Published in final edited form as:

J Autism Dev Disord. 2013 July ; 43(7): 1527–1538. doi:10.1007/s10803-012-1696-4.

Multisite Study of New Autism Diagnostic Interview-Revised (ADI-R) Algorithms for Toddlers and Young Preschoolers

So Hyun Kim, PhD, Audrey Thurm, PhD, Stacy Shumway, PhD, and Catherine Lord, PhD

S.K. is affiliated with Yale Child Study Center, Yale University School of Medicine (40 Temple St. Suite 7-D, New Haven, CT 06510) A.T. and S.S. are affiliated with the National Institute of Mental Health (Pediatrics and Developmental Neuroscience Branch, National Institute of Mental Health, 10 Center Drive, MSC 1255, Bethesda, Maryland 20892-1255). C.L. is affiliated with the Weill Cornell Medical College (Institute for Brain Development, 21 Bloomingdale Road, White Plains, NY 10605)

Abstract

Using two independent datasets provided by National Institute of Health funded consortia, the Collaborative Programs for Excellence in Autism and Studies to Advance Autism Research and Treatment ($n=642$) and the National Institute of Mental Health ($n=167$), diagnostic validity and factor structure of the new Autism Diagnostic Interview (ADI-R) algorithms for toddlers and young preschoolers were examined as a replication of results with the 2011 Michigan sample (Kim & Lord, 2011). Sensitivities and specificities and a three-factor solution were replicated. Results suggest that the new ADI-R algorithms can be appropriately applied to existing research databases with children from 12 to 47 months and down to nonverbal mental ages of 10 months for diagnostic grouping.

Keywords

Early Diagnosis; Autism Spectrum Disorders; Autism Diagnostic Interview-Revised

Kim & Lord (2012) recently proposed new algorithms for 12-47 month-old children intended to improve the diagnostic validity of the Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003) for toddlers and young preschoolers with autism spectrum disorders (ASD). The goal of the present study was to replicate the findings based on these new ADI-R algorithms for toddlers and young preschoolers in two different datasets: a multisite database provided by National Institute of Health funded consortia, the Collaborative Programs for Excellence in Autism and Studies to Advance Autism Research and Treatment (CPEA/STAART) and another provided by the Intramural Research Program of the National Institute of Mental Health (NIMH).

The ADI-R is a standardized, semi-structured, investigator-based interview for parents or caregivers of individuals with autism. It provides a diagnostic algorithm for the ICD-10 and DSM-IV definitions of autism (World Health Organization [WHO], 1992; American Psychiatric Association [APA], 1994). The interview is appropriate for the diagnostic assessment of persons in early childhood to adult life, provided that they have a non-verbal mental age above 2 years. The ADI-R includes 93 items in three domains of functioning –

Correspondence to So Hyun Kim, 40 Temple St. Suite 7-D, New Haven, CT 06510, Phone:(203) 764-5937, so-hyun.kim@yale.edu.

The results from the preliminary analyses were presented at the 2010 International Meeting for Autism Research (IMFAR). One of the authors, Catherine Lord, receives royalties for the ADI-R and ADOS; profits related to this study were donated to charity.

language/communication, reciprocal social interactions, and restricted, repetitive, and stereotyped behaviors and interests, as well as other aspects of behaviors. Up to 42 of the interview items are systematically combined to produce a formal, diagnostic algorithm for autism as specified by the authors, or a general diagnosis of ASD as used in several collaborative studies (Risi et al., 2006).

In addition to the standard version of the ADI-R, a ‘Toddler’ version of the ADI-R intended for children under 4 years has been developed for research purposes. It includes 32 new questions and codings about the onset of autism symptoms and general development for a total 125 items. Other items in both the standard and Toddler versions of the ADI-R are identical except that the Toddler ADI-R does not have codes for behaviors between 4 and 5 years of age (referred to as *most abnormal 4 to 5*). The newly developed ADI-R algorithms for toddlers and young preschoolers with ASD (Kim & Lord, 2012) include only items that appear in both standard and toddler versions. Thus existing datasets using either version of the ADI-R can be used for children under age 4 with nonverbal mental ages down to 10 months.

One of the goals of the new algorithms for toddlers and young preschoolers was to reduce the effect of child characteristics such as age and language levels on algorithm scores. This was achieved by dividing the children into smaller, more homogeneous cells by age and language level (Kim & Lord, 2012). The developmental cells in the original study, based on children recruited from Michigan (“Michigan sample”), were grouped as follows: (1) all children 12 to 20 months of age as well as nonverbal children 21 to 47 months of age (“12–20/NV21–47”); (2) children 21 to 47 months with single words (“SW21–47”); and (3) all children 21 to 47 months with phrase speech (“PH21–47”).

New ADI-R algorithms for toddlers and preschoolers have some changes, including factor structures and cutoff scores for the instrument classification of ASD, that are different from the original algorithms. Based on factor analyses using the Michigan sample, several items that had previously loaded separately onto Social or Communication domains of the previous algorithms were merged into a *Social Affect* (SA) domain for the “12-20/NV21-47” and “SW21-47” algorithms. An additional factor consisting of *Restricted and Repetitive Behaviors* (RRBs) items was also included in the algorithm because of its independent contribution to distinguishing ASD cases from non-spectrum cases. Another factor associated with *Imitation, Gesture, and Play* (IGP) emerged but was not predictive of diagnoses. Thus, for the “12-20/NV21-47” and “SW21-47” algorithms, items in the SA and RRB domains were combined to generate single cutoffs (rather than multiple cutoffs for different domains as in the original algorithm), for an instrument classification of ASD (rather than a classification of autism as in the original algorithm). The IGP domain is not part of diagnosis but was included because of its potential interest to researchers and clinicians. For the “PH21-47” algorithm, several items that had previously loaded separately onto the Social and Communication domains were merged into a single domain, *Social Communication* (SC). For this algorithm, items associated with SC, RRB and another domain, *Reciprocal and Peer Interaction* (RPI), also found to be predictive of diagnoses, were combined to generate single cutoffs.

