Autism Screening With Online Decision Support by Primary Care Pediatricians Aided by M-CHAT/F

Raymond Sturner, MD,^{a,b} Barbara Howard, MD,^{a,c} Paul Bergmann, MA,^{d,e} Tanya Morrel, PhD,^c Lindsay Andon, MPH,^f Danielle Marks, MPH, MSW,^g Patricia Rao, PhD,^h Rebecca Landa, PhD^h

abstract

BACKGROUND AND OBJECTIVE: Autism spectrum disorders (ASDs) often go undetected in toddlers. The Modified Checklist for Autism in Toddlers (M-CHAT) With Follow-up Interview (M-CHAT/F) has been shown to improve detection and reduce over-referral. However, there is little evidence supporting the administration of the interview by a primary care pediatrician (PCP) during typical checkups. The goal of this study was to evaluate the feasibility, validity, and reliability of the M-CHAT/F by PCPs with online prompts at the time of a positive M-CHAT screen.

DESIGN Forty-seven PCPs from 22 clinics completed 197 M-CHAT/Fs triggered by positive M-CHAT screens via the same secure Web-based platform that parents used to complete M-CHATs before an 18- or 24-month well-child visit. A second M-CHAT/F was administered live or by telephone by trained research assistants (RAs) at the Kennedy Krieger Institute Center for Autism and Related Disorders. The Autism Diagnostic Observation Schedule, Second Edition, and the Mullen Scales of Early Learning were administered as criterion measures. Measures of agreement between PCPs and RAs were calculated, and measures of test performance compared.

RESULTS: There was 86.6% agreement between PCPs and RAs, with a Cohen's κ of 0.72. Comparison of sensitivity, specificity, positive predictive value (PPV), and overall accuracy for M-CHAT/F between PCPs and RAs showed significant equivalence for all measures. Use of the M-CHAT/F by PCPs resulted in significant improvement in PPV compared with the M-CHAT alone.

CONCLUSIONS: Minimally trained PCPs can administer the M-CHAT/F reliably and efficiently during regular well-child visits, increasing PPV without compromising detection.

^aDepartment of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland; ^bCenter for Promotion of Child Development through Primary Care, Baltimore, Maryland; ^cTotal Child Health, Baltimore, Maryland; ^dPrairieCare Institute, Minneapolis, Minnesota; ^eForesight Logic, Inc, St Paul, Minnesota; ^fPopulation Health Research, Johns Hopkins HealthCare, Baltimore, Maryland; ^gWoman and Infant Health Program, Wyoming Department of Health, Cheyenne, Wyoming; and ^hKennedy Krieger Institute, Baltimore, Maryland

Drs Sturner and Howard conceptualized and designed the study, oversaw data collection during the study, and drafted the initial manuscript; Mr Bergmann conducted the statistical analyses, and reviewed and revised the manuscript; Dr Morrel, Ms Andon, and Ms Marks oversaw data collection, suggested protocol revisions, and assisted with manuscript preparation and review; Drs Rao and Landa contributed to study design, oversaw collection of criterion testing, and reviewed and revised the manuscript; and all authors approved the final manuscript as submitted.

DOI: 10.1542/peds.2015-3036

Accepted for publication Jun 17, 2016

WHAT'S KNOWN ON THIS SUBJECT: The widely used Modified Checklist for Autism in Toddlers autism screen now has a requirement for a structured follow-up clinician interview for positive parent reports. However, its feasibility and accuracy have not been determined when used by primary care pediatricians during routine health supervision visits.

WHAT THIS STUDY ADDS: The Modified Checklist for Autism in Toddlers Follow-up Interview is feasible for implementation by primary care pediatricians aided by electronic decision support during routine well-child care, yielding results that are as accurate as, and timelier than, those produced by specially trained clinicians.

To cite: Sturner R, Howard B, Bergmann P, et al. Autism Screening With Online Decision Support by Primary Care Pediatricians Aided by M-CHAT/F. *Pediatrics*. 2016;138(3):e20153036

Current prevalence estimates suggest that autism spectrum disorder (ASD) now affects 1 in 68 children.¹ Early, evidence-based intervention for children with ASDs is associated with improved developmental functioning and reduction of symptoms.^{2,3} Data suggest intervention before 3 years of age has the greatest impact,^{4,5} making early detection vital.

Guidelines from the American Academy of Pediatrics call for use of ASD-specific screening tools at both the 18- and 24-month checkup visits.⁶ Although the age at first diagnosis seems to be decreasing,⁷ recent assessments show the average age is still >4 years.¹ As screening becomes more widespread, over-referrals could tax system capacity for timely evaluation and early intervention with false-positive cases.

