

# 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 for autism assessment 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.

**Main Outcome Measures:** Consensus best-estimate diagnoses at 9 years of age.

**Results:** Percentage agreement between best-estimate diagnoses at 2 and 9 years of age was 67, with a weighted  $\kappa$  of 0.72. Diagnostic change was primarily accounted for by movement from pervasive developmental disorder not otherwise specified to autism. Each measure at age 2 years was strongly prognostic for autism at age 9 years, with odds ratios of 6.6 for parent interview, 6.8 for observation, and 12.8 for clinical judgment. Once verbal IQ ( $P = .001$ ) was taken into account at age 2 years, the ADI-R repetitive domain ( $P = .02$ ) and the ADOS social ( $P = .05$ ) and repetitive domains ( $P = .005$ ) significantly predicted autism at age 9 years.

**Conclusions:** Diagnostic stability at age 9 years was very high for autism at age 2 years and less strong for pervasive developmental disorder not otherwise specified. Judgment of experienced clinicians, trained on standard instruments, consistently added to information available from parent interview and standardized observation.

*Arch Gen Psychiatry.* 2006;63:694-701

**A**UTISM 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.<sup>1,2</sup> While parents typically report concerns in the first year of life,<sup>3</sup> 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,<sup>4-7</sup> and diagnoses made by age 5 years are stable up to late adolescence.<sup>8</sup> 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.<sup>9</sup>

Several intervention projects reported diagnostic changes and extraordinary levels of improvement in a substantial minority of young children with autism.<sup>10,11</sup> Other reports found little diagnostic change and fewer marked improvements.<sup>12,13</sup> Possible expla-

nations for these conflicting results are diagnostic instability or the lack of age-appropriate diagnostic criteria for very young children. In addition, epidemiological,<sup>14</sup> genetic,<sup>15</sup> and diagnostic studies<sup>16</sup> 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),<sup>17,18</sup> atypical autism, and Asperger syndrome.<sup>19,20</sup> Recently, investigators have begun to ask about the stability for the broader autism spectrum disorder (ASD) as well as for more narrowly defined autism.<sup>21</sup>

High stability has been found for clinical diagnoses between ages 2 and 3 years when health care professionals interpreted standard criteria for autism.<sup>4-6,22</sup> 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.<sup>9</sup> At age 2 years, children with severe retardation were over-

**Author Affiliations:** University of Michigan, Ann Arbor (Drs Lord and Risi); University of North Carolina, Chapel Hill (Dr DiLavore); Hebrew University, Jerusalem, Israel (Dr Shulman); National Institute of Mental Health, Bethesda, Md (Dr Thurm); University of Manchester, Manchester, England (Dr Pickles).

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

Variable	Diagnosis, 2 y			Diagnosis, 9 y		
	Autism (n = 84)	PDD-NOS (n = 46)	Nonspectrum (n = 42)*	Autism (n = 100)	PDD-NOS (n = 35)	Nonspectrum (n = 37)*
Female, %	14	11	40	14	14	41
White, %†	68	72	76	65	83	76
African American, %	30	26	21	31	17	24
Age, mo, at baseline assessment at 2 y, mean (SD)	29.1 (4.7)	29.1 (5.6)	28.8 (5.5)	29.0 (4.9)	30.3 (5.3)	27.8 (5.5)
Age, mo, at follow-up at 9 y, mean (SD)	110.1 (15.7)	113.8 (17.1)	114.9 (11.8)	111.5 (16.5)	111.1 (15.8)	115.6 (11.0)
Limited speech, 2 y, %‡	74	57	50	74	51	46
Limited speech, 9 y, %	30	4	7	28	0	6
VABC, <sup>23</sup> 2 y, mean (SD)	61.0 (12.3)	64.1 (10.3)	65.7 (9.4)	61.0 (12.3)	64.7 (8.2)	66.7 (9.7)
VABC, 9 y, mean (SD)	43.7 (22.8)	57.4 (26.6)	58.2 (27.4)	39.3 (18.6)	69.4 (22.6)	63.9 (29.1)
Ratio verbal IQ, 2 y, mean (SD)	26.4 (15.3)	45.6 (21.7)	57.9 (23.9)	28.5 (16.7)	49.6 (24.3)	58.5 (22.2)
Ratio verbal IQ, 9 y, mean (SD)	41.2 (36.5)	71.7 (36.9)	60.4 (31.1)	35.1 (26.8)	91.5 (32.2)	69.7 (33.4)
Ratio nonverbal IQ, 2 y, mean (SD)	63.3 (16.9)	74.0 (22.3)	72.7 (26.5)	62.7 (19.2)	80.0 (20.0)	73.1 (23.8)
Ratio nonverbal IQ, 9 y, mean (SD)	54.0 (30.9)	75.4 (33.3)	67.9 (33.5)	50.5 (28.9)	88.9 (25.5)	72.7 (34.3)