The development of the new ADI-R algorithms for toddlers and young preschoolers was prompted by findings of the preexisting algorithm’s lower diagnostic validity for toddlers and young preschoolers compared to older children (Ventola et al., 2006; Wiggins & Robins, 2008). The published algorithms for the ADI-R include a *current behavior algorithm* form (as distinguished from an empirically supported diagnostic algorithm) for children whose ages range from 2 years, 0 months to 3 years, 11 months. Using the current behavior algorithm, Lord and colleagues (1993) found satisfactory sensitivities and specificities for

children over 2 years in the original sample of 94 children (both over 90%). However, the authors found that diagnostic discrimination between nonverbal children with ASD and nonverbal children without ASD under 2 years of age, especially for those with mental ages under 18 months was poor, with low specificity (around 70%). Analyzing a larger sample (n=1529), Risi et al. (2006) also demonstrated high sensitivity (above 80%) of the ADI-R for the classification of children with ASD under 3 years of age, but similarly low specificity (around 70%) for these children in comparison to children with non-spectrum disorders (NS). Thus, despite the relatively good sensitivity and specificity of the ADI-R algorithms for older children, more appropriate algorithms were needed for younger children.

The goal of improved diagnostic validity for toddlers and preschoolers on the ADI-R was achieved in part by defining different cutoffs for research and clinical purposes (Kim & Lord, 2012). Research cutoffs were selected to maximize specificity while maintaining an adequate level of sensitivity (above 80%) for the comparison of Autism (more narrowly defined than ASD by excluding children with Pervasive Developmental Disorders-Not Otherwise Specified; PDD-NOS) vs. NS. As a result, research cutoffs can be used for researchers conducting expensive and time-consuming procedures such as neuroimaging researchers who may wish to limit samples to autism cases in order to more strictly exclude likely NS cases. Using the research cutoffs, the comparison of Autism vs. NS resulted in sensitivity ranging from 80 to 84% and specificity from 85 to 90%. As an alternative, clinical cutoffs were proposed that maximize sensitivity while maintaining an adequate level of specificity (above 70%) for the comparison of ASD vs. NS. Based on the Michigan sample, sensitivity for the clinical cutoffs for ASD vs. NS ranged from 80 to 94% and specificity from 70 to 81%, depending on the developmental cells (Kim & Lord, 2012).

In addition, the new ADI-R algorithms for toddlers and young preschoolers provide clinicians with *ranges of concern* that represent the severity of autism symptoms based on a dimensional approach. There are three ranges of concern, which reflect the severity of autism symptoms from little-to-no, mild-to-moderate, to moderate-to-severe. These *ranges of concern* can be used by a clinician or a researcher in deciding whether a child should be followed with further assessments or considered for treatment even when a diagnosis not established (Kim & Lord, 2012). Because the use of ranges of concern is relatively new, one of the goals for the present study was to replicate its use with different samples. The different options of research and clinical cutoffs and ranges of concern were created recognizing that diagnostic decisions about ASD in very young children are less stable and precise than for older children and adolescents (Lord et al., 2006; Sutter et al., 2007). Because there are always trade-offs between sensitivity and specificity in the diagnosis of ASD in toddlers and young preschoolers, these options allow researchers and clinicians to be transparent about the choices they make.

The previous study with the Michigan sample (Kim & Lord, 2012) showed that differentiating children with ASD from those with NS using the original ADI-R algorithms was most challenging for children at the extremes; those with the most limited verbal skills and older children with the most advanced verbal skills. For less verbally able children ("12-20/NV21-47") the new ADI-R algorithm showed substantial gains in specificity (37-42%) over the previous algorithm. For children with the highest verbal abilities, the "PH12-47" algorithm showed modest to moderate gains in specificity (2-14%) and consistent, moderate improvements in sensitivity (10-14%).

Replications across independent datasets with well-defined populations with and without ASD, particularly of diagnostic validity for research and clinical cutoffs as well as ranges of concern, are crucial before the new algorithms are widely used by researchers and clinicians. Replication of the factor structure from the Michigan sample was also important because

new factors (Imitation, Gestures, & Play and Reciprocal Social Interaction) emerged. Here we present data from two new independent samples. They are reported as two separate studies because of their unique characteristics.

Method: Study 1

Participants

Analyses were conducted on data provided by the CPEA, a network of 10 sites, and the STAART program, a network of eight research centers (some of which overlap with CPEA sites) throughout the United States and Canada. Multiple studies have been published based on combined datasets collected through CPEA and/or STAART (e.g., Dawson et al., 2010; Gotham et al., 2009; Ozonoff et al., 2004). Both networks were funded by the National Institutes of Health. The dataset represents 641 research participants from 11 different sites (children in CPEA from the Michigan sample that was used in the Kim and Lord 2012 study were excluded in the present study). Each case had a contemporaneous ADI-R, nonverbal IQ (NVIQ), and best estimate clinical diagnosis. Children had research diagnostic evaluations at the Boston University School of Medicine (n=188), University of Washington (n=140), University of Colorado Health System (n=109), University of Rochester (n=97), University of Utah (n=20), University of California, Los Angeles (n=33), University of California, Davis (n=30), University of California, Irvine (n=19), Mount Sinai Medical Center (n=3), and Yale University (n=2). The data collection process was conducted independently within each site.

A total of 641 children (496 males) with a mean age of 33 months (SD=8.01) were divided into three developmental cells. The “12-20/NV21-47” algorithm included 188 ASD, 16 NS cases and 19 cases with typical development (TD). The “SW21-47” algorithm included 207 ASD, 36 NS, and 8 TD cases. Finally, the “PH21-47” algorithm included 131 ASD, 18 NS, and 18 TD cases. The NS samples consisted of 70 cases with developmentally delays/intellectual disabilities. Children in the NS group were recruited either in response to advertisements targeted to children with possible symptoms of ASD or other developmental delay. The typical group was recruited in response to advertisements targeting typically developing children. Since differentiating the ASD from NS groups is more useful and challenging, TD cases were excluded in diagnostic comparisons but included in all other analyses. The racial/ethnic makeup of the total sample was 80% white, 9% multiracial, 4% African American, 2% Asian American, 0.5% Native American, 0.2% other races/ethnicities, and 8% unknown. See Table 1 for age and NVIQ distributions by diagnosis.

Measures and Procedure

For the CPEA/STAART dataset, a clinical diagnosis was made by a psychologist and/or psychiatrist after reviewing the ADI-R, Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999), and other psychometric assessments including neurocognitive testing. The ADI-R was administered by a clinical psychologist or trainee who met standard requirements for research reliability (Lord, Rutter, DiLavore, & Risi, 1999). A developmental hierarchy of psychometric measures, most frequently the Mullen Scales of Early Learning (Mullen, 1995) (n=612; MSEL), Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 1989) (n=15), and the Differential Abilities Scale (Elliott, 1990) (n=12), was used to determine nonverbal IQ (NVIQ) scores. For the MSEL, ratio NVIQs were used as estimates of ability based on the finding that ratio IQs derived from MSEL age equivalents had good convergent validity with the DAS (Bishop, Guthrie, Coffing, & Lord, 2011). MSEL ratio NVIQs were calculated based on a standard procedure by averaging the age equivalents of Visual Reception and Fine Motor subtests to obtain

mental age, and then dividing mental age by chronological age and multiplying by 100. For other tests, deviation IQs were used.

