Original Article

The Developmental Check-In: Development and initial testing of an autism screening tool targeting young children from underserved communities

Yvette M Janvier[1], Caroline N Coffield[2], Jill F Harris[3] [ID], David S Mandell[4] and Zuleyha Cidav[4]

Abstract

Children with autism spectrum disorder from low-income, minority families or those with limited English proficiency are diagnosed at a later age, or not at all, compared with their more advantaged peers. The Developmental Check-In is a new tool that could potentially be used to screen for autism that uses pictures to illustrate target behaviors. It was developed to enhance early identification of autism spectrum disorder in low literacy groups. The Developmental Check-In was tested in a sample of 376 children between the ages of 24 and 60 months, from underserved communities. It showed good ability to discriminate autism spectrum disorder from non-autism spectrum disorder (area-under-the-curve = 0.75) across the full age range represented in the sample. Twenty-six of the 28 Developmental Check-In items predicted the presence of autism spectrum disorder. Findings suggest that this pictorial tool may reduce linguistic and health literacy demands when screening for autism among vulnerable populations.

Keywords

autism, disparities, screening, underserved children, vulnerable children

Early detection and diagnosis of autism spectrum disorder (ASD) is critical for connecting children to intervention and maximizing developmental outcomes (Dawson et al., 2010). While an experienced clinician can diagnose ASD in children as young as 18 months (Ozonoff et al., 2015), the median age of diagnosis in the United States is about 50 months (Christensen et al., 2016). This is very close to the median age previously reported by the Centers for Disease Control and Prevention (CDC) in 2000 and 2002, suggesting that increased awareness has not resulted in earlier autism diagnosis (Christensen et al., 2016). Diagnosis is even later or missed altogether among children living in poverty and those from racial or ethnic minorities (Durkin et al., 2010, 2017; Fountain et al., 2010; Mandell et al., 2002, 2009).

Factors contributing to delayed and missed diagnosis among poor children and children from racial and ethnic minorities are complex and not well understood, but could, in part, be the result of cultural, language, social support, religious or literacy issues (Colbert et al., 2017). In a large prospective study of toddlers (Herlihy et al., 2014), minority children were neither more nor less likely to be diagnosed with ASD. However, parental reading level, understanding of questions, and frame of reference for comparison to other children affected screening and evaluation results. Tek and Landa (2012) suggest that culture influences parents' expectations for and interpretation of child behavior, as well as communication with healthcare providers. When minority children are identified early, their symptoms often are more severe than those of white children identified at the same age (Liptak et al., 2008; Tek and Landa, 2012), suggesting that minority parents are less likely than other parents to view more subtle symptoms of ASD as indicating atypical development. For parents for whom English is a second language, language differences

[1]Children's Specialized Hospital – Toms River, USA
[2]Rutgers Robert Wood Johnson Medical School, USA
[3]Children's Specialized Hospital – Mountainside, USA
[4]University of Pennsylvania, USA

Corresponding author:
Jill F Harris, Children's Specialized Hospital – Mountainside, 150 New Providence Road, Mountainside, NJ 07092, USA.
Email: jharris@childrens-specialized.org

may impede all phases of ASD identification, from awareness and communication of developmental concerns to monitoring and screening. Taken together, these findings support the need for targeted strategies to increase identification of ASD in minority communities (Christensen et al., 2016; Liptak et al., 2008).

An important part of screening for ASD involves the use of standardized screening tools. Two of the most commonly used autism screening tools are the Modified Checklist for Autism in Toddlers—Revised (M-CHAT-R; Robins et al., 1999) and the Social Communication Questionnaire (SCQ; Rutter et al., 2003). Some studies find that the M-CHAT and its revision, the M-CHAT-R (Robins et al., 2014), result in many false positives when completed by racial minority parents with low income or low level of education (Khowaja et al., 2015; Windham et al., 2014). Scarpa et al. (2013) also question the M-CHAT's accuracy among low-income, minority, and less educated parents or those from rural communities. While using the M-CHAT's recommended follow-up interview reduces false positives, implementing it may be unrealistic due to limited staff resources (Windham et al., 2014) and challenges reaching families for follow-up (Khowaja et al., 2015). Khowaja et al. (2015) further note that clarifying wording of items and providing written examples did not improve accuracy for families with lower maternal education, suggesting that literacy may be a factor, even though the M-CHAT-R requires only a fourth-grade reading level. They speculated that reading items aloud to the parents may help, but note that this also is labor intensive. Furthermore, some translated M-CHAT items were found to be confusing to Spanish-speaking parents (Kimple et al., 2014; Windham et al., 2014). These challenges are not limited to the M-CHAT. Moody et al. (2017) found that the specificity of the SCQ decreased for non-white mothers and for mothers with less education or lower family income. This decrease in specificity occurred regardless of whether the SCQ cut-off score was 15, as recommended in the SCQ manual, or a lower cut-off of 11 or 13 was used for children under 48 months. The authors concluded that there is a need for ASD screeners that better identify ASD across cultures and socio-economic strata.

