Combining Information From Multiple Sources in the Diagnosis of Autism Spectrum Disorders

SUSAN RISI, PH.D., CATHERINE LORD, PH.D., KATHERINE GOTHAM, B.A., CHRISTINA CORSELLO, PH.D., CHRISTINA CHRYSLER, B.A., PETER SZATMARI, M.D., EDWIN H. COOK, Jr., M.D., BENNETT L. LEVENTHAL, M.D., AND ANDREW PICKLES, PH.D.

ABSTRACT

Background: Standard case criteria are proposed for combined use of the Autism Diagnostic Interview-Revised and Autism Diagnostic Observation Schedule to diagnose autism and to define the broader category of autism spectrum disorders. **Method:** Single and combined Autism Diagnostic Interview-Revised and Autism Diagnostic Observation Schedule algorithms were compared to best estimate diagnoses in four samples: U.S. (n = 960) and Canadian (n = 232) participants 3 years and older, U.S. participants younger than 36 months (n = 270), and U.S. participants older than 36 months with profound mental retardation (n = 67). **Results:** Sensitivities and specificities of 80% and higher were obtained when strict criteria for an autism diagnosis using both instruments were applied in the U.S. samples, and 75% or greater in the Canadian sample. Single-instrument criteria resulted in significant loss of specificity. Specificity was poor in the sample with profound mental retardation. Lower sensitivity and specificity were also obtained when proposed criteria for broader spectrum disorders were applied. **Conclusions:** The Autism Diagnostic Interview-Revised and Autism Diagnostic Observation Schedule make independent, additive contributions to the judgment of clinicians that result in a more consistent and rigorous application of diagnostic criteria. *J. Am. Acad. Child Adolesc. Psychiatry*, 2006;45(9):1094–1103. **Key Words:** autism diagnosis, Autism Diagnostic Interview-Revised, Autism Diagnostic Observation Schedule.

Autism research has benefited from opportunities to define samples by diagnostic instruments such as the Autism Diagnostic Interview-Revised (ADI-R; Rutter et al., 2003) and the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1999). Standardized methods of collecting, coding, and summarizing information result in categorical diagnoses of autism or not autism on the ADI-R, a caregiver interview, and in classifications of autism, broader autism spectrum disorders (ASDs) or nonspectrum on the ADOS, a semistructured observation. These two instruments were intended to be used together, yet there has been no systematic attempt to evaluate how information from the instruments should be combined for diagnosis.

The core characteristics of autism are deficits in communication and social reciprocity accompanied by behavior that is restricted or repetitive (*DSM-IV-TR*; American Psychiatric Association, 2000). The ADI-R and ADOS were developed to operationalize these criteria to identify characteristics that differentiated autism from cases without autism that were equivalent in chronological age and language level. Individual items were not selected for the ADI-R and the ADOS

Accepted March 23, 2006.

Drs. Risi and Lord and Ms. Gotham are with the University of Michigan Autism and Communication Disorders Center, Ann Arbor; Dr. Corsello is with the Department of Psychiatry, University of California, San Diego; Ms. Chrysler and Dr. Szatmari are with the Offord Centre for Child Studies, McMaster University, Hamilton, Canada; Drs. Cook and Leventhal are with the Department of Psychiatry, University of Illinois at Chicago; Dr. Pickles is with the University of Manchester, United Kingdom.

This work was supported by grants NIMH R01 MH066496 and R01 MH46865 to Dr. Lord and was carried out as part of the NICHD/NIDCD Collaborative Programs for Excellence in Autism. The authors gratefully acknowledge the families who participated in these research projects and the staff at the TEACCH centers in North Carolina, the Department of Psychiatry at the University of Chicago, and the University of Michigan Autism and Communication Disorders Center.

Correspondence to Dr. Susan Risi, UMACC, 1111 East Catherine, AnnArbor, MI 48109-2054; e-mail: srisi@umich.edu.