The Modified Checklist for Autism in Toddlers (M-CHAT)⁸ is one of the most widely used autism-specific screens.^{9,10} It is among the tools approved for ASD screening by the American Academy of Pediatrics⁶ and advocacy groups, such as Autism Speaks.¹¹ The M-CHAT has replaced the original Checklist for Autism in Toddlers, which demonstrated adequate specificity but inadequate sensitivity.¹² Initial studies of the M-CHAT suggested that the modifications improved sensitivity but required a follow-up telephone interview for positive screens to clarify and/or correct parental responses to reduce over-referral rates.⁸

Another study in community practices reported that this follow-up procedure increased the positive predictive value (PPV) from 0.11 to 0.65, reinforcing the need for the follow-up interview.¹³ The interview consists of a script that prompts for specific examples of behaviors relevant to each failed item. Based on the parent's responses, the interviewer may change failed items to a pass.¹⁴ According to the M-CHAT authors, the follow-up

interview is required for most failed screens and "...the use of the M-CHAT without the follow-up interview is not recommended in low risk populations..." such as patients in primary care settings.¹⁵ The M-CHAT authors have also noted that when the M-CHAT score is ≥ 8 , the potential for false-positive screens in the absence of the follow-up interview is low enough that it can be omitted.¹⁶

The M-CHAT is therefore intended to be implemented as a 2-stage screening test consisting of completion by the parent and, if positive, conducting the follow-up interview to clarify/correct failed items. This process is referred to as the Modified Checklist for Autism in Toddlers Follow-Up Interview (M-CHAT/F). The M-CHAT/F authors also suggest that the second stage of the procedure has clinical advantages in facilitating discussion between parents and providers about a child's behaviors that may indicate ASD or other developmental delays.¹⁶

One limitation of the published research is that administration of the follow-up interview was almost exclusively completed by trained research assistants (RAs) during a telephone call completed an average of 2 or 3 months after the parent completed the M-CHAT^{8,13,17-19}; 1 study indicated that "most" of the follow-up interviews were conducted later by telephone, with an unspecified number conducted by primary care physicians.¹⁵ Another study documented use of the follow-up interview by a clinician during the same visit in which the M-CHAT was completed by the parent.²⁰ However, that particular 60-minute nurse/clinician visit occurred when the child was 2.5 years old.

It is therefore unclear whether the recommended follow-up interview can be completed during briefer well-child visits by a primary care pediatrician (PCP) or if a telephone follow-up interview conducted by

a specially trained interviewer is required.

Will PCPs find time for this follow-up interview and, if so, will they use the written interview script and adhere to the standardized wording reliably? Or, will PCPs (who do not traditionally use structured interviews) spontaneously formulate follow-up questions using a more liberal interpretation of the interview?

It is possible that parents, interviewed by telephone at a later time, may have had an opportunity to more accurately observe behaviors about which they were initially uncertain or the child's development may have changed during the interval. The results of a telephone follow-up interview might therefore differ from 1 completed the day of the positive screen even if both interviews were completed in the same way.

The goal of the present study was to determine whether PCPs will actually complete a follow-up interview to a positive M-CHAT during a routine check-up visit and how the results compare with a telephone follow-up interview later conducted by a trained RA. We reasoned that automated computer presentation of the exact M-CHAT/F questions for each failed M-CHAT item, as used here, and with scoring efficiencies, could facilitate the process.

METHODS

Screening

PCPs screened children at routine 18- and 24-month visits using a version of the M-CHAT completed by parents online either at home or on a device in the waiting room via a Web system (ie, the Child Health and Development Interactive System).²¹ The M-CHAT was automatically scored and, if positive, the PCP completed the follow-up interview also using the Child Health and

FIGURE 1

Sample flow. M-CHAT, screen completed by parent; M-CHAT/F, interview conducted by PCP; AC M-CHAT/F, M-CHAT/F interview conducted by AC.

Development Interactive System during the visit. The flow of the interview was determined by using the failed M-CHAT questions, with the system rescored after each initially failed item was clarified, and concluding when the initial screening result was either confirmed positive or reversed by the results of the follow-up interview. No extra time was allotted for these visits. 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 (73%) suburban, with some rural (18%) and urban (9%) locations. Practice estimation of office-level demographic characteristics 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 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 (Fig 1 provides a sample flow). Online access to the M-CHAT/F was available for all M-CHAT-positive findings for all offices for all children.

Diagnostic Testing

All children who initially screened positive on the M-CHAT were recruited for diagnostic testing at the Kennedy Krieger Institute Center for Autism and Related Disorders (Autism Center [AC]). M-CHAT results were monitored by research staff under a Health Insurance Portability and Accountability Act agreement and office approval. All children with positive M-CHAT screens (once the M-CHAT/F was complete) were offered a full evaluation blinded to all M-CHAT/F results without any PCP input apart from the M-CHAT/F itself. Consent was obtained using a protocol approved by The Johns Hopkins University School of

TABLE 1 Demographic Characteristics of Final Sample (N = 98)

Characteristic	Value
Age, mo	
Minimum	14.7
Median	20.3
Mean $\pm$ SD	22.9 $\pm$ 6.1
Maximum	40.8
Sex, (%)	
Male	25 (74.5)
Female	73 (25.5)
Race, (%)	
White	64 (65.3)
African American	11 (11.2)
Asian	5 (5.1)
Native American	2 (2.0)
Multiple	6 (9.2)
Other	9 (6.1)
Missing	1 (1.0)
Hispanic, (%)	
Yes	5 (5.1)
No	92 (93.9)
Missing	1 (1.0)
Insurance, (%)	
Private	74 (75.5)
Medicare/Medicaid/CHIPs	20 (20.4)
None	1 (1.0)
Missing	3 (3.1)
Respondent relationship, (%)	
Biological mother	88 (89.8)
Biological father	5 (5.1)
Grandmother	2 (2.0)
Other relative	1 (1.0)
Adoptive mother	1 (1.0)
Missing	1 (1.0)
Respondent education, (%)	
12th grade or lower	4 (4.1)
High school diploma or GED	4 (4.1)
Some college	86 (87.8)
College degree	1 (1.0)
Missing	3 (3.1)

CHIPs, Children's Health Insurance Program; GED, General Education Diploma.

