Validation of the Modified Checklist for Autism in Toddlers, Revised With Follow-up (M-CHAT-R/F)

AUTHORS: Diana L. Robins, PhD,[a,b] Karís Casagrande, BS,[a] Marianne Barton, PhD,[c] Chi-Ming A. Chen, PhD,[c] Thyde Dumont-Mathieu, MD, MPH,[c,d] and Deborah Fein, PhD[c,d]

[a]Department of Psychology, [b]Neuroscience Institute, Georgia State University, Atlanta, Georgia; and [c]Department of Psychology, [d]Department of Pediatrics, University of Connecticut, Storrs, Connecticut

KEY WORDS
autism, screening, toddlers

ABBREVIATIONS
ASD—autism spectrum disorder
CI—confidence interval
GSU—Georgia State University
M-CHAT—Modified Checklist for Autism in Toddlers
M-CHAT/F—Modified Checklist for Autism in Toddlers with Follow-up
M-CHAT-R—Modified Checklist for Autism in Toddlers, Revised
M-CHAT-R/F—Modified Checklist for Autism in Toddlers, Revised with Follow-up
PPV—positive predictive value
STAT—Screening Tool for Autism in Two-Year-Olds
UConn—University of Connecticut
WCC—well-child care

Dr Robins conceptualized and designed the study, contributed to the proposal for National Institutes of Health funding, participated in data collection, analyzed the data, and drafted the initial manuscript; Ms Casagrande participated in data collection and assisted with management of the study, data analyses, and drafting and editing of the initial manuscript; Dr Barton contributed to the conceptualization of the study, participated in data collection, and critically reviewed the manuscript; Dr Chen contributed to data analyses and critically reviewed the manuscript; Dr Dumont-Mathieu contributed to conceptualization of the study, participated in data collection, and critically reviewed the manuscript; Dr Fein conceptualized and designed the study, led the efforts to obtain National Institutes of Health funding, participated in data collection, contributed to the interpretation of data, and critically reviewed the manuscript; and all authors approved the final manuscript as submitted.

(Continued on last page)

WHAT'S KNOWN ON THIS SUBJECT: Screening for autism spectrum disorders (ASDs) using the Modified Checklist for Autism in Toddlers (M-CHAT) improves early detection and long-term prognosis of ASD. Reducing the false-positive rate may increase implementation of screening for ASDs.

WHAT THIS STUDY ADDS: The Modified Checklist for Autism in Toddlers, Revised with Follow-up (M-CHAT-R/F), simplifies wording of the original M-CHAT. The current validation study indicates that the M-CHAT-R/F improves the ability to detect autism spectrum disorders in toddlers screened during well-child care visits.

abstract

OBJECTIVE: This study validates the Modified Checklist for Autism in Toddlers, Revised with Follow-up (M-CHAT-R/F), a screening tool for low-risk toddlers, and demonstrates improved utility compared with the original M-CHAT.

METHODS: Toddlers (N = 16 071) were screened during 18- and 24-month well-child care visits in metropolitan Atlanta and Connecticut. Parents of toddlers at risk on M-CHAT-R completed follow-up; those who continued to show risk were evaluated.

RESULTS: The reliability and validity of the M-CHAT-R/F were demonstrated, and optimal scoring was determined by using receiver operating characteristic curves. Children whose total score was ≥3 initially and ≥2 after follow-up had a 47.5% risk of being diagnosed with autism spectrum disorder (ASD; confidence interval [95% CI]: 0.41–0.54) and a 94.6% risk of any developmental delay or concern (95% CI: 0.92–0.98). Total score was more effective than alternative scores. An algorithm based on 3 risk levels is recommended to maximize clinical utility and to reduce age of diagnosis and onset of early intervention. The M-CHAT-R detects ASD at a higher rate compared with the M-CHAT while also reducing the number of children needing the follow-up. Children in the current study were diagnosed 2 years younger than the national median age of diagnosis.

CONCLUSIONS: The M-CHAT-R/F detects many cases of ASD in toddlers; physicians using the 2-stage screener can be confident that most screen-positive cases warrant evaluation and referral for early intervention. Widespread implementation of universal screening can lower the age of ASD diagnosis by 2 years compared with recent surveillance findings, increasing time available for early intervention. *Pediatrics* 2014;133:37–45

Autism spectrum disorder (ASD) is a neurodevelopmental disorder identified by impairments in social interaction and communication and the presence of repetitive and restricted behaviors/interests.[1] The prevalence of ASD has increased in recent years and is now estimated at 1 in 88 children.[2] Aggressive early intervention leads to the best long-term prognosis.[3] Because ASD can often be detected before a child's third birthday, the American Academy of Pediatrics recommends autism-specific screening at 18- and 24-month well-child care (WCC) visits.[4] However, the median age of diagnosis is after the fourth birthday[2] and even later for children of low socioeconomic status or minority backgrounds.[5]

The Modified Checklist for Autism in Toddlers (M-CHAT)[6] is currently one of the most widely used ASD screening instruments both in the United States and internationally,[7,8] providing an accessible, low-cost[9] option for universal toddler screening. The M-CHAT with Follow-Up (M-CHAT/F) has been shown to have adequate sensitivity and specificity[10,11]; in a sample of nearly 19 000 toddlers aged 16 to 30 months,[12] 54% of children classified as at risk on the basis of the M-CHAT/F were diagnosed with ASD, and 98% of screen-positive cases presented with developmental delay or concerns. The purpose of revising the M-CHAT was to reduce the number of cases who initially screen positive and need the follow-up, while maintaining high sensitivity. The current study validates the M-CHAT, Revised with Follow-Up (M-CHAT-R/F), in a low-risk sample.

