Autism From 2 to 9 Years of Age

Catherine Lord, PhD; Susan Risi, PhD; Pamela S. DiLavore, PhD; Cory Shulman, PhD; Audrey Thurm, PhD; Andrew Pickles, PhD

Context: Autism represents an unusual pattern of development beginning in the infant and toddler years.

Objectives: To examine the stability of autism spectrum diagnoses made at ages 2 through 9 years and identify features that predicted later diagnosis.

Design: Prospective study of diagnostic classifications from standardized instruments including a parent interview (Autism Diagnostic Interview–Revised [ADI-R]), an observational scale (Pre-Linguistic Autism Diagnostic Observation Schedule/Autism Diagnostic Observation Schedule [ADOS]), and independent clinical diagnoses made at ages 2 and 9 years compared with a clinical research team’s criterion standard diagnoses.

Setting: Three inception cohorts: consecutive referrals to (1) state-funded community autism centers, (2) a private university autism clinic, and (3) case controls with developmental delay from community clinics.

Participants: At 2 years of age, 192 autism referrals and 22 developmentally delayed case controls; 172 children seen at 9 years of age.

Autism represents an unusual pattern of development beginning in infancy or the toddler years and defined by deficits in 3 areas: reciprocal social interaction, communication, and restricted and repetitive behaviors.1,2 While parents typically report concerns in the first year of life,3 many children do not receive diagnoses until much later. Several studies have suggested that diagnoses of autism made at age 2 years are stable through age 3 years.4-7 and diagnoses made by age 5 years are stable up to late adolescence.8 A recent study reported relatively good diagnostic stability but limited continuity in symptom severity to age 7 years for children given autism diagnoses at age 2 years.9 Several intervention projects reported diagnostic changes and extraordinary levels of improvement in a substantial minority of young children with autism.10,11 Other reports found little diagnostic change and fewer marked improvements.12,13 Possible explanations for these conflicting results are diagnostic instability or the lack of age-appropriate diagnostic criteria for very young children. In addition, epidemiological,14 genetic,15 and diagnostic studies16 have extended the conceptualization of autism to include a broader spectrum of disorders that range from autism to potentially milder forms of social deficits, including pervasive developmental disorder not otherwise specified (PDD-NOS),17,18 atypical autism, and Asperger syndrome.19,20 Recently, investigators have begun to ask about the stability for the broader autism spectrum disorder (ASD) as well as for more narrowly defined autism.21

High stability has been found for clinical diagnoses between ages 2 and 3 years when health care professionals interpreted standard criteria for autism.4,6,9,22 Diagnoses based on the Autism Diagnostic Interview–Revised (ADI-R), yielding an algorithm operationalizing DSM-IV and International Statistical Classification of Diseases, 10th Revision, were not as stable.9 At age 2 years, children with severe retardation were over-
diagnosed with autism and children who did not yet show repetitive behaviors or stereotyped speech were underdiagnosed.\(^4\) Charman and colleagues\(^5\) found that diagnostic thresholds from the ADI-R were crossed and recrossed between ages 2 to 7 years. Moore and Goodson,\(^6\) using the ADI-R modified to take into account clinical observations, found that 88% of children diagnosed with autism at age 2 years retained that diagnosis at ages 3 and 4 years. Increases during this period in repetitive behaviors and interests were also found. Stone and colleagues\(^7\) reported lower stability for children who initially received diagnoses of PDD-NOS than autism, though more than 90% of children remained within the autism spectrum 1 year later.

The present article reports prospective data from a relatively large sample of autism referrals and a comparison group of children with developmental delay seen at ages 2, 4 to 5, and 9 years, assessed using standardized instruments, including the ADI-R, a structured observation, and independent clinical diagnoses. Analyses first addressed the question of diagnostic stability of autism and PDD-NOS. Because the application of diagnostic measures to children younger than 3 years is not well established, we address the diagnostic utility of the instruments along with changes in the diagnoses of individual children. A second aim was to identify features at age 2 years that best predicted later diagnosis.

## METHODS

### SUBJECTS

One hundred ninety-two children were prospectively studied from the time they were referred for evaluation for possible autism before 36 months of age: 111 from North Carolina and 81 from Chicago, Ill. Sample children were consecutive referrals, seen before 38 months of age, to 4 regional state-funded autism centers in North Carolina and to a private university hospital in Chicago. Exclusion criteria included moderate to severe sensory impairments, cerebral palsy, or poorly controlled seizures. In addition, 22 children with developmental delays between ages 13 and 33 months who met the same exclusion criteria and who had never been referred for or diagnosed with autism were recruited from the sources of referral to the North Carolina autism centers. Mean (SD) chronological ages at the time of first assessment for the referred for evaluation groups (North Carolina, 29.2 [4.6] months; Chicago, 29.2 [3.4] months) and the developmental delay group (26.6 [6.7] months) were not significantly different (P = .09). A parent or guardian provided informed consent in accordance with institutional review boards of the University of North Carolina, Chapel Hill, and the University of Chicago. Assessments were free of charge; feedback and a report were provided after each assessment.