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.

‡Defined as a score of 2 on the ADI-R overall level of language (<5 words used on a daily basis).

diagnosed with autism and children who did not yet show repetitive behaviors or stereotyped speech were underdiagnosed.<sup>4</sup> Charman and colleagues<sup>9</sup> found that diagnostic thresholds from the ADI-R were crossed and recrossed between ages 2 to 7 years. Moore and Goodson,<sup>5</sup> 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<sup>6</sup> 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 sen-

sory impairments, cerebral palsy, or poorly controlled seizures. In addition, 22 children with developmental delays between ages 13 and 35 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 [5.4] 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,<sup>23</sup> 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.<sup>24</sup> 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<sup>25</sup>; 73 children, Differential Ability Scales<sup>26</sup>; 51 children, Mullen Scales of Early Learning; and 6 children, other. Because raw scores frequently fell outside stan-

**Table 2. ADI-R and ADOS Scores by Initial Best-Estimate Diagnoses at Ages 2 and 9 Years\***

Variable	Diagnosis, 2 y			Diagnosis, 9 y		
	Autism (n = 84)	PDD-NOS (n = 46)	Nonspectrum (n = 42)	Autism (n = 100)	PDD-NOS (n = 35)	Nonspectrum (n = 37)
ADI-R social domain, 2 y	19.7 (4.2)	14.7 (5.7)	9.7 (5.8)	18.6 (5.2)	15.0 (5.1)	9.6 (6.3)
ADI-R social domain, 9 y	25.0 (5.5)	20.8 (7.1)	13.5 (9.0)	25.4 (4.1)	18.7 (8.0)	11.5 (8.2)
ADI-R nonverbal communication domain, 2 y	10.0 (2.0)	8.3 (3.0)	5.8 (3.5)	9.8 (2.1)	7.5 (2.9)	5.9 (3.8)
ADI-R nonverbal communication domain, 9 y	11.8 (2.9)	8.8 (3.9)	5.4 (4.0)	11.8 (2.4)	7.4 (4.0)	4.7 (4.0)
ADI-R repetitive domain, 2 y	4.1 (1.5)	3.1 (2.3)	2.2 (1.7)	4.0 (1.8)	3.3 (1.9)	1.7 (1.3)
ADI-R repetitive domain, 9 y	5.9 (2.6)	5.5 (3.2)	3.8 (2.8)	6.3 (2.5)	4.9 (3.1)	4.7 (4.0)
ADOS social domain, 2 y	12.6 (1.7)	8.8 (3.4)	4.6 (3.6)	11.6 (3.1)	8.9 (3.4)	4.9 (3.7)
ADOS social domain, 9 y	10.3 (3.0)	7.1 (3.8)	5.0 (3.6)	10.7 (2.3)	5.6 (3.0)	3.6 (3.1)
ADOS communication domain, 2 y	6.5 (1.4)	4.4 (1.8)	2.5 (2.2)	5.9 (1.8)	5.0 (2.4)	2.5 (2.1)
ADOS communication domain, 9 y	6.4 (2.0)	4.8 (2.4)	3.5 (2.2)	6.8 (1.7)	3.9 (1.5)	2.6 (2.0)
ADOS repetitive domain, 2 y	4.0 (1.5)	2.5 (1.4)	0.8 (1.0)	3.6 (1.7)	2.3 (1.5)	1.0 (1.4)
ADOS repetitive domain, 9 y	2.9 (2.1)	1.7 (1.7)	1.3 (1.3)	3.1 (1.9)	1.0 (1.0)	1.0 (1.0)