METHODS

Participants

A total of 16 115 toddlers were screened (see Fig 1) in metropolitan Atlanta (Georgia State University [GSU]) or Connecticut (University of Connecticut [UConn]) (see Table 1). Participants with insufficient data (n =459) were excluded from analyses: 303 did not complete follow-up and 156 did not complete the evaluation. Additional participants (n = 44) were excluded for insufficient English proficiency (n = 15), previous ASD diagnosis (n = 4), a medical condition that precluded evaluation (n = 13), withdrawal from the study (n = 2), or being outside the study's screening age (n = 10).

The remaining 15 612 toddlers (mean age: 20.95 months; SD: 3.30 months; range: 16.00–30.95 months) included 7793 boys and 7570 girls (249 with gender unspecified). Twenty-two percent (22.7%) were screened twice before 30 months. The first screen was used in analyses unless the second screen triggered evaluation (n = 16; 0.1%). Of the 419 children invited for diagnostic evaluation, 263 completed evaluations (see Table 2).

Measures

The M-CHAT-R/F is a 2-stage screener (see www.mchatscreen.com and Supplemental Appendix), which is free for clinical, research, and educational use and requires little or no training for health care professionals. Initially, parents answer 20 yes/no questions, which takes <5 minutes; if children screen positive, parents are asked

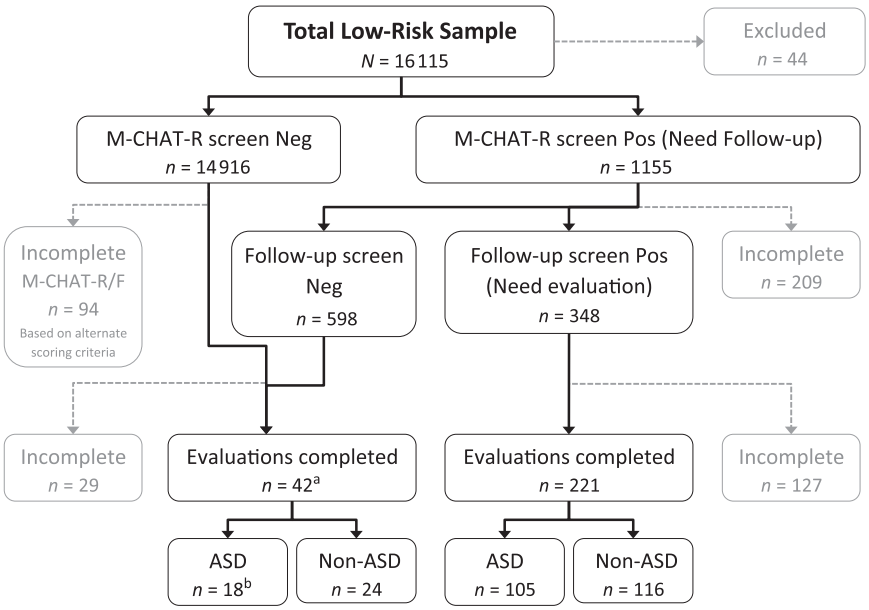


FIGURE 1
Flowchart indicating screening results. [a]Evaluations based on cases detected through STAT screen positive (n = 20), physician concern (n = 18), and alternate scoring (n = 4). [b]Detected through STAT (n = 6), physician concern (n = 9), and alternate scoring (n = 3). Neg, negative; Pos, positive.

TABLE 1 Total Sample Demographic Characteristics by Site and Level of Completion

	UConn		GSU		Total
	Complete	Incomplete	Complete	Incomplete	
Total sample	5932	115	9680	344	16 071
Sex					
Male	3048	71	4745	192	8056
Female	2877	43	4693	143	7756
Not reported	7	1	242	9	259
Race					
White/Caucasian	3712	47	4850	89	8698
Black/African American	426	15	2874	185	3500
Asian/Pacific Islander	316	15	412	9	752
Native American/ Alaskan Native	26	0	11	1	38
Bi- or multiracial	464	9	465	17	955
Other	46	3	56	5	110
Unknown	942	26	1012	38	2018
Ethnicity					
Hispanic	1361	35	435	7	1838
Non-Hispanic	4571	80	9245	337	14233
Maternal education					
Less than high school	429	15	379	36	859
High school/GED	1149	36	1389	82	2656
Some college	1356	25	2201	70	3652
College degree	1512	19	2477	60	4068
Advanced degree	1284	14	1764	29	3091
Unknown	202	6	1470	67	1745
ASD concerns[a]	13	6	32	13	65
Age at screening, mean (SD), mo	20.86 (3.20)	20.53 (3.04)	21.01 (3.36)	20.71 (3.57)	20.94 (3.30)
M-CHAT-R total score, mean (SD)[b]	0.68 (1.22)	4.63 (2.54)	0.64 (1.24)	4.10 (2.96)	0.76 (1.44)
M-CHAT-R/F total score, mean (SD)[b]	1.24 (2.09)	3.15 (1.75)	1.17 (2.15)	3.06 (2.20)	1.43 (2.20)

