Published in final edited form as:
J Am Acad Child Adolesc Psychiatry. 2008 June ; 47(6): 642–651. doi:10.1097/CHI.0b013e31816bffb7.

A Replication of the Autism Diagnostic Observation Schedule (ADOS) Revised Algorithms

Katherine Gotham, M.A., **Susan Risi, Ph.D.**, **Geraldine Dawson, Ph.D.**, **Helen Tager-Flusberg, Ph.D.**, **Robert Joseph, Ph.D.**, **Alice Carter, Ph.D.**, **Susan Hepburn, Ph.D.**, **William Mcmahon, M.D.**, **Patricia Rodier, Ph.D.**, **Susan L. Hyman, M.D.**, **Marian Sigman, Ph.D.**, **Sally Rogers, Ph.D.**, **Rebecca Landa, Ph.D.**, **M. Anne Spence, Ph.D.**, **Kathryn Osann, Ph.D.**, **Pamela Flodman, M.S.C.**, **Fred Volkmar, M.D.**, **Eric Hollander, M.D.**, **Joseph Buxbaum, Ph.D.**, **Andrew Pickles, Ph.D.**, and **Catherine Lord, Ph.D.**
Ms. Gotham and Drs. Risi and Lord are with the University of Michigan Autism and Communication Disorders Center; Dr. Dawson is with the University of Washington; Drs. Tager-Flusberg and Joseph are with Boston University School of Medicine; Dr. Carter is with the University of Massachusetts; Dr. Hepburn is with University of Colorado Health Sciences Center; Dr. McMahon is with the University of Utah; Drs. Rodier and Hyman are with University of Rochester Medical Center; Dr. Sigman is with University of California, Los Angeles; Dr. Rogers is with the University of California, Davis M.I.N.D. Institute; Dr. Landa is with Kennedy Krieger Institute; Drs. Spence and Osann and Ms. Flodman are with University of California, Irvine; Dr. Volkmar is with the Yale Child Study Center; Drs. Hollander and Buxbaum are with the Mount Sinai School of Medicine; and Dr. Pickles is with the University of Manchester.

Abstract

Objective—To replicate the factor structure and predictive validity of revised Autism Diagnostic Observation Schedule algorithms in an independent dataset (N = 1,282).

Method—Algorithm revisions were replicated using data from children ages 18 months to 16 years collected at 11 North American sites participating in the Collaborative Programs for Excellence in Autism and the Studies to Advance Autism Research and Treatment.

Results—Sensitivities and specificities approximated or exceeded those of the old algorithms except for young children with phrase speech and a clinical diagnosis of pervasive developmental disorders not otherwise specified.

Conclusions—Revised algorithms increase comparability between modules and improve the predictive validity of the Autism Diagnostic Observation Schedule for autism cases compared to the original algorithms.

Keywords

autism; pervasive developmental disorders not otherwise specified; Autism Diagnostic Observation Schedule; diagnosis

Correspondence to Katherine Gotham, UMACC, 1111 East Catherine Street, Ann Arbor, MI 48109-2054; kog@umich.edu.

Disclosure: Drs. Lord and Risi receive royalties for the ADOS. Prof. Pickles receives royalties from the SCQ and ADOS-G instruments. Dr. Carter receives royalties from Harcourt Assessment for the ITSEA/BITSEA. The other authors report no conflicts of interest.

In their 2007 article, Gotham et al.1 proposed revised algorithms intended to improve predictive validity of the Autism Diagnostic Observation Schedule (ADOS)[2] modules used with children (modules 1–3). Similar domain distributions in the original ADOS norming sample2 and the larger, more diverse 2007 sample (hereafter referred to as Michigan 2007, $N = 1,630$) suggested that new algorithms derived from Michigan 2007 data may be appropriately applied to existing research databases. The aim of this study was to replicate the revised ADOS algorithm findings in an independent dataset provided by National Institutes of Health (NIH)–funded consortia, the Collaborative Programs for Excellence in Autism (CPEA) and Studies to Advance Autism Research and Treatment (STAART). Particular attention was paid to the factor structure and predictive validity of the revised algorithms in this large independent dataset.

The ADOS is a semistructured, standardized assessment designed for use with individuals referred for possible autism spectrum disorders (ASDs). Four ADOS modules accommodate various developmental and language levels. In each, a protocol of activities or social presses is administered in approximately 45 minutes, and then items are scored on a 4-point scale, with 0 indicating “no abnormality of type specified” and 3 indicating “moderate to severe abnormality.” To receive an ADOS classification of autism or ASD, an individual’s scores on the original diagnostic algorithms must meet separate cutoffs in the Communication and Social domains, and a summation of the two. If any or all of these thresholds are not met, then a nonspectrum classification is assigned. Item scores of 2 and 3 are collapsed in the algorithms to reduce the impact of individual items.

ADOS algorithm revisions were prompted by questions of effects of impairment level on current totals. Gotham and colleagues[1] noted that module 1 totals in the Michigan 2007 sample exhibited a restricted range due to scoring communication items in nonverbal children. Joseph and colleagues[3] reported correlations between ADOS social domain totals and level of cognitive impairment for preschool children. De Bildt and colleagues[4] found that ADOS classifications appeared to be least valid for children with mild, compared to moderate or profound, mental retardation. Thus, algorithm revisions were undertaken to improve sensitivity and specificity while possibly reducing age and IQ effects of the ADOS.