Design and Analysis

The sample was divided by age and language level to yield the three developmental cells outlined in the Michigan study 1 (12-20/NV21-47; SW21-47; PH21-47). In order to interpret results from the Michigan study and the new datasets, we needed to determine the compatibility of the samples. Participant characteristics were compared to the Michigan sample. Domain totals and diagnostic classifications were generated for each case by adding the new algorithm item scores appropriate to the developmental cell of the participant and applying the new cutoffs. For statistical analyses, ADI-R item scores of 2 and 3 were collapsed into 2 as they are on the algorithms. Confirmatory Factor Analyses (CFA) for categorical data in Mplus 5.2 based on the geomin rotation (Muthén & Muthén, 2007) were used to confirm the factor structure of the ADI-R using the items in the new ADI-R algorithms. One of the items (*Current: Social Smiling*) included in the algorithms derived from the Michigan data was excluded from the factor analyses for the “12-20/NV21-47” and “SW21-47” algorithms because the proportion of children with data applicable for factor analyses on this score was not sufficiently large. For the same reason, *Current: Attention to Voice* was excluded from the factor analyses for the “PH21-47” algorithm. However, all of these items were included in other analyses. Receiver operating characteristic (ROC) curves were calculated (Siegel, Vukicevic, Elliott, & Kraemer, 1989) and the sensitivity and specificity of the new ADI-R algorithms were compared to the results from the Michigan sample. ROC results for single cutoffs of ASD (rather than separate autism and ASD cutoffs) are reported for children with a nonverbal mental age of 10 months or higher, as in the Michigan study. The percentages of children in each range of concern were calculated. Logistic regressions were performed to examine if domain totals predicted diagnoses. Because there were diagnostic group differences in age and NVIQ (See Table 1) for the CPEA/STAART dataset, analyses were performed controlling for these variables. Diagnostic comparisons included ASD vs. NS and autism alone (PDD-NOS) vs. NS.

Results: Study 1

Comparability of Michigan and Replication Samples

Differences between the Michigan sample (Kim & Lord, 2012) and the CPEA/STAART dataset occurred in the chronological age and IQ scores of specific developmental cells. For the “12-20/NV21-47,” the mean ages of the CPEA/STAART sample in all diagnostic groups (See Table 1) were significantly older ($F=8.56$, $p<.001$ for ASD, $F=29.3$, $p<.01$ for NS, $F=19.8$, $p<.001$ for TD), and the mean IQ scores for NS and TD were significantly lower ($F=12.41$, $p<.01$ for NS, $F=5.53$, $p<.05$ for TD) than the Michigan sample (Mean ages of 25.9, 31.5, and 15.6 mos.; Mean IQs of 73, 75.7, and 111 for ASD, NS and TD respectively). In contrast, the mean age of the CPEA/STAART sample for ASD was significantly younger for the “SW21-47” ($F=26.22$, $p<.001$) and the “PH21-47” ($F=11.3$, $p<.01$) than the Michigan sample (Mean ages of 36 and 40.9 mos. for “SW21-47” and “PH21-47” respectively). For the NS group in the “PH21-47,” the mean age of the CPEA/STAART sample was significantly older ($F=23$, $p<.001$), and the mean IQ score significantly lower ($F=36.3$, $p<.001$) than the Michigan sample (Mean age of 38.7 mos.; Mean IQ of 92.1).

Correlations with Participant Characteristics

Correlations between all ADI-R domain totals and participant characteristics including chronological age and NVIQ were minimal ($r<0.4$).

Confirmatory Factor Analyses

Using the CPEA/STAART sample, the three-factor model including the new factors (IGP and RPI) that emerged from the Michigan sample, replicated satisfactorily across all developmental cells, with root mean square error approximation (RMSEA) values ranging from 0.053 to 0.077 and Comparative Fit Indices (CFI) from 0.908 to 0.948 (a RMSEA of 0.08 or lower and a CFI between 0.9 and 1 indicate a good fit (Browne & Cudeck, 1993)). Correlations between factors ranged from 0.69 to 0.94 by cell. See Table 2 for the factor loadings under a three-factor solution.

Sensitivity and Specificity

As shown in Table 3, compared to the Michigan sample, the CPEA/STAART sample resulted in higher specificities for the NS cases (ranging from 92 to 94% using research cutoffs). Sensitivities for the CPEA/STAART sample were comparable to the Michigan sample (ranging from 85 to 96% using clinical cutoffs).

Ranges of Concern

The goal of the ranges of concern for the Michigan sample was that at least 80% of children with ASD and no more than about 5% of children with TD would fall in the two extreme risk ranges (mild-to-moderate and moderate-to-severe ranges). In the initial analyses, the percentage of children with ASD who fell in the risk ranges (mild-to-moderate and moderate-to-severe) varied from 80 to 85%, and the percentage of children with NS who fell in the risk ranges varied from 30 to 33%. Using the CPEA/STAART dataset, for all developmental cells, more than 85% of children with ASD fell in the risk ranges. The percentages of children with NS who were placed in the risk range varied from 6 to 16% depending on developmental cells.

Logistic Regression to Compare Samples on Predictors of Clinical Diagnosis

With the Michigan sample, logistic regressions indicated that both SA and RRB domains for the “12-20/NV21-47” and “SW21-47” groups and SC, RRB, and RPI domains for the “PH21-47” group made significant independent contributions to the prediction of autism and ASD diagnoses. With the CPEA/STAART sample, the SA/SC domains significantly predicted the ASD diagnosis for all developmental cells (“12-20/NV21-47,” odds ratio (OR)=1.58, $p<.01$; “SW21-47,” OR=2.04, $p<.01$, “PH21-47,” OR=1.5, $p<.01$). The contribution of the RRB domain to ASD diagnoses varied by developmental cells when age, NVIQ, and the other two domains were controlled (“12-20/NV21-47,” n/s; “SW21-47,” OR=1.31, $p=0.07$; “PH21-47,” OR=1.80, $p<0.01$). The Reciprocal Peer Interaction (RPI) domain for the “PH21-47” group did not separately predict diagnoses but when the domain was included in the total score, the total score significantly predicted ASD diagnosis (OR=1.46, $p<.01$). The total score for the other two developmental cells also significantly predicted ASD diagnoses (“12-20/NV21-47,” OR=1.45, $p<.01$; “SW21-47,” OR=1.66, $p<.01$). All three domains made significant independent contributions to autism (as opposed to the broader category of ASD) diagnoses for all developmental cells ($p<.01$).

