had been referred to a subspecialty clinic. Thus, the prevalence of ASD in our sample was high, and cases may have differed in unknown ways from unidentified cases in the general population. For example, because parents have already been referred and agreed to a developmental evaluation, they may be sensitized to ASD symptoms in a way that is not true in the general population. Thus, sensitivities of both tools in our population may be inflated and should be interpreted conservatively as an upper bound.

An additional limitation is that the recommended follow-up interview was not performed after determining a positive M-CHAT score. It should be noted that the purpose of a follow-up interview is to increase the positive predictive value by reducing false positives. A follow-up interview administered exclusively to children who screen positive can raise specificity, but it can only lower sensitivity. Thus, the inclusion of the M-CHAT follow-up interview would have increased the advantage of the M-CHAT over the POSI regarding specificity, but at the same time, it is likely to have magnified the advantage of the POSI over the M-CHAT regarding sensitivity. Furthermore, the revised M-CHAT format and scoring guidelines were not included in this study. Although we know of no direct comparisons between the original and revised versions of the M-CHAT, it is reasonable to believe that the use of the newer version might have resulted in somewhat greater accuracy.

Specificities in our study for both screening tools in both age groups were relatively low for similar reasons. Given that most of our study population had some form of developmental delay, with symptoms overlapping with those that are typical of ASD, the children with potential ASD were compared with a background population of children with delays rather than typically-developing children. In such cases, more children without ASD will score positive, thus lowering estimates of sensitivity. However, both screening tools should be influenced equally, so comparisons between tools should be unaffected. Future studies in primary care settings are well positioned to address this limitation because the lower limit of an instrument's specificity is equivalent to 100% minus the proportion who score positive in primary care.

Our findings have important implications when considered in the context of a growing body of data identifying children with ASD who were not detected by screening at 18 and 24 months old, including a subset of children later recognized as having an ASD who did not

meet criteria by age 2 years. Our results demonstrate that 2 readily available screening tools, the POSI and the M-CHAT, can be useful in evaluating these children beyond the recommended screening age of 30 months. Ongoing research evaluating these screening tools in a primary care context will provide data from a more generalizable sample to guide their use for routine screening for ASD at least up to the 4-year health supervision visit.

ACKNOWLEDGMENTS

The authors are indebted to Elizabeth Bayne for her energetic and thoughtful assistance in data collection and to Maire Diemer and Daryl Garfinkel for their assistance in the preparation of the manuscript.

REFERENCES

1. Johnson CP, Myers SM; American Academy of Pediatrics Council on Children With Disabilities. Identification and evaluation of children with autism spectrum disorders. *Pediatrics*. 2007;120:1183–1215.
2. Centers for Disease Control. *Learn the Signs. Act*; 2016. Available at: https://www.cdc.gov/ncbddd/actearly. Accessed November 30, 2017.
3. Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Pediatrics*. 2006;1:405–420.
4. Lipkin P, Macias M, Hyman S, et al. Identification of children <36 months at risk for developmental delay/autism: results of National survey of pediatricians. Presentation at: Society for Developmental Behavioral Pediatrics Annual Meeting; September 2016; Savannah, Georgia.
5. Daniels AM, Mandell DS. Explaining differences in age at autism spectrum disorder diagnosis: a critical review. *Autism*. 2014;18:583–597.
6. Prevalence of autism spectrum disorder among children aged 8 years: autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR Surveill Summ*. 2014;63:1–21.
7. Shattuck PT, Durkin M, Maenner M, et al. Timing of identification among children with an autism spectrum disorders: findings from a population-based surveillance study. *J Am Acad Child Adolesc Psychiatry*. 2009;5:474–483.
8. Sheldrick RC, Maye MP, Carter AS. Age at first identification of autism spectrum disorder: an analysis of two US Surveys. *J Am Acad Child Adolesc Psychiatry*. 2017;56:313–320.
9. Mandell DS, Wiggins LD, Carpenter LA, et al. Racial/ethnic disparities in the identification of children with autism spectrum disorders. *Am J Public Health*. 2011;99:493–498.
10. Selden TM. Compliance with well-child visit recommendations: evidence from the medical expenditure panel survey, 2000–2002. *Pediatrics*. 2006;118:e1766–e1778.
11. Wissow LS, Brown J, Fothergill KE, et al. Universal mental health screening in pediatric primary care: a systematic review. *J Am Acad Child Adolesc Psychiatry*. 2013;52:1134–1147.
12. Zwaigenbaum L, Bryson S, Lord, et al. Clinical assessment and management of toddlers with suspected autism spectrum disorder: insights from studies of high-risk infants. *Pediatrics*. 2009;123: 1383–1391.
13. Ozonoff S, Young GS, Landa RJ, et al. Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study. *J Child Psychol Psychiatry*. 2015;56: 988–998.
14. Maenner MJ, Schieve LA, Rice CE, et al. Frequency and pattern of documented diagnostic features and the age of autism identification. *J Am Acad Child Adolesc Psychiatry*. 2013;52:401–413.e8.
15. Reiersen AM. Early identification of autism spectrum disorder: is diagnosis by age 3 a reasonable goal? *J Am Acad Child Adolesc Psychiatry*. 2017;56:284–285.
16. Robins D, Fein D, Barton M, et al. The modified checklist for autism in toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *J Autism Developmental Disord*. 2001;31:131–144.
17. Kleinman JM, Robins DL, Ventola PE, et al. The modified checklist for autism in toddlers: a follow-up study investigating the early detection of autism spectrum disorders. *J Autism Developmental Disord*. 2008;38:827–839.
18. Chlebowski C, Robins DL, Barton ML, et al. Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics*. 2013; 131:1121–1127.
19. Stenberg N, Bresnahan M, Gunnes N, et al. Identifying children with autism spectrum disorder at 18 months in a general population sample. *Paediatr Perinat Epidemiol*. 2014;28:255–262.
20. McPheeters ML, Weitlauf AS, Vehorn A, et al. *Screening for Autism Spectrum Disorder in Young Children: A Systematic Evidence Review for the U.S. Preventive Services Task Force. Evidence Synthesis No. 129. AHRQ Publication No. 13-05185-EF-1*. Rockville, MD: Agency for Healthcare Research and Quality; 2016.
21. Smith N, Sheldrick RC, Perrin E. An abbreviated screening instrument for autism spectrum disorders. *Infant Ment Health J*. 2013;34:149–155.
22. Gotham K, Risi S, Pickles A, et al. The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. *J Autism Dev Disord*. 2007;37:613.