Data are presented as *n* unless otherwise indicated. Incomplete cases were eligible to complete additional steps based on M-CHAT-R scores but did not complete the study. Incomplete cases include those who initially screened positive and did not complete the follow-up as well as those who continued to screen positive on follow-up and did not attend the evaluation. Completed cases completed all eligible steps of the study, but in most cases, children who screened negative were not asked to continue to any additional assessment. GED, general educational diploma.

[a] Physicians noted ASD concerns by checking a box on the M-CHAT-R protocol.

[b] Note that the incomplete columns contain those cases who screened positive but did not complete the follow-up or evaluation. Therefore, it is expected that they show elevated scores relative to the completed cases.

structured follow-up questions to obtain additional information and examples of at-risk behaviors, which takes ~5 to 10 minutes with a professional (ie, nurse or physician's assistant).[13] The M-CHAT-R/F[14] incorporated 5 modifications to improve utility. Three items that performed poorly were dropped (peek-a-boo, playing with toys, and wandering without purpose). The remaining 20 items were reorganized to remove agreement bias. The items that comprised the Best7 score (see Supplemental Information) were placed within the first 10 items. Language was simplified to improve comprehension. For example, "Does your child ever use his/her index finger to point…" was rephrased as "Does your child point with one finger…." Finally, examples provided developmental context and clarity.

The original M-CHAT recommended a threshold of ≥3 items total or ≥2 critical items identified through discriminant function analysis.[15] However, analyses of larger samples indicated that the critical score did not improve sensitivity above the total score.[12] The current study tested several scoring methods. A threshold based on total score had strong psychometric properties and was more parsimonious than combinations of total and alternative scorings (see Supplemental Information).

Clinical measures included the Autism Diagnostic Observation Schedule,[16] Childhood Autism Rating Scale–2,[17] the Toddler Autism Symptom Interview,[18] Mullen Scales of Early Learning,[19] Vineland Adaptive Behavior Scales–II,[20] Behavioral Assessment System for Children–2,[21] and a developmental history form.

Procedures

Parents completed the M-CHAT, Revised (M-CHAT-R), and provided informed consent and demographic characteristics during their child's 18- or 24-month WCC visit (41 sites at GSU, 44 sites at UConn). Pediatricians were asked to indicate concern about ASD, based on their clinical judgment, by checking a box at the top of the screener. Completed M-CHAT-R forms were scored at GSU or UConn. Research staff contacted parents of screen-positive children to complete the follow-up by telephone; children who continued to screen positive on the M-CHAT-R/F or whose physician had concerns were offered a diagnostic evaluation. Evaluations were conducted by a team consisting of a licensed psychologist/developmental pediatrician supervising a graduate student and research assistants; team members were research reliable on all measures they administered.

The final diagnosis integrated all available information and used the psychologist/developmental pediatrician's clinical judgment to assess *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, text revision (DSM-IV-TR)[22] criteria for Autistic Disorder and Pervasive Developmental Disorder, Not Otherwise Specified. When ASD was ruled out, diagnoses of Global Developmental Delay, Language Delay, or other DSM-IV-TR disorders were considered. Children who did not meet criteria for any diagnosis were classified as typically developing or as having developmental concerns, which were operationally defined as

TABLE 2 Sample Characteristics of Children Who Completed Evaluations

	Evaluated Based on		Total
	Screen Positive (84%)	Screen Negative (16%)	
Total sample	221	42	263
ASD	105	18	123
Autistic	52	10	62
PDD-NOS	53	8	61
Non-ASD	116	24	140
GDD	55	6	61
Language	23	2	25
Other diagnosis	1	0	1
No diagnosis	25	5	30
Typical	12	11	23
Sex			
Male	146	29	175
Female	75	13	88
Race			
White/Caucasian	96	17	113
Black/African American	64	16	80
Asian/Pacific Islander	14	2	16
Native American/ Alaskan Native	0	0	0
Bi- or multiracial	19	2	21
Other	1	1	2
Unknown	27	4	31
Ethnicity			
Hispanic	36	5	41
Non-Hispanic	185	37	222
Maternal education			
Less than high school	28	3	31
High school/GED	41	8	49
Some college	59	8	67
College degree	53	13	66
Advanced degree	40	10	50

Data are presented as *n*. GDD, global developmental delay; GED, general educational diploma; PDD-NOS, pervasive developmental disorder, not otherwise specified.

subthreshold/mild weaknesses precluding a label of typical development.