In the authors' experience, garnered serving a low-income, urban population, the literacy demands of the M-CHAT and the SCQ may affect their accuracy in underserved groups. Specifically, initial parental response on the screening tools frequently is inconsistent with parent descriptions provided during follow-up interviews, particularly among parents with limited English proficiency or formal education. Many parents report confusion about the meaning of the items, despite the tools being provided in their primary language. Parents suggested that pictures of behaviors are easier to understand than written descriptions.

There are few examples of pictorial-based parent-report screening tools. Two tools were created for families outside of the United States and one tool focuses on only one domain that is associated with ASD. Perera et al. (2017) developed a visually-based ASD screening tool for use in a Sri Lankan population but the validation sample was relatively small, and it is unclear whether it is culturally valid in the United States. An Argentinian study described a general developmental screener completed by parents or healthcare staff that included illustrations (Lejarraga et al., 2013). Reported sensitivity was weak, and the tool was not ASD specific. The Pictorial Infant Communication Scale (PICS) is a parent-report tool to screen for joint attention that includes photos to aid caregivers in understanding and identifying behaviors of interest (Ghilain et al., 2017). PICS scores were correlated with ASD severity in a sample of primarily white preschoolers who had previously been diagnosed with ASD. Study limitations, however, included focus on joint attention alone, parents' awareness of their child's diagnosis which possibly influenced their responses on the tool, and photos that depicted only infants, even though the tool was used to screen preschool-aged children.

In response to the need for a visually-based ASD screening tool and the limitations of existing tools, we developed a visually-based ASD screening tool, the Developmental Check-In (DCI). In this study, we describe its development and its initial testing in a sample that combined children with developmental concerns at high risk for ASD and children recruited through community daycare centers at low risk for ASD.

Methods

This study received approval from the Rutgers Biomedical Health Sciences Institutional Review Board.

Development and design of the DCI

The DCI is an autism screening tool that uses pictures rather than written descriptions of behavior to reduce literacy demands and improve the clarity of the target behaviors. It includes visual depictions of key language, social-developmental milestones, and red flag behaviors.

To develop the DCI, a pool of 39 items was created based on (a) known red flags for ASD in young children and (b) items from the Autism Diagnostic Observation Schedule—Second Edition (ADOS-2) or the Screening Tool for Autism in Toddlers (STAT; Stone and Ousley, 2008) that differentiated between children with and without autism. Each item was then depicted as both a line drawing and a photograph. When necessary to promote clarity, brief labels were added to the items. For example, a bubble coming from the mouth of the adult with the name of a child was added to the item "Responds to name." Separate versions of the screening tool were created with English and Spanish item labels (see Figure 1 for a sample DCI item).

Figure 1. Sample DCI item ("Response to name").

Interviews were conducted with parents of children with ASD (n=16) and parents of typically developing children (n=8). Both English and Spanish-speaking parents from Puerto Rico and other Caribbean islands were included. Parents were shown both the line drawing and photograph versions of the DCI, with order of presentation randomized, and were asked to name the skill/behavior depicted. If the response was not on target for the intended skill/behavior, the interviewer asked "we were trying to show a child (performing skill/behavior). What would help it look like that?" Parents were also asked their preference for drawings or photos in terms of item clarity. Interviews were conducted until a consistent pattern of responses emerged. Parents demonstrated better understanding as well as a general preference for the photographs versus the line drawings. Items that were confusing to the parents or were considered duplicative (e.g. multiple items depicting imitation skills) were deleted. The resulting tool consisted of 28 photographs depicting communication (four items), play (seven items), social (eight items), and behavior (nine items). Items were grouped by domain. Brief descriptors were added for each item. For example, a photo of a young child looking away from adult who is holding her was labeled "Avoids eye contact."

Other screening measures: M-CHAT-R, SCQ

The M-CHAT-R (Robins et al., 2014) is a 20-item parent-report checklist used to examine children's developmental milestones. The M-CHAT-R has been validated as a screening tool for children ages 16–30 months. Previous studies have extended the M-CHAT age range to 48 months (Snow and Lecavalier, 2008). A score of 0–2 is considered low risk, a score of 3–7 is considered medium risk, and scores of 8 or higher are considered high risk. It is suggested that scores within the medium-risk range receive a follow-up interview for further clarification. In this study, the follow-up interview was not given to high-risk subjects because they were evaluated regardless of screening status. Community daycare children were administered the follow-up interview if their initial score was within the medium- or high-risk category.

SCQ (Rutter et al., 2003) is a 40-item parent-report screening measure for ASD based on the Autism Diagnostic Interview—Revised (ADI-R; Rutter et al., 2008). A cut-off score of 15 is considered at risk. The SCQ is designed for children ages 48 months and over; however, previous studies have extended the age range down to 24 months when screening for ASD and considered lower cut-off scores, although sensitivity and specificity were not strong (Allen et al., 2007; Corsello et al., 2007, 2013; Moody et al., 2017; Snow and Lecavalier, 2008).