Medicine Institutional Review Board, with initial screening considered standard of care requiring no separate consent. The Conflict of Interest Committee required that those individuals responsible for recruiting subjects, the diagnostic testers, and the statistician be conflict free. There was no charge for the diagnostic testing; parents received a \$100 subject fee. Follow-up interviews were administered by trained RAs by telephone at the time appointments were scheduled for diagnostic testing. Evaluators were naive to other test results, and none

of the AC staff knew results of the PCPs' follow-up interviews.

A total of 99 parents consented (29% of initial positive screens) and completed all relevant evaluation measures; 1 full ASD evaluation could not be completed because of "untestable" behavior.

Diagnostic testing included administration of the Autism Diagnostic Observation Schedule (ADOS)²² by certified research-reliable speech and language pathologists. ADOS is a semi-structured behavior observation assessment of social and communication skills. It is recognized as the gold standard test for diagnosis of ASDs across age, developmental level, and language skills. The assessment relies on a series of planned occasions designed to elicit specific social and communicative processes in a standardized way. According to the author of the test, the criterion for diagnosis of autism in this study was the overall clinical diagnosis²³ of an experienced and reliable ADOS tester who also conducted a standard developmental assessment (Mullen Scales of Early Learning [MSEL]). 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. Testers provided data on those children judged to show some atypical features but not enough for an ASD diagnosis, in a category of "suspected ASD/new phenotype."

The MSEL²⁴ was also administered during diagnostic testing. The MSEL is a standardized developmental test for children 3 to 69 months of age. It has 5 subscales: gross motor, fine motor, visual reception, receptive language, and expressive language. An Early Learning

TABLE 2 Comparison of M-CHAT Scores According to Level of Study Participation

Variable	N	Mean (95% CI)	<i>t</i> Test, <i>P</i>	TOST	
				<i>t</i> ₁ <i>P</i>	<i>t</i> ₂ <i>P</i>
Overall ^a			<.001	<.001	>.99
M-CHAT fail	341	5.0 (4.7–5.3)			
M-CHAT pass	4730	0.4 (0.4–0.4)			
M-CHAT fail ^b			.78	.003	<.001
Without M-CHAT/F	144	5.0 (4.5–5.5)			
With M-CHAT/F	197	5.0 (4.6–5.4)			
M-CHAT/F completed ^a			.03	<.001	.40
Not in study	98	4.5 (4.0–5.0)			
In study	99	5.4 (4.8–6.0)			
M-CHAT/F fail ^c			.67	.26	.07
Not in study	21	8.1 (6.5–8.7)			
In study	40	7.7 (6.6–8.8)			
M-CHAT/F pass ^b			.17	<.001	.002
Not in study	77	3.5 (3.3–3.7)			
In study	59	3.8 (3.4–4.3)			

Comparisons were made with 2-tailed *t* test and TOST, Δ = 1 point. CI, confidence interval.

^a Relevant difference.

^b Equivalence.

^c Indeterminate (underpowered).

TABLE 3 Comparison of M-CHAT Scores of Study Cases According to ASD Diagnosis

Variable	N	Mean (95% CI)	<i>t</i> Test <i>P</i> Value	TOST	
				<i>t</i> ₁ <i>P</i>	<i>t</i> ₂ <i>P</i>
Overall ^a			.03	<.001	.76
ASD positive	39	6.5 (5.2–7.7)			
ASD negative	59	5.0 (4.2–5.8)			
M-CHAT/F fail ^b			.56	.08	.39
ASD positive	23	8.4 (6.7–10.1)			
ASD negative	17	7.7 (6.1–9.3)			
M-CHAT/F pass ^b			.71	.06	.01
ASD positive	16	3.7 (3.2–4.2)			
ASD negative	42	3.9 (3.3–4.5)			

Comparisons were made with 2-tailed *t* test and TOST, Δ = 1 point. CI, confidence interval.

^a Relevant difference.

^b Indeterminate (underpowered).

Composite score is generated from the fine motor, receptive language, expressive language, and visual reception scale scores. The MSEL is used to determine the expected levels of communicative and social functioning for comparison with the ADOS results, as well as to identify children with developmental delays.

Analyses

Demographic characteristics of patients and respondents were tabulated. Potential selection bias was evaluated by using *t* tests and 2 one-sided tests of equivalence (TOST)^{25,26} of mean M-CHAT scores. Measures of test performance of MCHAT/F using PCP and AC

follow-up interviews were calculated and compared with 2 proportion *z* tests and TOST.^{25,27} Combining the results of the *t* test/*z* test and TOST enables us to make an inference regarding significant difference or equivalence and whether significant differences are clinically/practically relevant.^{26,27} A general index of overall percent agreement and Cohen's κ ²⁸ were used to measure the degree to which PCP follow-up interviews agreed with AC follow-up interviews.