DOI:10.1097/01.chi.0000227880.42780.0e

algorithms because of their ability to discriminate autism from other ASDs or to discriminate more broadly defined ASD from other disorders. Genetics research has indicated that the boundaries of what is transmitted familially, however, extend beyond autism as operationally defined on the ADI-R and ADOS (International Molecular Genetics Study of Autism Consortium [IMGSAC], 2001; Le Couteur et al., 1996). This has led to increasing interest in including in research individuals who do not meet criteria for autism, but who share many of the same characteristics (Constantino et al., 2003). Several different operational definitions of "almost autism," ASD, and pervasive developmental disorder not otherwise specified exist (Buitelaar et al., 1999; International Molecular Genetics Study of Autism Consortium [IMGSAC], 2001). Unfortunately, studies have indicated that interrater reliability in distinguishing nonautism ASD from autism and in distinguishing ASD from nonspectrum disorders is often poor (Lord et al., 1999; Szatmari et al., 2002).

The purpose of this article is to propose standard criteria for the combined use of the ADI-R and ADOS to diagnose cases of autism and to identify a broader category of ASD cases that may have less pervasive or milder symptoms. Data are also presented about the instruments' performance with very young children and for individuals with profound mental retardation. In a clinical setting, wrongly denying a child access to services may be the greatest concern,whereas in genetic analyses, wrongly designating unaffected individuals as affected may have more negative consequences. Consequently, data for alternative methods are presented so that these considerations can be taken into account when selecting criteria.

An inherent difficulty in establishing caseness criteria is determining the gold standard to which classifications derived from the diagnostic instruments are compared. Because reporting clinical diagnoses based on information other than the ADI-R or ADOS when those instruments were used is impractical, our solution was to seek replication across different sites. In this study, data are reported from U.S. and Canadian centers that used different strategies for determining consensus best estimate (BE) diagnoses. Standardized administrations of the diagnostic instruments were performed at both sites, but how the information was used to determine diagnoses differed. In the U.S. samples, consensus BE diagnoses were not independent of the diagnostic instruments. In almost all cases, a psychologist conducted or observed both the ADI-R and the ADOS and summary information from the ADI-R and ADOS was available to physicians who participated in the diagnoses. In the Canadian sample, consensus BE diagnoses were made by physicians and psychologists who had not been directly involved in the ADI-R or ADOS administration but who had access to the clinical information from these instruments.

METHOD: STUDY 1

Participants

Data were collected from 1,039 participants who completed a diagnostic evaluation at the University of Chicago Developmental Disorders Clinic ($N = 627$; 497 males, 130 females), the University of Michigan Autism and Communication Disorders Center ($N = 150$; 115 males, 35 females), as part of a longitudinal study conducted through TEACCH Centers at the University of North Carolina, Chapel Hill ($N = 129$; 100 males, 29 females) and the University of Chicago ($N = 80$; 69 males, 11 females), or in a study of individuals with disorders other than ASD ($N = 53$; 37 males, 16 females). Only participants with known developmental, cognitive, or behavioral diagnoses were included. One hundred twenty-six participants (12%) were affected siblings. The sample was 82% white, 13% African American, 4% Asian American, and 1% other or multiracial. Participants with visual, hearing, or motor impairments that precluded standard administration of an instrument were excluded. Parents signed an institutional review board–approved informed consent form to participate in research before actual participation.

The majority of the 1,039 participants received ASD diagnoses. However, 158 (15%) of the participants had diagnoses other than ASD (41% nonspecific mental retardation, 25% language disorder, 14% oppositional defiant disorder and/or attention-deficit/hyperactivity disorder, 12% Down syndrome, 7% mood and/or anxiety disorder, and 1% Tourette's syndrome). About 60% of nonspectrum participants had been referred for possible ASD; the remaining 40% were recruited into research comparison groups.

