

Combining information from multiple sources for the diagnosis of autism spectrum disorders for toddlers and young preschoolers from 12 to 47 months of age

So Hyun Kim and Catherine Lord

University of Michigan Autism and Communication Disorders Center (UMACC), Ann Arbor, MI, USA

Background: Purpose of this study was to systematically examine combined use of the Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule (ADOS) for children under age 4 using newly developed and revised diagnostic algorithms. **Methods:** Single and combined use of the ADI-R and ADOS algorithms were compared to clinical best estimate diagnoses for 435 children with autism spectrum disorders (ASD), 113 children with nonspectrum disorders, and 47 children with typical development from 12 to 47 months of age. Sequential strategies to reach a diagnostic decision by prioritizing administrations of instruments were also evaluated. **Results:** Well-balanced sensitivities and specificities above 80% were obtained for ASD diagnoses using both instruments. Specificities significantly improved when both instruments were used compared to one. Scores that can be used to systematically prioritize administrations of instruments were identified. **Conclusions:** The ADI-R and ADOS make independent, additive contributions to more accurate diagnostic decisions for clinicians evaluating toddlers and young preschoolers with ASD. Sequential assessment strategies using the scores identified may be appropriate for some children. **Keywords:** Autism spectrum disorders, autism diagnostic interview-revised, autism diagnostic observation schedule, early diagnosis.

Introduction

In the past few years, research has flourished concerning detection of autism spectrum disorders (ASD) symptoms in the first 3 years of life due to the belief that earlier provision of services and treatments is associated with better outcomes. With increasing needs for early detection of ASD, new diagnostic algorithms for toddlers and young preschoolers from 12 to 47 months of age have recently been developed for the Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003b), a standardized, semi-structured, investigator-based interview for caregivers (Kim & Lord, 2011). In addition, algorithms for the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999) have been revised to improve the diagnostic validity of the instrument, a standardized, semi-structured, clinician-administered observation of communication, social interaction, and play (Gotham, Risi, Pickles, & Lord, 2007), including two revised algorithms for children who use five or more single words or even less language and for children under age 5 using phrase speech. New algorithms have also been developed for the ADOS-Toddler module for children under 30 months of age (Luyster et al., 2009). Past studies showed enhanced diagnostic validity when information from both the ADI-R

and ADOS are used together. There has not yet been a systematic attempt to examine the combined use of these instruments using the newly developed and revised algorithms. Thus, the present study focuses on the validity of the combined use of the ADI-R and ADOS using the new and revised algorithms with children under age 4.

With a best estimate clinical (BEC) diagnosis treated as the gold standard, previous work primarily with older children and adolescents found that using data from multiple sources (i.e. clinicians, caregivers, and teachers) enhances accuracy for the diagnosis of ASD. For example, the Social Responsiveness Scale resulted in high diagnostic specificity for children and adolescents with ASD when information from both parent and teacher reports were combined (Constantino et al., 2007). Bishop and Baird (2001) reported improved validity of the Children's Communication Checklist when information from both parents and professionals were used for 151 children with pervasive developmental disorders (PDD) or other developmental disorders between 5 and 17 years of age. Corsello et al. (2007) reported enhanced diagnostic validity by combining information across instruments, either the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003a) or the ADI-R with the ADOS for the diagnosis of children with ASD between age 2 and 16 years.

Risi et al. (2006) found a better balance of sensitivity and specificity when the ADI-R and ADOS were used in combination compared to when each

Conflict of interest statement: C.L. receives royalties for the ADI-R; profits from this study were donated to charity.

instrument was used alone. For example, the combined use of these instruments resulted in sensitivity and specificity of 82% and 86% for children with autism compared to children with nonspectrum disorders over age 3 years. For younger children, sensitivity and specificity for the same diagnostic comparison using both instruments were 81% and 87%, respectively. In contrast, when each instrument was used alone, specificities ranged from 59% to 72%, with sensitivities remaining above 80%. Most of the results were obtained primarily from older children, even though a study by Le Couteur, Haden, Hammal, and McConachie (2008) also found that combining information from both instruments using preexisting algorithms provided improved diagnostic accuracy for preschoolers with ASD compared to either instrument in isolation. However, using newly developed and revised algorithms, the present study is the first study examining the validity of the combined use of the ADI-R and ADOS for toddlers as young as 12 months of age.

In very young children, diagnostic differentiation between nonautism ASD (e.g. PDD-NOS) and autism is less stable than for older children and adolescents (Lord et al., 1999; Szatmari et al., 2002; Wiggins, Robins, Adamson, Bakeman, & Henrich, 2011). Consequently, as the newest ADI-R and ADOS-T algorithms for toddlers and young preschoolers have been developed (Luyster et al., 2009; Kim & Lord, 2011), a shift has occurred from having separate autism and other ASD cutoffs to using only a single classification of ASD. In addition, in order to formally acknowledge the less clear stability of diagnoses in younger children, these algorithms provide *ranges of concern* (little-to-no, mild-to-moderate, or moderate-to-severe concern), to be used in clinical monitoring and follow-up. However, groupings are necessary for several different purposes. Thus, the new ADI-R algorithms also provide two cutoffs, one for research (more restrictive; higher specificity with lower sensitivity) and one for clinical purposes (more inclusive; higher sensitivity with lower specificity).