Another goal of the Michigan 2007 revisions was to modify the existing ADOS domain structure of distinct domains and cutoffs for Social and Communication items, based on several studies that found a single factor best described social and communication domain items.[5–7] In response to findings that observation of repetitive behaviors may make an independent contribution to diagnostic stability,[8] restricted, repetitive behavior (RRB) items were included in the total to which classification thresholds are applied. Finally, algorithm revisions were intended to increase comparability across modules by creating algorithms with a fixed number of items of similar conceptual content.[1]

Revised algorithms originally were created by dividing the Michigan 2007 sample by age and language level within modules to yield five developmental cells.[1] These cells reduced the strength of association between ADOS totals, age, and verbal IQ. Module 1 was divided into “some words” and “no words” on the basis of single words used within the administration (item A1); this reduced ceiling effects in module 1 Communication totals. Module 2 was separated into children younger than 5, and those ages 5 and older to reduce the difference between younger, more rapidly developing children and older children. Module 3 represented a distinct developmental cell. Each item distribution was examined by cell, and a pool of preferred items that maximized differentiation between clinical diagnoses was generated. These items were organized into domains based on multifactor item response analysis, and the sensitivity and specificity of the new algorithms were compared to the existing model. For revised algorithm item composition and thresholds, see Table A in the

supplementary material on the *Journal's* Web site (www.jaacap.com) via the Article Plus feature.

In the Michigan 2007 sample, the revised algorithms increased specificity particularly in classifying nonautism ASDs in lower functioning populations and generally maintained the high predictive validity of the ADOS.[1] The Social and Communication domains of the previous algorithms were merged into a Social Affect (SA) domain to increase construct validity. RRB items included toward algorithm cutoffs were found to aid in distinguishing pervasive developmental disorders not otherwise specified (PDD-NOS, or nonautism ASDs) from nonspectrum cases. Items with similar or identical content were selected from each developmental cell to allow for easier comparison of ADOS scores within and between individuals, setting the stage for future efforts to adapt the ADOS for use as a severity measure in ASDs.

Replication with a large independent dataset is crucial before the new algorithms are widely used by researchers and clinicians. The 2007 authors noted that, although the revisions improved on the existing models in classifying PDD-NOS, sensitivity in this group continued to be lower than desired.[1] The present study aims not only to replicate the psychometric properties of the new algorithms but also to generate more data on the diagnosis of nonautism ASDs within the field.[9]

METHOD

Participants

Analyses were conducted on data provided by the CPEA, a network of 10 sites funded by the National Institute of Child Health and Human Development and the National Institute of Deafness and Other Communication Disorders, and the STAART program, an NIH-funded network of eight research centers (some of which overlap with CPEA sites) throughout the United States and Canada. This dataset represents 1,259 different participants from 11 different sites, excluding children from Michigan (who were included in the previous article1). In the Michigan 2007 sample, analyses were unchanged by inclusion of repeat assessment data, therefore 23 participants with assessments at two different time points were included in this replication sample, yielding a total of 1,282 cases (a case is defined by a contemporaneous ADOS, verbal IQ, and best estimate clinical diagnosis). As in the Michigan 2007 sample, these participants were clinic referrals or research participants. They received diagnostic evaluations at the University of Washington (n = 472), Boston University School of Medicine (n = 316), University of Colorado Health System (n = 85), University of Utah (n = 79), University of Rochester (n = 78), University of California, Los Angeles (n = 59), University of California, Davis (n = 52), Kennedy Krieger Institute (n = 50), University of California, Irvine (n = 47), Yale University (n = 30), and Mount Sinai Medical Center (n = 14).

The sample was limited to participants ages 12 years or younger for modules 1 and 2 and 16 and younger for module 3, resulting in an age range of 18 months to 16 years. Because older adolescents and adults were thought to merit individual study, ADOS module 4 recipients were excluded from both the Michigan 2007 sample and the present sample.

The final dataset included 970 cases with clinical diagnoses of autism (76%), 98 with a nonautism ASD (7%), and 214 with non-ASD developmental delays (17%). Within the nonspectrum sample of 214 cases, 90 children had nonspecific mental retardation, 64 had language disorders, 16 had fragile X syndrome, 6 were developmentally delayed family members of probands, and 38 had unspecified developmental disorders. Seventy-two percent of the sample was male. The racial/ethnic makeup was 3% African American, 3%

Asian American, 1% Native American, 7% multiracial, 84% white, and 2% other races, with 3% of the sample identified as Hispanic. Table 1 provides a detailed sample description (for additional information, see Table B in the supplementary material on the *Journal's* Web site (www.jaacap.com) via the Article Plus feature.

Measures and Procedure

The most common research protocol across CPEA/STAART sites was the administration of the ADI-R[10] to a parent or caregiver, followed by a child assessment including the ADOS and psychometric testing. A clinical diagnosis then was made by a psychologist and/or psychiatrist after review of all of the available data. Eighty-one participants were recruited from a study in which eligibility was dependent on meeting ADOS criteria. These cases were excluded from analyses of the predictive value of the ADOS but retained for analyses of the factor structure of the measure. The ADOS was administered by a clinical psychologist or trainee who met standard requirements for research reliability.[5] One site used the Pre-Linguistic ADOS,[11] for which identical items were recoded to module 1 scores. A developmental hierarchy of psychometric measures, most frequently the Mullen Scales of Early Learning[12] and the WISC,[13] determined IQ scores. The ADI-R was available for 1,063 cases. This research was approved by the institutional review boards at the respective universities and the University of Michigan.