At approximately age 5 years, 103 North Carolina and 11 Chicago children referred for evaluation and 22 children with developmental delay were reassessed. At age 9 years, 87 North Carolina and 68 Chicago children referred for evaluation and 17 children with developmental delay were reassessed, representing an 80.4% follow-up rate. Attrition was unrelated to original diagnosis, sex, verbal or nonverbal IQ, adaptive functioning, or language level but was significantly higher for nonwhite ethnicity. The 172 children with data at both ages 2 and 9 years form the basis of this report (Table 1).

### MEASURES

Children received a 2-part standard assessment at each point in the study. Most frequently, parents were interviewed at home and then the child and family were seen for a second session at the child’s school or clinic. The Vineland Adaptive Behavior Scales,\(^21\) a standardized measure of adaptive functioning based on a parent interview, were administered immediately following the ADI-R at each age. At age 2 years, all but 1 child (given the Stanford-Binet), were administered the Mullen Scales of Early Learning.\(^24\) At age 9 years, the selection of cognitive tests followed a standard hierarchy designed for use when children could not achieve a basal score or achieved ceiling scores: 39 children, Wechsler Intelligence Scale for Children;\(^25\) 73 children, Differential Ability Scales;\(^26\) 51 children, Mullen Scales of Early Learning; and 6 children, other. Because raw scores frequently fell outside stan-

#### Table 1. Descriptive Characteristics by Best-Estimate Diagnoses at Ages 2 and 9 Years in 172 Children

<table>
<thead>
<tr>
<th>Variable</th>
<th>Autism (n = 84)</th>
<th>PDD-NOS (n = 46)</th>
<th>Nonspectrum (n = 42)*</th>
<th>Autism (n = 100)</th>
<th>PDD-NOS (n = 35)</th>
<th>Nonspectrum (n = 37)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, %</td>
<td>56</td>
<td>52</td>
<td>59</td>
<td>50</td>
<td>50</td>
<td>51</td>
</tr>
<tr>
<td>White, %†</td>
<td>60</td>
<td>55</td>
<td>62</td>
<td>60</td>
<td>62</td>
<td>60</td>
</tr>
<tr>
<td>African American, %</td>
<td>26</td>
<td>23</td>
<td>22</td>
<td>27</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>Age, mo, at baseline assessment at 2 y, mean (SD)</td>
<td>29.1 (4.7)</td>
<td>29.1 (5.6)</td>
<td>28.8 (5.5)</td>
<td>29.0 (4.9)</td>
<td>30.3 (5.3)</td>
<td>27.8 (5.5)</td>
</tr>
<tr>
<td>Age, mo, at follow-up at 9 y, mean (SD)</td>
<td>110.1 (15.7)</td>
<td>113.8 (17.1)</td>
<td>114.9 (11.8)</td>
<td>111.5 (16.5)</td>
<td>111.1 (15.8)</td>
<td>115.6 (11.0)</td>
</tr>
<tr>
<td>Limited speech, 2 y, %‡</td>
<td>74</td>
<td>57</td>
<td>50</td>
<td>74</td>
<td>51</td>
<td>46</td>
</tr>
<tr>
<td>Limited speech, 9 y, %</td>
<td>30</td>
<td>4</td>
<td>7</td>
<td>28</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>VABC, 2 y, mean (SD)</td>
<td>61.0 (12.3)</td>
<td>64.1 (10.3)</td>
<td>65.7 (9.4)</td>
<td>61.0 (12.3)</td>
<td>64.7 (8.2)</td>
<td>68.7 (9.7)</td>
</tr>
<tr>
<td>VABC, 9 y, mean (SD)</td>
<td>43.7 (22.6)</td>
<td>57.4 (26.6)</td>
<td>58.2 (27.4)</td>
<td>39.3 (18.6)</td>
<td>69.4 (22.6)</td>
<td>63.9 (29.1)</td>
</tr>
<tr>
<td>Ratio verbal IQ, 2 y, mean (SD)</td>
<td>26.4 (15.3)</td>
<td>46.6 (21.7)</td>
<td>57.9 (23.9)</td>
<td>28.5 (16.7)</td>
<td>49.6 (24.3)</td>
<td>58.5 (22.2)</td>
</tr>
<tr>
<td>Ratio verbal IQ, 9 y, mean (SD)</td>
<td>41.2 (36.5)</td>
<td>71.7 (36.9)</td>
<td>60.4 (31.1)</td>
<td>35.1 (26.8)</td>
<td>91.5 (32.2)</td>
<td>69.7 (33.4)</td>
</tr>
<tr>
<td>Ratio nonverbal IQ, 2 y, mean (SD)</td>
<td>63.3 (16.9)</td>
<td>74.0 (22.3)</td>
<td>72.7 (26.5)</td>
<td>62.7 (19.2)</td>
<td>80.0 (20.0)</td>
<td>73.1 (23.8)</td>
</tr>
<tr>
<td>Ratio nonverbal IQ, 9 y, mean (SD)</td>
<td>54.0 (30.9)</td>
<td>75.4 (33.3)</td>
<td>67.9 (33.5)</td>
<td>50.5 (28.9)</td>
<td>88.9 (25.5)</td>
<td>72.7 (34.3)</td>
</tr>
</tbody>
</table>