For 182 participants (18%), more than one full assessment was available. No differences were found for separate analyses of all data compared with analyses with only the most recent assessment for each participant. Thus, all 1,297 assessments with contemporaneous ADI-R and ADOS administrations were included. Age at assessment ranged from 14 months to 18 years, with a median of 58 months of age. All sites in this study primarily evaluate individuals referred for possible ASD.

The largest data set consisted of 960 assessments of participants at least 36 months old who had a nonverbal mental age of at least 18 months (Table 1). This data set included cases with clinical diagnoses of autism (540 assessments), pervasive developmental disorder not otherwise specified (PDD-NOS; 252 assessments), Asperger disorder (5 assessments; most cases with a diagnosis of Asperger disorder had a research diagnosis of autism, which was given precedence), or a nonspectrum disorder (163 assessments). In addition, performance of diagnostic criteria for autism and ASD compared with nonspectrum disorders was examined separately for children younger than 36 months' chronological age (270 assessments) and for participants older than 3 years of age with profound mental

TABLE 1
Participant Characteristics: Three U.S. Samples and One Canadian Sample (Means and SDs)

	Study 1: U.S.									Study 2: Canada		
Sample	≥3 Yr (*n* = 960)			<36 Mo (*n* = 270)			Profound Mental Retardation (*n* = 67)			≥3 Yr (*n* = 232)		
BE Diagnosis	Autism (*n* = 540)	Nonautism ASDs (*n* = 257)	Nonspectrum (*n* = 163)	Autism (*n* = 162)	Nonautism ASDs (*n* = 65)	Non-spectrum (*n* = 43)	Autism (*n* = 45)	Nonautism ASDs (*n* = 12)	Non-spectrum (*n* = 10)	Autism (*n* = 184)	Nonautism ASDs (*n* = 19)	Non-spectrum (*n* = 29)
Demographics												
Gender (M/F)	449/91	199/58	106/57	137/25	57/8	21/22	36/9	8/4	4/6	152/32	15/4	22/7
Age, mo	77.4 (34.5)	85.8 (38.8)	86.1 (41.9)	29.5 (4.4)	28.0 (5.7)	27.8 (6.1)	74.1 (37.7)	46.9 (10.1)	66.6 (23.3)	94.4 (37.6)	87.4 (41.5)	96.3 (35.7)
Verbal IQ	47.9 (28.6)	83.2 (28.9)	78.6 (24.9)	31.5 (19.7)	51.2 (23.7)	63.8 (24.8)	17.8 (12.3)	21.0 (11.0)	24.1 (14.2)	—	—	—
Nonverbal IQ	66.5 (26.1)	88.7 (24.1)	79.3 (26.1)	67.0 (18.2)	79.4 (22.5)	75.6 (23.7)	25.3 (10.2)	29.7 (10.5)	24.4 (7.4)	80.3 (19.8)	94.1 (23.9)	92.8 (16.1)
VABC Composite (SS)	52.0 (16.7)	65.5 (16.3)	65.5 (19.7)	61.4 (6.2)	66.5 (10.9)	69.0 (9.2)	39.9 (13.2)	44.1 (7.3)	41.6 (15.4)	63.6 (17.4)	72.0 (14.2)	78.8 (21.7)
ADI-R												
Social	21.1 (6.2)	14.3 (7.2)	9.1 (6.4)	19.3 (4.4)	14.7 (5.8)	8.4 (6.4)	23.6 (5.4)	18.3 (5.3)	17.8 (7.8)	—	—	—
Nonverbal communication	10.4 (3.3)	6.5 (3.8)	4.6 (3.8)	11.7 (2.3)	9.7 (3.3)	6.9 (4.1)	12.1 (3.2)	10.8 (3.6)	8.9 (5.1)	—	—	—
Verbal communication	17.0 (4.1)	12.1 (5.2)	8.1 (5.5)	14.6 (5.3)	13.5 (3.5)	5.4 (3.4)	14.5 (7.3)	—	7.0	—	—	—
Repetitive behaviors	6.1 (2.4)	4.7 (2.8)	3.1 (2.5)	4.2 (1.7)	3.3 (2.2)	1.7 (1.3)	5.2 (1.6)	4.0 (1.9)	2.8 (2.0)	—	—	—
ADOS												
Communication and social	16.2 (3.5)	9.8 (4.1)	5.0 (3.6)	17.7 (2.5)	12.8 (4.3)	6.1 (4.6)	16.9 (2.5)	15.2 (4.3)	13.9 (5.1)	—	—	—
Repetitive behaviors	3.3 (1.8)	1.6 (1.5)	0.9 (1.2)	3.7 (1.6)	2.4 (1.5)	0.6 (0.8)	4.6 (1.4)	3.8 (2.0)	2.5 (2.3)	—	—	—