Past studies examining validity of the ADI-R and ADOS have found that parent reports and clinician observations do not always agree. Agreement between these instruments has varied across samples and analytic techniques. In a sample of 797 ASD and 163 nonspectrum cases over 36 months of age, Risi et al. (2006) found that the Pearson *r* correlation between ADI-R and ADOS algorithm totals was 0.57. Correlations differed by domains in the study by Le Couteur et al. (2008), ranging from 0.51 to 0.71 for a sample of 77 preschoolers with ASD and 24 with other developmental disorders. Agreement between the instruments using Kappa ranged from 0.48 to 0.62. In another study (de Bildt et al., 2004), correlations ranged from 0.52 to 0.54 between the ADI-R and ADOS algorithm totals for 123 children aged 5–20 years with ASD and intellectual disability and 62 with intellectual disability *only*. In contrast,

Ventola et al. (2006) compared the application of full ADI-R and ADOS diagnostic criteria to each other and clinical diagnosis in a sample of 36 ASD and 9 nonspectrum cases aged 16–31 months. Significant levels of agreement were found between the ADOS and clinical judgment ($\kappa = 0.59$, $p < .001$) but agreement between the ADI-R and clinical judgment ($\kappa = 0.15$, ns) and between the ADI-R and the ADOS ($\kappa = 0.07$, ns) was poor.

Because the combined use of the ADI-R and ADOS has shown better diagnostic validity than either individual instrument, it is recommended that clinicians and researchers use information from both instruments when making diagnoses. However, due to constraints in time, cost, or expertise, often only one of the instruments is actually used. Relatively little is known about ways to maximize validity in this case. One approach would be to determine scores on the instruments associated with a very high (or low) probability of receiving the classification of ASD on the ‘alternative instrument’ (referred to as ‘*positive (or negative) screening estimate*’ hereafter). For instance, if a child’s score reaches a *positive screening estimate* on the ADI-R, a clinician could presumably omit the ADOS assuming that the probability of the child receiving an ASD classification on the ADOS would be very high. The same strategy could be used with a *negative screening estimate*.

Another approach is to conduct similar analyses using BEC diagnoses based on all available information as the gold standard and then to determine if there are scores on each instrument that result in 100% specificity for ASD. That is, we can examine what score on each instrument successfully excludes all cases determined to not have ASD (henceforth referred to as ‘*high specificity case scores*’) and then describe the sensitivities of these scores. For example, if a child meets or exceeds a *high specificity case score* on the ADOS, a clinician evaluating the child could assume that the chance of the child receiving a BEC diagnosis of ASD would be very high and choose to omit the ADI-R.

In sum, the purpose of this study is to examine the combined use of the ADI-R and ADOS for children under age 4 using the new and revised algorithms. Often, a misdiagnosis that results in a child failing to receive necessary services is the greatest concern. On the other hand, over-diagnosis has negative consequences for individual children, public health strategies and research. Consequently, we present data supporting alternative methods for using both research and clinical cutoffs from the new ADI-R algorithms. Agreement between the two instruments is also evaluated by examining the overlap between the ADI-R and ADOS-T *ranges of concern* and correlations between algorithm totals. A final goal was to evaluate sequential assessment strategies using positive/negative screening estimates and high specificity case scores that could allow the

empirically-validated use of a single instrument when information from a second instrument would very likely be redundant.

Methods

Participants

All 604 children with complete data from a contemporaneous ADOS, ADI-R, nonverbal IQ (NVIQ), and BEC diagnosis were included from two projects, early diagnosis of autism (EDX) and first words and toddlers (FW/T) and for clinic patients at the University of Michigan Autism and Communication Disorders Center (UMACC).

Children in the FW/T projects entered the study between 12 to 18 months and were administered the ADI-R and ADOS-T. The remaining children were administered the ADI-R and either the Pre-Linguistic ADOS (PL-ADOS; DiLavore, Lord, & Rutter, 1995), or ADOS Module 1–3 depending on their age and language level. Out of 604 children, 195 children, who were nonverbal or had single words only, received the PL-ADOS, which was recoded to the ADOS Module 1.

All participants, aged 12–47 months, were walking at the time of assessment. Mean age was 31.8 months ($SD = 9.6$), and 435 children had ASD (345 males), 113 children nonspectrum disorders (NS; 81 males), and 47 children typical development (TD; all younger than 21 months; 31 males). NS participants had a range of diagnoses, including language disorders (53%), intellectual disability of unknown etiology (18%), Down syndrome (6.4%), externalizing disorders (5.5%), internalizing disorders (2.7%), and general, mild developmental delays (14.4%). Ethnicity was not associated with diagnosis; 74% of participants were Caucasian, 15% African American, 3% Asian American, 3% biracial, and 5% Native American or other races. The sample in the present study was a subset of children (about 30%) from the sample used to develop the new ADI-R algorithms for toddlers and young preschoolers (Kim & Lord, 2011). In addition, approximately at least 30% and 15% of the sample was also used for the develop-

ment of revised ADOS algorithms and new ADOS-T algorithms, respectively (Gotham et al., 2007; Luyster et al., 2009).