Abbreviations: ADI-R, Autism Diagnostic Interview–Revised; ASD, autism spectrum disorder; PDD-NOS, pervasive developmental disorder not otherwise specified; VABC, Vineland Adaptive Behavior Composite.

*The nonspectrum group includes all of the children with developmental delay as well as children referred for evaluation who did not receive ASD diagnoses.

†Four children were of mixed or Hispanic ethnicity.
The Autism Diagnostic Observation Schedule (ADOS)30,31 and an adaptation for younger children, the Pre-Linguistic Autism Observation Schedule (PL-ADOS),32 provided standardized observation of social and communicative behavior. In 1999, the PL-ADOS and the former ADOS31 were combined into a single instrument with separate modules for children at different language levels. The algorithm for the ADOS uses thresholds in social reciprocity and communication domains, as well as an overall cutoff. Reliability and validity have been established for children as young as 2 years.33 Cutoffs for autism provide clear differentiation between children with autism and verbally matched children with nonspectrum disorders. However, the overlap between the narrower classification of autism and the broader classification of ASD is considerable.30 We refer to the administered test as the PL-ADOS because it included additional tasks and scores not retained in the ADOS module 1, but the ADOS algorithm was used for analyses.

At initial assessment, a PL-ADOS (n = 172) was administered to all subjects referred for evaluation for autism and with developmental delay. At age 5 years, the PL-ADOS (n = 119) or ADOS module 2 (n = 11) was administered. At age 9 years, the ADOS modules 1 (n = 64), 2 (n = 46), and 3 (n = 60) were administered. The ADI-R and PL-ADOS/ADOS items were scored during administration; algorithms were completed after the clinical diagnosis was made and did not yet exist when the children were age 2 years. Both the ADI-R and PL-ADOS provide item totals for social, communication (for the ADI-R, nonverbal communication was used here), and repetitive-behavior domains.

Clinical diagnoses were made at ages 2, 5, and 9 years, using somewhat different procedures. For the 2-year-olds, following psychological assessment, 2 clinicians reviewed all test results and the ADI-R summary, discussed the content of the PL-ADOS, and proposed a binary clinical diagnosis (autism, not autism) to which they applied a certainty rating that generated an autism spectrum score from 1 (certain not autism) to 10 (certain autism). There was no attempt to train the clinicians, who were clinical and educational psychologists, in making standard diagnoses of 2-year-olds. Certainty scores were initially introduced because clinicians were uncomfortable making diagnostic decisions for such young children. For purposes of analysis, certainty scores were grouped into definite nonspectrum (1 and 2), ASD including PDD-NOS and less certain cases of atypical autism (3–7), and definite autism (8–10). This approach confounds certainty with severity in that PDD-NOS by definition involves less comprehensive and/or less intense symptoms. As presented