Parents received oral and written feedback, including local intervention resources. When parents declined to complete follow-up or evaluation, the physician was informed of M-CHAT-R results; parents were welcome to rejoin the study at any time. Institutional review boards at both sites approved this study.

At GSU, a stratified random sample of children who screened negative on the M-CHAT-R/F were invited to complete the Screening Tool for Autism in Two-Year-Olds (STAT),[23] a brief autism-specific play-based screening. To maximize the chance of finding missed cases, children who initially screened positive on M-CHAT-R but then screened negative on follow-up were most heavily recruited for the STAT; among those who screened negative on the initial M-CHAT-R, stratification overrecruited those who scored 2 compared with those who scored 1 or 0. Of 375 children who completed the STAT, 20 were evaluated on the basis of a screen-positive STAT.

RESULTS

Reliability

Across all M-CHAT-R items, internal consistency was below the threshold for adequate (Cronbach's α = 0.63), which is not surprising given that the M-CHAT-R items do not assess a unitary dimension, and some motor items were created to be foils. When the 2-stage screen was examined, internal consistency for M-CHAT-R/F was adequate (Cronbach's α = 0.79).

Outcomes for Screen-Positive Cases

The majority of cases (92.6%) who completed an M-CHAT-R screened negative. More than half (n = 598; 63.2%) of children whose parents completed the second stage of the M-CHAT-R/F (follow-up), no longer screened positive. The mean age at evaluation was 26.23 months (SD: 5.45 months).

Optimal Scoring for M-CHAT-R/F

To evaluate scoring for the M-CHAT-R/F in a low-risk sample, receiver operating characteristic curves verified optimal cutoff scores for the 2-stage M-CHAT-R/F. Sensitivity was calculated as the proportion of all ASD children identified by any means (M-CHAT-R/F, physician concern, STAT) who screened positive. Specificity was calculated as the proportion of all presumed non-ASD cases who screened negative.

Initial M-CHAT-R Scoring

Area under the curve was 0.977. The threshold for which both sensitivity and specificity exceeded 0.90 was 3, supporting the established cutoff score; increasing or decreasing the cutoff led to a notable drop in sensitivity or specificity.

Two-Stage M-CHAT-R/F Scoring

Initially, the same cutoff score of ≥3 on the total score (Total3) was used for the follow-up. However, as a result of increased efforts to ascertain missed cases (ie, physician concerns and STAT screening), 7 screen-negative cases were diagnosed with ASD, 5 of whom scored 2 on the M-CHAT-R/F. This result led to a change in the threshold for those who were offered evaluations on the basis of results of the follow-up to Total2 (cutoff score of ≥2 on the total score). Results indicated that using

Total2 as the threshold on the follow-up halves the number of missed cases, significantly increasing sensitivity (McNemar's test, $P < .001$) and demonstrates an area under the curve of 0.907 (see Table 3 and Fig 2). Psychometrics were verified on the subsample ascertained after the score change was implemented ($n = 7579$; 60 diagnosed with ASD).

To increase utility of a 2-stage screening tool in busy pediatric settings, it would be helpful to bypass the follow-up in cases who are likely to continue to screen positive. The sample was examined to determine (1) the number of cases who reverted from screen positive to screen negative during the follow-up and (2) the initial M-CHAT-R scores for those children diagnosed with ASD to arrive at the following risk classifications: low risk (total score: 0–2; requires no further evaluation unless other risk factors are present), medium risk (total score: 3–7; requires administration of the follow-up to determine whether referrals are warranted), and high risk (total score: 8–20; warrants immediate referral for evaluation and intervention) (see Fig 3). In the current sample, 75 children scored in the high-risk range on M-CHAT-R and completed the evaluation, all of whom were diagnosed with developmental disorders or concerns (44 ASDs, 27 non-ASD disorders, 4 developmental concerns). Compared with a positive predictive value (PPV) of the initial questionnaire of 0.26 for any developmental delay or concern (confidence interval [95% CI]: 0.20–0.32), the PPV for high-risk scores is 1.0. It is

TABLE 3 Psychometric Properties of M-CHAT-R and M-CHAT-R/F Scores

	TP (hit)	FN (miss)	FP	TN	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV	LR+	LR−
M-CHAT-R initial scoring										
Total3[a]	112	11	700	14 798	0.911 (0.86–0.96)	0.955 (0.95–0.96)	0.138	0.999	20.160	0.094
M-CHAT-R/F scoring										
Total3 w/follow-up Total3[b]	82	41	79	15 419	0.667 (0.58–0.75)	0.995 (0.99–0.99)	0.509	0.997	130.785	0.335
Total3 w/follow-up Total2[c]	105	18	116	15 382	0.854 (0.79–0.92)	0.993 (0.99–0.99)	0.475	0.999	114.052	0.147
Subsample of Total3 w/follow-up Total2[d]	50	10	61	7458	0.833 (0.73–0.93)	0.992 (0.98–0.99)	0.450	0.999	103.087	0.168

FN, false-negative cases; FP, false-positive cases; LR+, likelihood ratio of positive screen; LR−, likelihood ratio of negative screen; NPV, negative predictive value; TN, true-negative cases, presumed based on screening negative without other indicators of ASD risk; TP, true-positive cases.

