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.

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. J. Am. Acad. Child Adolesc. Psychiatry, 2008;47(6):642-651.

Key Words: autism, pervasive developmental disorders not otherwise specified, Autism Diagnostic Observation Schedule, diagnosis.

In their 2007 article, Gotham et al. proposed revised algorithms intended to improve predictive validity of the Autism Diagnostic Observation Schedule (ADOS) modules used with children (modules 1–3). Similar domain distributions in the original ADOS norming sample 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.\(^5\) 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.\(^5\) 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 article\(^1\)). 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 (\(n = 36\)),
and Mount Sinai Medical Center (\(n = 14\)).

**TABLE 1**

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

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Module 1, No Words</th>
<th>Module 1, Some Words</th>
<th>Module 2, Younger</th>
<th>Module 2, Older</th>
<th>Module 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism</td>
<td>295</td>
<td>41.8</td>
<td>24.3</td>
<td>183</td>
<td>47.7</td>
</tr>
<tr>
<td>VIQ</td>
<td>295</td>
<td>36.4</td>
<td>17.3</td>
<td>183</td>
<td>59.4</td>
</tr>
<tr>
<td>NVIQ</td>
<td>295</td>
<td>60.4</td>
<td>22.7</td>
<td>183</td>
<td>71.4</td>
</tr>
<tr>
<td>ADI social</td>
<td>240</td>
<td>20.3</td>
<td>5.0</td>
<td>169</td>
<td>17.7</td>
</tr>
<tr>
<td>ADI comm-V</td>
<td>18</td>
<td>17.8</td>
<td>3.6</td>
<td>59</td>
<td>16.0</td>
</tr>
<tr>
<td>ADI comm-NV</td>
<td>208</td>
<td>11.7</td>
<td>2.1</td>
<td>109</td>
<td>10.1</td>
</tr>
<tr>
<td>ADI-R</td>
<td>240</td>
<td>5.0</td>
<td>2.1</td>
<td>169</td>
<td>5.2</td>
</tr>
<tr>
<td>ADOS social</td>
<td>295</td>
<td>11.3</td>
<td>2.5</td>
<td>183</td>
<td>9.5</td>
</tr>
<tr>
<td>ADOS comm</td>
<td>295</td>
<td>5.7</td>
<td>1.7</td>
<td>183</td>
<td>5.0</td>
</tr>
<tr>
<td>ADOS SA</td>
<td>295</td>
<td>16.1</td>
<td>3.6</td>
<td>183</td>
<td>13.8</td>
</tr>
<tr>
<td>ADOS RR</td>
<td>295</td>
<td>4.2</td>
<td>2.3</td>
<td>183</td>
<td>3.3</td>
</tr>
<tr>
<td><strong>PDD-NOS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>6</td>
<td>58.0</td>
<td>28.3</td>
<td>21</td>
<td>47.7</td>
</tr>
<tr>
<td>VIQ</td>
<td>6</td>
<td>40.3</td>
<td>15.7</td>
<td>21</td>
<td>59.0</td>
</tr>
<tr>
<td>NVIQ</td>
<td>6</td>
<td>59.2</td>
<td>20.5</td>
<td>21</td>
<td>70.1</td>
</tr>
<tr>
<td>ADI social</td>
<td>6</td>
<td>19.5</td>
<td>6.6</td>
<td>19</td>
<td>13.3</td>
</tr>
<tr>
<td>ADI comm-V</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>8</td>
<td>11.0</td>
</tr>
<tr>
<td>ADI comm-NV</td>
<td>4</td>
<td>10.5</td>
<td>4.0</td>
<td>9</td>
<td>9.0</td>
</tr>
<tr>
<td>ADI-R</td>
<td>6</td>
<td>1.2</td>
<td>7.7</td>
<td>18</td>
<td>5.2</td>
</tr>
<tr>
<td>ADOS social</td>
<td>6</td>
<td>10.3</td>
<td>2.4</td>
<td>21</td>
<td>6.5</td>
</tr>
<tr>
<td>ADO S comm</td>
<td>6</td>
<td>6.0</td>
<td>1.8</td>
<td>21</td>
<td>4.3</td>
</tr>
<tr>
<td>ADOS SA</td>
<td>6</td>
<td>15.2</td>
<td>2.9</td>
<td>21</td>
<td>10.3</td>
</tr>
<tr>
<td>ADO S RR</td>
<td>6</td>
<td>1.7</td>
<td>1.5</td>
<td>21</td>
<td>2.1</td>
</tr>
<tr>
<td><strong>Nonspectrum</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>51</td>
<td>36.6</td>
<td>14.5</td>
<td>64</td>
<td>44.4</td>
</tr>
<tr>
<td>VIQ</td>
<td>51</td>
<td>48.7</td>
<td>17.8</td>
<td>64</td>
<td>66.9</td>
</tr>
<tr>
<td>NVIQ</td>
<td>51</td>
<td>62.8</td>
<td>19.2</td>
<td>64</td>
<td>71.4</td>
</tr>
<tr>
<td>ADI social</td>
<td>43</td>
<td>9.4</td>
<td>7.4</td>
<td>53</td>
<td>6.8</td>
</tr>
<tr>
<td>ADI comm-V</td>
<td>1</td>
<td>14.0</td>
<td>—</td>
<td>21</td>
<td>8.4</td>
</tr>
<tr>
<td>ADI comm-NV</td>
<td>38</td>
<td>6.8</td>
<td>4.4</td>
<td>27</td>
<td>3.7</td>
</tr>
<tr>
<td>ADI-R</td>
<td>43</td>
<td>3.1</td>
<td>2.1</td>
<td>53</td>
<td>2.5</td>
</tr>
<tr>
<td>ADOS social</td>
<td>51</td>
<td>5.4</td>
<td>4.1</td>
<td>64</td>
<td>2.7</td>
</tr>
<tr>
<td>ADOS comm</td>
<td>51</td>
<td>3.1</td>
<td>2.3</td>
<td>64</td>
<td>1.5</td>
</tr>
<tr>
<td>ADOS SA</td>
<td>51</td>
<td>7.9</td>
<td>5.6</td>
<td>64</td>
<td>3.9</td>
</tr>
<tr>
<td>ADOS RR</td>
<td>51</td>
<td>2.2</td>
<td>2.1</td>
<td>64</td>
<td>1.5</td>
</tr>
</tbody>
</table>