<table>
<thead>
<tr>
<th>Variable</th>
<th>Diagnosis, 2 y (n = 46)</th>
<th>Diagnosis, 2 y (n = 42)</th>
<th>Diagnosis, 2 y (n = 42)</th>
<th>Diagnosis, 9 y (n = 100)</th>
<th>Diagnosis, 9 y (n = 35)</th>
<th>Diagnosis, 9 y (n = 37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADI-R social domain, 2 y</td>
<td>19.7 (4.2)</td>
<td>14.7 (5.7)</td>
<td>9.7 (5.8)</td>
<td>18.6 (5.2)</td>
<td>15.0 (5.1)</td>
<td>9.6 (6.3)</td>
</tr>
<tr>
<td>ADI-R social domain, 9 y</td>
<td>25.0 (5.5)</td>
<td>20.8 (7.1)</td>
<td>13.5 (9.0)</td>
<td>25.4 (4.1)</td>
<td>18.7 (8.0)</td>
<td>11.5 (8.2)</td>
</tr>
<tr>
<td>ADI-R nonverbal communication domain, 2 y</td>
<td>10.0 (2.0)</td>
<td>8.3 (3.0)</td>
<td>5.8 (3.5)</td>
<td>9.8 (2.1)</td>
<td>7.5 (2.9)</td>
<td>5.9 (3.8)</td>
</tr>
<tr>
<td>ADI-R nonverbal communication domain, 9 y</td>
<td>11.8 (2.9)</td>
<td>8.8 (3.9)</td>
<td>5.4 (4.0)</td>
<td>11.8 (2.4)</td>
<td>7.4 (4.0)</td>
<td>4.7 (4.0)</td>
</tr>
<tr>
<td>ADI-R repetitive domain, 2 y</td>
<td>4.1 (1.5)</td>
<td>3.1 (2.3)</td>
<td>2.2 (1.7)</td>
<td>4.0 (1.8)</td>
<td>3.3 (1.9)</td>
<td>1.7 (1.3)</td>
</tr>
<tr>
<td>ADI-R repetitive domain, 9 y</td>
<td>5.9 (2.6)</td>
<td>5.5 (3.2)</td>
<td>3.8 (2.8)</td>
<td>6.3 (2.5)</td>
<td>4.9 (3.1)</td>
<td>4.7 (4.0)</td>
</tr>
<tr>
<td>ADOS social domain, 2 y</td>
<td>12.6 (1.7)</td>
<td>8.8 (3.4)</td>
<td>4.6 (3.6)</td>
<td>11.6 (3.1)</td>
<td>8.9 (3.4)</td>
<td>4.9 (3.7)</td>
</tr>
<tr>
<td>ADOS social domain, 9 y</td>
<td>10.3 (3.0)</td>
<td>7.1 (3.8)</td>
<td>5.0 (3.6)</td>
<td>10.7 (2.3)</td>
<td>5.6 (3.0)</td>
<td>3.6 (3.1)</td>
</tr>
<tr>
<td>ADOS communication domain, 2 y</td>
<td>6.3 (1.4)</td>
<td>4.4 (1.8)</td>
<td>2.5 (2.2)</td>
<td>5.9 (1.8)</td>
<td>5.0 (2.4)</td>
<td>2.5 (2.1)</td>
</tr>
<tr>
<td>ADOS communication domain, 9 y</td>
<td>6.4 (2.0)</td>
<td>4.8 (2.4)</td>
<td>3.5 (2.2)</td>
<td>6.8 (1.7)</td>
<td>3.9 (1.5)</td>
<td>2.6 (2.0)</td>
</tr>
<tr>
<td>ADOS repetitive domain, 2 y</td>
<td>4.0 (1.5)</td>
<td>2.5 (1.4)</td>
<td>0.8 (1.0)</td>
<td>3.6 (1.7)</td>
<td>2.3 (1.5)</td>
<td>1.0 (1.4)</td>
</tr>
<tr>
<td>ADOS repetitive domain, 9 y</td>
<td>2.9 (2.1)</td>
<td>1.7 (1.7)</td>
<td>1.3 (1.3)</td>
<td>3.1 (1.9)</td>
<td>1.0 (1.0)</td>
<td>1.0 (1.0)</td>
</tr>
</tbody>
</table>

Abbreviations: ADI-R, Autism Diagnostic Interview–Revised; ADOS, Autism Diagnostic Observation Schedule; PPD-NOS, pervasive developmental disorder not otherwise specified.

*Values are expressed as mean (SD). The ADOS scores for age 2 years used the module 1 algorithm. At age 9 years, for comparability across modules, all ADOS scores were converted to module 2 (see Lord et al30 for ranges). The ADI-R totals include “ever” scores. The nonspectrum group included all children with developmental delay.
intensity and number of symptoms; clinical certainty ratings were taken into account but it was left to the clinicians to decide how to use information about a particular child. Parallel information for age 9 years was used to generate a consensus best-estimate diagnosis by an independent psychologist and child psychiatrist blind to earlier diagnoses.

Reliability was initially obtained on the diagnostic measures (ADI-R, PL-ADOS, and ADOS) after intensive training until each pair of examiners reached more than 90% exact agreement ($\kappa > 0.70$) on individual items for the ADI-R and 80% exact agreement ($\kappa > 0.60$) on codes for the PL-ADOS/ADOS for 3 consecutive administrations. Approximately every sixth administration of each instrument was scored by 2 raters, yielding $\kappa$ between 0.60 and 0.80. Reliability for clinical diagnoses at age 2 years was measured in 1 in 6 cases with 92% agreement for autism/not autism. The intraclass correlation for certainty ratings was 0.89. For clinical diagnoses at ages 5 and 9 years, agreement between the examiners was established on cases outside this study and monitored once a month (overall agreement > 90% for best-estimate autism cases, and 83% for children with PDD-NOS and nonspectrum disorders).

Diaries completed by parents summarized educational and other treatments their children had received during each year. Two raters coded the diaries, having first established reliability on general classifications (eg, 1 to 1 vs group). There was considerable variation in type and amount of treatment. For the purposes of this article, treatment intensity was defined very crudely by hours of treatment (including education and formal home programming).