Participants were divided into three developmental cells by the child's age and language level following the structure of the developmental groupings of the new ADI-R algorithms: (a) all children between 12 and 20 months, 31 days of age and nonverbal children between 21 and 47 months, 31 days of age ('12–20/NV21–47'), (b) children between 21 and 47 months, 31 days of age with single words ('SW21–47'), and (c) children between 21 and 47 months, 31 days of age with phrase speech ('PH21–47').

As shown in Table 1, children with TD and NS were significantly younger and had significantly higher NVIQ and Vineland Adaptive Behavior Composite scores (Sparrow, Balla, & Cicchetti, 1984) than children with ASD for the '12–20/NV21–47' group ($p < .001$). For both 'SW21–47' and 'PH21–47' groups, Vineland composite scores were significantly higher for children with NS than ASD ($p < .001$). A significant age difference emerged for the 'SW21–47' group (children with ASD were older than children with NS, $p < .05$). For the 12–20/NV21–47 group, 156 children received the PL-ADOS, 122 children module 1, and 60 children the ADOS-T. For the SW21–47 group, 105 children received module 1, 39 children the PL-ADOS, and 7 children the ADOS-T. For the PH21–47 group, 106 children received module 2 and 4 children module 3. All of these children received appropriate modules based on their level of language.

Measures

In the new ADI-R algorithms for toddlers and young preschoolers, item scores in *Social Affect (SA)* and *Restricted and Repetitive Behaviors (RRBs)* for the '12–20/NV21–47' and 'SW21–47' groups and *Social Communication (SC)*, *RRBs*, and *Reciprocal and Peer Interaction (RPI)* for the 'PH21–47' group are combined to generate cutoffs for the classification of ASD. Thirteen to 20 items comprise the new ADI-R algorithms depending on children's ages and language levels. For the revised

Table 1 Description of sample

Diagnosis	12–20			NV21–47		SW21–47		PH21–47	
	ASD	NS	TD	ASD	NS	ASD	NS	ASD	NS
Gender (male, female)	30, 6	13, 5	31, 16	175, 35	19, 8	86, 37	17, 11	56, 13	33, 8
Age in months, mean (standard deviation)	17.6 (2.2)	16.8 (2.6)	15.6 (2.6)	32.7 (7)	28.8 (5.8)	36.1 (6.6)	33 (6.1)	41.1 (5.1)	39.3 (5.4)
Mean NVIQ* (standard deviation)	91.3 (15.6)	97.4 (14.4)	112.6 (15.3)	65.8 (19.9)	74.7 (27.6)	69.9 (17.5)	69.2 (18)	91.4 (20.8)	97.8 (16.5)
Mean VABC (standard deviation)	77.1 (12.9)	83.3 (10.4)	96.5 (10.1)	59.7 (8.1)	65.9 (11.3)	65.3 (10.3)	72.6 (11.2)	76.3 (13)	83.9 (10.7)

NVIQ = nonverbal IQ, VABC = Vineland Adaptive Behavior Composite standard score; 12–20 all children 12–20 months, NV21–47 = nonverbal children 21–47 months, SW21–47 = children 21–47 months with single words, PH21–47 = children 21–47 months with phrase speech.

*For some children, NVIQ scores were not available, thus replaced by full scale IQ scores: 37 TD cases in '12–20' group; 1 ASD and 2 NS cases in 'PH21–47' group.

ADOS and new ADOS-T algorithms, the total number of items in the algorithms is 14, with the composition of items in each algorithm differing by children's ages and language levels.

Procedure

Each caregiver was administered the ADI-R and the Vineland. The ADOS and cognitive testing were then completed by the same or by a different clinical psychologist or a trainee within a few days' time. A standard hierarchy of cognitive measures, most frequently the Mullen Scales of Early Learning ($n = 438$; Mullen, 1995) or the Differential Ability Scales ($n = 61$; Elliott, 1990) was used to determine IQ scores. Examiners in the study had completed research training and met standard requirements for research reliability for the ADI-R and ADOS. Inter-rater reliability was monitored through periodic observations and scoring by two examiners and scoring of videotapes. Caregivers signed an Institutional Review Board approved informed consent to participate in research before participation.

Consensus BEC Diagnosis

For children in the EDX study, an experienced clinical researcher used the videotaped ADOS and ADI-R scores and observations made during the testing to generate an independent BEC diagnosis of autism, PDD-NOS, or nonspectrum disorders (American Psychiatric Association., 1994). For children in the FW/T project, scores on the ADI-R, ADOS, and clinical observations were used by two clinicians to make a BEC diagnosis operationalizing DSM-IV criteria (American Psychiatric Association., 1994; See Luyster et al., 2009). For clinic cases, a diagnosis was made by a psychologist and/or psychiatrist after review of all information.