Design and Analysis

The sample first was divided by age and language level within each module to yield the five developmental cells outlined in the 2007 article[1] (module 1, fewer than five words cell; module 1, five or more words cell; module 2, younger than 5 years cell; module 2, 5 years or older cell; and module 3). Domain totals and diagnostic classification were generated for each case by adding the new algorithm item scores appropriate to the developmental cell of the participant and applying the revised threshold cutoffs.

For statistical analyses, ADOS item scores of 3 were recoded to 2 as they are on the algorithms. Exploratory multifactor item response analysis was performed to compare the factor structure of revised algorithm items by cell to those of the Michigan 2007 sample. Receiver operating characteristic (ROC) curves[14] were calculated, and the sensitivity and specificity of the existing and revised ADOS algorithms were contrasted by developmental cells within the replication dataset and compared to the revised algorithms in the Michigan 2007 sample.

RESULTS

Comparability of Revision and Replication Samples

The Michigan 2007 sample included more data from children with clinical diagnoses of PDD-NOS than did the CPEA/STAART dataset for all developmental cells (Michigan 2007 $N = 439$; CPEA/STAART $N = 98$). In the 2007 sample, the majority of children with nonspectrum diagnoses had been specifically recruited from populations with Down syndrome, fetal alcohol syndrome, and non-ASD language delays to provide a control group against which to assess the predictive validity of the ADOS and ADI-R. In contrast, many CPEA/STAART nonspectrum cases were initial ASD referrals who did not meet criteria. The patterns of impairments seen in these children pose different measurement challenges, especially concerning specificity, than those from the purposefully recruited control groups.

Another salient difference between samples was the chronological age and Verbal IQ of specific cells. In the module 1, no words autism cell, the verbal IQ of the CPEA/STAART sample (mean 36.4, SD 17.3) was significantly higher on average ($t[579] = -9.3$, $p < .01$),

and the mean chronological age younger (mean 3.5 years, SD 2.0 years; $t[660.2] = 4.9$, $p < .01$), than the Michigan 2007 sample (Verbal IQ mean 24.6, SD 14.8; age mean 4.3 years; SD 2.3 years). In module 3, the CPEA/STAART sample had mean chronological ages 12 to 17 months younger than the Michigan 2007 sample for all diagnostic groups (Michigan 2007 mean 8.4 years, SD 2.5 years; CPEA/STAART mean 9.8 years; SD 2.6 years, $t[853] = -7.7$, $p < .001$).

Division by Developmental Cells

Data were configured into the developmental cells described above. Because of its greatly limited distribution across diagnostic groups (nonautism ASD, $n = 9$; nonspectrum, $N = 8$), the module 2, older cell was excluded from analyses of factor structure and sensitivity and specificity. ROC curve results are reported separately for children with a nonverbal mental age (NVMA) of 15 months or lower, as was done in the Michigan 2007 study to examine the specificity of the measure in extremely low functioning populations. Insufficient data also precluded the ROC analysis of low-NVMA module 1, no words comparison groups with nonspectrum diagnoses ($n = 5$) and PDD-NOS ($n = 0$), as well a higher NVMA module 1, no words PDD-NOS group ($n = 6$).

Correlations With Participant Characteristics

Correlations between domain totals and participant characteristics were examined for the ASD sample to identify relationships between ADOS scores and chronological age and Verbal or Nonverbal IQ. These correlations were minimal ($r < .30$), with the exception of SA domain and Verbal IQ for module 1, no words group ($r[251] = -0.51$) and module 2, older group ($r[109] = -0.43$).

Factor Analysis

In a replication of Gotham et al.[1] methods, exploratory factor analyses for categorical data (Mplus software version 3.0)[15] was run for the 14 revised algorithm items in each developmental cell, using the ULS estimator and promax rotation.

In the Michigan 2007 analyses, a two-factor solution fitted well, with items loading onto clear SA and RRB factors that were positively correlated (Table 2 in Gotham et al.1). Confirmatory factor analysis of the Michigan 2007 sample showed the two-factor model to fit substantially better than the one-factor model. When a third factor was allowed, a joint attention factor composed of pointing (module 1, some words cell; module 2, younger cell; module 2, older cell) or response to joint attention (module 1, no words cell), as well as gesturing, showing, initiation of joint attention, and unusual eye contact items emerged in children without verbal fluency. The two-factor model (SA and RRB) was chosen for classification purposes due to its greater consistency across the five cells.

A root mean square error approximation (RMSEA) of ≤0.08 is considered a satisfactory fit in exploratory factor analysis.[16] Under this criterion, the two-factor model replicated satisfactorily in all CPEA/STAART developmental cells, with RMSEA values ranging from 0.05 in the module 1, no words cell, to 0.08 in the module 2, younger cell. Correlations between the two-factor–based domains ranged from 0.34 to 0.57 by cell. See Table 2 for eigenvalues and factor loadings under a two-factor solution. Complete two- and three-factor solution item loading information from these analyses can be found in Tables C and D in the supplementary material on the *Journal's* Web site (www.jaacap.com) via the Article Plus feature.