Method: Study 2

Participants

A few studies have been published based on the NIMH dataset in the past (Shumway et al., 2011; Shumway et al., in press). Because this dataset had relatively few non-ASD participants, subset of analyses from the original study was conducted on these data (sample size ranging from 12-21 for NS and 0-8 for TD groups depending on the developmental cells). Each case included a contemporaneous Toddler ADI-R, NVIQ, and best estimate

clinical diagnosis. There were 166 children (134 males) from 12 to 47 months ($M=36$, $SD=8.04$). The “12-20/NV21-47” analyses included 40 ASD, 12 NS, and 2 TD cases. The “SW21-47” analyses included 37 ASD and 19 NS cases. The “PH21-47” analyses included 27 ASD, 21 NS, and 8 TD cases. See Table 4 for more details.

The NS groups consisted of children with general developmental delays and language delays. Children in the NS group were recruited in two ways. Most children were referred for possible autism symptoms but did not meet the criteria for ASD. Some children were recruited in response to advertisements targeted to children with developmental delay. The typical group was recruited in response to advertisements targeting typically developing children. The racial/ethnic makeup was 15% African American, 5% Asian American, 7% multiracial, and 73% Caucasian.

Measures and Procedure

For the cases in the NIMH dataset, a clinical diagnosis was made by doctoral level clinicians after reviewing all available information from the Toddler ADI-R, ADOS, and other psychometric testing including neurocognitive testing. The MSEL (Mullen, 1995) was used to determine ratio NVIQ scores.

Design and Analysis

Analyses of the NIMH dataset focused on examining the factor structure and diagnostic validity of the new algorithms. After the sample was divided by age and language level into the three developmental cells outlined in the Michigan study, domain totals and diagnostic classifications were generated for each case in the same way described in Study 1. Confirmatory factor analyses were performed and ROC curves calculated to obtain sensitivity and specificity. Ranges of concern were not examined due to the limited sample sizes within cells.

Results: Study 2

Comparability of Michigan and Replication Samples

Compared to the Michigan sample where children with NS were specifically recruited to provide a control group against which to assess the diagnostic validity of the diagnostic instruments, many of children in the NS group in the NIMH dataset included children who were initially ASD referrals who did not meet criteria for ASD. Other differences between samples were chronological age and NVIQ scores of specific developmental cells. For the “12-20/NV21-47,” the mean age of the NIMH sample for NS and TD (See Table 4) was significantly older ($F=9$, $p<.01$ for NS; $F=4.17$, $p<.05$ for TD) than the Michigan sample (Mean ages of 31.5 and 15.6 mos. for NS and ASD respectively). For “SW21-47,” the mean IQ score of the NIMH data for NS was significantly higher ($F=5.52$, $p<.05$) than the Michigan sample (Mean IQ of 69.1). For “PH21-47,” the mean IQ score of the NIMH data for NS was significantly lower ($F=5.67$, $p<.05$) and the mean age of the NIMH data for TD was significantly older ($F=12.43$, $p<.01$) than the Michigan sample (Mean IQ of 92.1 for NS; Mean age of 28.4 mos. for TD).

Correlations with Participant Characteristics

All correlations between ADI-R domain totals and participant characteristics including chronological age and NVIQ were minimal ($r \leq 0.4$).

Confirmatory Factor Analyses

Given the smaller sample size for the NIMH data, results from factor analyses were less stable compared to the Michigan and CPEA/STAART samples. The three-factor model replicated well with the NIMH sample for the 12-20/NV21-47 and SW21-47 groups. However, the goodness of fit was less satisfactory for the PH21-47 group (See Table 2). Using the CFA, correlations between factors ranged from 0.55 to 0.99 by cell. The RMSEA values ranged from 0.066 to 0.093 (a RMSEA of 0.08 or lower is considered good), and the CFI ranged from 0.806 to 0.892 (a CFI between 0.9 and 1 indicate a good fit). See Table 2 for more details.

Sensitivity and Specificity

As shown in Table 3, compared to the Michigan sample, sensitivities for the NIMH sample (89-97%) were higher using the new algorithms for the clinical cutoffs. Specificities for the NIMH sample (86-89% using research cutoffs) were comparable to the Michigan data except the specificity for the “12-20/NV21-47” group, which was 64%. To explore this further, we examined children in the low specificity group in more detail. There were 3 false positives using the research cutoff and 1 more false positive using the clinical cutoff out of 12 children. All of these children were also classified as false positives using the original ADI-R algorithm. Thus, the small sample size and complex clinical samples posed a challenge.

Discussion

Overall, the diagnostic validity of new ADI-R algorithms (Kim & Lord, 2012) for toddlers and young preschoolers from a Michigan sample was satisfactorily replicated using two independent datasets. In the previous study, the new ADI-R algorithms showed improved diagnostic validity for children from 12 to 47 months compared to the pre-existing algorithm. The total number of items used to generate a cutoff for an ASD classification was reduced to 13-20 items compared to the pre-existing algorithm of 33-39 items. Here, in Study 1, the sensitivities obtained from 641 children in the CPEA/STAART dataset were comparable and the specificities markedly improved compared to the Michigan sample. Results from 167 children from NIMH in Study 2 varied more; compared to the Michigan sample, sensitivities were higher and specificities were comparable except for the “12-20/NV21-47” group, which, based on a small number of cases showed a slightly lower specificity.

One of the goals of the new ADI-R algorithms for toddlers and young preschoolers was to reduce the effect of participant characteristics other than severity of ASD symptoms on algorithm scores. By dividing the children into smaller groups with more homogenous age and language levels, correlations between participant characteristics and algorithm scores remained at or below r of 0.4 for the CPEA/STAART and NIMH datasets, as they had been for the Michigan sample. Even though the correlations were not large, since there is tremendous heterogeneity in the manifestations of autism symptoms within ASD by age, language level, and IQ, clinicians and researchers interpreting the scores should still note that these factors could have potential influence on the algorithm scores.