Analyses

Sensitivities and specificities for single and combined use of the ADI-R and ADOS algorithms were compared with BEC diagnoses. Sensitivities and specificities (Siegel, Vukicevic, Elliott, & Kraemer, 1989) were considered in each of these conditions: (a) Meeting ADI-R criteria, (b) Meeting ADOS criteria, (c) Meeting *either* ADI-R or ADOS criteria when both were administered, and (d) Meeting criteria on *both* the ADI-R and ADOS. For the sensitivities and specificities, 95% confidence intervals were also calculated using the Wilson score method (Newcombe, 1998). Characteristics of children correctly or incorrectly classified were examined. Correlations were used to assess the agreement between the ADI-R and ADOS algorithm totals as well as between domain totals for three different developmental cells ('12–20/NV21–47', 'SW21–47', 'PH21–47'). Correlation coefficients were compared using Fisher's Z transformations (Steiger, 1980). Seventy children who received both new ADI-R and ADOS-T algorithms were selected to examine the overlap between the *ranges of concern* from both instruments. Odds ratios were calculated to assess the likelihood of receiving a diagnosis of ASD when a child was classified by the ADI-R and/or ADOS in these ranges.

Positive/negative screening estimates were identified for each instrument by selecting scores associated with very high/low percentages of cases that received a classification of ASD on the other instrument. Sensitivities and specificities for these scores were then evaluated. In addition, *high specificity case scores* were selected for each instrument by examining total scores that resulted in high specificities (100%, 90%, and 80%) of the BEC diagnoses for the comparison of ASD versus NS cases. Sensitivities for these scores were also examined.

Results

Sensitivities and specificities for the comparison of ASD versus NS

Not surprisingly, as shown in Table 2, the most satisfactory results were obtained when the most stringent condition, requiring a child to meet criteria on *both* the ADI-R (using clinical cutoffs) and the ADOS was used. In these cases, sensitivity and specificity for ASD versus NS were consistently above 80%. For example, using both instruments yielded comparable sensitivities and significant improvements in specificities (4%–22%) beyond when only ADI-R algorithms were used. Compared to when ADOS algorithms were used alone, using both instruments resulted in significant gains in specificities (10%–31%) and slightly lower sensitivities, though they were still above 80% when ADI-R clinical cutoffs were used. As noted in previous papers (Gotham et al., 2007; Risi et al., 2006), when children with nonverbal mental ages below 15 months were included, specificities were slightly lower. Because evaluating children with low nonverbal mental age is a reality in clinical practice, these specificities are reported in parentheses (See Table 2). However, we also present separate results from data without children whose nonverbal mental ages fell below 15 months for researchers who wish to restrict their samples for better diagnostic accuracy.

As expected, the least restrictive condition, requiring a child to meet *either* the ADI-R or ADOS criteria, resulted in excellent sensitivities for ASD cases (97%–99%), but poor specificities (45%–85%). As in past studies, for all developmental cells, sensitivities improved when children whose BEC diagnoses were PDD-NOS were excluded. Although comparisons between the ASD and TD cases are not very informative clinically, because much research with younger children contrasts ASD and mixed TD and NS samples (as in studies with baby siblings), it is useful to know that not surprisingly, specificities also improved when TD cases were included. Likelihood ratios for the comparison of ASD versus NS were most satisfactory when both instruments were used in combination using the conventional criteria (Likelihood ratio above 5 is considered satisfactory; Jaeskchke, Guyatt, & Lijmer, 2002).

Table 2 Validity of all conditions tested

	Sensitivity 95% CI		Specificity 95% CI	ASD vs. NS
	Autism	ASD	NS	Likelihood ratio positive
12-21/ NV21-47				
ADI-R (CLI) and ADOS	92% 87-95	84% 79-88	88% 72-95 (80% 66-89)	6
ADI-R (RES) and ADOS	85% 79-89	76% 70-81	94% 80-98 (89% 77-95)	19
ADI-R (CLI) alone	92% 87-95	85% 80-89	72% 55-84 (69% 54-80)	3
ADI-R (RES) alone	85% 79-89	78% 72-83	88% 72-95 (84% 71-92)	7
ADOS alone	99% 96-100	95% 92-97	75% 58-87 (64% 50-77)	4
ADI-R (CLI) or ADOS	100% 98-100	98% 96-99	53% 36-69 (49% 35-63)	2
ADI-R (RES) or ADOS	100% 98-100	98% 96-99	69% 51-82 (60% 45-73)	3
SW21-47				
ADI-R (CLI) and ADOS	98% 92-99	92% 86-96	92% 76-98	11
ADI-R (RES) and ADOS	79% 69-86	70% 61-77	96% 81-99	18
ADI-R (CLI) alone	99% 94-100	94% 89-97	85% 66-94	6
ADI-R (RES) alone	80% 70-87	72% 63-79	92% 76-98	9
ADOS alone	99% 94-100	98% 93-99	65% 46-81	3
ADI-R (CLI) or ADOS	100% 96-100	99% 96-100	62% 43-78	3
ADI-R (RES) or ADOS	99% 94-100	97% 92-98	85% 66-94	6
PH21-47				
ADI-R (CLI) and ADOS	90% 74-97	82% 71-89	80% 65-90	4
ADI-R (RES) and ADOS	83% 66-93	71% 59-81	88% 74-95	6
ADI-R (CLI) alone	90% 74-97	83% 73-90	58% 42-71	2
ADI-R (RES) alone	83% 66-93	73% 61-82	75% 60-85	3
ADOS alone	100% 89-100	97% 90-99	68% 52-80	3
ADI-R (CLI) or ADOS	100% 89-100	99% 92-99	45% 31-60	2
ADI-R (RES) or ADOS	100% 89-100	98% 92-99	55% 40-69	2

12-20/NV21-47 all children 12-20 months and nonverbal children 21-47 months, SW21-47 children 21-47 months with single words, PH21-47 children 21-47 months with phrase speech, ADI-R Autism Diagnostic Interview-Revised, ADOS Autism Diagnostic Observation Schedule, CLI Clinical Cutoff, RES Research Cutoff, CI Confidence Interval.