Developmental assessment measures: ADOS-2, MSEL

The ADOS-2 (Lord et al., 2012) is a semi-structured, play-based assessment for ASD with standardized administration and scoring schema. Child behavior is sampled through the use of systematic probes for ASD symptoms in communication, social interaction, play, and restricted and repetitive behaviors. Algorithm criteria for ASD and autism classification are provided, based on Social Affect as well as Restricted and Repetitive Behavior summary scores. The measure of interest from the ADOS-2 used in analyses was overall algorithm classification (autism and autism spectrum were combined as "ASD"). This algorithm classification also helped inform clinical diagnosis.

Mullen Scales of Early Learning (MSEL; Mullen, 1995) were administered to measure cognitive and language functioning. The Mullen is a comprehensive assessment of development normed for children ages birth to 5 years, 8 months that includes the following scales: Visual Reception, Expressive and Receptive Language, and Fine and Gross Motor. The Mullen yields both standard T-scores and age equivalents for each scale. For this study, standard scores for all scales except Gross Motor were the measures of interest used to establish differential diagnosis.

Participants and procedures

The DCI data were collected from two different samples: a high-risk sample of children referred for developmental evaluation and a low-risk sample of children recruited from community daycare centers. The high-risk sample was selected from children seen at an outpatient developmental evaluation program either at a non-profit pediatric hospital or at Federally Qualified Health Centers (FQHCs) located in five medically underserved communities in a mid-Atlantic state. These children were referred for evaluation because the healthcare clinic staff or parent had developmental concerns about the child. Low-risk children were those who had not been referred for developmental concerns and were recruited through daycare centers located in underserved communities in the same state. The purpose of this sampling strategy was to create a sample that had a high proportion of children who were

poor and for whom English was a second language, and to enrich the sample for children without autism. Including a low-risk sample is similar to the strategy used to validate the M-CHAT (Robins et al., 2001). Children were included in the sample if they were (a) 24–60 months of age and (b) met one or more of the following inclusion criteria: they were Medicaid enrolled or had no insurance, were non-white, were Hispanic/Latino, or English was not the primary language spoken in the home.

Participants recruited from the developmental evaluation clinic were screened for study criteria when they called to make an evaluation appointment. By definition, children receiving healthcare at FQHCs met demographic criteria for this study (Centers for Medicare & Medicaid Services, 2017). Participating FQHCs were asked to refer all children between the ages of 24 and 60 months for whom parents or healthcare staff had noted developmental concern. FQHC staff described the study to the parents (i.e. that parents would be asked to complete several forms, and that the child would receive a free developmental evaluation) and arranged an evaluation appointment for those parents who expressed interest.

Full evaluations were provided regardless of scores on screening tools for the high-risk children recruited from the developmental evaluation clinics and the FQHCs. Evaluations were provided at the hospital clinic site for those children seen through the pediatric hospital and at the FQHC site for those children referred from FQHCs. A packet was mailed to parents to complete and bring with them to the evaluation appointment. Packets were available in English or Spanish. The Spanish-Western Hemisphere version of the M-CHAT-R was used because those families with primary language Spanish were primarily from Puerto Rico or Caribbean countries. The packet consisted of the written consent form, the parent-report screening tools, and a demographic form. All parents were asked to complete the M-CHAT-R, the SCQ, and the DCI.

When families arrived for the evaluation, a bilingual research assistant reviewed the consent with the parent. Informed consent was obtained for all subjects in the study. The research assistant then conducted a structured parent interview to gather child and family health information and gave the completed demographic and health information to the evaluator for review. The evaluators administered the Mullen Scales of Early Learning (MSEL; Mullen, 1995) and ADOS-2 (Lord et al., 2012), scored the measures, and provided the parent with clinical diagnostic impressions and recommendations. Evaluators were developmental pediatricians, advanced practice nurses, or licensed psychologists and all were research reliable on the ADOS-2 and experienced in diagnosis of ASD and related disorders. The ADOS-2 classification helped inform the assigned ASD status, but the final ASD diagnosis was based on clinical judgment of the experienced evaluator

who administered the ADOS-2 and directly observed the child. Clinical diagnosis of ASD was closely aligned to ADOS-2 classification: 99% of those diagnosed with ASD were classified as ASD on the ADOS-2 and 81% of those diagnosed as non-ASD were classified as non-ASD on the ADOS-2. Families received a gift card for their participation. In this group of children, final classification (ASD vs non-ASD) was determined by expert clinicians conducting an in-person evaluation assisted by the instruments described above.

Families recruited from daycare centers received the same packet described above; teachers sent the packet home with each child. Clinical evaluations were not provided for these low-risk children. In this sample, children were determined not to have ASD if they screened negative on both the M-CHAT-R and SCQ. Children who screened positive on either measure were excluded from analyses. Those who did not complete both screening measures were also excluded from analyses. Those who screened positive on any measure and who provided contact information were referred for evaluation and given resources.

One strategy that has been suggested to overcome the problem of absence of a gold standard (i.e. clinical diagnosis) in diagnostic research is to combine the outcomes of multiple tests for disease verification. Results from multiple tests can be combined through a fixed rule as in a composite reference standard (i.e. diagnostic decision rule (Alonzo and Pepe, 1999; Hadgu et al., 2012)). We think the strategy of using negative screens on multiple screening tools to determine the absence of diagnosis comprises a more rigorous strategy than that used in validation studies of other measures, which considered subjects to be non-ASD or "typically developing" based on lack of a diagnosis in a health chart (Perera et al., 2017) or parent report of lack of a diagnosis (e.g. Smith et al., 2013; Volker et al., 2010).