**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-R = 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.

Copyright © 2008 American Academy of Child and Adolescent Psychiatry. Unauthorized reproduction of this article is prohibited.
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 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 Learning12 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 article1 (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) curves14 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)\(^1\) 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
## TABLE 2
Revised Algorithm Factor Loadings in Collaborative Programs for Excellence in Autism/Studies to Advance Autism Research and Treatment Dataset

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

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 for complete names. RMSEA = root mean square error approximation. Factor loadings are listed by the Social Affect / Restricted, Repetitive Behavior domain assignment from Gotham et al. 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).
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

---

**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

<table>
<thead>
<tr>
<th>AUT vs. NS (n = 949)</th>
<th>Michigan 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CPEA/STAART</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Meets Original Comm-Soc for AUT</strong></td>
<td><strong>Meets SA + RRB Total</strong></td>
</tr>
<tr>
<td><strong>Sensitivity</strong></td>
<td><strong>Specificity</strong></td>
</tr>
<tr>
<td>Module 1, no words; NVMA &gt;15 (AUT = 203, NS = 46)</td>
<td>89</td>
</tr>
<tr>
<td>Module 1, some words (AUT = 154, NS = 46)</td>
<td>73</td>
</tr>
<tr>
<td>Module 2, Younger (AUT = 52, NS = 18)</td>
<td>85</td>
</tr>
<tr>
<td>Mod 3 (AUT = 339, NS = 73)</td>
<td>72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>CPEA/STAART</strong></th>
<th><strong>Michigan 2007</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Meets Original Comm-Soc for ASD</strong></td>
<td><strong>Meets SA + RRB Total</strong></td>
</tr>
<tr>
<td><strong>Sensitivity</strong></td>
<td><strong>Specificity</strong></td>
</tr>
<tr>
<td>Module 1, some words (PDD-NOS = 21, NS = 64)</td>
<td>100</td>
</tr>
<tr>
<td>Module 2, younger (PDD-NOS = 17, NS = 18)</td>
<td>88</td>
</tr>
<tr>
<td>Module 3</td>
<td>49</td>
</tr>
</tbody>
</table>

*Note: Numbers in parentheses indicate results computed with the SA factor alone. Michigan 2007 = dataset used in Gotham et al.; 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; SA = Social Affect domain; RRB = Restricted, Repetitive Behaviors domain; NVMA = nonverbal mental age in months; ASD = autism spectrum disorder; PDD-NOS = pervasive developmental disorders not otherwise specified.*

---

648 WWW.JAACAP.COM J. AM. ACAD. CHILD ADOLESC. PSYCHIATRY, 47:6, JUNE 2008
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 mis-
classifications: 56% of the CPEA/STAART misclassi-
fied 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  
eexisting 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 non-  
autism 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. 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 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. 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.

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, 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.

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.

REFERENCES


Prevalence of Correlates, Disability, and Comorbidity of DSM-IV Drug Abuse and Dependence in the United States

Compton WM, Thomas YF, Stinson FS, Grant BF

Background: Current and comprehensive information on the epidemiology of DSM-IV 12-month and lifetime drug use disorders in the United States has not been available. Objectives: To present detailed information on drug abuse and dependence prevalence, correlates, and co-morbidity with other Axis I and II disorders. Design, Setting, and Participants: Face-to-face interviews using the Alcohol Use Disorder and Associated Disabilities Interview Schedule of the National Institute on Alcohol Abuse and Alcoholism in a large representation sample of US adults (N = 43,093). Main Outcome Measures: Twelve-month and lifetime prevalence of drug abuse and dependence and the associated correlates, treatment rates, disability, and comorbidity with other Axis I and II disorders. Results: Prevalences of 12-month and lifetime drug abuse (1.4% and 7.7%, respectively) exceeded rates of drug dependence (0.6% and 2.6%, respectively). Rates of abuse and dependence were generally greater among men, Native Americans, respondents aged 18 to 44 years, those of lower socioeconomic status, those in the West, and those who were never married or widowed, separated, or divorced (all P < .05). Associations of drug use disorders with other substance use disorders and anti-social personality disorder were diminished but remained strong when we controlled for psychiatric disorders. Dependence associations with most mood disorders and generalized anxiety disorder also remained significant. Lifetime treatment- or help-seeking behavior was uncommon (8.1%, abuse; 37.9%, dependence) and was not associated with sociodemographic characteristics but was associated with psychiatric comorbidity. Conclusions: Most individuals with drug use disorders have never been treated, and treatment disparities exist among those at high risk, despite substantial disability and comorbidity. Comorbidity of drug use disorders with other substance use disorders and antisocial personality disorder, as well as dependence with mood disorders and generalized anxiety disorder, appears to be due in part to unique factors underlying each pair of these disorders studied. The persistence of low treatment rates despite the availability of effective treatments indicates the need for vigorous educational efforts for the public and professionals. Reprinted with permission from Arch Gen Psychiatry 2007;64:566–576. Copyright © 2007, American Medical Association. All rights reserved.