Numbers in parentheses are when children with nonverbal mental age lower than 15 were included.

Characteristics of misclassified children

We then compared the characteristics of true positives (TPs) and false negatives (FNs) for each instrument as well as false positives (FPs) and true negatives (TNs). The most common trend was that FPs (NS cases misclassified as ASD) were significantly older and had significantly lower NVIQ and Vineland scores than TNs (correctly classified NS cases). On the other hand, FNs (ASD cases misclassified as NS) were younger and showed higher NVIQ and Vineland scores than TPs (correctly classified ASD cases). See Table S1 in Electronic Appendix.

Overlap between the ADI-R and ADOS-T ranges of concern

Most children (71%) whose scores were in the little-to-no range of concern in the ADI-R fell in the same range in the ADOS-T. Similarly, 64% of children whose scores fell in the moderate-to-severe range in the ADI-R fell in the same range in the ADOS-T. If a child was classified as at risk (mild-to-moderate and moderate-to-severe ranges) by only one instrument (23%), the odds ratio for the child to be placed in a risk group by the other instrument was 12.69 ($\chi^2 = 19.2$, $p < .001$, See Figure 1). When children were placed in a risk group by *both* instruments (50%), the odds ratio of having a BEC ASD diagnosis was 56.19 ($\chi^2 = 19.2$, $p < .001$).

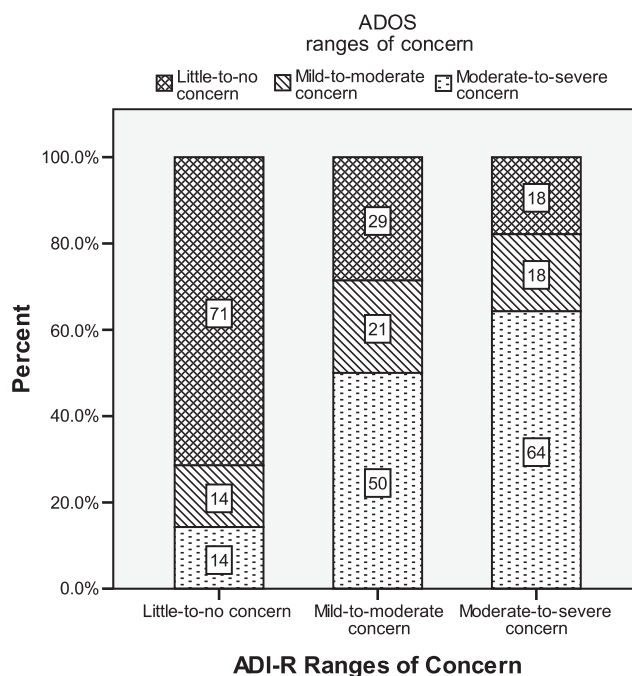


Figure 1 Overlap between the ADI-R and ADOS ranges of concern. ADI-R = Autism diagnostic interview-revised, ADOS = Autism diagnostic observation schedule

Agreement across the instruments

A correlation between the ADI-R and ADOS algorithm totals for the '12-20/NV21-47' group ($r = .75$)

was also significantly greater than those of the 'SW21–47' and 'PH21–47' groups ($r = .47$, $Z = 4.7$; $r = .59$, $Z = 2.7$, both $p < .01$). The correlation between the ADI-R and ADOS SA domains for the '12–20/NV21–47' group ($r = .69$) was significantly greater than that of the 'SW21–47' group ($r = .49$, $Z = 3.1$, $p < .01$). The correlation between the ADI-R and ADOS RRB domains for the '12–20/NV21–47' group ($r = .62$) was significant greater than those of the 'SW21–47' and 'PH21–47' groups ($r = .44$, $Z = 2.5$, $p < .05$; $r = .29$, $Z = 3.9$, $p < .01$, respectively).

Positive and negative screening estimates

Total scores on the ADI-R and ADOS algorithms which resulted in very *high* probabilities (100%) of receiving an ASD classification on the other instrument (*positive screening estimates*) for all ASD cases, ranged from 18 to 25 and 18 to 22 respectively. Total algorithm scores resulting in very *low* probabilities (less than 5%) of receiving an ASD classification on the other instrument (*negative screening estimates*) ranged from 4 to 5 and 8 to 11 respectively. See Table S2 in Electronic Appendix for sensitivities and specificities for these scores.