Note: BE = best estimate; ASDs = autism spectrum disorders; ADI-R = Autism Diagnostic Interview-Revised; ADOS = Autism Diagnostic Observation Schedule; VABC = Vineland Adaptive Behavior Composite; M = male; F = female; SS = standard score.

retardation (i.e., nonverbal mental ages at or younger than 18 months [67 assessments]).

A standard developmental hierarchy of measures, most frequently the Mullen Scales of Early Learning (Mullen, 1995) and the Differential Ability Scales (Elliot, 1990), were used to determine IQ scores. For 1,251 assessments, a caregiver also completed the Vineland Adaptive Behavior Scales (Sparrow et al., 1984).

Procedure

The ADI-R was administered to a caregiver by a clinical psychologist or a trainee and followed within a few days by the same psychologist and/or a trainee completing psychometric testing and the ADOS. Several days later, physicians were given a one-page summary of information from the ADI-R, ADOS, and other testing before seeing the participants and their families for a 1-hour unstandardized interview and observation. Immediately following the final visit, a consensus BE diagnosis was determined by all of the clinicians after reviewing all of the information. Clinic-referred participants who participated received oral feedback and a report without financial compensation. Participants who were seen only for research purposes received compensation and an evaluation summary.

Measures

All of the examiners in the study had completed research training and met standard requirements for reliability (see Lordet al., 1999; Rutter et al., 2003). Interrater reliability was monitored through joint administration and scoring by two examiners for at least one in 10 cases and through scoring of videotapes every 6 months. Agreement remained at >85%. For the U.S. samples, 26 examiners collected the data from the ADI-R and ADOS over a 10-year period.

The ADI-R algorithm consists of 42 items organized according to *DSM-IV-TR* (American Psychiatric Association, 2000). A classification of autism on the ADI-R requires that a participant meet or exceed thresholds in social reciprocity, verbal or nonverbal communication, and repetitive behaviors, as well as have evidence of onset before 36 months. There are no standard cutoffs on the ADI-R for ASD, atypical autism, PDD-NOS, or Asperger disorder. Several possible thresholds for nonautism ASDs proposed by different investigators were examined (e.g., International Molecular Genetic Study of Autism Consortium [IMGSAC], 2001; Sung et al., 2005). We report data from two of these cutoffs.

The ADOS diagnostic algorithms and thresholds differ across four modules, although there is substantial item overlap. Selection of module is determined by the language level of the participant. On the ADOS, an autism classification requires meeting or exceeding thresholds in the social reciprocity and communication domains and total. Unlike the ADI-R, the ADOS provides an algorithm for nonautism ASDs, which consists of lower thresholds for each domain as well as the total. In Table 1, ADOS domain scores were converted to equivalent scores for modules 3 (n = 282) and 4 (n = 45) to be comparable with modules 1 and 2 (module 1: n = 742; module 2: n = 251).