ANALYSIS

All analyses were undertaken in Stata 8.0. Agreement among contemporaneous diagnostic measures and between baseline and follow-up diagnosis was assessed using $\kappa$ statistics that correct for chance agreement for nominal measures. Prediction of autism and ASD used logistic regression. To compare odds ratios (ORs) we used Wald tests of interactions from a 2-response generalized estimating equations logistic model with an exchangeable working correlation matrix and robust parameter covariance matrix.

To assess the effect of treatment, there was a need to take account of children’s differential access to treatment. For such selective treatment assignment, an instrumental variable approach was used, requiring identification of a variable that, while associated with treatment received, was assumed, given treatment (and confounders), unrelated to outcome. Recruitment site (North Carolina or Chicago) was used as an instrumental variable approach.

RESULTS

BASELINE ASSESSMENT

Table 1 and Table 2 describe the sample by initial and follow-up best-estimate diagnoses. Rates of diagnosis of autism (and autism plus PDD-NOS) were 55% (81%) for the ADI-R, 65% (83%) for the PL-ADOS, 38% (69%) for the clinicians, and 49% (76%) according to the best-estimate diagnosis. Percentage agreement ($\kappa$) was 85.5% (0.53) for interview-observation, 81.1% (0.47) for interview-clinician, and 84.3% (0.53) for observation-clinician.

In contrast to the ADI-R and the PL-ADOS, Figure 1 shows that clinicians rarely (2 in 172 cases or 1%) classified children as having autism who had not been classified in the same way by 1 of the other measures. On the other hand, clinicians relatively frequently (26 in 172 cases or 15%) indicated autism as not present when both interview and observation classified it as present, though in 19 (73%) of these cases the clinician indicated PDD-NOS. Notwithstanding, best-estimate autism prevalence was consistently high among children identified by clinicians.

For ASD diagnoses, Figure 1 and Table 3 show that the ADI-R and PL-ADOS had similar levels of inclusion, with both more inclusive than clinical judgment. Levels of agreement with the contemporaneous best-estimate diagnosis, reflecting the relative weight attached to each measure in coming to the best-estimate diagnosis at age 2 years, were 0.40 for the interview, 0.54 for the observation, and 0.67 for the clinical judgment (of 1.00 maximum).
Table 3. Cross-tabulation of Initial Diagnostic Measures and Best-Estimate Diagnoses at Ages 2 and 9 Years*

<table>
<thead>
<tr>
<th>Age 2 Diagnostic Measure</th>
<th>Best-Estimate Diagnosis, 2 y</th>
<th>Best-Estimate Diagnosis, 9 y</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. (%)†</td>
<td>Autisim</td>
</tr>
<tr>
<td>ADI-R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism</td>
<td>94 (55)</td>
<td>67</td>
</tr>
<tr>
<td>PDD-NOS</td>
<td>45 (26)</td>
<td>15</td>
</tr>
<tr>
<td>Nonspectrum</td>
<td>33 (19)</td>
<td>2</td>
</tr>
<tr>
<td>ADOS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism</td>
<td>111 (65)</td>
<td>80</td>
</tr>
<tr>
<td>PDD-NOS</td>
<td>31 (18)</td>
<td>3</td>
</tr>
<tr>
<td>Nonspectrum</td>
<td>30 (17)</td>
<td>1</td>
</tr>
<tr>
<td>Clinical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism</td>
<td>65 (38)</td>
<td>63</td>
</tr>
<tr>
<td>PDD-NOS</td>
<td>53 (31)</td>
<td>19</td>
</tr>
<tr>
<td>Nonspectrum</td>
<td>54 (31)</td>
<td>2</td>
</tr>
<tr>
<td>Best-estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism</td>
<td>84 (49)</td>
<td>71</td>
</tr>
<tr>
<td>PDD-NOS</td>
<td>46 (27)</td>
<td>27</td>
</tr>
<tr>
<td>Nonspectrum</td>
<td>42 (24)</td>
<td>2</td>
</tr>
<tr>
<td>No. (%)‡†</td>
<td></td>
<td>102 (49)</td>
</tr>
</tbody>
</table>

Abbreviations: See Table 2.
*Values are expressed as number of children unless otherwise specified. The nonspectrum group consists of all children with diagnoses other than autism spectrum disorder. This includes all of the children initially seen in the developmental delay group, as well as some children referred for evaluation.
†Number and percentages of children seen at age 2 years and at age 9 years.
‡Number and percentages of all children seen at age 2 years and number and percentages of children seen at age 9 years.

Figure 2. Frequency of diagnostic combinations at age 2 years and prevalence of best-estimate diagnosis (in parentheses) at age 9 years. A, Autism. B, Autism spectrum. PL-ADOS indicates Pre-Linguistic Autism Diagnostic Observation Schedule; ADI-R, Autism Diagnostic Interview–Revised.