Abbreviations: ADI-R, Autism Diagnostic Interview–Revised; ADOS, Autism Diagnostic Observation Schedule; PDD-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 al<sup>30</sup> for ranges). The ADI-R totals include “ever” scores. The nonspectrum group included all children with developmental delay.

dard ranges for deviation scores, ratio IQs were calculated separately for verbal and nonverbal subtests.

Three measures of diagnosis were obtained at ages 2 and 5 years. Before the direct assessment, a research associate administered to parents a toddler version of the ADI-R, which included additional questions about early development and symptom onset.<sup>27</sup> The toddler ADI-R is a standardized semistructured interview of 132 questions. It yields a diagnostic algorithm for autism by providing scores in 3 domains, social reciprocity, communication, and restricted, repetitive behaviors, and has items about age at onset. Adequate validity and interrater and test-retest reliability have been established for children from age 3 years to adults.<sup>28</sup> For the purpose of these initial analyses, PDD-NOS was defined post hoc as not meeting autism criteria on the ADI-R but falling within 1 to 2 points of autism cutoffs for algorithm criteria in the social and/or communication domains, with no requirement for repetitive behavior.<sup>29</sup> Immediately after conducting the interview, the research associate dictated a 2-page summary, without scoring the algorithm or referring to individual scores. This text was used in the consensus diagnosis at ages 2 and 5 years.

The Autism Diagnostic Observation Schedule (ADOS)<sup>30,31</sup> and an adaptation for younger children, the Pre-Linguistic Autism Observation Schedule (PL-ADOS),<sup>32</sup> provided standardized observation of social and communicative behavior. In 1999, the PL-ADOS and the former ADOS<sup>31</sup> 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.<sup>32</sup> 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.<sup>30</sup> 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.

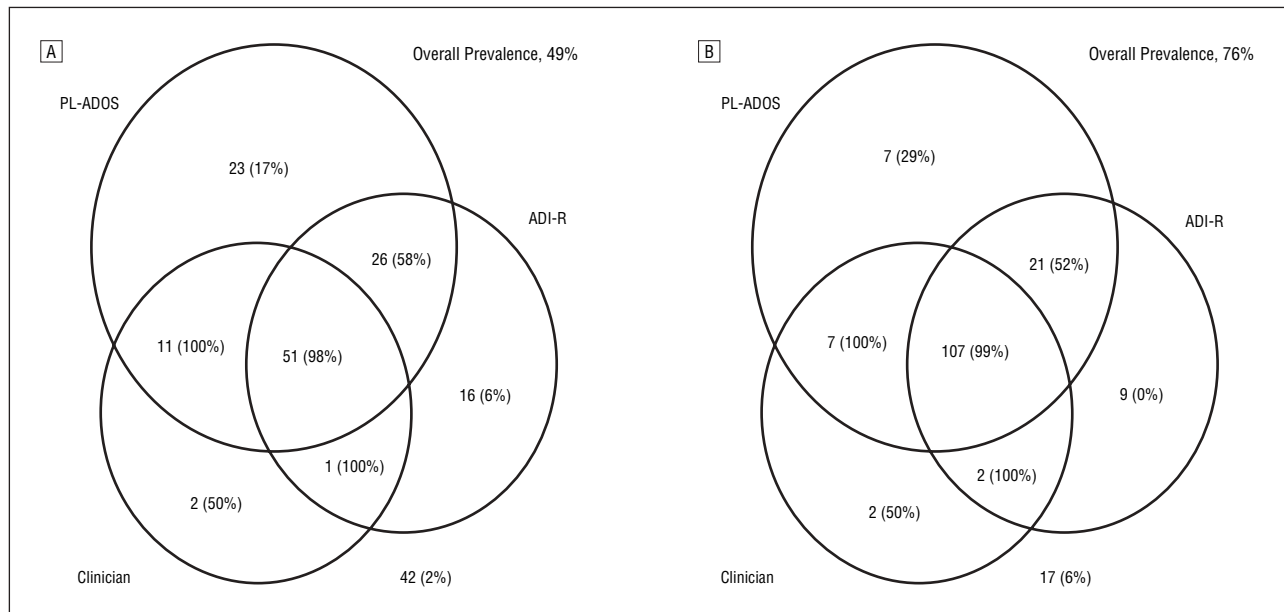
At initial assessment, a PL-ADOS (n=172) was administered to all subjects referred for evaluation for autism and with developmental delay. At age 5 years, the PL-ADOS (n=119) or ADOS module 2 (n=11) was administered. At age 9 years, the ADOS