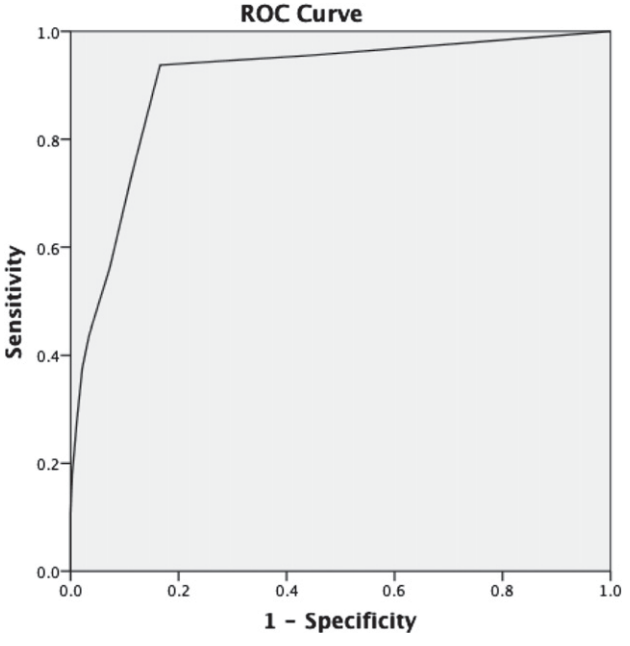
[a] Total3: classification as TP, FN, FP, or TN based on a threshold of 3 on the total score.

[b] w/follow-up: these scores are represented by the threshold(s) applied to the initial screening with the additional threshold(s) applied to the follow-up.

[c] Total2: classification as TP, FN, FP, or TN based on a threshold of 2 on the follow-up total score.

[d] Subsample after score change for follow-up implemented to verify Total3 – with/follow-up Total2 scoring in a prospective sample.

A



B

AUC = 0.907

Cutoff on M-CHAT-R/F Total Score	Sensitivity	Specificity
1	0.96	0.56
2	0.94	0.83
3	0.73	0.89
4	0.56	0.93
5	0.44	0.97
6	0.38	0.98
7	0.28	0.99
8	0.18	0.99
9	0.14	0.99

FIGURE 2

A: ROC curve for 2-stage M-CHAT-R/F. AUC, area under the curve; ROC, receiver operating characteristic. B: Table showing sensitivity and specificity for each M-CHAT-R/F total score.

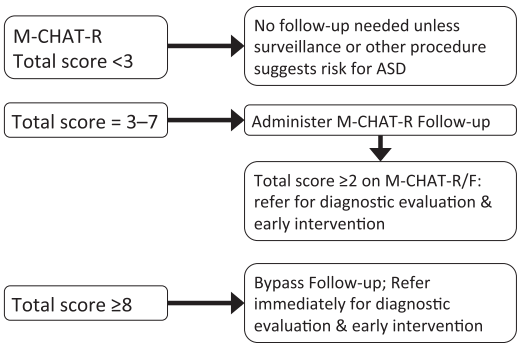


FIGURE 3
Recommended algorithm based on 2-stage M-CHAT-R/F screening.

important to note that many children with ASD will score lower than these higher cutoffs, emphasizing the need to complete the follow-up with medium-risk cases.

Finally, examining specific diagnostic outcomes for screen-positive cases on the 2-stage M-CHAT-R/F (Total3 initially + Total2 on follow-up) indicated 47.5% (n = 105) diagnosed with ASD, yielding a likelihood ratio for positive screens of 114.052. Among the remaining cases, 35.7% (n = 79) had other delays, 11.3% (n = 25) had developmental concerns but no formal diagnosis, and only 5.4% (n = 12) were judged to be typically developing; the PPV for any developmental delay or concern was 0.946 (95% CI: 0.92–0.98).

Among cases flagged by the physician for ASD concerns (n = 64), 45 attended the evaluation; of these cases, 42 had delays or concerns, 30 of whom were diagnosed with ASD; this finding indicates that physician concern alone has a sensitivity of 0.244 (30 of 123 ASD cases; 95% CI: 0.17–0.32). Notably, physicians were more likely to express ASD concerns when parents were highly educated. See Table 4 for clinical characterization of the sample by screening status.

Comparison of M-CHAT-R/F to the Original M-CHAT/F

To investigate whether revision to the M-CHAT improved the tool for use in low-risk samples, the outcomes from the current validation study for M-CHAT-R/F were compared with the original M-CHAT/F sample as reported in Chlebowski et al.[12] There was a significant reduction in the initial screen-positive rate (from 9.15% to 7.17%; χ^2 [1, n = 35 060] =39.62; $P < .001$); the PPV for the 2-stage screening was not significantly different across versions (P = .492). The rate of ASD detection was significantly higher for the M-CHAT-R/F, which detected 67 cases per 10 000 compared with the original M-CHAT/F, which detected 45 cases per 10 000 (χ^2 [1, n = 35 060] = 8.63; P = .003).