The revised algorithms also better represent the observed diagnostic features of ASD within the context of anticipated changes in the next edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM; APA, 1994) in several ways; 1) social and communication behaviors fall under one domain and 2) RRBs contribute to the ADI-R classification of ASD. These results are consistent with past studies using the ADI-R with older children that showed that items associated with social and communication loaded onto

a single factor (Frazier, Youngstrom, Kubu, Sinclair, & Rezai, 2008; Snow, Lecavalier, & Houts, 2009; van Lang et al., 2006) as well as recent changes to the ADOS algorithms (Gotham, Risi, Pickles, & Lord, 2007) including those for the ADOS-Toddler module (Luyster et al., 2009). The factor structure of a combined social-communication domain plus two other domains was replicated in both the CPEA/STAART and NIMH samples. The goodness of fit was especially good for the CPEA/STAART sample due in part to the larger sample size.

As in the previous study with the Michigan sample, differentiating children with ASD from those referred for ASD but who received NS diagnoses was more challenging both for children with minimal language and those with the most advanced language skills. For less verbally able children, the “12-20/NV21-47” algorithm using the CPEA/STAART sample showed substantial improvement in specificity (2-24%) while maintaining similar levels of sensitivity compared to the Michigan sample. In addition, consistent with the Michigan sample, more than 80% of children with ASD using the CPEA/STAART sample fell in the two ranges of clinical concern (mild-to-moderate and moderate-to-severe ranges). In fact, the proportion of the NS cases in the risk ranges was smaller for the CPEA/STAART sample than the Michigan sample, in part because the CPEA/STAART sample was purely research-based whereas the Michigan sample included less straightforward, consecutive clinical cases. The NS group in the NIMH sample represented an even more extreme example of complex cases because an even higher proportion of children who received the NS diagnosis were originally referred for possible autism. The NS groups included in the NIMH sample tended to show higher scores on most of the domains compared to those in the CPEA/STAART sample. As expected, their specificities were lower than the Michigan and CPEA/STAART samples, which both included a higher percentages of NS cases specifically recruited as research controls.

Results from logistic regressions confirmed an independent contribution of the SA/SC domains to diagnosis of ASD. Results for the RRB domain varied by developmental cell. RRBs made a significant contribution to the diagnosis ASD for older and more able children and less for younger and/or more impaired children. In contrast, using a sample of 455 toddlers and preschoolers from 8 to 56 months of age, Kim and Lord (2010) found that all children with ASD had at least one RRB at the time of assessment when RRB scores from both the ADI-R and ADOS were examined. This is similar to past studies showing that not all parents of children with ASD, especially when their children are very young, report RRBs (e.g., Wiggins and Robins, 2008). Thus, it seems possible that some of these children had RRBs but their parents did not yet recognize them as such, especially when children were very young (under 21 months of age) or when RRBs were accompanied by more severe impairments in social and communication skills.

Past studies suggested that RRBs add to stability of diagnoses over time and to diagnostic predictability across measures (Kim & Lord, 2012; Risi et al., 2006; Lord et al., 2006). Similarly, the present study showed that when RRBs were added to the totals, total scores were significantly predictive of diagnosis of ASD in all developmental cells. The results from these past studies and these new results are in keeping with the proposed DSM-5 ASD criteria which include a requirement of RRBs for all ASD.

The results from the Michigan sample showed that the Reciprocal Peer Interaction (RPI) domain predicted the diagnosis of ASD for the “PH21-47” algorithm. However, using the CPEA/STAART sample, the domain did not on its own predict a diagnosis of ASD while controlling for the other domains, age, and NVIQ, even though it was reliably predictive of autism diagnosis alone (without children with PDD-NOS). Nevertheless, when the scores from all three domains (SC, RRB, and RPI) were combined to generate single cutoffs, the

diagnostic validity improved for ASD and autism compared to when the RPI domain was excluded. The Imitation, Gesture, and Play (IGP) domain for the “12-20/NV21-47” and “SW21-47” groups was not included in the algorithm total nor for the Michigan sample because it did not contribute to diagnosis though the domain score will remain on the algorithm form as a source of information.

With the new algorithms, clinicians and researchers now have an option of using ranges of concern. As Kim and Lord (2012) indicated, ranges of concern and clinical cutoffs (more inclusive than research cutoffs) can be used by clinicians who wish to avoid denying a child access to services especially at young ages where diagnoses are less stable (Lord et al., 2006). Researchers who may wish to include a broader group of ASD in their samples can use the clinical cutoffs. On the other hand, researchers who wish to restrict their samples to children who have clearer symptoms of ASD and exclude as many NS cases may choose to use the research cutoffs. Unlike the original ADI-R algorithms, single cutoffs are used for the instrument classification of ASD for the new toddler and preschooler algorithms. Even though domain totals are not used for the instrument classification of ASD on these toddler and preschooler algorithms, they can provide useful information. For example, since there is tremendous heterogeneity in manifestation of symptoms, domain scores can be used to quantify separate symptom presentations in social and communication skills as well as RRBs. Some children may have significant social and communication impairments but only few RRBs. Others may have different profiles such as relatively intact social and communication skills but strong fixated interests. These different profiles may potentially help researchers stratify subtypes within ASD and examine their associations with etiological, neurocognitive, and molecular genetic factors. In addition, the quantified profiles obtained based on different profiles (social communication, RRBs, and IGP/RPI) can be used for designing individualized goals in intervention and educational programs.

Limitations

Though the three-factor solution was satisfactory with the CPEA/STAART dataset, the goodness of fit using the NIMH dataset varied by developmental cells. Specificities were lower than expected, especially for the NIMH dataset in which many of the NS cases were referred for possible ASD. Recruitment differences and possible treatment effects may have affected the discriminability of children in these different groups. Small sample size, especially for the control groups, was a limitation. Two items were excluded for the factor analyses of the CPEA/STAART dataset due to the limited number of valid scores on these items even though they were included in the analyses for the NIMH dataset. Therefore, further evaluation of these algorithms with larger population based and surveillance samples using more systematic data collection procedures would also be helpful to examine the diagnostic validity and factor structure of the new ADI-R algorithms for toddlers and preschoolers. There were some differences that emerged among the CPEA/STAART, NIMH, and Michigan samples. For example, for CPEA/STAART, there was a little difference in sensitivity between AUT and ASD cases, whereas the Michigan sample had a big drop in sensitivity from AUT to ASD (with the NIMH is in the middle). When the score distributions of these three samples were examined, the mean domain scores and totals were fairly comparable across these groups. Therefore, the changes may be due to site differences in using different diagnostic categories within ASD (e.g., Lord et al., 2012). In addition, diagnostic validity of the measure is likely influenced by reliability of administration across sites and studies. Each site was associated with an ADI-R and ADOS administrator who had originally achieved reliability with ADI-R and ADOS trainers (Lord et al., 1995), but the degree to which reliability was maintained within sites was not documented from these archived datasets.