RESULTS

The demographic description of the subjects completing diagnostic

testing with complete data is given in Table 1. The mean age of participating children was 22.9 months. All but 4 of the 98 children in the study were in the recommended age range; one was younger (14.7 months), and 3 were older (37, 37, and 40.8 months). As typical for screen-positive children, 74.5% were male.¹⁵ Respondents were nearly always mothers (89.8%) and had at least some college education (88.8%). Tables 2 and 3 present comparisons of mean M-CHAT scores according to level of study participation and ASD diagnosis. Among M-CHAT failures, M-CHAT scores were equivalent between those who received the M-CHAT/F and those who did not.

TABLE 4 Performance of M-CHAT and M-CHAT/F Screens Predicting ASD Diagnosis

Variable	N ^a	Ratio (95% CI)	z Test, <i>P</i>	TOST	
				<i>z</i> ₁ <i>P</i>	<i>z</i> ₂ <i>P</i>
PPV ^b			.01	.86	<.001
M-CHAT	98	0.40 (0.30–0.50)			
PCP M-CHAT/F	98	0.58 (0.48–0.67)			
PPV ^c			.49	.24	.02
PCP M-CHAT/F	98	0.58 (0.48–0.67)			
AC M-CHAT/F	97	0.53 (0.43–0.63)			
Sensitivity ^c			.37	.30	.01
PCP M-CHAT/F	98	0.59 (0.49–0.69)			
AC M-CHAT/F	97	0.53 (0.43–0.63)			
Specificity ^c			.80	.10	.04
PCP M-CHAT/F	98	0.71 (0.62–0.80)			
AC M-CHAT/F	97	0.69 (0.60–0.79)			
Accuracy ^c			.62	.17	.02
PCP M-CHAT/F	98	0.67 (0.57–0.76)			
AC M-CHAT/F	97	0.63 (0.53–0.73)			

Comparisons were made with 2-tailed z test and TOST, $\Delta = 0.1$ unit of proportion. Only patients with a positive screen on the M-CHAT are included in this sample. CI, confidence interval.

^a One patient did not have a complete AC M-CHAT/F (Fig 1).

^b Relevant difference.

^c Equivalence.

TABLE 5 Performance of M-CHAT and M-CHAT/F Screens Predicting ASD or Suspected ASD Diagnosis

Variable	N ^a	Ratio (95% CI)	z Test, <i>P</i>	TOST	
				<i>z</i> ₁ <i>P</i>	<i>z</i> ₂ <i>P</i>
PPV ^b			.001	.94	<.001
M-CHAT	98	0.57 (0.47–0.67)			
PCP M-CHAT/F	98	0.78 (0.69–0.86)			
PPV ^c			.30	.29	.004
PCP M-CHAT/F	98	0.78 (0.69–0.86)			
AC M-CHAT/F	97	0.71 (0.62–0.80)			
Sensitivity ^c			.38	0.30	.01
PCP M-CHAT/F	98	0.55 (0.46–0.65)			
AC M-CHAT/F	97	0.49 (0.39–0.59)			
Specificity ^c			.43	.20	.008
PCP M-CHAT/F	98	0.79 (0.70–0.87)			
AC M-CHAT/F	97	0.74 (0.65–0.83)			
Accuracy ^c			.24	.40	.005
PCP M-CHAT/F	98	0.60 (0.56–0.75)			
AC M-CHAT/F	97	0.57 (0.47–0.67)			

Table includes suspected ASD/new phenotype. Comparisons were made with 2-tailed z test and TOST, $\Delta = 0.1$ unit of proportion. Only patients with a positive screen on the M-CHAT are included in this sample. CI, confidence interval.

^a One patient did not have a complete AC M-CHAT/F (Fig 1).

^b Relevant difference.

^c Equivalence.

Children with completed M-CHAT/Fs who were not included in the study had an average of 1 fewer failed M-CHAT item than those included. However, mean scores were not meaningfully different across study participation within the M-CHAT/F pass and M-CHAT/F fail subgroups.

Overall mean M-CHAT scores for those who were ASD-positive were higher than for those who were ASD-negative. However, differences

in mean scores diminished when comparing ASD-positive cases versus ASD-negative cases among those positive according to the M-CHAT/F. The same is true for those labeled as M-CHAT/F negative (Table 3).

Table 4 presents measures of performance of the M-CHAT and M-CHAT/F against the final AC ASD diagnosis. The PPV of the M-CHAT was 0.40. Use of the M-CHAT/F

by PCPs significantly improved the performance of the screen, producing a PPV of 0.58 ($P = .01$). When comparing the M-CHAT/F results of PCPs versus those of the RAs, we observed moderately high agreement (86.6%) with a κ of 0.72 ($P < .001$). PPV, sensitivity, specificity, and overall accuracy were all statistically equivalent between PCPs and RAs based on the combined results of the z tests and TOST.