DISCUSSION

The current study validates the M-CHAT-R/F, a 2-stage, level 1 ASD screening tool that requires little time and cost[9] to administer to toddlers attending 18- and 24-month WCC visits. Analyses indicated that optimal scoring relies only on total, rather than alternate, scoring.

The recommended algorithm classifies children into 3 risk ranges on the basis of the initial questionnaire. Children who score in the low-risk range (93% of cases) are not in need of M-CHAT-R follow-up or additional evaluation unless surveillance indicates ASD risk. Children should be rescreened if they are younger than 24 months, as recommended by the American Academy of Pediatrics.[4] Children whose scores are in the medium-risk range (6% of cases) require administration of the follow-up, which gathers additional detail about at-risk items. Approximately one-third of children whose parents complete the second stage of M-CHAT-R/F continue to show ASD risk and require referrals for evaluation and possible early intervention. Children who score in the high-risk range (1% of cases)

TABLE 4 Clinical Data by Screening Status

	True Positive (n = 105)		False Negative (n = 18)		False Positive (n = 116)		True Negative (n = 24)	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Age at screen, mo	21.87	(3.83)	22.74	(4.08)	20.80	(3.43)	22.27	(3.63)
Age at evaluation, mo	26.22	(4.91)	29.57	(3.58)	25.34	(6.11)	28.04	(4.17)
M-CHAT-R total	7.23	(3.24)	3.50	(3.19)	5.96	(2.35)	2.50	(2.50)
M-CHAT-R/F total	5.22	(3.17)	0.80	(0.92)	3.55	(1.57)	0.40	(0.70)
ADOS-2 comparison[a]	6.54	(1.98)	6.89	(1.88)	1.85	(1.29)	2.13	(1.26)
CARS2[b]	32.66	(5.02)	33.19	(4.36)	21.11	(3.39)	20.23	(4.09)
MSEL VR[c]	29.64	(10.86)	31.94	(12.99)	38.18	(13.77)	44.21	(15.65)
MSEL FM	27.59	(10.30)	31.33	(11.55)	35.16	(11.06)	39.21	(14.05)
MSEL EL	25.44	(8.63)	25.44	(7.24)	31.68	(10.65)	36.79	(10.64)
MSEL RL	23.70	(8.97)	26.17	(9.41)	33.14	(12.44)	39.75	(11.64)
VABS-II Communication[d]	71.85	(12.57)	79.89	(12.32)	85.11	(12.76)	95.46	(9.56)
VABS-II DL	80.00	(14.36)	85.06	(10.40)	90.11	(15.04)	98.29	(13.03)
VABS-II Socialization	77.23	(10.10)	80.78	(12.36)	86.79	(11.29)	96.58	(10.44)
VABS-II Motor	83.93	(11.27)	86.11	(11.17)	88.44	(13.67)	94.96	(9.82)

ADOS-2, Autism Diagnostic Observation Schedule, Second Edition; CARS2, Childhood Autism Rating Scale, Second Edition; DL, Daily Living Skills; EL, expressive language; FM, fine motor; MSEL, Mullen Scales of Early Learning; RL, receptive language; VABS-II, Vineland Adaptive Behavior Scales, Second Edition; VR, visual reception.
[a] ADOS-2 comparison score indicates severity of ASD-related symptoms.
[b] CARS2 range:15–60; threshold for autism = 30.
[c] MSEL T scores (mean = 50, SD = 10).
[d] VABS-II standard scores (mean = 100, SD = 15).

may bypass the follow-up. On the basis of initial screening only, the total sample of screen-positive children have a 27% risk of any developmental delay or concern (95% CI: 0.20–0.32), whereas all cases in the high-risk range were diagnosed with delays or concerns, justifying immediate referrals for evaluation and possible early intervention.

Children who screened positive on M-CHAT-R/F were 114 times more likely to receive an ASD diagnosis than children who screened negative. In addition, 94.6% of children evaluated for ASD risk on the 2-stage M-CHAT-R/F showed developmental delay or concern that warranted referrals to early intervention (95% CI: 0.92–0.98). Although one might interpret this finding to mean that the M-CHAT-R/F may be screening more broadly than for ASD, it is not justified to use the screener for that purpose, given that the sensitivity of the tool for non-ASD delays is not known. One study directly comparing M-CHAT to the Parents' Evaluation of Developmental Status (PEDS)[24] found that 25% of children demonstrated risk for a broad range of developmental concerns on the PEDS, far exceeding the screen-positive rate of the M-CHAT.[25] An important finding from the study is that the average age of diagnosis was just after the second birthday, which is 2 years earlier than the median age of diagnosis[2]; this finding suggests that implementing standardized screening and expeditious evaluation for positive cases can greatly increase the time that children are eligible for early-intervention services and therefore improve the outcome. However, it is important to note that ASD screening continues to be challenging. Because no screening tool can have perfect sensitivity and specificity, providers should continue to perform developmental surveillance in addition to using validated screening tools.