Data analysis

Our analytic goal was to assess the performance of the DCI in discriminating between children with ASD and those without ASD in a traditionally underserved population. First, we compared total DCI score and sub-scores for speech, play, social engagement, and behavior between groups using the Mann–Whitney U test. Non-parametric tests were used given the skewed distribution of the scores. Next, we performed an item analysis of the DCI to examine discriminant ability of each item using logistic regression. The odds of scoring at risk on each item were estimated for children with ASD compared with children without ASD.

We then examined diagnostic accuracy of the DCI in the overall sample of children and in demographic subgroups. A receiver-operating characteristic (ROC) area-under-the-curve (AUC) analysis was performed to assess the

discriminant ability of the DCI in distinguishing children with ASD from those without ASD. ROC analysis does not assume a priori cut-off point, but describes the tool's performance (sensitivity and specificity) for all possible cut-off values. The AUC value lies between 0.5 and 1; higher AUC suggests better discriminant ability. An accepted rule of thumb is that an AUC of 0.5 indicates no discrimination, an AUC of 0.7 indicates an acceptable discrimination, and an AUC of 0.9 indicates an excellent discrimination (Hosmer and Lemeshow, 2000).

In evaluating the diagnostic accuracy of a screener, a covariate can be a confounder if it is associated with both condition status (ASD, non-ASD) and screener results (DCI score) (Alonzo and Pepe, 2002; Janes et al., 2009; Janes and Pepe, 2008, 2009). For example, if the distribution of age differs between children with and without ASD and age is associated with the DCI score, then age could confound the association between DCI score and ASD. To account for potential confounding, covariate adjustment by regression model in ROC analysis was introduced (Hajian-Tilaki, 2013; Janes and Pepe, 2008; Pepe, 1998; Pepe et al., 2009; Tosteson and Begg, 1988). While adjustment for confounding is widely used in etiologic studies in epidemiology, it is rarely used to control for confounding in ROC analysis in medical literature of diagnostic studies (Hajian-Tilaki, 2013). If a variable confounds the association between DCI score and ASD status, different thresholds for different levels of that confounder should be established (Pepe, 2003).

In this study, we followed the guidelines (Janes et al., 2009; Janes and Pepe, 2008, 2009) on the ROC-regression methods to account for confounders and effect modifiers in the analysis of the diagnostic accuracy of the DCI. First, we explored associations between all measured covariates and the diagnostic status and DCI total score. If any significant associations were observed, the estimated ROC curve was adjusted for these covariates. The AUC for both the unadjusted and adjusted ROC curve was reported.

To determine if covariate-specific ROC analysis was needed, we examined the effect of the measured covariates on the estimated ROC curve. Covariates that affect the ROC curve are analogous to effect modifiers in the association setting (Janes and Pepe, 2008). The discrimination between ASD and non-ASD children varies with covariate value. A significantly positive effect of a covariate on the ROC curve suggests better discriminative power of the DCI, and vice versa. Such analyses enabled us to examine whether the performance of the DCI varied across demographic groups. For example, if the estimated coefficient for the indicator of female sex is significantly positive, it suggests that the DCI performs differently in girls than in boys. In such a case, separate ROC curves should be estimated for girls and boys (Janes et al., 2009; Janes and Pepe, 2008, 2009).

Finally, we estimated optimal cut-off points for the DCI using the Youden index, the Liu index and using the nearest point on the ROC curve closest to the (0, 1) point that corresponded with 100% sensitivity and 100% specificity (Fluss et al., 2005; Liu, 2012; Perkins and Schisterman, 2006). Since all three methods yielded almost identical results, we only reported cut-offs obtained via Youden index. For covariates identified as potential confounders, covariate-specific cut-off points were estimated (Pepe, 2003). We report specificity and sensitivity of the DCI at the optimal cut-off points.

Results

Of the 288 children evaluated in the high-risk group, 214 were diagnosed with ASD. Of the 88 children in the daycare group, 74 were considered not to have ASD based on passing both M-CHAT-R and SCQ. The non-ASD sample therefore consisted of 148 children. Demographic characteristics for the children with and without ASD are shown in Table 1.

The mean age of the children in the sample was 38 months. The sample was primarily low socio-economic status. While we did not collect income data, 92% were Medicaid enrolled, suggesting an income close to or below the poverty level. Most children were male and Hispanic; 42% came from families in which English was not the primary language spoken at home.

Compared with children without ASD, children with ASD were younger, more likely to be male and white, and less likely to be black or uninsured. The groups differed in age, gender, and race. Age and gender were significantly associated with DCI score, so we accounted for the confounding effect of age and gender in the evaluation of the DCI performance.