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 in **Table 2**, unsurprisingly, children described as having PDD-NOS received lower scores on diagnostic measures, indicating fewer or less severe symptoms.

One examiner carried out the assessment at age 5 years for each child and followed the procedures described earlier to make a clinical diagnosis. In about two thirds of cases, examiners were unfamiliar with the child. For the 9-year-olds, most cases were seen by 2 examiners, both unfamiliar with the child: 1 for the ADI-R/Vineland Adaptive Behavior Scales and one for the ADOS and psychometrics. The clinical diagnosis was made jointly.

For the best-estimate diagnoses at both 2 and 5 years of age, 2 psychologists considered the independent clinical diagnosis, the ADI-R and ADOS algorithm scores, and the cognitive, language, and adaptive test scores. They read the ADI-R notes, watched the PL-ADOS/ADOS videotape, and discussed all the findings from that age until they reached a consensus. Following *DSM-IV*, distinctions between autism and PDD-NOS were made on the basis of number of domains affected as well as the



**Figure 1.** Frequency of diagnostic combinations and contemporaneous best-estimate diagnosis prevalence (in parentheses) at age 2 years. A, Autism. B, Autism spectrum. PL-ADOS indicates Pre-Linguistic Autism Diagnostic Observation Schedule; ADI-R, Autism Diagnostic Interview–Revised.

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.<sup>33</sup> 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.<sup>34</sup> 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.<sup>35</sup>

To assess the effect of treatment, there was a need to take account of children's differential access to treatment.<sup>35</sup> To con-

trol 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.<sup>36</sup> 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, 81.7% (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\***