Of note was that in the module 2, younger cell, most items assigned to the RRB factor by Gotham et al.1 did not load onto this factor (i.e., loadings were <0.40). Rather, the second

factor was composed of pointing and initiation of joint attention items in this cell, recalling the third factor noted previously.[1] Under a three-factor model, the expected RRB items did load together, along with a clear SA factor and an approximate joint attention factor. The module 2, younger cell had a low subject-to-item ratio (1:6.3), and communalities (the percentage of variance in a given item explained by all of the factors) were <0.50 for six of the 14 items analyzed, indicating an underpowered analysis for this developmental group.17

Exploratory factor analysis was rerun by cell for ASD subjects only with results similar to the all-diagnoses-combined analyses described above. Two-factor RMSEAs ranged from 0.06 (module 1, no words cell) to 0.10 (module 2, younger cell). Across the ASD sample, the SA and RRBs domains were not highly correlated (0.12 to 0.35 by cell).

Sensitivity and Specificity

Predictive validity was assessed with ROC curves to obtain the sensitivity and specificity of both the old and the new algorithms by cell. When a diagnostic group included fewer than 15 cases, that group and its comparison cases were dropped from the analysis. Cases included in a specific study sample contingent on meeting ADOS criteria also were removed. In Table 3, sensitivity and specificity are listed by diagnostic group and developmental cell first for the original ADOS algorithm in the CPEA/STAART dataset, then for the revised algorithm, and finally from the revised algorithm applied to the Michigan 2007 sample.

Specificity remained relatively stable using the old and new algorithms for autism and nonautism ASDs. For autism versus nonspectrum, the revised algorithms showed approximately equivalent sensitivity in module 1, no words cell, and improved sensitivity in every other developmental cell (from a 9% increase in module 2, younger cell, to a 16% increase in module 1, some words cell) compared to the original algorithms.[2] Although small sample size precluded formation of comparison groups (and thus inclusion in Table 3) for the following groups, the 41 module 1, no words autism cases with nonverbal mental age younger than 16 months had stable sensitivity of 95% in the original and revised algorithms, and sensitivity improved across the 100 module 2, older autism cases from 85% under the original algorithm to 95% with the revised algorithm.

For nonautism ASD versus nonspectrum, sensitivity remained approximately equivalent in the module 1, some words group, increased by 11% in the module 3 group, and dropped by 23% in the module 2, younger group compared to the earlier algorithm.

Sensitivity and specificity of the SA domain on its own are given in parentheses below the two-domain results in Table 3. Overall, the first factor by itself tended to perform less well than the two-domain model, as found in the Michigan 2007 sample. This was not true in the module 2, younger PDD-NOS cell, in which the SA factor alone was markedly superior.

Logistic Regression to Compare Samples on Predictors of Clinical Diagnosis

Logistic regressions had indicated that both the SA and RRBs domains made significant independent contributions to the prediction of autism and PDD-NOS diagnoses in the Michigan 2007 sample. The present analyses were run entering age and Verbal IQ, developmental cell, ADI-R domains, and ADOS domains as predictors of best estimate clinical diagnoses. For autism versus nonspectrum children, the ADOS SA domain was a consistent predictor of diagnosis beyond the ADI-R (odds ratio 1.46, confidence interval 1.27, 1.67; $p < .001$), and the ADOS RRBs domain was a significant predictor of diagnosis when ADI-R domains were excluded from the model (odds ratio 1.33, confidence interval 1.16–1.54; $p < .001$). Neither the SA nor RRB ADOS domains predicted PDD-NOS diagnoses when ADI-R domains were included in analyses. When ADI-R domains were

excluded, SA predicted PDD-NOS (odds ratio 1.42, confidence interval 1.29–1.57; $p < .001$). Developmental cell predicted autism versus nonspectrum diagnosis (likelihood ratio test statistic(4) = 31.22; $p < .001$) only when ADI-R domains were not included in the model. Developmental cell predicted PDD-NOS diagnoses in models excluding (likelihood ratio test statistic(4) = 48.21; $p < .001$) and including (likelihood ratio test statistic(4) = 22.01; $p < .001$) ADI-R domains.

Item and Domain Total Differences by Site in Low-Sensitivity Cells

To further investigate the decrease in sensitivity under the revised algorithm for PDD-NOS cases within the module 2, younger than 5 cell, new and old domain totals and item totals for datasets were compared. The new SA total was not significantly higher ($t[28.8] = -0.43$; $p = .67$) in the replication sample (mean 8.4, SD 4.1) than in the Michigan 2007 sample (mean 7.9, SD 4.2) for this developmental cell and diagnostic group; the RRB total was significantly lower (CPEA/STAART mean 1.3, SD 1.4, versus Michigan 2007 sample mean 3.4, SD 2.1; $t[42.7] = 4.7$; $p < .001$). The low sensitivity (65%) in this comparison was based on six children who did not meet new cutoffs but were diagnosed with PDD-NOS. Five of these six cases were from one site; they had low RRB scores (one score of 2 being the highest). Each missed the classification cutoff by 1 point only. When mean scores on RRB items were compared between datasets, no one item stood out as contributing more to the domain total discrepancy. The general pattern within cells was one of consistently lower RRB scores in the CPEA/STAART dataset than in the Michigan 2007 dataset. Domain scores on the RRBs domain of the ADI-R for this cell, however, were not significantly different between the samples (mean 4.3, SD 2.5, in the Michigan 2007 dataset; mean 4.0, SD 3.0, in CPEA/STAART; $t[23.9] = 0.33$, $p = .75$).