Table 2 shows DCI total and sub-scores for children with and without ASD. On average, children with ASD had a significantly higher total DCI score than children without ASD. Children with ASD scored positive on an average of 11 DCI items; those without ASD scored positive on an average of 5 DCI items. Speech, play, social engagement, and behavior scores for ASD children were all significantly higher than those for non-ASD children.

Table 3 presents the DCI item analysis. All items were statistically significantly different between children with ASD and without ASD, except the items *Stacks blocks* and *Lines up toys*. Items with the highest odds ratio for ASD were *Smiles back, Points, Shows toys, Comforts others, Shares interests and Plays with children*.

Discriminative performance of the DCI

The AUC for unadjusted ROC curve was 0.78 (95% CI: (0.73, 0.83)), and 0.75 (95% CI: (0.70, 0.81)) after adjusting for age and gender. Figure 2 shows ROC curve for the DCI adjusted for age and gender.

The analysis of the effects of specific covariates listed in Table 1 on the discriminative performance of the DCI

Table 1. Sample characteristics by ASD status.

	ASD (N = 214)	Non-ASD (N = 148)	p value
Mean age in months	36.5 (SD 8.6)	38.7 (SD 9.4)	0.025
Male	77%	57%	<0.001
Black	19%	28%	0.040
White	22%	12%	0.016
Hispanic	51%	53%	0.648
Other race	8%	6%	0.501
English not primary language at home	39%	46%	0.167
Parent with high school or less education	84%	85%	0.834
Medicaid coverage	94%	89%	0.050
Other health insurance coverage	6%	8%	0.313
No health insurance coverage	0%	3%	0.014

ASD: autism spectrum disorder; SD: standard deviation.
Sample means, proportions, and corresponding standard deviations, and p values for the test of difference in variables between ASD and non-ASD groups.

Table 2. Comparison of DCI scores by ASD status.

	ASD (N = 214)	Non-ASD (N = 148)	p value[a]
Tot DCI score (total 28 items) (std. dev.)	10.8 (5.7)	5.3 (4.6)	<0.001
Tot DCI-speech score (total 4 items)	1.7 (1.4)	0.6 (1.1)	<0.001
Tot DCI-play score (total 7 items)	2.5 (2.0)	1.0 (1.6)	<0.001
Tot DCI-social engagement score (total 8 items)	2.1 (2.1)	0.6 (1.2)	<0.001
Tot DCI-behavior score (total 9 items)	4.5 (2.2)	2.9 (2.1)	<0.001

ASD: autism spectrum disorder; DCI: Developmental Check-In.
[a]For Mann–Whitney U test statistic for comparison of instrument scores among ASD and non-ASD children.

indicated no statistically significant effect for any covariate, meaning there was no evidence of varying accuracy of the DCI based on the specific covariate groups in this sample (data not shown).

Cut-off points for the DCI

Table 4 shows optimal cut-off points for the DCI in the overall sample as well as by age (for three age groups) and gender based on the ROC analysis. The Youden index yielded an empirical cut-off score of 7.5 for the DCI (sensitivity = 0.66, specificity = 0.76) for the overall sample. The optimal cut-off score declined with age and was higher for males than for females.

Discussion

Among this sample of predominantly low-socio-economic status, minority children, for many of whom English was not their primary language, the DCI demonstrated good ability to discriminate between children with and without ASD (AUC = 0.75). This is encouraging as the tool was designed specifically to enhance identification of young children at risk for ASD in underserved populations. The DCI performed well across all ages from 24 to 60 months,

including children ages 36–47 months, for whom neither the M-CHAT-R nor SCQ has been validated. Although previous studies have suggested that both the M-CHAT and SCQ could be used in this age group, the sensitivity and specificity of both instruments may be compromised when used with children in this age bracket (Allen et al., 2007; Corsello et al., 2013; Snow and Lecavalier, 2008).

All DCI items except two significantly predicted the presence of ASD. While this is expected, given that items were chosen based on their predictive value on the ADOS-2 or other ASD screeners such as the STAT, our findings confirm that these items are predictive in pictorial form and with this underserved sample. *Lines up toys* was one of the items that did not discriminate ASD in this sample of underserved children. Our initial cognitive interviewing with parents suggested that this behavior can be depicted visually. Thus, the poor discriminative performance for this item may be related to different patterns of repetitive behavior for young underserved children with ASD. It also is possible that children without ASD also line up toys and this item may not discriminate meaningfully, particularly on parent-report measures.

Cut-off points for the DCI were determined empirically to maximize sensitivity and specificity. Ultimate cut-offs should be informed by clinical aim. For example, if it is

Table 3. Percentage of children in each group who scored positive on each DCI item.