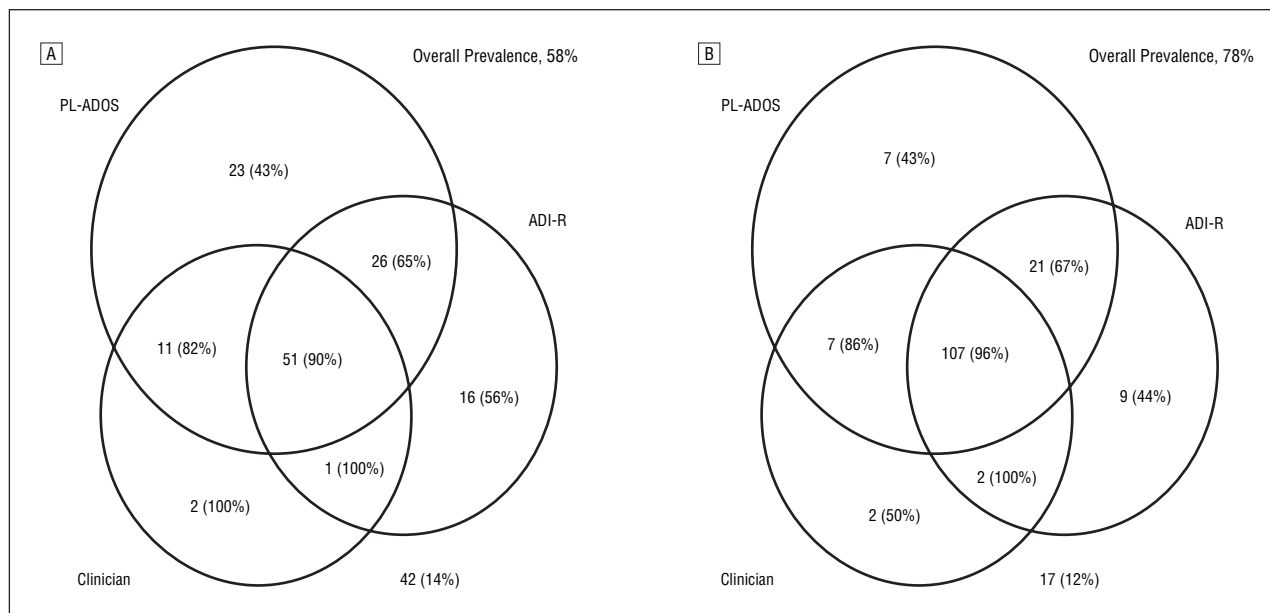
Age 2 Diagnostic Measure	No. (%)‡	Best-Estimate Diagnosis, 2 y			Best-Estimate Diagnosis, 9 y		
		Autism	PDD-NOS	Nonspectrum	Autism	PDD-NOS	Nonspectrum
ADI-R							
Autism	94 (55)	67	18	9	73	16	5
PDD-NOS	45 (26)	15	19	11	20	14	10
Nonspectrum	33 (19)	2	9	22	7	5	21
ADOS							
Autism	111 (65)	80	24	7	82	22	7
PDD-NOS	31 (18)	3	19	9	14	8	9
Nonspectrum	30 (17)	1	3	26	4	5	21
Clinical							
Autism	65 (38)	63	2	0	58	6	1
PDD-NOS	53 (31)	19	32	2	32	16	5
Nonspectrum	54 (31)	2	12	40	10	13	31
Best-estimate							
Autism	84 (49)				71	12	1
PDD-NOS	46 (27)				27	14	5
Nonspectrum	42 (24)				2	9	31
No. (%)‡		102 (49)	59 (28)	53 (25)	100 (58)	35 (20)	37 (22)

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% confidence interval [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 (social,  $P = .07$ ; communication,  $P = .01$ ; repetitive,  $P = .03$ ). When verbal IQ ( $P < .001$ ) and nonverbal IQ ( $P < .60$ ) 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 ( $P = .01$ ) and nonverbal ( $P = .90$ ) 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

diagnoses of PDD-NOS at age 2 years received nonspectrum best-estimate diagnoses (ie, not autism or ASD) by age 9 years, and nearly 30% continued to receive diagnoses of PDD-NOS, indicating mild symptoms at age 9 years. A significant minority of children with milder difficulties within ASD at age 2 years showed only mild deficits in the clinical ASD range at age 9 years. Classifications changed substantially more often from ages 2 to 5 years than from ages 5 to 9 years. The bulk of change in diagnosis occurring in early years is consistent with another recent study.<sup>9</sup> At age 2 years, diagnostic groups were more similar in functioning and IQ than the diagnostic groups identified at age 9 years, when the autistic group showed very poor adaptive functioning and the PDD-NOS group, much less abnormal 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 clinical 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.<sup>16</sup>

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 psychologist or developmental pediatrician. The use of standardized measures seems likely to have improved the stability 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 classifications yielded by both instruments were not supported by the clinicians at age 2 years, 40% were children with severe mental retardation (and not autism) or children 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 consistently underinclusive at age 2 years, both for narrow diagnoses of autism and for broader classifications of ASD at age 9 years. Thus, scores from standardized instruments also made real contributions beyond their influence on informing and structuring clinical judgment. Overall, 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 measures, was of limited prognostic value. Nonetheless, both instruments and clinical judgment added to the prediction 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 repetitive domains, which made independent contributions, similar to the findings of Charman and colleagues.<sup>9</sup> These and other findings support the conceptualization of ADI-R and