BEST-ESTIMATE PROGNOSTIC PERFORMANCE

Table 3 shows that, according to the best-estimate diagnosis, between ages 2 and 9 years the proportion with autism increased from 49% to 58%, mainly because fewer children were classified as having PDD-NOS. The best-estimate diagnosis improved between ages 2 and 9 years for 18 children (8%) (only 1 from autism to nonspectrum disorder), compared with 38 (18%) with worse classification. Overall exact agreement between the best-estimate diagnoses at ages 2 and 9 years was 67% ($\kappa = 0.47$), 76% for autism vs nonautism ($\kappa = 0.51$), and 90% for autism spectrum vs nonspectrum ($\kappa = 0.72$). For 112 children assessed at age 5 years, stability was 72% ($\kappa = 0.72$) from ages 2 to 5 years and 88% ($\kappa = 0.92$) from ages 5 to 9 years.

ADI-R, PL-ADOS, AND CLINICIAN PROGNOSTIC PERFORMANCE

Figure 2 and Table 3 also show the relative performance of individual and combinations of measures at age 2 years in predicting the best-estimate diagnosis at age 9.
years. Classifications of autism were frequent for all clinician-positive combinations of measures. The measure of clinical diagnostic uncertainty at age 2 years was strongly associated with change. While just 10% of children with definitely nonspectrum diagnoses and 18% of the children with definite autism changed diagnosis, 43% of the children with less certain diagnoses changed classification. Each instrument was strongly prognostic for autism with an OR of 6.6 (95% CI, 3.3-12.9) and sensitivity of 73% and specificity of 71% for the ADI-R; OR of 6.8 (95% CI, 3.4-13.5) with sensitivity of 82% and specificity of 60% for the PL-ADOS/ADOS; and OR of 12.8 (95% CI, 5.3-30.8) with sensitivity of 58% and specificity of 90% for clinical judgment.

In a simple additive logistic regression for best-estimate autism diagnosis at age 9 years, all 3 diagnostic measures at age 2 years made an independent contribution to prediction, with a partial OR of 3.4 (95% CI, 1.6-7.3) (P = .001) for the ADI-R; partial OR of 2.4 (95% CI, 1.0-5.3) (P = .04) for the PL-ADOS/ADOS, and partial OR of 6.2 (95% CI, 2.4-16.2) (P = .001) for clinical diagnosis, giving an overall sensitivity of 75% and specificity of 78%. Similar analyses showed the ADI-R domain scores at age 2 years made independent contributions to social (P = .07), repetitive (P = .03), and communication (P = .01) IQ at age 2 years were covaried (lower verbal IQ increased the odds of autism), only the ADI-R repetitive domain remained significant (social, P = .30; communication, P = .40; repetitive, P = .02). For the PL-ADOS at age 2 years, independent prediction from social and repetitive domains (social, P = .003; communication, P = .90; repetitive, P = .002), while reduced, remained significant (social, P = .05; communication, P = .30; repetitive, P = .005) in the presence of verbal (P = 0.1) and nonverbal (P = 0.90) IQ.

Tests comparing the ORs for predicting autism and ASDs showed some specific relationships with instruments and domains. While nonverbal IQ at age 2 years did not predict autism at age 9 years, higher nonverbal IQ and higher PL-ADOS/ADOS communication scores predicted ASD diagnoses (interactions, P = .006 and P < .03, respectively). The ADI-R repetitive score at age 2 years predicted ASD at age 9 years more strongly than it predicted autism (interaction, P = .006).

**BASELINE MEASURES AND PREDICTED CHANGE**

As expected by their definition, the mean “most abnormal 4 to 5” or “ever” lifetime ADI-R algorithm scores in Table 2 are higher at age 9 years than age 2 years. By contrast, the mean ADI-R total score based on current items (excluding verbal items) indicated a marked reduction (8.1 points [95% CI, 6.4-9.7]; P < .001) in abnormality, and PL-ADOS/ADOS scores (corrected for the number of possible items in the algorithm and the distribution of social and communication items) also fell (2.1 points [95% CI, 3.2-1.0]; P < .001). Change-score analysis of ADI-R and PL-ADOS/ADOS item totals gave similar findings, with no significant associations with sex (P = .70 and .30), ethnicity (P = .30 and .50), mother’s education (P = .40 and .30) nor baseline verbal (P = .10 and .07) or nonverbal (P = .20 and .50) IQs or adaptive behavior (P = .50 and .70).