		ASD (N = 214), %	Non-ASD (N = 148), %	Odds ratio[a]	p value
Items about speech	Asks for help	42	14	4.4	<0.001
	Says 1 word	18	9	2.3	0.014
	Says 2 words)	46	17	4.2	<0.001
	Says 3 words	61	26	4.5	<0.001
Items about play	Plays peek-a-boo	16	8	2.0	0.046
	Stacks blocks	11	7	1.6	0.234
	Feeds toy	56	23	4.4	<0.001
	Imitates	22	7	3.9	<0.001
	Plays pretend	41	19	2.9	<0.001
	Plays make believe	40	16	3.6	<0.001
	Plays pretend with children	61	24	4.9	<0.001
Items about social engagement	Smiles back	8	1	6.7	0.012
	Makes eye contact	19	8	2.5	0.008
	Points	18	1	31.7	0.001
	Shows toys	14	3	5.9	0.001
	Comforts others	45	12	5.9	<0.001
	Shares interest	29	7	5.6	<0.001
	Shares toy	49	24	3.0	<0.001
	Plays with children	30	8	5.0	<0.001
Items about behavior	Lines up toys	71	73	0.9	0.612
	Walks on toes often	44	28	2.0	0.002
	Bothered by noise	47	33	1.8	0.009
	Tantrums often	70	55	1.9	0.005
	Avoids other children	45	22	3.0	<0.001
	Avoids eye contact	42	16	3.9	<0.001
	Spins wheels over and over	46	30	2.0	0.002
	Flaps hands a lot	44	20	3.2	<0.001
	Responds to name	45	22	2.9	<0.001

ASD: autism spectrum disorder; DCI: Developmental Check-In.
[a]Logistic regression results for the odds of scoring positive on each item for children with ASD compared to children without ASD; and the p value for the significance of the estimated odds.
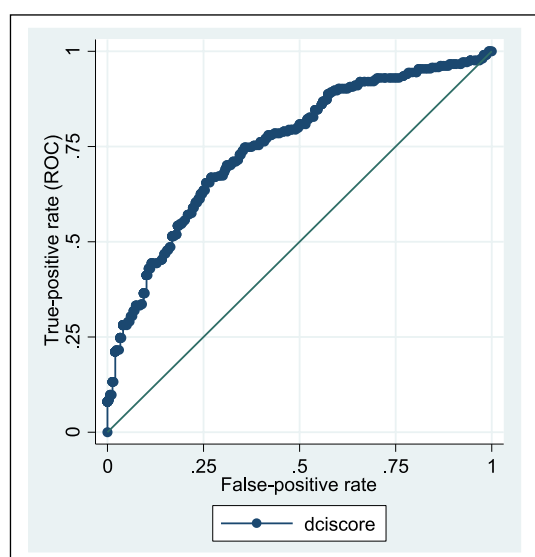
Figure 2. ROC curve for the DCI adjusted for age and gender.

important that no children at risk for ASD be missed, then specificity may be less important than sensitivity and a lower cut-off should be used. Different cut-offs also may be important to use for different demographics. For example, a lower DCI cut-off score was found to have better discriminant ability for older children and females in this sample.

The DCI adds to the limited examples of visually enhanced developmental or ASD screeners. Unlike those used only in non-US samples (Lejarraga et al., 2013; Perera et al., 2017) or that focus exclusively on one skill area (Ghilain et al., 2017), the DCI covers more developmental domains relevant to ASD and has demonstrated utility with underserved US populations. Our future research will cross-validate the DCI in different settings and with different samples.

Common to epidemiological studies, but often not done in ASD screener validation, we adjusted for potential confounders and explored effect modifiers, that is, whether specific covariates had an effect on the

Table 4. ROC analysis–optimal cut-offs for the DCI.

	N	AUC (std. err.)	95% CI	Empirical optimal cut-off	Se	Spe
All samples	362	0.75 (0.029)	(0.691, 0.806)	7.5	66%	76%
Age of 24–35 months	164	0.77 (0.038)	(0.688, 0.841)	7.5	78%	66%
Age of 36–47 months	142	0.77 (0.039)	(0.689, 0.847)	6.5	71%	69%
Age of 48–60 months	56	0.83 (0.051)	(0.733, 0.935)	2.5	92%	61%
Male	113	0.73 (0.034)	(0.659, 0.793)	7.5	67%	69%
Female	249	0.84 (0.037)	(0.755, 0.911)	5.5	73%	78%

ROC: receiver-operating characteristic; DCI: Developmental Check-In; AUC: area-under-the-curve.

discriminant ability of the DCI. This enabled us to look at the accuracy of the DCI by demographic characteristics. While significant differences by demographic groups were not found, our sample was purposely homogenous, as we focused on children who were underserved. It is not known whether the DCI performs best for underserved groups, or if it may perform equally well across all populations. Future work with the DCI is planned in demographically heterogeneous at risk and low-risk groups in order to address these questions.

There are some study limitations. It is possible that some parents who participated in the pre-test interviews were not form underserved communities because although race, ethnicity and language were assessed, insurance status was not. The study did not monitor the rate of parent refusal of the research evaluation or the no-show rate. Families who agreed to participate in this study may not fully represent underserved young children at risk for autism. Formal assessment of parent literacy was considered, but ultimately not conducted due to the burden of an additional measure, as well as the concern that measuring parent literacy would promote mistrust toward the research team. Full clinical evaluation was not available for children in daycare centers. We believe using performance on two validated ASD screeners to determine ASD status is a more rigorous estimate of ASD status than relying on parent report that the child is typically developing or non-ASD (National Health Interview Survey, 2017; study two in Smith et al., 2013; Volker et al., 2010) or based on absence of ASD listed in health records (Perera et al., 2017). It should also be noted that the M-CHAT-R and SCQ have previously been used outside their originally intended age ranges as they were in this study for the community children, but accuracy of results may be compromised when this is done (Allen et al., 2007; Corsello et al., 2007, 2013; Moody et al., 2017; Snow and Lecavalier, 2008). We used a SCQ cut-off score of 15 for all ages in the daycare sample. It is possible that a lower cut-off for daycare children under 4 years old would have identified some children as at risk who were otherwise considered non-ASD.