Conclusion

The new ADI-R algorithms can be appropriately applied to existing research databases with toddlers and young preschoolers from 12 to 47 months down to a nonverbal mental age of 10 months. Within the more developmentally appropriate frameworks provided by the new algorithms, researchers can address questions about etiology and early signs of autism more effectively using existing databases. Given that the new algorithms are shorter than the existing algorithm and seem to better reflect both clinical judgment, at least in the U.S., and proposed revisions to DSM criteria that combine social and communication domains into one domain and retains a domain for restricted interests and repetitive behaviors, it may be time to consider a more efficient ADI-R for the use of the instrument across lifespan as well as for children under age 4. This would need to be empirically tested. Simply using the algorithm items in isolation would not necessarily yield the same results because item scores would be based on different information than when all current questions are administered. The development of revised algorithms for children over 47 months of age is also underway. Providing both research and clinical cutoffs allows clinicians and researchers to use different thresholds for different purposes. The ranges of concern also offer the possibility of a less categorical, more dimensional approach to the early identification of children with ASD in both clinical and research settings. Finally, the ADI-R has not been designed to be used alone. Past studies have consistently shown that combining information from multiple sources of information such as clinician observations and cognitive and developmental testing, in addition to parent reports, result in more valid diagnostic classifications for children across all age groups (Risi et al., 2006). Thus, the ADI-R should be used in conjunction with other instruments in diagnosing ASD.

Acknowledgments

We gratefully acknowledge the help of Kaite Gotham and Alice Kau as well as the families that participated in this research. This study was supported by National Institute of Mental Health (NIMH) R01MH066469, NIMH R25MH067723, National Institute of Health Intramural Research Program, NIMH, National Institute on Deafness and Other Communication Disorders, National Institute of Neurological Disorders and Stroke, and Eunice Kennedy Shriver National Institute of Child Health and Human Development.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-IV-TR. 4. Washington, DC: Author; 1994.
- Bishop S, Guthrie W, Coffing M, Lord C. Convergent validity of the Mullen Scales of Early Learning and the Differential Ability Scales in children with autism spectrum disorders. *American Journal on Intellectual and Developmental Disabilities*. 2011; 116:5, 331–343.
- Browne MW, Cudeck R. Alternative ways of assessing model fit. *Testing structural equation models*. 1993; 154:136–162.
- Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, Donaldson A, Varley J. Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model. *Pediatrics*. 2010; 125(1):17–23.
- Elliott, CD. DAS administration and scoring manual. San Antonio, TX: Psychological Corporation; 1990.
- Frazier TW, Youngstrom EA, Kubu CS, Sinclair L, Rezai A. Exploratory and confirmatory factor analysis of the Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders*. 2008; 38(3):474–480.10.1007/s10803-007-0415-z [PubMed: 17619129]
- Gotham K, Risi S, Dawson G, Tager-Flusberg H, Joseph R, Carter A, et al. A replication of the Autism Diagnostic Observation Schedule (ADOS) revised algorithms. *Journal of American Academy of Adolescent Psychiatry*. 2009; 47(6):642–651.

- Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*. 2007; 37(4):613–627.10.1007/s10803-006-0280-1 [PubMed: 17180459]
- Kim SH, Lord C. New Autism Diagnostic Interview-Revised algorithms for toddlers and young preschoolers from 12 to 47months of age. *Journal of Autism and Developmental Disorders*. 2012; 42(1):82–93. [PubMed: 21384244]
- Lord C, Petkova E, Hus V, Gan W, Lu F, Martin DM, et al. A multisite study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry*. 2012; 69(3):306–13.
- Lord, C.; Rutter, M.; DiLavore, PC.; Risi, S. Autism Diagnostic Observation Schedule: Manual. Los Angeles: Western Psychological Services; 1999.
- Lord C, Storoschuk S, Rutter M, Pickles A. Using the ADI-R to diagnose autism in preschool children. *Infant Mental Health Journal*. 1993; 14(3):234–252.
- Lord, Catherine; Risi, S.; DiLavore, PS.; Shulman, C.; Thurm, A.; Pickles, A. Autism from 2 to 9 years of age. *Archives of General Psychiatry*. 2006; 63(6):694–701.10.1001/archpsyc.63.6.694 [PubMed: 16754843]
- Luyster R, Gotham K, Guthrie W, Coffing M, Petrak R, Pierce K, Bishop S, et al. 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*. 2009; 39(9):1305–1320.10.1007/s10803-009-0746-z [PubMed: 19415479]
- Mullen, EM. Mullen Scales of Early Learning: AGS edition. Circle Pines, MN: American Guidance Service; 1995.
- Muthén, LK.; Muthén, BO. M-plus User's Guide, Version 5. Los Angeles, CA: Muthen and Muthen; 2007.
- Ozonoff S, Cook I, Coon H, Dawson G, Joseph RM, Klin A, McMahon WM, Minshew NJ, Munson JA, Pennington BF, Rogers SJ, Spence MA, Tager-Flusberg H, Volkmar FR, Wrathall D. Performance on Cambridge Neuropsychological Test Automated Battery Subtests sensitive to frontal lobe function in people with autistic disorder: Evidence from the Collaborative Programs of Excellence in Autism Network. *Journal of Autism and Developmental Disorders*. 2004; 34(2): 139–150. [PubMed: 15162933]
- Risi, Susan; Lord, C.; Gotham, K.; Corsello, C.; Chrysler, C.; Szatmari, P.; Cook, EH., Jr, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2006; 45(9):1094–1103.10.1097/01.chi.0000227880.42780.0e [PubMed: 16926617]
- Rutter, M.; Le Couteur, A.; Lord, C. Autism diagnostic interview, revised. Los Angeles: Western Psych Services; 2003.
- Shumway S, Thurm A, Swedo SE, Deprey L, Barnett LA, Amaral DG, Rogers SJ, Ozonoff S. Brief report: Symptom onset patterns and functional outcomes in young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2011; 41(12):1727–1732. [PubMed: 21360021]
- Shumway, S.; Farmer, C.; Thurm, A.; Joseph, L.; Black, D.; Golden, C. The ADOS calibrated severity score: Relationship to phenotypic variables and stability over time. *Autism Research*. May 24. (in press)(epub ahead of print), retrieved July, 6, 2012, from <http://www.ncbi.nlm.nih.gov/pubmed/22628087>
- Siegel B, Vukicevic J, Elliott GR, Kraemer HC. The use of signal detection theory to assess DSM-III-R criteria for autistic disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1989; 28(4):542–548. [PubMed: 2768150]
- Snow AV, Lecavalier L, Houts C. The structure of the Autism Diagnostic Interview-Revised: diagnostic and phenotypic implications. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*. 2009; 50(6):734–742.10.1111/j.1469-7610.2008.02018.x
- Sutera S, Pandey J, Esser EL, Rosenthal MA, Wilson LB, Barton M, et al. Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2007; 3(1):98–107. [PubMed: 17206522]
- van Lang NDJ, Boomsma A, Sytema S, de Bildt AA, Kraijer DW, Ketelaars C, Minderaa RB. Structural equation analysis of a hypothesised symptom model in the autism spectrum. *Journal of*