High specificity case scores

We then identified scores on the ADI-R and ADOS that resulted in high specificities for BEC diagnoses of ASD. The lowest scores on both instruments that resulted in 100% specificities were first identified for each developmental cell. For the ADOS, when scores were selected by 100% specificity, sensitivities ranged from 17% to 80% depending on developmental cells. Clinical cutoffs on the ADI-R are reported. As expected, when the ADI-R research cutoffs were used, a similar pattern emerged but with lower sensitivities and higher specificities. For high specificity case scores (100%) on the ADI-R, sensitivities ranged from 14% to 41%. Scores that resulted in specificities around 90% and 80% were also identified. An example of sequential assessment strategies using the Positive Screening Estimates (PSE), Negative Screening Estimates (NSE) and high specificity case scores are described in Figure 2 and the Discussion.

Discussion

Consistent with findings from older children (Risi et al., 2006), use of information from both the new ADI-R algorithms for toddlers and young preschoolers and the revised ADOS and new ADOS-T algorithms together better reflected clinical best estimate diagnoses of ASD than when either single instrument was used. The ADI-R includes a developmental history and a detailed description of individual's functioning in a variety of social contexts as well as caregivers' perceptions of the level of impairment

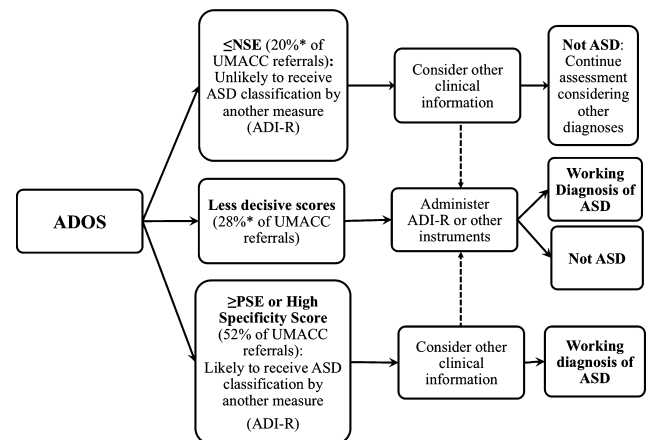


Figure 2 Sequential assessment strategies using positive/negative screening estimates (PSE/NSE) and high specificity case scores. ADI-R = autism diagnostic interview-revised, ADOS = autism diagnostic observation schedule. *In general developmental disorders clinics, autism cases would comprise a smaller proportion of likely diagnoses, thus the percent of cases with scores below or equal to NSE and/or possibly the less decisive range would increase

and/or frequency of different behaviors. The ADOS provides a summary of an experienced clinician's standardized observations of individual's behaviors within contexts that elicit social initiations and responses as well as communication interchanges. As suggested by low to moderate correlations between the ADI-R and ADOS in this study and in previous research (de Bildt et al., 2004; Ventola et al., 2006), the instruments provide overlapping but not identical information. Though the lack of high agreement between the instruments is frustrating in terms of each instrument's diagnostic validity, it increases their additive value. In fact, the combination of new and revised algorithms revealed even higher validity for toddlers and preschoolers than expected from studies using the original algorithms (Risi et al., 2006).

These newly developed and revised algorithms were created in a way that the influence of age and IQ scores on the algorithm scores was minimized. Nevertheless, we found differences in age, IQ, and adaptive functioning between children who were correctly identified and those misclassified by the instruments. For example, ASD cases misclassified as NS tended to be younger toddlers who had higher nonverbal intellectual and adaptive functioning. On the other hand, NS cases misclassified as ASD were older preschoolers with lower intellectual and adaptive functioning. These results are consistent with past studies showing that differentiating children with ASD from other developmental disorders is more difficult for very young children, children with severe delays (with lower IQ scores and/or who are nonverbal), and the most able toddlers and young preschoolers (with very high IQ scores and/or phrase speech; Gotham et al., 2007; Kim & Lord, 2011; Storoschuk, Rutter, & Pickles, 1993).

More able children, in this case, primarily older preschoolers, showed lower correlations between the ADI-R and ADOS than the younger and/or nonverbal group. In addition, mean ADI-R domain and algorithm total scores were lower for the 'PH21-47' group than the 'SW21-47' and '12-20/NV21-47' groups whereas mean ADOS scores were similar across all three groups. This may indicate that parents of preschoolers with more advanced levels of language, children who almost always also have stronger nonverbal skills, perceive their children's symptoms as less severe than clinicians evaluating the same children based on direct observations. This supports the usefulness of integrating perspectives from both caregivers and experienced clinicians especially when evaluating more complex cases.

Different sequential strategies could be used to determine when use of a single instrument might be sufficient. Each strategy has a distinct process in terms of obtaining a diagnostic classification. For example, as in Figure 2, if a clinician first administered the ADOS and the child's score on the ADOS was above the PSE (or the high specificity case score), unless other information suggested otherwise, the clinician could reasonably assume that the child would be likely to receive an ASD diagnosis without administering the ADI-R. Based on the clinic referrals from the dataset used in the present study, such an approach was appropriate for about 72% of the clinic referrals (with 52% very likely ASD and 20% likely not ASD). However, about 28% of the referrals obtained less decisive scores, showing that such an approach may not be appropriate for all children (See Table 3 and Table S2 in Electronic Appendix). It is also important to note that UMACC is an autism clinic; in general developmental disorders clinics, autism cases would comprise a smaller

proportion of likely diagnoses, so that the percent of cases with scores below or equal to NSE and possibly those in the less decisive range would increase. Studies of baby siblings of children with autism also suggest that the proportion of less decisive cases may be higher when children are not specifically referred for an autism assessment (Landa & Garrett-Mayer, 2006; Zwaigenbaum et al., 2005).