This improvement contrasted with a marked worsening during the same period in mean adaptive-behavior standard scores from 63 to 51 (-12.1 points [95% CI, 15.9-8.4]; P < .001). The decline was associated with low verbal (P < .001) and nonverbal (P < .001) IQ at age 2 years and high ADI-R symptom severities in the social (P < .001) and nonverbal communication (P < .001) domains at age 2 years but not with restricted and repetitive behavior (P = .30). Change in adaptive behavior was not associated with ethnicity (P = .10), sex (P = .30), or mother’s education (P = .60). Vineland correlations from ages 2 to 5 years were 0.72; from age 5 to 9 years, 0.85; and from ages 2 to 9 years, 0.62. This decline in functioning is also evident from Table 1. While all 3 groups had similar functioning at age 2 years, the autism group at 9 years of age had markedly lower scores. Table 1 suggests a quite distinctive profile for the PDD-NOS group at age 9 years, with markedly higher verbal IQ and, to a lesser extent, nonverbal IQ compared with differences in group means at age 2 years.

**CROSS-DOMAIN PREDICTION**

For each ADI-R and PL-ADOS domain score, regression prediction of each domain score at age 9 years by the set of 3 domain scores at age 2 years showed significant continuity within the same domain. The 1 exception was the ADOS communication score at age 9 years that was predicted by the ADOS social (P = .01) and repetitive (P = .002) domains at age 2 years, with no significant independent contribution from communication (P = .70). Other independent cross-domain predictions occurred for the PL-ADOS social score at age 2 years, predicting the repetitive domain score at age 9 years (P = .008), and for the ADI-R, where nonverbal communication score at age 2 years independently predicted social scores at age 9 years (P = .02) and social scores at age 2 years independently predicted nonverbal communication scores at age 9 years (P = .003).

**ASSOCIATION WITH TREATMENT**

Our rather crude measure of hours of treatment was associated with worsening of the ADI-R total score (P = .01), adaptive behavior (P < .001), and PL-ADOS/ADOS total score (P = .06). However, this did not take into account selective treatment exposure, which was strongly associated with region of referral (P = .003). Using region as an instrument for treatment exposure in a 2-stage least squares regression did not alter the estimated direction of effects, but all effects were then nonsignificant (P = .08, .10, and .08, respectively).

**COMMENT**

Diagnosis of autism in 2-year-olds was quite stable up through 9 years of age, with the majority of change associated with increasing certainty of classifications moving from ASD/PDD-NOS to autism. Only 1 of 84 children with best-estimate diagnoses of autism at age 2 years received a nonspectrum diagnosis at age 9 years, and more than half of children initially diagnosed with PDD-NOS later met autism criteria. Nevertheless, more than 10% of children with
other findings support the conceptualization of ADI-R and ADOS social and both ADI-R and PL-ADOS/ADOS repeti-
tion at age 9 years. The independent predictive power of
instruments and clinical judgment added to the predic-
tive functioning and the PDD-NOS group, much less ab-
normal verbal and nonverbal IQ.

Among this specialized group of clinicians, clinical
judgment of autism at age 2 years was a better predictor
of later diagnosis than either standardized interview or
observation. Contemporaneous agreement between clini-
cal judgment and best-estimate judgment for 2-year-
olds was equal to that found between experienced raters
in the DSM-IV field trials for older children and adults.10

Though the clinical diagnoses at age 2 years were made
without knowledge of the ADI-R and ADOS algorithm
scores, each clinician had administered either the PL-
ADOS or the ADI-R and had the opportunity to discuss his
or her impressions with the experienced clinician who had
administered the other instrument. Thus, the information
available to them was very different from the information
obtained during a typical single office visit to a clinical psy-
chologist or developmental pediatrician. The use of stan-
dardized measures seems likely to have improved the sta-
bility of diagnosis both directly through straightforward use
of algorithms for autism and ASD and also indirectly through
structuring clinical judgment. Of cases in which the clas-
sifications yielded by both instruments were not sup-
ported by the clinicians at age 2 years, 40% were children
with severe mental retardation (and not autism) or chil-
dren with very difficult behavior (and not autism), while
the remainder were mild cases of autism characterized as
uncertain. On the other hand, clinical judgments were con-
sistently underinclusive at age 2 years, both for narrow di-
agnoses of autism and for broader classifications of ASD
at age 9 years. Thus, scores from standardized instru-
ments also made real contributions beyond their influ-
ence on informing and structuring clinical judgment. Over-
all, while standardized research instruments at age 2 years
did not fully capture the insight in the form of certainty
ratings made by experienced, well-trained clinicians, this
insight was not by itself sufficient.