TABLE 6 Performance of M-CHAT and M-CHAT/F Screens Predicting ASD and/or Developmental Delay Diagnosis

Variable	N ^a	Ratio (95% CI)	z Test, <i>P</i>	TOST	
				<i>z</i> ₁ <i>P</i>	<i>z</i> ₂ <i>P</i>
PPV ^b			.001	.94	<.001
M-CHAT	98	0.68 (0.59–0.78)			
PCP M-CHAT/F	98	0.88 (0.81–0.94)			
PPV ^c			.51	.09	.004
PCP M-CHAT/F	98	0.88 (0.81–0.94)			
AC M-CHAT/F	97	0.84 (0.62–0.80)			
Sensitivity ^c			.60	.19	.03
PCP M-CHAT/F	98	0.52 (0.42–0.62)			
AC M-CHAT/F	97	0.48 (0.39–0.58)			
Specificity ^c			.56	.11	.008
PCP M-CHAT/F	98	0.84 (0.77–0.91)			
AC M-CHAT/F	97	0.81 (0.73–0.89)			
Accuracy ^c			.62	.18	.03
PCP M-CHAT/F	98	0.62 (0.53–0.72)			
AC M-CHAT/F	97	0.59 (0.49–0.69)			

Developmental Delay Positive was defined as >1.5 SDs below the mean on ≥ 2 Mullen scales or >2 SDs below the mean on ≥ 1 Mullen scale. Comparisons were made with 2-tailed z test and TOST, $\Delta = 0.1$ unit of proportion. Only patients with a positive screen on the M-CHAT are included in this sample. CI, confidence interval.

^a One patient did not have a complete AC M-CHAT/F (Fig 1).

^b Trivial difference.

^c Equivalence.

TABLE 7 Performance of M-CHAT and M-CHAT/F Screens Predicting ASD or Suspected ASD Diagnosis and/or Developmental Delay Diagnosis

Variable	N ^a	Ratio (95% CI)	z Test, <i>P</i>	TOST	
				<i>z</i> ₁ <i>P</i>	<i>z</i> ₂ <i>P</i>
PPV ^b			.01	.74	<.001
M-CHAT	98	0.77 (0.68–0.85)			
PCP M-CHAT/F	98	0.90 (0.84–0.96)			
PPV ^c			.90	.01	.008
PCP M-CHAT/F	98	0.90 (0.84–0.96)			
AC M-CHAT/F	97	0.89 (0.83–0.96)			
Sensitivity ^c			.77	.13	.05
PCP M-CHAT/F	98	0.48 (0.38–0.58)			
AC M-CHAT/F	97	0.46 (0.36–0.56)			
Specificity ^c			.99	.03	.03
PCP M-CHAT/F	98	0.83 (0.75–0.90)			
AC M-CHAT/F	97	0.83 (0.75–0.90)			
Accuracy ^c			.84	.12	.05
PCP M-CHAT/F	98	0.56 (0.46–0.66)			
AC M-CHAT/F	97	0.55 (0.45–0.65)			

Predictions include suspected ASD/new phenotype. Developmental Delay Positive was defined as >1.5 SDs below the mean on ≥ 2 Mullen scales or >2 SDs below the mean on ≥ 1 Mullen scale. Comparisons were made with 2-tailed z test and TOST, $\Delta = 0.1$ unit of proportion. Only patients with a positive screen on the M-CHAT are included in this sample. CI, confidence interval.

^a One patient did not have a complete AC M-CHAT/F (Fig 1).

^b Trivial difference.

^c Equivalence.

Because children with false-positive screens may have atypical features but not enough for an ASD diagnosis, we included children with suspected ASD/new phenotype but no ASD diagnosis to form a combined outcome (Table 5). Differences in PPV between the M-CHAT/F and M-CHAT remained significant when including the “suspect” criterion (0.78 vs 0.57). Because children with false-positive screens for ASD may nevertheless

have developmental difficulties that would benefit from early intervention, we examined prediction of a combined criterion of scores >1.5 SD below the mean on ≥ 2 Mullen scales or >2 SDs on any 1 Mullen scale (a common state eligibility criterion²⁹) and/or receiving an ASD diagnosis (Table 6). Using this combined criterion increases the PPV to 0.88 for M-CHAT/F and 0.68 for M-CHAT, reducing the numbers of

M-CHAT/F false-positive screens by 71%. Including both developmental failures and “suspected ASD” cases improved the PPV to 0.90 with M-CHAT/F and 0.77 for M-CHAT (Table 7). PPV, sensitivity, specificity, and accuracy of the M-CHAT/F were all statistically equivalent between PCPs and RAs.

Mean ADOS and Mullen scores are presented in Tables 8 and 9 according to screening result and

TABLE 8 Comparison of Mean ADOS Scores

Variable	Score	<i>N</i>	Mean (95% CI)	<i>P</i>
PCP M-CHAT/F				
Total	Positive	39	13.2 (11.1–15.2)	<.001
	Negative	58	7.4 (6.1–8.8)	
Communication	Positive	39	4.6 (3.9–5.3)	<.001
	Negative	58	2.7 (2.2–3.2)	
Social interaction	Positive	39	8.6 (7.1–10.1)	<.001
	Negative	58	4.8 (3.8–5.7)	
AC M-CHAT/F				
Total	Positive	37	12.6 (10.4–14.7)	<.001
	Negative	59	7.8 (6.4–9.3)	
Communication	Positive	37	4.3 (3.6–5.1)	.001
	Negative	59	2.8 (2.3–3.4)	
Social interaction	Positive	37	8.2 (6.6–9.8)	<.001
	Negative	59	5.0 (4.0–6.0)	
ASD diagnosis				
Total	Positive	38	15.3 (14.2–16.5)	<.001
	Negative	59	6.1 (4.8–7.5)	
Communication	Positive	38	5.3 (4.7–5.9)	<.001
	Negative	59	2.2 (1.8–2.7)	
Social Interaction	Positive	38	10.0 (9.3–10.8)	<.001
	Negative	59	3.9 (2.9–4.9)	

Comparisons were made with 2-tailed *t* test. CI, confidence interval.