Analyses

We report results for single and combined ADI-R and ADOS algorithms compared with BE clinical diagnosis as the gold standard (Dunn, 2000) using standard signal detection methods as implemented in the diagt procedure for Stata 8 (Seed and Tobias, 2001). This procedure evaluates binary tests and provides standard measures of performance including sensitivity, specificity, likelihood ratios, and predictive values with appropriate confidence intervals. Because some participants were assessed more than once, 95% confidence intervals were calculated using a bootstrap procedure that recognized the clustering of the data at the participant level. A test is evaluated by the extent to which it identifies individuals with the disorder (high sensitivity) and excludes those without the disorder (high specificity; Kraemer, 1992). A high positive predictive value indicates that a positive test result is strongly suggestive of the disorder. A high negative predictive value indicates that a negative test result is strongly suggestive of not having the disorder. These values are influenced by sample prevalence. Both for comparison and because the instruments were intended for use in contexts similar to the U.S. samples, that is, specialized referral and research centers for ASDs, the predictive values for the Canadian sample were calculated using the U.S. sample prevalence of autism (56.3%).

RESULTS: STUDY 1

Autism Diagnosis Case Criteria

Preliminary correlations between the diagnostic instruments showed the ADI-R total scores correlated 0.57 with ADOS total scores. Table 2 shows the results from applying five criteria using the ADI-R, the ADOS, or a combination of both for establishing an autism diagnosis, ordered from the most stringent (requiring autism diagnoses from both instruments) to the least stringent (requiring an autism diagnosis from either instrument).

Cases 3 Years and Older. In the sample of subjects 3 years and older, meeting autism criteria on both instruments resulted in the fewest false positives and the most false negatives (i.e., cases that did not meet instrument criteria but had BE clinical diagnoses of autism), with correspondingly high PPVs and lower NPVs. Of the false-positive cases that resulted from applying the two most stringent criteria, the majority had clinical diagnoses of nonautism ASDs (90% and 89%, respectively). Of the false-negative cases using the most stringent criterion, 35% were identified as positive when the criterion was relaxed to use ADOS (ASD). Thus, the primary difference between the BE judgments and the combined ADI-R/ADOS autism criterion was the exclusion or inclusion of individuals with nonautism ASD diagnoses.

Neither instrument worked as well singly as it did in combination. Sensitivity was similar for the ADI-R and ADOS with somewhat higher specificity for the ADOS. Using a criterion of meeting autism criteria on either the ADI-R or ADOS resulted in specificity <50% and identified ≈29% of the nonspectrum participants as having autism.

Cases Younger Than 36 Months. Preliminary correlations between the diagnostic instruments showed the ADI-R total scores correlated 0.60 with ADOS total