In the low-sensitivity module 3, PDD-NOS group, the replication dataset had lower mean scores in both domains than did the Michigan 2007 sample. Eighteen misclassified cases fell short of the cutoff by a range of 1 to 5 points. Inclusion of the RRB total in the revised algorithm thresholds less clearly contributed to misclassifications: 56% of the CPEA/STAART misclassified cases had RRB totals of 1 or 2 compared to 47% in 2007.

DISCUSSION

Recently proposed improvements to the algorithm[1] resulted in increased comparability across ADOS modules; now each algorithm includes 14 items of similar content. The revised algorithms also better represent observed diagnostic features of ASD in that social, communication, and RRBs contribute to both a measure classification and *DSM-IV* diagnosis of autism. Predictive value of the ADOS for autism cases generally increased under the revised algorithms in this large independent multisite sample. Sensitivity to classify PDD-NOS cases was improved in some subsamples (verbally fluent children) with the new algorithm, but decreased in another (children younger than 5 with phrase speech only), although this was based on a limited amount of data.

Because most of the CPEA/STAART nonspectrum sample represented children referred for possible ASD and siblings of probands with similar developmental impairments, specificity had been expected to be lower than the Michigan 2007 results, which included nonspectrum children specifically recruited as controls. In fact, specificity in the CPEA/STAART dataset was high in many of the developmental cells under both the existing and revised algorithms. Sensitivity was markedly improved by the revised algorithms for autism cases in this dataset, despite the fact that the diagnoses in most samples were influenced by the original ADOS criteria, and a drop in revised algorithm sensitivity therefore may have been expected. Children with PDD-NOS were not actively recruited at most of the sites, possibly

leading to a more idiosyncratic, less representative nonautism ASD sample with less noticeable improvements under these algorithms than those pertaining to autism cases.

The two exceptions to replicating the Michigan 2007 results in the CPEA/STAART dataset both involved the module 2, younger than 5 cell. Here, the joint attention factor reported by Gotham et al.[1] was evident in two- and three-factor models for this cell, with a three-factor solution fitting best. In addition, the young module 2 cell exhibited a marked decrease in sensitivity for nonautism ASD under the new algorithm. This anomalous cell was composed of just 17 PDD-NOS (14 from one site) and 18 nonspectrum cases. All of the other results, representing far greater amounts of data, replicated the Michigan 2007 findings.

Factor structure may be expected to vary across samples given the small sample size and low subjects-to-item ratio of module 2, younger cell. Because joint attention behaviors are especially salient for younger children, another possible explanation for the difference in factor structure between the samples could be the younger average age of ASD children in the module 2, younger cell of the CPEA/STAART dataset compared to the 2007 article.[1] Eventually, the joint attention factor may prove to be a developmentally useful descriptor for an even younger module 2 group.

In response to the sensitivity decrease in the module 2 younger PDD-NOS cell, we explored whether these cases were receiving lower scores overall or exhibiting fewer RRBs that now counted toward classification thresholds. ADOS total scores were distributed differently between the two domains in each dataset, with RRBs domain scores significantly lower in the replication sample. Five of the six misclassified cases in this cell came from one research site, and each missed the classification cutoff by 1 point. These cases showed no difference in mean ADI-R RRBs domain score from the equivalent Michigan 2007 data, suggesting that site differences in observation and scoring of these behaviors on the ADOS may have influenced the predictive validity reported here. Low RRBs totals no doubt also influenced the factor structure of the module 2, young cell observed in this replication. Because these items did not contribute to the original ADOS algorithm totals, it is possible that they were not scored as vigilantly as possible. A crucial factor in clinicians' ability to observe and score repetitive behaviors is the pace of the ADOS administration and the deliberate inclusion of some less structured time between tasks. Reliability in RRBs scoring needs to be emphasized more clearly in future training and manuals if it is a source of variation.

In the module 3, PDD-NOS group, sensitivity actually improved under the revised algorithm, but was still undesirably low (60%). The lower domain mean scores in the replication dataset indicate that two sites were diagnosing ASD in children with milder symptoms (ADI-R scores were not available to verify this). The relative similarity across datasets in distributions of module 3, PDD-NOS domain scores indicates that inclusion of the RRBs domain in the revised algorithms is not likely to be the primary explanation for low sensitivity. No pattern was apparent to explain the problematic sensitivity of this group, suggesting that the differences may lie in the clinical threshold for diagnosing PDD-NOS, which depends on often rather arbitrary interpretation of the *DSM-IV*[18] criteria. Such thresholds may have been affected by recruitment source (e.g., clinic referrals) or affected sibling status. Improvements are needed both in the *DSM-IV* criteria for nonautism ASD and in module 3 tasks and codes.