TABLE 9 Comparison of Mean Mullen T Scores

Variable	Score	N	Mean (95% CI)	P
PCP M-CHAT/F				
Gross motor	Positive	32	36.0 (31.8–40.2)	.002
	Negative	50	43.2 (40.8–45.6)	
Visual reception	Positive	40	34.5 (29.3–39.8)	<.001
	Negative	58	45.2 (42.1–48.2)	
Fine motor	Positive	40	29.1 (25.0–33.3)	<.001
	Negative	58	42.0 (39.5–44.5)	
Expressive language	Positive	40	24.6 (18.3–31.0)	<.001
	Negative	58	39.8 (36.4–43.2)	
Receptive language	Positive	40	23.4 (17.3–29.5)	<.001
	Negative	58	38.4 (33.8–42.9)	
AC M-CHAT/F				
Gross motor	Positive	30	36.4 (32.1–40.6)	.005
	Negative	51	43.0 (40.5–45.5)	
Visual reception	Positive	38	35.2 (29.7–40.8)	.002
	Negative	59	44.6 (41.5–47.7)	
Fine motor	Positive	38	30.6 (26.0–35.1)	<.001
	Negative	59	40.6 (37.9–43.3)	
Expressive language	Positive	38	27.4 (21.0–33.7)	.002
	Negative	59	38.2 (34.4–42.0)	
Receptive language	Positive	38	25.9 (19.2–32.7)	.005
	Negative	59	36.9 (32.4–41.4)	
ASD diagnosis				
Gross motor	Positive	32	35.2 (32.1–38.4)	<.001
	Negative	50	43.7 (40.9–46.5)	
Visual reception	Positive	39	37.2 (32.5–41.9)	.046
	Negative	59	43.2 (39.5–47.0)	
Fine motor	Positive	39	32.0 (28.9–35.2)	.003
	Negative	59	39.8 (36.3–43.3)	
Expressive language	Positive	39	29.0 (23.5–34.5)	.03
	Negative	59	36.7 (32.1–41.3)	
Receptive language	Positive	39	21.9 (16.4–27.4)	<.001
	Negative	59	39.1 (34.4–43.8)	

Comparisons were made with 2-tailed *t* test. CI, confidence interval.

the overall diagnostic sample.

Detailed differences in the 13 cases for whom the M-CHAT/F results differed between the 2 raters are described in Tables 10, 11, and 12. Three of 7 children who screened positive on the PCP M-CHAT/F and negative on the AC M-CHAT/F were diagnosed as having ASD; of the 6 children who screened positive on the AC M-CHAT/F but negative on the PCP M-CHAT/F, only 1 child was diagnosed with ASD.

DISCUSSION

This study confirms the previous finding of improved PPV of the M-CHAT⁸ using the M-CHAT/F.¹³ It also confirms previous studies showing that most children with false-positive screens have developmental difficulties¹⁷ of a degree that would make them eligible for early intervention. Some children with false-positive screens had atypical features not meeting criteria for ASD.

The online M-CHAT/F enabled PCPs to clarify positive parent responses to M-CHAT items during well-child visits, rather than requiring another visit or call by a trained interviewer. This study found that the performance of the M-CHAT/F by a PCP was equivalent to one administered by trained AC staff. This report is the first demonstrating feasibility of administration of the M-CHAT/F during the time of well-child visit in community practices.

The PCPs in this study received only brief orientation and access to a 10-minute interactive Web-based demonstration, considerably less than previous studies providing 1.5 days of training in autism screening.²⁰

Although time-motion data were not collected, PCP visits were not scheduled to be longer than that of their colleagues. The process of clarifying initial parent M-CHAT

TABLE 10 Mean ADOS Scale Scores According to M-CHAT/F Mismatch Category

Variable	N	Mean (95% CI)
Total		
PCP pass/AC fail	6	7.3 (3.0–11.6)
PCP fail/AC pass	7	10.9 (4.7–17.0)
Communication		
PCP pass/AC fail	6	3.0 (1.2–4.8)
PCP fail/AC pass	7	4.6 (2.1–7.1)
Social interaction		
PCP pass/AC fail	6	4.3 (1.1–7.6)
PCP fail/AC pass	7	6.3 (2.3–10.2)

CI, confidence interval.

TABLE 11 Mean Mullen T Scores According to M-CHAT/F Mismatch Category

Variable	N	Mean (95% CI)
Gross motor		
PCP pass/AC fail	5	42.0 (25.0–59.0)
PCP fail/AC pass	6	40.0 (23.8–56.2)
Visual reception		
PCP pass/AC fail	6	45.5 (36.0–55.0)
PCP fail/AC pass	7	40.4 (29.2–51.7)
Fine motor		
PCP pass/AC fail	6	44.0 (32.6–55.4)
PCP fail/AC pass	7	32.0 (19.8–44.2)
Expressive language		
PCP pass/AC fail	6	38.0 (30.8–45.2)
PCP fail/AC pass	7	24.9 (8.1–41.6)
Receptive language		
PCP pass/AC fail	6	40.2 (22.6–57.7)
PCP fail/AC pass	7	27.6 (17.1–38.0)

CI, confidence interval.