Since a major value of using ASD screeners is to identify young children in order to implement effective early intervention, it would have been ideal to include children younger than 24 months of age. However, it was uncommon for children who met inclusion for the study and who were younger than 24 months to be referred to either the hospital developmental evaluation clinic or to be identified as at developmental risk within the FQHCs, suggesting that work is still needed to decrease the age of identification and referral of very young, underserved children.

Despite these limitations, there are important implications related to our findings. Reducing racial, ethnic, income and language disparities in early diagnosis of ASD is key to improving access to services and functional outcomes. The DCI may improve access to diagnosis by reducing the literacy and linguistics demand of a parent-report ASD screening tool. This is likely to be especially effective for families of young children from underserved groups who experience disparities in age of diagnosis of ASD (e.g. Durkin et al., 2017). While future work is needed to cross-validate the DCI in different settings, with different samples, and determine utility of the DCI in comparison to other validated ASD screening tools and in a general population of young children, these results are promising for this new measure.

Ethical approval

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

ORCID iD

Jill F Harris https://orcid.org/0000-0001-6757-3426

References

Allen CW, Silove N, Williams K, et al. (2007) Validity of the Social Communication Questionnaire in assessing risk of autism in preschool children with developmental problems. *Journal of Autism and Developmental Disorders* 37: 1272–1278.

Alonzo TA and Pepe MS (1999) Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Statistics in Medicine* 18(22): 2987–3003.

Alonzo TA and Pepe MS (2002) Distribution-free ROC analysis using binary regression techniques. *Biostatistics* 3: 421–432.

Centers for Medicare & Medicaid Services (2017) Federally Qualified Health Center. Available at https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/downloads/fqhcfactsheet.pdf (accessed 11 January 2016).

Christensen DL, Baio J, Braun KV, et al. (2016) Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *Morbidity and Mortality Weekly Report: Surveillance Summaries* 65(3): 1–23.

Colbert AM, Webber J and Graham R (2017) Factors that influence autism knowledge in Hispanic cultures: a pilot study. *Journal of Racial and Ethnic Disparities* 4: 156–164.

Corsello C, Akshoomoff N and Stahmer A (2013) Diagnosis of autism spectrum disorders in 2-year-olds: a study of community practice. *Journal of Child Psychology and Psychiatry* 54(2): 178–185.

Corsello C, Hus V, Pickles A, et al. (2007) Between a ROC and a hard place: decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry* 48: 932–940.

Dawson G, Rogers S, Munson J, et al. (2010) Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics* 125(1): e17–e23.

Durkin MS, Maenner MJ, Baio J, et al. (2017) Autism spectrum disorder among US children (2002–2010): socioeconomic, racial, and ethnic disparities. *American Journal of Public Health* 107: 1818–1826.

Durkin MS, Maenner MJ, Meaney FJ, et al. (2010) Socioeconomic inequality in the prevalence of autism spectrum disorder: evidence from a US cross-sectional study. *PLoS ONE* 5(7): e11551.

Fluss R, Faraggi D and Reiser B (2005) Estimation of the Youden Index and its associated cutoff point. *Biometrical Journal* 47: 458–472.

Fountain C, King MD and Bearman PS (2010) Age of diagnosis for autism: individual and community factors across 10 birth cohorts. *Journal of Epidemiology and Community Health* 65: 503–510.

Ghilain CS, Parladel MV, McBee MT, et al. (2017) Validation of the Pictorial Infant Communication Scale for preschool-aged children with autism spectrum disorder. *Autism* 21: 203–216.

Hadgu A, Dendukuri N and Wang L (2012) Evaluation of screening tests for detecting Chlamydia trachomatis: bias associated with the patient-infected-status algorithm. *Epidemiology* 23(1): 72–82.

Hajian-Tilaki K (2013) Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine* 4(2): 627–635.

Herlihy LE, Brooks B, Dumont-Mathieu T, et al. (2014) Standardized screening facilitates timely diagnosis of autism spectrum disorders in a diverse sample of low-risk toddlers. *Journal of Developmental and Behavioral Pediatrics* 35: 85–92.

Hosmer D and Lemeshow S (2000) *Applied Logistic Regression*. New York: John Wiley & Sons.

Janes H and Pepe MS (2008) Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. *American Journal of Epidemiology* 168: 89–97.

Janes H and Pepe MS (2009) Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. *Biometrika* 96: 371–382.

Janes H, Longton M and Pepe MS (2009) Accommodating covariates in receiver operating characteristic analysis. *Stata Journal* 9: 17–39.