Results from the logistic regressions indicate that the ADOS adds to the validity of an autism diagnosis beyond the ADI-R, supporting earlier findings that data from both measures make independent contributions to diagnoses and predictions of diagnoses years later.[8] Moreover, the goal of reducing age and Verbal IQ effects on ADOS totals was largely achieved. The degree of correlation remaining between SA and Verbal IQ in the module 1,

no words cell and module 2, older cell supports the fact that cognitive impairment is correlated with degree of autistic impairment (not simply developmental level). For the nonverbal children, greater association between ADOS and cognitive scores was expected due to the role of social communication in measuring cognitive skills at this age and ability level.[1]

Limitations of this study include small sample sizes, which precluded analysis of algorithm performance for the module 1, no words PDD-NOS cell and module 2, older cell and contributed to underpowered factor analysis of the module 2, younger cell. There is a continued need for replication in these areas. Recruitment differences and possible treatment effects may have affected characteristics of children in specific cells. In addition, predictive validity of the measure is likely influenced by reliability of administration across sites. Each site was associated with an ADOS administrator that originally achieved reliability with central ADOS trainers,[5] but the degree to which reliability was maintained within sites was not known.

In summary, Gotham et al. 2007 revised ADOS algorithms better represent observed diagnostic features through new domains, increase comparability between modules in algorithm item content and number, and improve ADOS predictive validity for autism compared to previous algorithms. The ADOS, along with other diagnostic measures, ideally will continue to contribute to understanding and discussion of ASDs. This is best accomplished through data sharing to create large samples, as is reflected by the consortia efforts described herein.

Acknowledgments

This study was supported by NIMH RO1 MH066469 and NIMH R25MH067723, and CPEA/STAART grants from NIMH, NIDCD, NINDS, and NICHD.

We gratefully acknowledge the help of Shanping Qiu, Kathryn Larson, and Mary Yonkovit. We thank the families in all CPEA/STAART sites.

REFERENCES

1. Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule (ADOS): revised algorithms for improved diagnostic validity. J Autism Dev Disord 2007;37:400–408.
2. Lord C, Risi S, Lambrecht L, et al. The Autism Diagnostic Observation Schedule–Generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord 2000;30:205–223. [PubMed: 11055457]
3. Joseph RM, Tager-Flusberg H, Lord C. Cognitive profiles and social-communicative functioning in children with autism spectrum disorders. J Child Psychol Psychiatry 2002;43:807–821. [PubMed: 12236615]
4. de Bildt A, Sytema S, Ketelaars C, et al. 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. J Autism Dev Disord 2004;34:129–137. [PubMed: 15162932]
5. Lord, C.; Rutter, M.; DiLavore, P.; Risi, S. Autism Diagnostic Observation Schedule Manual. Los Angeles: Western Psychological Services; 1999.
6. Constantino JN, Gruber CP, Davis C, Hayes S, Passanante N, Przybeck T. The factor structure of autistic traits. J Child Psychol Psychiatry 2004;45:719–726. [PubMed: 15056304]
7. Lecavalier L, Aman MG, Scahill L, McDougle CJ, McCracken JT, Vitiello B. Validity of the autism Diagnostic Interview–Revised. Am J Ment Retard 2006;111:199–215. [PubMed: 16597187]
8. Lord C, Risi S, DiLavore R, Shulman C, Thurm A, Pickles A. Autism from two to nine. Arch Gen Psychiatry 2006;63:694–701. [PubMed: 16754843]

9. Walker DR, Thompson A, Zwaigenbaum L, Goldberg J, Bryson SE, Mahoney WJ. Specifying PDD-NOS: a comparison of PDD-NOS, Asperger syndrome, and autism. J Am Acad Child Adolesc Psychiatry 2004;43:172–180. [PubMed: 14726723]
10. Rutter, M.; LeCouteur, A.; Lord, C. Autism Diagnostic Interview–Revised–WPS. WPS ed.. Los Angeles: Western Psychological Services; 2003.
11. DiLavore P, Lord C, Rutter M. Pre-Linguistic Autism Diagnostic Observation Schedule (PL-ADOS). J Autism Dev Disord 1995;25:355–379. [PubMed: 7592249]
12. Mullen, E. Mullen Scales of Early Learning. AGS ed.. Circle Pines, MN: American Guidance Service; 1995.
13. Wechsler, D. Wechsler Intelligence Scale for Children. 4th ed.. San Antonio, TX: Psychological Corporation; 2003.
14. Siegel B, Vukicevic J, Elliott G, Kraemer H. The use of signal detection theory to assess *DSM-III-R* criteria for autistic disorder. J Am Acad Child Adolesc Psychiatry 1989;28:542–548. [PubMed: 2768150]
15. Muthen, LK.; Muthen, BO. M-plus User's Guide. Los Angeles: Muthen & Muthen; 1998.
16. Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. Testing Structural Equation Models. Newbury Park, CA: Sage Publications; 1993. p. 136-162.
17. MacCallum R, Widaman K, Preacher K, Hong S. Sample size in factor analysis: the role of model error. Multivariate Behav Res 2001;36:611–637.
18. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th ed (DSM-IV). Washington, DC: American Psychiatric Association; 1994.