TABLE 12 Diagnoses According to M-CHAT/F Mismatch Category

Variable	N	ASD	ASD/Delay
PCP pass/AC fail	6	1	4
PCP fail/AC pass	7	3	6

responses prompted by the online follow-up interview would be expected to require less time than discussing autism referrals with parents of children who screened positive. In addition, MCHAT/F–negative families avoided the distress of waiting for a diagnostic evaluation.

We agree with the M-CHAT authors that the M-CHAT/F provides clinical utility by facilitating discussion with parents regarding their child’s behaviors and potential developmental challenges.¹⁵ It is important to use a validated method, such as the M-CHAT/F, to ascertain whether the parent is reporting behaviors consistent with signs

of ASD or has misunderstood the questions.

Once a child has been identified as at risk for ASD through screening, challenges remain for persuading the family that a diagnostic evaluation is necessary.³⁰ In addition to reducing false-positive findings, use of the follow-up interview facilitates a structured clarification process and discussion of parent perceptions that led to the positive screen, encouraging parents to follow through with a referral for further evaluation.

This study differed from that of the M-CHAT authors by including

M-CHAT/F negative findings as well as positive findings for full evaluation. This study also found that even with trained M-CHAT/F interviewers, one cannot assume that all children with negative interview results do not have autism; even the recommended 2-stage procedure missed cases. Further study of the sensitivity of the M-CHAT in community populations with systematic evaluation of M-CHAT–negative cases and M-CHAT/F–negative cases is needed to fully evaluate the cost/benefit of various toddler screening strategies.

Further research is needed to determine the types of conversation, education, and monitoring required to ensure timely referral to early intervention. Successful referral begins with a shared understanding of how the child’s behaviors relate to a potentially serious, although sometimes subtle, condition requiring attention. Although discussion could be informal, the validated, structured (and billable) M-CHAT/F provides evidence-based documentation for or against a serious condition. Because most positive M-CHAT screens are false-positive findings, a positive M-CHAT/F screen can help clinicians be more persuasive about referral, knowing its higher likelihood of significance.

We therefore disagree with the authors of the M-CHAT/F who advise that a M-CHAT total score ≥ 8 does not require the follow-up interview.¹⁶ Thoughtful review of parental perceptions of their child’s behaviors for these 1% of cases using M-CHAT/F acknowledges the importance of parental observations, supports the need for referral, and activates follow-through. Use of a computerized algorithm to guide the follow-up interview also shortens the interview for those endorsing large

number of behaviors because the process stops automatically when results either confirm or refute the initial M-CHAT screen result.

One limitation of this study, shared by other studies, is that children passing the M-CHAT were not systematically recruited and, of those completing the M-CHAT/F who were recruited, many declined full evaluation. However, unlike the initial M-CHAT studies, this sample was derived completely from a community, primary care population with a participation rate similar to the latest M-CHAT community sample.¹⁶ This limitation is not restricted to research studies, as similar suboptimal rates of successful referral from screening occur in practice.³⁰ The requirement to travel to an AC may have been an impediment because our subsequent study offered in-home assessments and had higher participation rates. The proportion of African-American children in the evaluation sample was only one-third of the proportion of African-American subjects

estimated in the overall clinic population. Because we did not have individual demographic data for children who did not enroll in the evaluation, we do not know if the disparity was at the level of screening, screening positive, or study enrollment. The proportion of African-American subjects in the clinics used, however, was based on clinician estimates. Known disparities in age of diagnosis in African-American children³¹ make this factor important to understand.

This study was conducted before publication of the revised M-CHAT¹⁶; the original 23-item version was used. The revised M-CHAT/F contains slight wording changes and eliminates 3 items. Due to the follow-up algorithm, we could not reanalyze our data as though the revised M-CHAT was used.

The authors of the M-CHAT assert that the follow-up interview is a requirement for its use in well-child visits.¹⁶ This study provides the first evidence that PCPs can do so feasibly and reliably. Although an online system is not

necessary, it provides important efficiencies. Adherence to formally structured interview protocols is not the tradition of PCPs. Electronic decision support may provide the structure needed when adherence to an algorithm is appropriate and clinically desirable.

ABBREVIATIONS

AC:	Center for Autism and Related Disorders
ADOS:	Autism Diagnostic Observation Schedule
ASD:	autism spectrum disorder
M-CHAT:	Modified Checklist for Autism in Toddlers
M-CHAT/F:	Modified Checklist for Autism in Toddlers Follow-up Interview
PPV:	positive predictive value
MSEL:	Mullen Scales of Early Learning
PCP:	primary care pediatrician
RA:	autism center research assistant
TOST:	2 one-sided tests of equivalence

Address correspondence to Raymond Sturner, MD, Center for Promotion of Child Development through Primary Care, 6017 Altamont Place, Baltimore, MD 21210. E-mail: rsturner@childhealthcare.org

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2016 by the American Academy of Pediatrics

FINANCIAL 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. The Child Health and Development Interactive System (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. Mr Bergmann has consulted for TCH through his company Foresight Logic but received no funding for this study. Dr Morrel is an employee of TCH and a stockholder in the company. The other authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: All phases of this study were supported by the National Institute of Mental Health grant R44MH085399. Funded by the National Institutes of Health (NIH).