A positive ADI-R or PL-ADOS/ADOS classification of
autism or PDD-NOS, when contradicted by the other mea-
sures, was of limited prognostic value. Nonetheless, both
instruments and clinical judgment added to the predic-
tion at age 9 years. The independent predictive power of
the communication domain in the PL-ADOS/ADOS and
both the social and communication domains in the ADI-R
was modest, standing in contrast with the PL-ADOS/
ADOS social and both ADI-R and PL-ADOS/ADOS repeti-
tive domains, which made independent contributions, simi-
lar to the findings of Charman and colleagues.9 These and
other findings support the conceptualization of ADI-R and
ADOS social and nonverbal communication items as re-
flecting 1 factor. The limitations of the repetitive domain
score of the PL-ADOS/ADOS, based on a brief sample of
behavior, are well understood,29,30 and several studies have
found that a significant number of children who receive
autism diagnoses in later preschool years are not de-
scribed as having repetitive behaviors before 30 months of
age.4,6,22 To find the repetitive domain score from the ADI-R
and the PL-ADOS/ADOS so strongly predictive of progno-
sis for autism and ASD 7 years later, both before and after
verbal IQ was taken into account, was surprising. As ex-
pected, low verbal IQ was also associated with increased
probability of an outcome of autism or ASD.8 As a group,
children with uncertain clinical diagnoses and high ver-
bal and nonverbal IQs at age 2 years who showed more
prosocial behavior (a relatively low social score on the
ADOS) and little or no repetitive behavior during the ADOS
and ADI-R were the most likely to change diagnosis from au-
tism to PDD-NOS and PDD-NOS to nonspectrum catego-
ries at age 9 years and were least likely to show losses in
adaptive behavior at age 9 years (and so have relatively bet-
ter outcome in everyday skills).

As reported elsewhere,8 the overall totals on the ADI-R
and ADOS were not systematically related to change in au-
tistic symptoms from age 2 to 9 years. The lack of evidence
for a true association between the amount of therapeutic
intervention and amount of diagnostic change is not encou-
raging for very time-intensive treatments but may reflect our
rather gross quantitative measure of hours of intervention,
which had no control for kind or quality of treatment.

This study has the usual strengths and limitations of a
prospective cohort study. Children were identified at young
ages, which allowed for prospective study but also meant
that these cohorts are not necessarily representative of chil-
dren referred for autism at older ages. The oldest of these
children was referred 14 years ago, which also means that
a cohort of 2-year-olds today might be rather different. The
clinicians providing the clinical judgments were very ex-
perienced clinicians, though not with 2-year-olds, who made
up a relatively small proportion of routine referrals at that
time. This lack of familiarity with 2-year-olds likely con-
tributed to the clinicians' consistently underinclusive judg-
ments, a finding replicated by others,8 which deserves special
attention at a time when most concern is about
overdiagnosis of ASD for older children.

Overall, referrals of 2-year-olds for possible autism to
2 very different programs in different regions (North Caro-
olina and Chicago) included many more children who ac-
tually had ASD than we expected, with just less than half
of the referred children receiving autism diagnoses and
75%, ASD diagnoses. This attests to the ability of com-

community physicians, and the parents who for the most part
initiated the process, to make appropriate referrals when a
free evaluation was easily accessible, though it is impor-
tant to remember that we cannot determine how many
children were not referred who should have been.

In turn, clinicians in the study, using standardized in-
struments and their own judgments to integrate informa-
tion into a best-estimate diagnosis at age 2 years, were able
to make classifications that predicted diagnosis within the
autism spectrum at age 9 years for almost all cases. There
are real questions about the usefulness of PDD-NOS as a
categorical diagnosis. However, especially for very young children, having a way for experienced clinicians to acknowledge their uncertainty about some 2-year-olds was ultimately helpful as a means of flagging children who by age 9 years had a range of difficulties from autism to very mild social deficits. On a more somber note, because more than half of the children with PDD-NOS clinical diagnoses at age 2 years received best-estimate diagnoses of autism by age 9 years, health care professionals should be wary of telling parents that their young children do not have autism, only PDD-NOS. In the end, the development of meaningful measures of continuous dimensions of behavior in ASD should improve research and practice.

Submitted for Publication: June 6, 2005; final revision received November 23, 2005; accepted December 21, 2005. Correspondence: Catherine Lord, PhD, University of Michigan Autism and Communication Disorders Center, 1111 E Catherine St, Ann Arbor, MI 48109 (celord@umich.edu).

Financial Disclosure: Drs Lord and Risi receive royalties from the publication of the Autism Diagnostic Interview—Revised and Pre-Linguistic Autism Diagnostic Observation Schedule/Autism Diagnostic Observation Schedule, though at the time of this study the instruments were distributed free of charge.

Funding/SUPPORT: This work was supported by grants MH57167 and MH006469 from the National Institute of Mental Health and HD 35482-01 from the National Institute of Child Health and Human Development (Dr Lord).

Disclaimer: This work was not written as part of Dr Thurum’s official duties as a government employee. Views expressed in this article do not necessarily represent those of the National Institutes of Health or the US government.

Previous Presentations: Parts of this work were presented at the Society for Research in Child Development; April 23, 2003; Tampa, Fla; and April 17, 2001; Minneapolis, Minn.

Acknowledgment: We thank D. Deborah Anderson, PhD, Debra Combs, BA, E. Glenna Osborne, MA, Rebecca Niehus, MA, Shapning Qiu, MA, and Lyn Carpenter, PhD, for data collection and management assistance.

REFERENCES