TABLE 2
Autism Criteria

Diagnostic Criteria		No. of True Positives	No. of True Negatives[a]	No. of False Positives[b]	No. of False Negatives	Sensitivity (95% CI)[c]	Specificity (95% CI)[c]	PPV (95% CI)[c]	NPV (95% CI)[c]
ADI-R and ADOS (AUT)	U.S.: 3+	443	361 (204)	59 (53)	97	82.0% (78–85)	86.0% (83–89)	88.3% (85–91)	78.8% (75–83)
	CAN: 3+	142	36 (10)	12 (9)	42	77.2% (70–83)	75.0% (60–86)	79.9%	72.0%
	U.S.: <36	131	94 (53)	14 (12)	31	80.9% (74–86)	87.0% (81–93)	90.3% (86–95)	75.2% (68–82)
	U.S.: MR	41	11 (6)	11 (6)	4	91.1% (83–98)	50.0% (31–75)	78.9% (67–90)	73.3% (50–93)
ADI–R and ADOS (ASDs)	U.S.: 3+	476	305 (155)	115 (102)	64	88.2% (85–91)	72.6% (68–76)	80.5% (77–84)	82.7% (79–87)
	CAN: 3+	162	32 (7)	16 (12)	22	88.0% (83–92)	66.7% (52–80)	77.3%	81.2%
	U.S.: <36	134	82 (41)	26 (24)	28	82.7% (77–88)	75.9% (67–83)	83.8% (77–89)	74.6% (66–82)
	U.S.: MR	41	10 (5)	12 (7)	4	91.1% (81–98)	45.5% (24–68)	77.4% (65–88)	71.4% (43–92)
ADI-R	U.S.: 3+	480	248 (123)	172 (134)	60	88.9% (86–91)	59.1% (54–64)	73.6% (70–77)	80.5% (76–85)
	CAN: 3+	175	27 (7)	21 (12)	9	95.1% (91–98)	56.3% (41–71)	73.7%	89.9%
	U.S.: <36	134	78 (40)	30 (25)	28	82.7% (77–88)	72.2% (63–80)	81.7% (75–87)	73.6% (65–82)
	U.S.: MR	41	9 (4)	13 (8)	4	91.1% (82–98)	40.9% (20–61)	75.9% (64–86)	69.2% (41–92)
ADOS (AUT)	U.S.: 3+	497	309 (161)	111 (96)	43	92.0% (90–94)	73.6% (69–78)	81.7% (78–85)	87.8% (84–91)
	CAN: 3+	148	27 (5)	21 (14)	36	80.4% (74–86)	56.3% (41–71)	70.3%	69.2%
	U.S.: <36	158	64 (28)	44 (37)	4	97.5% (95–99)	59.3% (49–68)	78.2% (72–83)	94.1% (87–99)
	U.S.: MR	45	4 (2)	18 (10)	0	100%	18.2% (5–35)	71.4% (61–82)	100%
ADI-R or ADOS (AUT)	U.S.: 3+	534	196 (80)	224 (177)	6	99.0% (98–100)	46.7% (42–52)	70.5% (67–73)	97.0% (94–99)
	CAN: 3+	181	18 (2)	30 (17)	3	98.4% (95–100)	37.5% (24–53)	67.0%	94.8%
	U.S.: <36	161	48 (15)	60 (50)	1	99.4% (98–100)	44.4% (36–55)	72.9% (67–78)	98.0% (92–100)
	U.S.: MR	45	2 (0)	20 (12)	0	100%	9.1% (0–28)	69.2% (58–80)	100%

Note: CI = confidence interval; PPV = positive predictive value; NPV = negative predictive value; ADI-R = Autism Diagnostic Interview-Revised; ADOS = Autism Diagnostic Observation Schedule; AUT = autism; MR = mental retardation; ASDs = autism spectrum disorders.

[a] Numbers in parentheses are cases with clinical diagnoses of pervasive developmental disorder not otherwise specified that did not meet autism criteria.

[b] Numbers in parentheses are cases with clinical diagnoses of pervasive developmental disorder not otherwise specified that met autism criteria.

[c] Ranges from bootstrap analyses.

scores. As shown in Table 2, the two most stringent criteria yielded the same pattern as found in the older sample. Relaxing the criterion to use the ADOS (ASD) cutoff resulted in three additional true-positive cases, but increased false-positive cases by 12. However, all of the new false-positive cases had BE diagnoses of ASDs, suggesting that the broader ADOS algorithm may be sufficient in many circumstances. Again, the majority of children who were false positive for autism (86% and 92%, respectively) had a clinical diagnosis of nonautism ASD. The majority of false negatives (27 of 31) missed meeting criteria on the ADI-R; only one case missed meeting criteria on both measures. Reduction of false-negative cases (with a correrspond-ing increase in true positives) could only be achieved at the cost of lower specificity.

As with the older sample, neither the ADI-R nor the ADOS worked as well singly as they did in combination, although in the younger sample, the ADI-R had higher specificity than the ADOS. Specificity fell to just 44% when criteria allowed either the ADI-R OR the ADOS (AUT), with ≈23% of the nonspectrum participants classified with autism.