Khowaja MK, Hazzard AP and Robins DL (2015) Sociodemographic barriers to early detection of autism: screening and evaluation using the M-CHAT, M-CHAT-R, and Follow-Up. *Journal of Autism and Developmental Disorders* 45(6): 1797–1808.

Kimple KS, Bartelt EA, Wysocki KL, et al. (2014) Performance of the Modified Checklist for Autism in Toddlers in Spanish-speaking patients. *Clinical Pediatrics* 53(7): 632–638.

Lejarraga H, Kelmansky D, Lejarraga C, et al. (2013) Validation of a questionnaire for the detection of children at high risk of developmental disorders: the PRUNAPE Pre-Screening Test. *Archivos Argentinos de Pediatria* 111: 476–483.

Liptak GS, Benzoni LB, Mruzek DW, et al. (2008) Disparities in diagnosis and access to health services for children with autism: data from the National Survey of Children's Health. *Journal of Developmental and Behavioral Pediatrics* 29: 152–160.

Liu X (2012) Classification accuracy and cut point selection. *Statistics in Medicine* 31(23): 2676–2686.

Lord C, Rutter M, DiLavore PC, et al. (2012) *Autism Diagnostic Observation Schedule (ADOS-2)*. 2nd ed. Torrance, CA: Western Psychological Services.

Mandell DS, Listerud J, Levy S, et al. (2002) Race differences in the age at diagnosis among Medicaid-eligible children with autism. *Journal of the American Academy of Child and Adolescent Psychiatry* 41(12): 1447–1453.

Mandell DS, Wiggins LD, Carpenter LA, et al. (2009) Racial/ethnic disparities in the identification of children with autism spectrum disorders. *American Journal of Public Health* 99(3): 493–498.

Moody EJ, Reyes N, Ledbetter C, et al. (2017) Screening for autism with the SRS and SCQ: variations across demographic, developmental and behavioral factors in preschool children. *Journal of Autism and Developmental Disorders* 47: 3550–3561.

Mullen EM (1995) *Mullen Scales of Early Learning: AGS Edition*. Circle Pines, MN: American Guidance Service.

National Health Interview Survey (2017) Data collection description. Available at: cdc.gov/nchs/nhis/about_nhis.htm#sample_design

Ozonoff S, Young GS, Landa RJ, et al. (2015) Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study. *Journal of Child Psychology and Psychiatry* 56(9): 988–998.

Pepe MS (1998) Three approaches to regression analysis of receiver operating characteristic curves for continuous test results. *Biometrics* 54: 124–135.

Pepe MS (2003) *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford University Press.

Pepe MS, Longton GM and Janes H (2009) Estimation and comparison of receiver operating characteristic curves. *Stata Journal* 9: 1–16.

Perera H, Jeewandara KC, Seneviratne S, et al. (2017) Culturally adapted pictorial screening tool for autism spectrum disorder: a new approach. *World Journal of Clinical Pediatrics* 6: 45–51.

Perkins NJ and Schisterman EF (2006) The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *American Journal of Epidemiology* 163: 670–675.

Robins DL, Casagrande K, Barton M, et al. (2014) Validation of the Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M-CHAT-R/F). *Pediatrics* 133: 37–45.

Robins DL, Fein D and Barton ML (1999) *The Modified Checklist for Autism in Toddlers (M-CHAT)*. Self-Published.

Robins DL, Fein D, Barton ML, et al. (2001) The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders* 31: 131–144.

Rutter M, Bailey A and Lord C (2003) *Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services.

Rutter M, LeCouteur A and Lord C (2008) *Autism Diagnostic Interview-Revised (ADI-R)*. Los Angeles, CA: Western Psychological Services.

Scarpa A, Reyes NM, Patriquin MA, et al. (2013) The Modified Checklist for Autism in Toddlers: reliability in a diverse rural American sample. *Journal of Autism and Developmental Disorders* 43: 2269–2279.

Smith NJ, Sheldrick C and Perrin EC (2013) An abbreviated screening instrument for autism spectrum disorders. *Infant Mental Health Journal* 34: 149–155.

Snow AV and Lecavalier L (2008) Sensitivity and specificity of the Modified Checklist for Autism in Toddlers and the Social Communication Questionnaire in preschoolers suspected of having pervasive developmental disorders. *Autism* 12(6): 627–644.

Stone WL and Ousley OY (2008) *Screening Tool for Autism in Toddlers and Young Children: STAT User's Manual*. Nashville, TN: Vanderbilt University.

Tek S and Landa RJ (2012) Differences in autism symptoms between minority and non-minority toddlers. *Journal of Autism and Developmental Disorders* 42(9): 1967–1973.

Tosteson AN and Begg CB (1988) A general regression methodology for ROC curve estimation. *Medical Decision Making* 8: 204–215.

Volker MA, Lopata C, Smerbeck AM, et al. (2010) BASC-2 PRS profiles for students with high-functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders* 40: 188–199.

Windham GC, Smith KS, Rosen N, et al. (2014) Autism and developmental screening in a public, primary care setting primarily serving Hispanics: challenges and results. *Journal of Autism and Developmental Disorders* 44(7): 1621–1632.