POTENTIAL 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. The Child Health and Development Interactive System (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. Dr Morrel is an employee of TCH and a stockholder in the company. The other authors have indicated they have no potential conflicts of interest to disclose.

REFERENCES

1. Christensen DL, Baio J, Braun KV, et al Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2012. *MMWR Surveill Summ.* 2016;65(3):1–23
2. National Research Council. In: Lord C, McGee JP, eds. *Educating Children With Autism*. Washington, DC: National Academies Press; 2001
3. Howlin P, Magiati I, Charman T. Systematic review of early intensive behavioral interventions for children

- with autism. *Am J Intellect Dev Disabil*. 2009;114(1):23–41
4. Harris S, Handelman J. Age and IQ at intake as predictors of placement for young children with autism: a four- to six-year follow-up. *J Autism Dev Disord*. 2000;30(2):137–142
 5. Wetherby AM, Woods J, Allen L, Cleary J, Dickinson H, Lord C. Early indicators of autism spectrum disorders in the second year of life. *J Autism Dev Disord*. 2004;34(5):473–493
 6. American Academy of Pediatrics, Council on Children With Disabilities. Policy statement: 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
 7. Mandell D, Novak M, Zubritsky C. Factors associated with age of diagnosis among children with autism spectrum disorders. *Pediatrics*. 2005;116(6):1480–1486
 8. Robins D, Fein D, Barton M, Green J. The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *J Autism Dev Disord*. 2001;31(2):131–144
 9. Canal-Bedia R, García-Primo P, Martín-Cilleros MV, et al. Modified Checklist for Autism in Toddlers: cross-cultural adaptation and validation in Spain. *J Autism Dev Disord*. 2010;41(10):1342–1351
 10. Inada N, Koyama T, Inokuchi E, Kuroda M, Kamio Y. Reliability and validity of the Japanese version of the Modified Checklist for Autism in Toddlers (M-CHAT). *Res Autism Spectr Disord*. 2011;5(1):330–336
 11. Learn the Signs of Autism. Autism Speaks. Available at: <https://www.autismspeaks.org/what-autism/learn-signs>. Accessed March 13, 2015
 12. Baron-Cohen S, Allen J, Gillberg C. Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *Br J Psychiatry*. 1992;161(6):839–843
 13. 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 Dev Disord*. 2008;38(5):827–839
 14. Robins DL. Official M-CHAT Website. Available at: <http://mchatscreen.com>. Accessed March 13, 2015
 15. 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). Available at: www.pediatrics.org/cgi/content/full/131/4/e1121
 16. 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
 17. Robins DL. Screening for autism spectrum disorders in primary care settings. *Autism*. 2008;12(5):537–556
 18. Pandey J, Verbalis A, Robins DL, et al. Screening for autism in older and younger toddlers with the Modified Checklist for Autism in Toddlers. *Autism*. 2008;12(5):513–535
 19. 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. *J Autism Dev Disord*. 2014;44(1):194–203
 20. Nygren G, Sandberg E, Gillstedt F, Ekeröth G, Arvidsson T, Gillberg C. A new screening programme for autism in a general population of Swedish toddlers. *Res Dev Disabil*. 2012;33(4):1200–1210
 21. Welcome to CHADIS—Child Health & Development Interactive System. CHADIS: Child Health & Development Interactive System. Available at: www.chadis.com/. Accessed March 13, 2015
 22. Lord C, Luyster R, Gotham K, Guthrie W. *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part II): Toddler Module*. Torrance, CA: Western Psychological Services; 2012
 23. Lord C, Risi S, Lambrecht L, et al. The autism diagnostic observation schedule—generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord*. 2000;30(3):205–223
 24. Mullen E. *The Mullen Scales of Early Learning (AGS Ed.)*. Circle Pines, MN: American Guidance Service; 1995
 25. Schuirmann DA. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm*. 1987;15(6):657–680
 26. Tryon WW, Lewis C. An inferential confidence interval method of establishing statistical equivalence that corrects Tryon's (2001) reduction factor. *Psychol Methods*. 2008;13:272–277
 27. Tryon WW, Lewis C. Evaluating independent proportions for statistical difference, equivalence, indeterminacy, and trivial difference using inferential confidence intervals. *J Educ Behav Stat*. 2009;34(2):171–189
 28. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37–46
 29. National Early Childhood TA Center. State and jurisdictional eligibility definitions for infants and toddlers with disabilities under IDEA. Available at: <http://ectacenter.org/~pdfs/pubs/nnotes21.pdf>. Accessed March 28, 2016
 30. King TM, Tandon SD, Macias MM, et al. Implementing developmental screening and referrals: lessons learned from a national project. *Pediatrics*. 2010;125(2):350–360
 31. Jo H, Schieve LA, Rice CE, et al. Age at autism spectrum disorder (ASD) diagnosis by race, ethnicity, and primary household language among children with special health care needs, United States, 2009–2010. *Matern Child Health J*. 2015;19(8):1687–1697