Cases 3 Years and Older with Profound Mental Retardation. Preliminary correlations between the diagnostic instruments showed the ADI-R total scores correlated 0.28 with ADOS total scores and ranged between 0.23 and 0.95 for specific domains. Table 2 also shows the results from applying autism case criteria to cases with profound mental retardation older than 3 years of age. The most stringent criterion, ADI-R and ADOS (AUT), resulted in sensitivity of 91% with specificity of 50%. Sensitivities for the remaining criteria and for nonautism ASDs were equal or higher; with specificities even lower.

ASD Diagnosis Case Criteria

Prior analyses indicated that performance of ASD criteria is often obscured by the greater number of autism cases; thus, we elected to analyze cases diagnosed by clinicians with nonautism ASD, excluding cases with the narrower classification of autism. ASD criteria for the ADI-R required careful consideration because no empirically validated or generally accepted rules apply (Rutter et al., 2003). We provide results for two criteria that performed comparatively well.

Cases 3 Years and Older. The first criterion in Table 3 required that ASD criteria be met on the ADOS as well as autism criteria in the domain of social reciprocity, and either communication or restricted, repetitive behavior on the ADI-R (ASD1). The second criterion required meeting ASD criteria on the ADOS and coming within one point on ADI-R social and communication domains, or meeting the ADI-R autism cutoff on one domain and coming within two points on the other. Because results for these strategies were similar, only the second criterion is reported. The second criterion improved sensitivity (61% versus 55%) while losing <1% in specificity.

The two-instrument criterion resulted in greater specificity than either the ADI-R or ADOS singly. Sensitivity and specificity were similar for the ADOS, in contrast to the ADI-R alone, which had better sensitivity but reduced specificity. Either instrument alone resulted in a high number of false positives, with ≈53% of the nonspectrum sample meeting the ASD caseness criterion.

Cases Younger Than 36 Months. Performance with younger cases appeared substantially better than those with older children (Table 3). Both the ADOS (ASD) and the ADI-R (S+C within two points), as single instruments, had high sensitivity and lower specificity. For all the ADI-R and ADOS combined criteria, the discrepancy between sensitivity and specificity was smaller for younger than older participants, with the and-combination of ADI-R (S+C within two points) and ADOS (ASD) yielding relatively equal sensitivity and specificity and good positive and negative predictive values.

METHOD: STUDY 2

Participants

The Canadian sample of 232 mainly white participants was ascertained as part of a genetic study of ASD. Families with more than one child with ASD were recruited; singleton participants were randomly selected for comparison purposes. As shown in Table 1, 79% of the participants had autism, 8% had ASD (PDD-NOS or Asperger disorder), and 13% had nonspectrum diagnoses, most commonly language disorder or learning disability.

Procedure

The ADI-R and ADOS were administered in variable order to all participants. Included in this sample were administrations of older versions of the ADI-R (Lord et al., 1994) and ADOS (Lord et al., 1989). These scores are not directly comparable to the study 1 samples and therefore are not presented. These data were collected by eight examiners during a 15-year period. An independent diagnosis was made on the basis of *DSM-IV-TR* criteria (American Psychiatric Association, 2000) using all available information except previous diagnosis. (See Mahoney et al., 1998 for criteria for