ADOS social and nonverbal communication items as reflecting 1 factor. The limitations of the repetitive domain score of the PL-ADOS/ADOS, based on a brief sample of behavior, are well understood,<sup>29,30</sup> and several studies have found that a significant number of children who receive autism diagnoses in later preschool years are not described as having repetitive behaviors before 30 months of age.<sup>4,6,22</sup> To find the repetitive domain score from the ADI-R and the PL-ADOS/ADOS so strongly predictive of prognosis for autism and ASD 7 years later, both before and after verbal IQ was taken into account, was surprising. As expected, low verbal IQ was also associated with increased probability of an outcome of autism or ASD.<sup>9</sup> As a group, children with uncertain clinical diagnoses and high verbal 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 most likely to change diagnosis from autism to PDD-NOS and PDD-NOS to nonspectrum categories at age 9 years and were least likely to show losses in adaptive behavior at age 9 years (and so have relatively better outcome in everyday skills).

As reported elsewhere,<sup>9</sup> the overall totals on the ADI-R and ADOS were not systematically related to change in autistic 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 encouraging 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 children 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 experienced 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 contributed to the clinicians' consistently underinclusive judgments, a finding replicated by others,<sup>9</sup> 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 Carolina and Chicago) included many more children who actually 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 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 important to remember that we cannot determine how many children were not referred who should have been.

In turn, clinicians in the study, using standardized instruments and their own judgments to integrate information 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 MH066469 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 Thurm'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, Shanping Qiu, MA, and Lyn Carpenter, PhD, for data collection and management assistance.