Although distributions of children by *ranges of concern* did not overlap perfectly between the ADI-R and ADOS-T, the majority of children classified as those needing follow-up evaluations and treatments by one instrument were also classified as at risk by the other instrument. In addition, the high likelihood ratio of receiving a BEC diagnosis of ASD for children classified into the risk groups by both instruments supports the validity of ASD risk categories even in very young children.

Limitations

Compared to the samples in past studies, the sample used in the present study was smaller because we selected only children with a contemporaneous ADI-R and ADOS for each case. Thus, restricted size and possible recruitment biases of more complex cases in the control groups (children with NS and TD) may have resulted in lower specificities in the present study compared to the original studies (Gotham et al., 2007; Kim & Lord, 2011; Luyster et al., 2009). In addition, because samples included a subset of children from previous studies mentioned above, replications from different sites will be critical.

The ADI-R and ADOS were administered by the same clinician for 75% of children; for 66% of these children, the ADI-R was administered before the

Table 3 High specificity (100%, 90%, and 80%) case scores and sensitivities

ADI-R	12-20/NV21-47		SW21-47		PH21-47	
<i>N</i> (ASD vs. NS)	289 (244, 45)		115 (85, 30)		105 (65,40)	
Specificity	Score	Sensitivity	Score	Sensitivity	Score	Sensitivity
100%	22	22%	18	41%	28	14%
90%	17	55%	13	70%	23	34%
80%	12	77%	8	83%	21	43%

ADOS	ADOS-T		Module 1		Module 2	
<i>N</i> (ASD vs. NS)	46 (31, 18)		113 (90, 13)		61 (52, 39)	
Specificity	Score	Sensitivity	Score	Sensitivity	Score	Sensitivity
100%	18	35%	14	80%	20	17%
90%	-	-	13	80%	11	68%
80%	17	45%	12	90%	9	79%

12-20/NV21-47 all children 12-20 months and nonverbal children 21-47 months, SW21-47 children 21-47 months with single words, PH21-47 children 21-47 months with phrase speech, ADI-R Autism Diagnostic Interview-Revised, ADOS Autism Diagnostic Observation Schedule, ADOS-T ADOS-Toddler

High specificity case scores are available from the ADOS-T for 12-20/NV21-47 group, from Module 1 for SW21-47 group, from Module 2 for PH21-47 group.

ADOS. Thus, in about half of the cases, clinicians were not blind to developmental history and the caregiver's descriptions, which might have affected their ADOS administration and coding. However, the correlation between the algorithm total scores for the two instruments was slightly higher when different clinicians versus the same clinician administered the instruments (r of .66 vs. .59).

Even though about 280 children (46% of the entire sample) were clinic referrals at UMACC, the results were not entirely based on clinic patients but also derived from research participants. In general developmental disorders clinics, characteristics of referrals might differ from the children in the present study. Thus, inferences about the generalization of results from the present study should be carefully considered.

Conclusion

The ADI-R and ADOS provide both unique and overlapping information important for clinicians and researchers making diagnostic decisions. When both instruments were used in combination, well-balanced sensitivities and specificities were obtained. Using the newly developed ADI-R algorithms and the revised ADOS and new ADOS-T algorithms, even with young children, validity for combined use of the instruments in the present study was comparable or higher than in past research (Risi et al., 2006). Taking into account information from both a skilled clinician and a caregiver contributes to diagnostic differentiations especially for more complex cases. Alternative combinations with other instruments besides or in addition to the ADI-R and/or ADOS, such as the SCQ (Rutter et al., 2003a), SRS (Constantino & Gruber, 2005), CCC-2 (Bishop,

2003), and the Screening Tool for Autism in Two-Year-Olds (STAT; Stone, Coonrod, & Ousley, 2000) may be equally effective. In addition, sequential assessment strategies may be appropriate for some children allowing cost- and time- effective research and clinical practice.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 Characteristics of misclassified children

Table S2 Sensitivities and specificities of positive and negative screening estimates (PSE/NSE)

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Acknowledgements

We gratefully acknowledge the help of Susan Risi, Pamela Dixon Thomas, Suzi Naguib, Fiona Miller, Rhiannon Luyster, Whitney Guthrie, Kaite Gotham, Kathryn Larson, Kathy Hatfield, and Shanping Qiu, as well as the families that participated in this research. This study was funded by NIMH RO1 MH066469, MH57167, and HD 35482-01.

One of the authors, Catherine Lord, receives royalties for the ADI-R and ADOS; profits related to this study were donated to charity.