TABLE 1

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

	Module 1, No Words			Module 1, Some Words			Module 2, Younger			Module 2, Older			Module 3		
Diagnosis	***n***	**Mean**	**SD**	***n***	**Mean**	**SD**	***n***	**Mean**	**SD**	***n***	**Mean**	**SD**	***n***	**Mean**	**SD**
Autism															
Age	295	41.8	24.3	183	47.7	25.4	53	44.2	9.2	100	91.0	24.8	339	118.2	32.0
VIQ	295	36.4	17.3	183	59.4	19.2	53	77.2	20.1	100	62.0	18.9	339	87.5	24.0
NVIQ	295	60.4	22.7	183	71.4	20.1	53	85.4	21.5	100	81.9	24.5	339	97.9	25.1
ADI social	240	20.3	5.0	169	17.7	5.7	50	15.7	5.0	97	22.8	5.4	307	21.0	5.2
ADI comm-V	18	17.8	3.6	59	16.0	4.0	47	14.8	4.1	89	17.9	4.0	304	16.9	4.7
ADI comm-NV	208	11.7	2.1	109	10.1	2.8	2	11.5	3.5	10	11.2	2.4	0	—	—
ADI-RR	240	5.0	2.1	169	5.2	2.3	50	6.3	2.3	97	6.9	2.6	307	6.5	2.7
ADOS social	295	11.3	2.5	183	9.5	2.5	53	9.2	2.7	100	10.1	2.9	339	9.0	2.9
ADOS comm	295	5.7	1.7	183	5.0	2.1	53	6.6	1.9	100	6.8	2.0	339	4.1	1.8
ADOS SA	295	16.1	3.6	183	13.8	3.6	53	12.4	3.9	100	13.3	4.2	339	11.4	4.3
ADOS RR	295	4.2	2.3	183	3.3	1.9	53	3.5	1.9	100	3.4	1.8	339	2.3	1.8
PDD-NOS															
Age	6	58.0	28.3	21	47.7	23.5	17	44.7	5.3	9	84.4	20.2	45	117.5	28.3
VIQ	6	40.3	15.7	21	59.0	21.2	17	85.2	16.1	9	65.8	23.3	45	101.0	24.4
NVIQ	6	59.2	20.5	21	70.1	25.4	17	85.9	13.3	9	106.6	25.6	45	105.1	16.7
ADI social	6	19.5	6.6	19	13.3	6.2	16	10.7	4.9	9	15.2	9.0	44	16.0	6.9
ADI comm-V	0	—	—	8	11.0	5.9	14	10.9	4.1	8	13.2	7.2	44	13.4	6.4
ADI comm-NV	4	10.5	4.0	9	9.0	2.7	2	9.5	0.7	1	12.0	—	0	—	—
ADI-RR	6	1.2	.7	18	5.2	2.9	16	4.0	3.0	9	3.4	1.9	44	5.4	2.8
ADOS social	6	10.3	2.4	21	6.5	1.7	17	6.2	2.9	9	8.2	2.7	45	5.3	2.9
ADOS comm	6	6.0	1.8	21	4.3	1.5	17	4.6	1.5	9	3.8	1.9	45	2.7	1.5
ADOS SA	6	15.2	2.9	21	10.3	2.9	17	8.4	4.1	9	9.7	2.5	45	6.6	3.8
ADOS RR	6	1.7	1.5	21	2.1	1.8	17	1.3	1.4	9	1.8	1.4	45	1.4	1.3
Nonspectrum															
Age	51	36.6	14.5	64	44.4	22.9	18	50.1	5.6	8	74.5	7.3	73	115.4	32.3
VIQ	51	48.7	17.8	64	66.9	17.6	18	76.1	14.8	8	76.6	18.7	73	98.5	21.2

J Am Acad Child Adolesc Psychiatry. Author manuscript; available in PMC 2011 March 15.

	Module 1, No Words			Module 1, Some Words			Module 2, Younger			Module 2, Older			Module 3		
Diagnosis	***n***	**Mean**	**SD**	***n***	**Mean**	**SD**	***n***	**Mean**	**SD**	***n***	**Mean**	**SD**	***n***	**Mean**	**SD**
NVIQ	51	62.8	19.2	64	71.4	16.8	18	84.0	20.2	8	85.5	24.2	73	100.8	20.9
ADI social	43	9.4	7.4	53	6.8	7.3	12	2.8	2.4	4	5.5	1.3	59	4.2	4.4
ADI comm-V	1	14.0	—	21	8.4	6.2	11	4.6	3.3	4	6.8	2.6	59	3.1	3.3
ADI comm-NV	38	6.8	4.4	27	3.7	3.7	1	1.0	—	0	—	—	0	—	—
ADI-RR	43	3.1	2.1	53	2.5	2.7	12	2.0	2.1	4	2.0	2.2	59	1.2	1.4
ADOS social	51	5.4	4.1	64	2.7	2.8	18	1.5	1.0	8	3.6	3.5	73	2.4	2.5
ADOS comm	51	3.1	2.3	64	1.5	1.8	18	2.2	1.2	8	3.0	2.5	73	1.4	1.4
ADOS SA	51	7.9	5.6	64	3.9	4.0	18	2.6	1.1	8	5.3	4.9	73	2.8	3.2
ADOS RR	51	2.2	2.1	64	1.5	1.7	18	0.3	0.5	8	1.1	1.6	73	0.2	0.7

Note: All ages in months. VIQ = Verbal IQ; NVIQ = Nonverbal IQ; ADI social = ADI-R Social total; ADI-R comm-V = Autism Diagnostic Interview-Revised Communication Total for verbal subjects; ADI-R comm-NV = ADI-R Communication total for nonverbal subjects; ADI-RR = ADI-R Restricted, Repetitive Behaviors total; ADOS social = Autism Diagnostic Observation Schedule Social total; ADOS comm = ADOS Communication total; ADOS SA = revised algorithm Social Affect domain; ADOS RR = revised algorithm Restricted, Repetitive Behavior domain; PDD-NOS = pervasive developmental disorder-not otherwise specified.