The performance of the M-CHAT-R/F was compared with the published studies using the M-CHAT/F in low-risk samples.[12] Overall, the revision significantly reduced the initial screen-positive rate, which means that fewer children require follow-up. Also indicating improvement of the tool, the rate of ASD detection increased for M-CHAT-R/F. Although it is impossible to rule out increasing ASD prevalence as contributing to this finding, this finding suggests that the reduction in the initial screen-positive rate is not negatively affecting sensitivity. When M-CHAT-R/F was compared with physician clinical judgment, sensitivity was significantly higher for M-CHAT-R/F; when these methods were combined, ASD detection was very high, indicating that standardized screening in conjunction with routine developmental surveillance optimizes early detection for ASD.

It is important to examine the prevalence of ASD detected in the current sample to evaluate utility of the screening tool. The M-CHAT-R/F detected ASD at a rate of 1 per 149 cases. This rate is notably below the published prevalence of ASD as 1 in 88[2]; however, the Centers for Disease Control and Prevention's prevalence data were ascertained on the basis of review of school and health records for 8-year-old children, and it is not expected that all ASD cases will be detectable in toddlers. Furthermore, with enhanced methods to detect missed cases, such as following up on physician concern and sampling screen-negative cases, 123 ASD cases were detected in the sample, which is 1 in 127. It is likely that many of the remaining children who will later be diagnosed with ASD, such as those with Asperger disorder without early developmental delays, are not showing significant symptoms at this young age[26]; in addition, later detection may occur in mild cases only in the school setting where peer interactions can be seen. Therefore, the rate of detection of 1 in 127 may not be far off from the actual prevalence in 2-year-olds.

Limitations and Future Directions

One limitation of level 1 screening research is that it is impossible to evaluate all screen-negative cases to identify misses and calculate true sensitivity. An additional challenge is that many parents of children who initially screened positive did not complete additional steps in the study.[27] The current study had a disproportionately high number of African-American families who did not complete the study (eg, follow-up or evaluation), indicating that barriers continue to exist even under standardized protocols. In addition, maternal education was significantly higher in the GSU sample (mean: 14.93 years; SD: 2.53 years) than the UConn sample (mean: 14.57 years; SD: 2.46 years) (t [13 938] = −8.37; $P < .001$), although the effect size was very small (η^2 = 0.005). These variables are complex and are addressed in other articles.[28]

The current study used multiple approaches to detect possible false-negative cases, improving accuracy of sensitivity estimates. Both sites asked physicians to identify cases of possible ASD, and these families were offered evaluation regardless of M-CHAT-R/F score. In some cases, when physicians had ASD concerns but the child screened negative on the M-CHAT-R/F, the children were found to have other developmental delays; however, 9 ASD cases were detected with physician concern but had negative M-CHAT-R/F. Not all physicians applied this surveillance component equally in their practices, and further research may identify factors that predict use of surveillance, screening, and their integration. A second approach to find false negatives conducted at 1 site (GSU) invited a sample of screen-negative cases for play-based screening. The sample was

stratified to oversample cases who just missed screening positive. A barrier to this approach was that many families declined to schedule or attend this session. However, of the 375 completed, 7 ASD cases were detected, suggesting that this is a successful strategy for finding cases missed by the M-CHAT-R/F. In fact, these cases contributed to the change in threshold for the follow-up. A final approach to finding missed cases is under way, rescreening participants by mail when they are 3.5 to 4 years old.

Future research should validate the M-CHAT-R/F in high-risk samples. Our group is screening children with older siblings already diagnosed with ASD, but additional high-risk groups include children referred for early intervention but not yet diagnosed and children with risk factors for developmental delay, such as prematurity. It will be essential for such validation studies to determine diagnosis, to evaluate the utility of the M-CHAT-R/F in these high-risk samples, and to assess whether the published thresholds for low-risk, or level 1, samples are the same or different in these high-risk groups.

Additional examination of thresholds for low-, medium-, and high-risk scores on the M-CHAT-R will also be fruitful. As the threshold for bypassing the follow-up and referring immediately is lowered, the risk of unnecessary referrals increases. A risk-benefit analysis may help balance the cost of unnecessary referrals against the benefit of immediate referral for those who need intensive early intervention.

Conclusions

The M-CHAT-R/F is an effective tool to screen for ASDs in low-risk pediatric samples. Integration of screening and surveillance strategies reduces the age of ASD diagnosis by 2 years, facilitating early intervention and optimizing long-term prognosis. The simplified scoring of the M-CHAT-R/F, paired with specific algorithms based on outcome, should ease implementation.

ACKNOWLEDGMENTS

We thank all of the toddlers and their families for participating in the screening study. In addition, we thank the pediatricians and health care providers for distributing the M-CHAT-R during well-child care visits and all of the members of the research team who participated in data collection. We also thank our funding source, the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development.