Child Psychology and Psychiatry, and Allied Disciplines. 2006; 47(1):37–44.10.1111/j.1469-7610.2005.01434.x

Ventola PE, Kleinman J, Pandey J, Barton M, Allen S, Green J, Robins D, Fein D. Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders*. 2006; 36(7):839–47. [PubMed: 16897398]

Wechsler, D. Wechsler Preschool and Primary Scale of Intelligence. New York: The Psychological Corporation; 1989. (Rev. ed.)

Wiggins LD, Robins DL. Excluding the ADI-R behavioral domain improves diagnostic agreement in toddlers. *Journal of Autism and Developmental Disorders*. 2008; 38(5):972–976. [PubMed: 17879150]

World Health Organization. The ICD-IO classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines. Geneva: Author; 1992.

Collaborative Programs for Excellence in Autism/Studies to Advance Autism Research and Treatment Sample Description

Table 1

Diagnosis	12-20/NV21-47			SW21-47			PH21-47		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
ASD	188			207			131		
Age (months)		32.7 ^b	7.8		33.4	6.7		39.0 ^b	5.8
NVIQ		65.7 ^b	21.5		73.4	17.3		83.6 ^b	20.2
ADI SA/SC		10.6 ^{ab}	3.0		9.9 ^{ab}	3.4		11.1 ^{ab}	4.2
ADI RRBs		5.0 ^{ab}	2.1		6.3 ^{ab}	3.0		6.3 ^{ab}	2.8
ADI IPG/RPI		10.9 ^{ab}	1.8		7.7 ^{ab}	2.0		3.3 ^{ab}	1.5
Non-spectrum Disorders (NS)	16			36			18		
Age		36.0 ^c	8.6		34.3	7.0		39.7 ^c	6.1
NVIQ		58.9 ^c	12.8		69.9	14.5		74.7 ^c	20.8
ADI SA/SC		2.6 ^a	2.9		2.1 ^a	2.6		3.7 ^a	2.4
ADI RRBs		2.3 ^a	2.4		2.2 ^a	2.5		1.7 ^a	1.8
ADI IPG/RPI		5.0 ^a	3.7		3.2 ^a	2.5		1.3 ^a	1.4
Typical Development (TD)	19			8			18		
Age		21.5 ^{bc}	8.5		29.8	9.7		31.4 ^{bc}	8.2
NVIQ		97.8 ^{bc}	30.8		88.1	28.5		105.1 ^{bc}	23.6
ADI SA/SC		4.7 ^b	6.2		5.4 ^b	6.5		5.1 ^b	5.7
ADI RRBs		2.0 ^b	2.5		2.1 ^b	1.7		3.2 ^b	3.2
ADI IPG/RPI		5.7 ^b	4.7		4.1 ^b	3.8		1.4 ^b	1.6

SA/SC Social Affect total for the "12-20/NV21-47" and "SW21-47" algorithms/Social Communication total for the "PH21-47" algorithm; RRBs Restricted and Repetitive Behaviors total; IPG/RPI Initiation, Play and Gestures total for the "12-20/NV21-47" and "SW21-47" algorithms/Reciprocal and Peer Interaction total for the "PH21-47" algorithm; 12-20/NV21-47 All children from 12-20 months and nonverbal children from 21-47 months; SW21-47 Children from 21-47 months with single words; PH21-47 Children from 21-47 months with phrase speech. Significant group differences marked at $p < 0.01$ with Bonferroni corrections.

^a ASD > NS;

^b ASD > TD;

NS>TD.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

New ADL-R Algorithm Factor Loadings in Collaborative Programs for Excellence in Autism/Studies to Advance Autism Research and Treatment (CPEA/STAART) and National Institute of Mental Health (NIMH) Datasets