Correspondence to

So Hyun Kim, 2236 East Hall, 530 Church St., Ann Arbor, MI 48109, USA; Tel: (734) 764-2580; Fax: (734) 764-3520; Email: sohkim@umich.edu

Key points

- The ADI-R and ADOS make independent, additive contributions to more consistent and accurate diagnostic decisions for clinicians evaluating toddlers and young preschoolers with ASD.
- Well-balanced sensitivities and specificities are obtained for ASD diagnoses using information from both parent interview and clinician observation.
- Sequential assessment strategies using high specificity case scores and negative/positive screening estimates may allow clinicians to use a single measure for the diagnosis of ASD in some cases.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). Washington, DC: Author.
- de Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Mulder, E., Volkmar, F., & Minderaa, R. (2004). Interrelationship between Autism Diagnostic Observation Schedule-Generic (ADOS-G), Autism Diagnostic Interview-Revised (ADI-R), and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders*, 34, 129–137.
- Bishop, D. (2003). *The children's communication checklist-2*. London: Psychological Corporation.
- Bishop, D., & Baird, G. (2001). Parent and teacher report of pragmatic aspects of communication: Use of the children's communication checklist in a clinical setting. *Developmental Medicine and Child Neurology*, 43, 809–818.
- Constantino, J., & Gruber, C. (2005). *Social responsiveness scale*. Los Angeles, CA: Western Psychological Services.
- Constantino, J., LaVesser, P., Zhang, Y., Abbacchi, A., Gray, T., & Todd, R. (2007). Rapid quantitative assessment of autistic social impairment by classroom teachers. *Journal of American Academy of Child and Adolescent Psychiatry*, 46, 1668–1676.

- Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E., Leventahl, B., & Lord, C. (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.
- DiLavore, P., Lord, C., & Rutter, M. (1995). The pre-linguistic autism diagnostic observation schedule (PL-ADOS). *Journal of Autism and Developmental Disorders*, *25*, 355–379.
- Elliott, C.D. (1990). *Differential abilities scale (DAS)*. San Antonio, TX: Psychological Corporation.
- Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule (ADOS): Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*, *37*, 613–627.
- Jaeschke, R., Guyatt, G., & Lijmer, J. (2002). Diagnostic tests. In G. Guyatt & D. Rennie (Eds.), *User's guide to the medical literature* (pp. 121–140). Chicago: AMA Press.
- Kim, S., & Lord, C. (2011). New Autism Diagnostic Interview-Revised (ADI-R) algorithms for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Autism and Developmental Disorders*, in press. Epub ahead of print. Retrieved August 10, 2011, from <http://www.ncbi.nlm.nih.gov/pubmed/21384244>.
- Landa, R., & Garrett-Mayer, E. (2006). Development in infants with autism spectrum disorders: A prospective study. *Journal of Child Psychology and Psychiatry*, *47*, 629–638.
- Le Couteur, A., Haden, G., Hammal, D., & McConachie, H. (2008). Diagnosing autism spectrum disorders in preschoolers using two standardised assessment instruments: The ADI-R and the ADOS. *Journal of Autism and Developmental Disorders*, *38*, 362–372.
- Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). *Autism diagnostic observation schedule: Manual*. Los Angeles: Western Psychological Services.
- Lord, C., Storoschuk, S., Rutter, M., & Pickles, A. (1993). Using the ADI-R to diagnose autism in preschoolers. *Infant Mental Health Journal*, *14*, 234–252.
- Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K.,... & Lord, C. (2009). The Autism Diagnostic Observation Schedule—Toddler module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*, 1305–1320.
- Mullen, E. (1995). *Mullen scales of early learning (AGS ed.)*. Circle Pines, MN: American Guidance Service.
- Newcombe, R.G. (1998). Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Statistics in Medicine*, *17*, 873–890.
- Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P.,... & Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*, 1094.
- Rutter, M., Bailey, A., & Lord, C. (2003a). *The social communication questionnaire*. Los Angeles: Western Psychological Services.
- Rutter, M., Le Couteur, A., & Lord, C. (2003b). *Autism diagnostic interview-revised*. Los Angeles: Western Psychological Services.
- Siegel, B., Vukicevic, J., Elliott, G., & Kraemer, H. (1989). The use of signal detection theory to assess DSM-III-R criteria for autistic disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *28*, 542–548.
- Sparrow, S., Balla, D., & Cicchetti, D. (1984). *Vineland adaptive behavior scales*. Circle Pines, MN: American Guidance Service.
- Steiger, J. (1980). Test for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251.
- Stone, W., Coonrod, E., & Ousley, O. (2000). Screening tool for autism in two-year-olds (stat): Development and preliminary data. *Journal of Autism and Developmental Disorders*, *30*, 607–612.
- Szatmari, P., Merette, C., Bryson, S.E., Thivierge, J., Roy, M.A., Cayer, M., & Maziade, M. (2002). Quantifying dimensions in autism: A factor-analytic study. *Journal of American Academy of Child and Adolescent Psychiatry*, *41*, 467–474.
- Ventola, P.E., Kleinman, J., Pandey, J., Barton, M., Allen, S., Green, J.,... & Fein, D. (2006). Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders*, *36*, 839–847.
- Wiggins, L., Robins, D., Adamson, L., Bakeman, R., & Henrich, C. (2011). Support for a dimensional view of autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders*, in press. Epub ahead of print. Retrieved August 10, 2011, from <http://www.ncbi.nlm.nih.gov/pubmed/21448751>.
- Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestation of autism in the first year of life. *International Journal of Developmental Neuroscience*, *23*, 143–152.

Accepted for publication: 8 July 2011

Published online: 1 September 2011