## REFERENCES

- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision*. Washington, DC: American Psychiatric Association; 2000.
- World Health Organization. *The ICD-10 Classification of Mental and Behavioral Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva, Switzerland: World Health Organization; 1992.
- De Giacomo A, Fombonne E. Parental recognition of developmental abnormalities in autism. *Eur Child Adolesc Psychiatry*. 1998;7:131-136.
- Lord C. Follow-up of two-year-olds referred for possible autism. *J Child Psychol Psychiatry*. 1995;36:1365-1382.
- Moore V, Goodson S. How well does early diagnosis of autism stand the test of time? follow-up study of children assessed for autism at age 2 and development of an early diagnostic service. *Autism*. 2003;7:47-63.
- Stone WL, Lee EB, Ashford L, Brissie J, Hepburn SL, Coonrod EE, Weiss BH. Can autism be diagnosed accurately in children under 3 years? *J Child Psychol Psychiatry*. 1999;40:219-226.
- Gillberg C, Ehlers S, Schaumann H, Jakobsson G, Dahlgren SO, Lindblom R, Bagenholm A, Tjuus T, Blidner E. Autism under age 3 years: a clinical study of 28 cases referred for autistic symptoms in infancy. *J Child Psychol Psychiatry*. 1990; 31:921-934.
- McGovern CW, Sigman M. Continuity and change from early childhood to adolescence in autism. *J Child Psychol Psychiatry*. 2005;46:401-408.
- Charman T, Taylor E, Drew A, Cockerill H, Brown JA, Baird G. Outcome at 7 years of children diagnosed with autism at age 2: predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *J Child Psychol Psychiatry*. 2005;46:500-513.
- Lovaas OI. Behavioral treatment and normal educational and intellectual functioning in young autistic children. *J Consult Clin Psychol*. 1987;55:3-9.
- Strain PS. Generalization of autistic children's social behavior change: effects of developmentally integrated and segregated settings. *Analysis Intervent Dev Disabilities*. 1983;3:23-34.
- Sheinkopf SJ, Siegel B. Home-based behavioral treatment of young children with autism. *J Autism Dev Disord*. 1998;28:15-23.
- Smith T, Groen AD, Wynn JW. Randomized trial of intensive early intervention for children with pervasive developmental disorder. *Am J Ment Retard*. 2000; 105:269-285.
- Wing L, Gould J. Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. *J Autism Dev Disord*. 1979;9:11-29.
- Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, Rutter M. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med*. 1995;25:63-77.
- Volkmar FR, Klin A, Siegal B, Szatmari P, Lord C, Campbell M, Freeman BJ, Cicchetti DV, Rutter M, Kline W, Buitelaar J, Hattab Y, Fombonne E, Fuentes J, Werry J, Stone W, Kerbeshian J, Hoshino Y, Bregman J, Loveland K, Szymanski L, Towbin K. Field trial for autistic disorder in DSM-IV. *Am J Psychiatry*. 1994;151:1361-1367.
- Buitelaar JK, Van der Gaag R, Klin A. Exploring the boundaries of pervasive developmental disorder not otherwise specified: analyses of data from the DSM-IV autistic disorder field trial. *J Autism Dev Disord*. 1999;29:33-43.
- Towbin KE. Pervasive developmental disorder not otherwise specified. In: Cohen DJ, Volkmar FR, eds. *The Handbook of Autism and Other Pervasive Developmental Disorders*. New York, NY: John Wiley & Sons; 1997:123-147.
- Volkmar FR, Klin A, Schultz R, Bronen R, Marans WD, Sparrow S, Cohen DJ. Asperger's syndrome. *J Am Acad Child Adolesc Psychiatry*. 1996;35:118-123.
- Gillberg C. Asperger syndrome and high-functioning autism. *Br J Psychiatry*. 1998; 172:200-209.
- Tanguay PE. Pervasive developmental disorders: a 10-year review. *J Am Acad Child Adolesc Psychiatry*. 2000;39:1079-1095.
- Cox A, Klein K, Charman T, Baird G, Baron-Cohen S, Swettenham J, Drew A, Wheelwright S. Autism spectrum disorders at 20 and 42 months of age: stability of clinical and ADI-R diagnosis. *J Child Psychol Psychiatry*. 1999;40:719-732.
- Sparrow S, Balla D, Cicchetti D. *Vineland Adaptive Behavior Scales*. Circle Pines, Minn: American Guidance Service; 1984.
- Mullen E. *The Mullen Scales of Early Learning*. Circle Pines, Minn: American Guidance Service Inc; 1995.
- Wechsler D. *Wechsler Intelligence Scale for Children*. 3rd ed. San Antonio, Tex: Psychological Corp; 1991.
- Elliott CD. *Differential Ability Scales (DAS)*. San Antonio, Tex: Psychological Corp; 1990.
- Lord C, Shulman C, DiLavore P. Regression and word loss in autistic spectrum disorders. *J Child Psychol Psychiatry*. 2004;45:936-955.
- Lord C, Rutter M, LeCouteur A. The Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord*. 1994;24:659-685.
- Risi S, Lord C, Corsello C, Chrysler C, Szatmari P, Cook EH, Leventhal BL, Pickles A. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry*. In press.
- Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, Pickles A, Rutter M. 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.
- Lord C, Rutter ML, Goode S, Heemsbergen J, Jordan H, Mawhood L, Schopler E. Autism Diagnostic Observation Schedule: a standardized observation of communicative and social behaviour. *J Autism Dev Disord*. 1989;19:185-212.
- DiLavore PC, Lord C, Rutter M. The Pre-Linguistic Autism Diagnostic Observation Schedule. *J Autism Dev Disord*. 1995;25:355-379.
- StataCorp. *Stata Statistical Software, Release 8.0*. College Station, Tex: StataCorp; 2003.
- Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement of partial credit. *Psychol Bull*. 1968;70:213-220.
- Pickles A. Generalized estimating equations. In: Armitage P, Colton T, eds. *The Encyclopedia of Biostatistics*. Vol 2. New York, NY: John Wiley & Sons; 1998:1626-1637.
- Hogan JW, Lancaster T. Instrumental variable and inverse probability weighting for causal inference from longitudinal observational studies. *Stat Methods Med Res*. 2004;13:17-48.