12-20/NV21-47		Factor Loadings		SW21-47		Factor Loadings		PH21-47		Factor Loadings	
<u>Social Affect</u>		CPEA	NIMH	<u>Social Affect</u>		CPEA	NIMH	<u>Social Communication</u>		CPEA	NIMH
C. Attention to Voice *		0.67	0.58	C. Attention to Voice *		0.64	0.73	C. Attention to Voice *		N/A	0.62
C. Direct Gaze *		0.71	0.61	C. Direct Gaze *		0.62	0.70	C. Direct Gaze *		0.61	0.76
C. Social Smiling †		N/A	0.68	C. Social Smiling †		N/A	0.58	C. Nodding to mean yes		0.42	0.53
C. Seeking to Share Enjoyment *		0.70	0.59	C. Seeking to Share Enjoyment *		0.51	0.66	C. Seeking to Share Enjoyment *		0.62	0.55
C. Range of Facial Expression *		0.68	0.55	C. Range of Facial Expression *		0.66	0.64	C. Range of Facial Expression *		0.58	0.54
C. Inappropriate Facial Expression †		0.49	0.45	C. Inappropriate Facial Expression †		0.53	0.35	C. Offers Comfort		0.63	0.59
C. Appropriateness of Social Response *		0.69	0.57	C. Appropriateness of Social Response *		0.71	0.74	C. Pointing *		0.52	0.63
C. Interest in Children *		0.73	0.78	C. Interest in Children *		0.35	0.64	C. Directing Attention *		0.67	0.74
C. Response to Approaches of Children *		0.75	0.72	C. Response to Approaches of Children *		0.69	0.71	C. Quality of Social Overtures †		0.65	0.70
				C. Quality of Social Overtures †		0.71	0.63	C. Social Chat		0.66	0.45
								C. Use of Other's Body		0.22	0.20
<u>Repetitive & Restricted Behaviors</u>		CPEA	NIMH	<u>Repetitive & Restricted Behaviors</u>		CPEA	NIMH	<u>Repetitive & Restricted Behaviors</u>		CPEA	NIMH
E. Repetitive Use of Objects *		0.75	0.67	E. Repetitive Use of Objects *		0.79	0.81	C. Stereotyped Language *		0.50	0.50
E. Hand Mannerisms *		0.45	0.58	E. Hand Mannerisms *		0.77	0.55	E. Hand Mannerisms *		0.54	0.72
E. Complex Mannerisms *		0.51	0.65	E. Complex Mannerisms *		0.74	0.51	E. Complex Mannerisms *		0.59	0.71
E. Unusual Sensory Interests *		0.57	0.59	E. Unusual Sensory Interests *		0.71	0.54	E. Unusual Sensory Interests *		0.67	0.61
				E. Unusual Preoccupations †		0.41	0.24	E. Unusual Preoccupations †		0.37	0.39
				E. Compulsions/Rituals †		0.49	0.29	E. Compulsions/Rituals †		0.38	0.55
<u>Imitation, Gestures & Play</u>		CPEA	NIMH	<u>Imitation, Gestures & Play</u>		CPEA	NIMH	<u>Reciprocal and Peer Interaction</u>		CPEA	NIMH
C. Pointing *		0.76	0.61	C. Pointing *		0.47	0.66	C. Appropriateness of Social Response *		0.52	0.72
C. Gestures †		0.80	0.78	C. Gestures †		0.67	0.82	C. Interest in Children *		0.68	0.84
C. Imitation of Actions †		0.73	0.75	C. Imitation of Actions †		0.34	0.52	C. Response to Approaches of Children *		0.76	0.84

12-20/NV21-47	Factor Loadings	SW21-47	Factor Loadings	PH21-47	Factor Loadings
<i>C. Offering to Share</i> [†]	0.75	0.69	<i>C. Offering to Share</i> [†]	0.55	0.58
<i>C. Imaginative Play</i> [†]	0.62	0.42	<i>C. Imaginative Play</i> [†]	0.70	0.50
<i>C. Directing Attention</i> [*]	0.89	0.72			
	CFI		CFI		CFI
	0.948	0.852	0.908	0.892	0.913
					0.806
	RMSEA		RMSEA		RMSEA
	0.057	0.084	0.077	0.066	0.053
					0.093

^{*} Items that overlap across all three developmental cells;

[†] Items that overlap across two cells.

12-20/NV21-47 All children from 12-20 months and nonverbal children from 21-47 months; SW21-47 Children from 21-47 months with single words; PH21-47 Children from 21-47 months with phrase speech; Factors that are *not* included in the algorithm cutoffs are *italicized*. CCurrent; EEver; CFIComparative Fit Index; RMSEARoot Mean Square Error Approximation.

Sensitivity and Specificity for the New ADI-R Algorithms in Collaborative Programs for Excellence in Autism/Studies to Advance Autism Research and Treatment (CPEA/STAART) and National Institute of Mental Health (NIMH) Datasets

Table 3

		Michigan				CPEA/STAART				NIMH			
		sensitivity		specificity		sensitivity		specificity		sensitivity		specificity	
		AUT	ASD	NS	ASD	AUT	ASD	NS	ASD	AUT	ASD	NS	ASD
12-20/NV21-47	Research Cutoff=13	84	77	85	77	76	94	88	85	64			
	Clinical Cutoff=11	91	85	70	86	85	94	94	90	64			
SW21-47	Research Cutoff=13	80	71	90	73	72	92	97	89	89			
	Clinical Cutoff=8	99	94	81	96	96	83	100	97	58			
PH21-47	Research Cutoff=16	84	70	82	86	80	94	69	67	86			
	Clinical Cutoff=13	93	80	70	93	89	94	92	89	76			

12-20/NV21-47 All children from 12-20 months and nonverbal children from 21-47 months; SW21-47 Children from 21-47 months with single words; PH21-47 Children from 21-47 months with phrase speech

Table 4

National Institute of Mental Health Sample Description

Diagnosis	12-20/NV21-47			SW21-47			PH21-47		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
ASD	40			37			27		
Age (months)		33.1	1.1		36.6	1.0		41.7	1.3
NVIQ		63.8 ^b	9.1		70.4	12.0		78.4 ^b	14.2
ADI SA/SC		11.7 ^a	0.5		11.2 ^a	0.6		10.8 ^{ab}	0.6
ADI RRBs		5.4	0.3		6.1 ^a	0.4		5.5 ^{ab}	0.6
ADI IPG/RPI		11.0 ^{ab}	1.5		8.0 ^a	1.7		3.4 ^{ab}	0.3
Non-spectrum Disorders (NS)	12			19			21		
Age		31.4	2.6		36.6	1.6		41.1	1.1
NVIQ		66.5	26.4		75.7	16.3		81.8 ^c	14.8
ADI SA/SC		6.8 ^a	1.3		4.5 ^a	0.7		5.0 ^a	0.7
ADI RRBs		3.7	0.8		2.5 ^a	0.5		2.8 ^a	0.5
ADI IPG/RPI		8.3 ^a	3.5		3.8 ^a	2.4		1.5 ^a	0.3
Typical Development (TD)	2						8		
Age		19.5	4.0					36.7	2.4
NVIQ		106.4 ^b	9.1					108.3 ^{bc}	9.1
ADI SA/SC		6.0	2.0					0.5 ^b	0.3
ADI RRBs		3.0	2.0					0.4 ^b	0.2
ADI IPG/RPI		6.5 ^b	7.7					0.3 ^b	0.2

SA/SC Social Affect total for the "12-20/NV21-47" and "SW21-47" algorithms/Social Communication total for the "PH21-47" algorithm; RRBs Restricted and Repetitive Behaviors total; IPG/RPI Imitation, Play and Gestures total for the "12-20/NV21-47" and "SW21-47" algorithms/Reciprocal and Peer Interaction total for the "PH21-47" algorithm; 12-20/NV21-47 All children from 12-20 months and nonverbal children from 21-47 months; SW21-47 Children from 21-47 months with single words; PH21-47 Children from 21-47 months with phrase speech. Significant group differences are marked at $p < 0.05$ level with Bonferroni corrections.

^a ASD > NS,

^b ASD > TD,

^c NS > TD