REFERENCES

1. Autism Spectrum Disorder. *American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013:50–59
2. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2008 Principal Investigators; Centers for Disease Control and Prevention. Prevalence of autism spectrum disorders—Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *MMWR Surveill Summ*. 2012;61(3):1–19
3. Myers SM, Johnson CP; American Academy of Pediatrics Council on Children With Disabilities. Management of children with autism spectrum disorders. *Pediatrics*. 2007;120(5):1162–1182
4. 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(5):1183–1215
5. Mandell DS, Listerud J, Levy SE, Pinto-Martin JA. Race differences in the age at diagnosis among Medicaid-eligible children with autism. *J Am Acad Child Adolesc Psychiatry*. 2002;41(12):1447–1453
6. Robins DL, Fein D, Barton M. *The Modified Checklist for Autism in Toddlers (M-CHAT)*. Self-published; 1999
7. 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*. 2011;41(10):1342–1351
8. Inada N, Koyama T, Inokuchi E, et al. 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
9. Gura GF, Champagne MT, Blood-Siegfried JE. Autism spectrum disorder screening in primary care. *J Dev Behav Pediatr*. 2011;32(1):48–51
10. Dumont-Mathieu T, Fein D. Screening for autism in young children: the Modified Checklist for Autism in Toddlers (M-CHAT) and other measures. *Ment Retard Dev Disabil Res Rev*. 2005;11(3):253–262
11. 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
12. Chlebowski C, Robins DL, Barton ML, Fein D. Large-scale use of the Modified Checklist for Autism in low-risk toddlers. *Pediatrics*. 2013;131(4):e1121–e1127
13. Nygren G, Sandberg E, Gillstedt F, Ekeroth 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
14. Robins DL, Fein D, Barton M. *The Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M-CHAT-R/F)*. Self-published; 2009
15. Robins DL, Fein D, Barton ML, Green JA. The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *J Autism Dev Disord*. 2001;31(2):131–144
16. 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
17. Schopler E, Reichler RJ, Renner BR. *The Childhood Autism Rating Scale-2*. Los

Angeles, CA: Western Psychological Services; 2010

18. Barton M, Boorstein H, Herlihy L, Dumont-Mathieu T, Fein D. *Toddler ASD Symptom Interview*. Self-published; 2012
19. Mullen EM. *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service, Inc; 1995
20. Sparrow SS, Cicchetti DV, Balla DA. *Vineland Adaptive Behavior Scales*. 2nd ed. (Vineland II) Survey Interview Form/Caregiver Rating Form. Livonia, MN: Pearson Assessments; 2005
21. Reynolds CR, Kamphaus RW. *Behavior Assessment System for Children (BASC-2)*. 2nd ed. Bloomington, MN: Pearson Assessments; 2004
22. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed, text revision. Washington, DC: American Psychiatric Association; 2000
23. Stone WL, Coonrod EE, Ousley OY. Brief report: Screening Tool for Autism in Two-year-olds (STAT): development and preliminary data. *J Autism Dev Disord*. 2000;30(6):607–612
24. Glascoe FP. *Parents' Evaluation of Developmental Status (PEDS)*. Nolensville, TN: PEDSTest.com, LLC; 2010. Available at: www.pedstest.com. Accessed September 9, 2013
25. Wiggins LD, Piazza V, Robins DL. Comparison of a broad-band screen versus disorder-specific screen in detecting young children with an autism spectrum disorder [published online December 21, 2012]. *Autism*. 2012. Available at: http://aut.sagepub.com/content/early/2012/12/21/1362361312466962.full.pdf. (Accessed September 4, 2013)
26. Marks K, Glascoe FP, Aylward GP, Shevell MI, Lipkin PH, Squires JK. The thorny nature of predictive validity studies on screening tests for developmental-behavioral problems. *Pediatrics*. 2008;122(4):866–868
27. Pierce K, Carter C, Gallagher N, et al. Detecting, studying, and treating autism early: the one-year well-baby check-up approach. *J Pediatr*. 2011;159(3):458–465
28. Herlihy L, Brooks B, Dumont-Mathieu T, et al. Standardized screening facilitates timely diagnosis of autism spectrum disorder in a diverse sample of low-risk toddlers. *J Dev Behav Pediatr*. 2013, In press

(Continued from first page)

www.pediatrics.org/cgi/doi/10.1542/peds.2013-1813

doi:10.1542/peds.2013-1813

Accepted for publication Oct 21, 2013

Address correspondence to Diana L. Robins, PhD, Department of Psychology, Georgia State University, PO Box 5010, Atlanta, GA 30302-5010. E-mail: drobins@gsu.edu

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

Copyright © 2014 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: Dr Robins is co-owner of M-CHAT LLC, which receives royalties from parties that license use of the M-CHAT in electronic products. Drs Fein and Barton are co-owners of M-CHAT LLC; they receive royalties, which are entirely allocated to research and clinical training expenditures. No royalties were received for any of the data presented in the current study. The other authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: This study was supported by *Eunice Kennedy Shriver* National Institute of Child Health and Human Development grant R01HD039961. Funded by the National Institutes of Health (NIH).

POTENTIAL CONFLICT OF INTEREST: Dr Robins is co-owner of M-CHAT LLC, which licenses use of the M-CHAT in electronic products. The other authors have indicated they have no potential conflicts of interest to disclose.