TABLE 2

Revised Algorithm Factor Loadings in Collaborative Programs for Excellence in Autism/Studies to Advance Autism Research and Treatment Dataset

Domains	Module 1, No Words, N = 352 (1:25.1)[a]	Factor Loadings	Module 1, Some Words, N = 268 (1:19.1)[a]	Factor Loadings	Module 2, Younger, N = 88 (1:6.3)[a]	Factor Loadings	Module 3, N = 457 (1:32.6)[a]	Factor Loadings
Social Affect	Unusual eye contact	0.76	Unusual eye contact	0.83	Unusual eye contact	0.73	Unusual eye contact	0.56
	Gaze and other behaviors	0.82	Gaze and other behaviors	0.81	Amount of social communication	0.73	Amount of social communication	0.78
	Facial expressions	0.83	Facial expressions	0.84	Facial expressions	0.82	Facial expressions	0.64
	Frequency of vocalization	0.82	Frequency of vocalization	0.87	Quality of rapport	0.78	Quality of rapport	0.77
	Shared enjoyment	0.76	Shared enjoyment	0.68	Shared enjoyment	0.81	Shared enjoyment	0.84
	Quality of social overtures	0.94	Quality of social overtures	0.83	Quality of social overtures	0.73	Quality of social overtures	0.72
	Response to joint attention	0.47	Pointing	0.62	Pointing	−0.06 (1.0)	Conversation	0.63
	Gestures	0.78	Gestures	0.69	Gestures	0.48	Gestures	0.66
	Showing	0.72	Showing	0.77	Showing	0.56	Quality of social response	0.72
	Initiation of joint attention	0.72	Initiation of joint attention	0.64	Initiation of joint attention	0.25 (0.58)	Reporting of events	0.37
Restricted, repetitive behaviors	Intonation	0.20 (0.54)	Stereotyped language	0.32 (0.47)	Stereotyped language	−0.02 (0.57)	Stereotyped language	0.66
	Unusual sensory interest	0.67	Unusual sensory interest	0.71	Unusual sensory interest	−0.10 (0.72)	Unusual sensory interest	0.43
	Repetitive interests	0.67	Repetitive interests	0.55	Repetitive interests	−0.01 (0.63)	Highly specific topics	0.54
	Hand mannerisms	0.68	Hand mannerisms	0.51	Hand mannerisms	−0.16 (0.53)	Hand mannerisms	0.57
Eigenvalues	7.4	1.7	7.5	1.4	7.1	1.3	6.8	1.4
RMSEA	0.05		0.06		0.08		0.06	

Note: Item names have been abbreviated from the Western Psychological Services Autism Diagnostic Observation Schedule item names. Refer to the key from Figure 6 in the *Autism Diagnostic Observation Schedule Manual*[5] for complete names. RMSEA = root mean square error approximation. Factor loadings are listed by the Social Affect/Restricted, Repetitive Behavior domain assignation from Gotham et al.[1] Where the current loading of an item is <0.40 for the assigned factor, the alternate loading is given in parentheses.

[a] $N = n$(Items:Subjects).

J Am Acad Child Adolesc Psychiatry. Author manuscript; available in PMC 2011 March 15.

TABLE 3

Sensitivities and Specificities of Present and Revised Algorithms in Collaborative Programs for Excellence in Autism/Studies to Advance Autism Research and Treatment and Michigan 2007 Datasets

	AUT vs. NS (*n* = 949)					
	CPEA/STAART				**Michigan 2007**	
	Meets Original Comm-Soc for AUT		**Meets SA + RRB Total**		**Meets SA + RRB Total**	
	Sensitivity	**Specificity**	**Sensitivity**	**Specificity**	**Sensitivity**	**Specificity**
Module 1, no words; NVMA >15 (AUT = 203, NS = 46)	89	78	86 (78)	80 (80)	95	94
Module 1, some words (AUT = 154, NS = 46)	73	94	89 (81)	91 (94)	97	91
Module 2, Younger (AUT = 52, NS = 18)	85	100	94 (81)	100 (100)	98	93
Mod 3 (AUT = 339, NS = 73)	72	96	82 (79)	92 (90)	91	84
	Nonautism ASD vs. NS (*n* = 238)					
	CPEA/STAART				Michigan 2007	
	Meets Original Comm-Soc for ASD		Meets SA + RRB Total		Meets SA + RRB Total	
	Sensitivity	Specificity	Sensitivity	75 (75)	Sensitivity	Specificity
Module 1, some words (PDD-NOS = 21, NS = 64)	100	80	95 (100)	100 (94)	77	82
Module 2, younger (PDD-NOS = 17, NS = 18)	88	100	65 (88)	88 (86)	84	77
Module 3	49	89	60 (53)	75 (75)	72	76

Note: Numbers in parentheses indicate results computed with the SA factor alone. Michigan 2007 = dataset used in Gotham et al.[1]; CPEA/STAART = Collaborative Programs for Excellence in Autism/Studies to Advance Autism Research and Treatment; AUT = autism; NS = cases with nonspectrum diagnoses; Comm-Soc = Communication + Social cutoffs from 2000 norms[2]; SA = Social Affect domain[1]; RRB = Restricted, Repetitive Behaviors domain[1]; NVMA = nonverbal mental age in months; ASD = autism spectrum disorder; PDD-NOS = pervasive developmental disorders not otherwise specified.