TABLE 3
Autism Spectrum Disorder Criteria

Diagnostic Criteria		No. of True Positives	No. of True Negatives	No. of False Positives[a]	No. of False Negatives[b]	Sensitivity (95% CI)[c]	Specificity (95% CI)[c]	PPV (95% CI)[c]	NPV (95% CI)[c]
ADI-R[d] and ADOS (ASD1)	3+	142	144	19	115	55.3% (50–62)	88.3% (84–93)	88.2% (83–93)	55.6% (49–62)
	<36	46	36	7	19	70.8% (60–82)	83.7% (71–93)	86.8% (76–95)	65.5% (52–79)
	MR	11	2	8	1	91.7% (73–100)	20.0% (0–56)	57.9% (35–83)	66.7% (0–100)
ADI-R[e] and ADOS (ASD2)	3+	157	143	20	100	61.1% (54–67)	87.7% (82–93)	88.7% (83–93)	58.9% (53–66)
	<36	54	34	9	11	83.1% (73–92)	79.1% (66–90)	85.7% (77–94)	75.6% (62–88)
	MR	10	2	8	2	83.3% (63–100)	20.0% (0–57)	55.6% (33–82)	50.0% (0–100)
ADI-R[e](ASD2)	3+	198	92	71	59	77.0% (72–82)	56.4% (49–64)	73.6% (68–79)	60.9% (52–69)
	<36	56	27	16	9	86.2% (77–94)	62.8% (48–77)	77.8% (66–86)	75.0% (59–88)
	MR	11	1	9	1	91.7% (70–100)	10.0% (0–43)	55.0% (33–79)	50.0% (0–100)
ADOS (ASDs)	3+	186	127	36	71	72.4% (67–78)	77.9% (72–84)	83.8% (78–88)	64.1% (57–71)
	<36	63	29	14	2	96.9% (91–100)	67.4% (52–82)	81.8% (72–90)	93.6% (83–100)
	MR	11	1	9	1	91.7% (70–100)	10.0% (0–40)	55.0% (35–79)	50.0% (0–100)
ADI-R[e] or ADOS (ASD2)	3+	227	76	87	30	88.3% (84–92)	46.6% (39–54)	72.3% (67–77)	71.7% (63–81)
	<36	65	22	21	0	100%	51.2% (35–67)	75.6% (65–84)	100%
	MR	12	0	10	0	100%	0.0%	54.6%	—

Note: CI = confidence interval; PPV = positive predictive value; NPV = negative predictive value; ADI-R = Autism Diagnostic Interview-Revised; ADOS = Autism Diagnostic Observation Schedule; MR = mental retardation; ASD = autism spectrum disorder.

[a] All cases have best estimate diagnoses of pervasive developmental disorder not otherwise specified; cases with autism diagnoses were excluded.

[b] All cases have nonspectrum diagnoses.

[c] Ranges from bootstrap analyses.

[d] ADI-R (ASD1): Meets criteria on Social domain *and* meets criteria on either Communication or Behavior domain.

[e] ADI-R (ASD2): Meets criteria on Social *and* Communication domains *or* meets criteria on Social *and* within 2 points of Communication criteria *or* meets criteria on Communication *and* within 2 points of Social criteria *or* within 1 point on both Social and Communication domains.

nonautism ASD diagnoses.) Nonverbal IQ scores were obtained using the Leiter International Performance Scale (Levine, 1986). The Vineland Adaptive Behavior Scale (Sparrow et al., 1984) was completed in face-to-face interviews or by telephone.

RESULTS: STUDY 2

Because only 19 cases in this sample had BE diagnoses of nonautism ASDs, analyses were conducted for autism criteria only. As in study 1, the combination of ADI-R and ADOS autism classifications yielded the most balanced sensitivity and specificity. As shown in Table 2, the positive predictive value of the stringent criteria was high, but because of the small number of nonautism participants, the negative predictive value was much lower than for the U.S. sample. Relaxing the ADOS (AUT) criteria to ADOS (ASD), again increased the sensitivity and decreased the specificity, the latter somewhat more than in the U.S. sample. As in study 1, specificity for any single instrument, and either/or combination of instruments, was poor. The ADOS, unlike for the U.S. participants, did not accurately discriminate cases with autism in the Canadian sample.

DISCUSSION

Autism spectrum diagnostic case criteria that use combined information from the ADI-R and ADOS better reflect consensus clinical judgments of autism and ASD than any single instrument. We do not know whether these results are specific to the ADI-R and ADOS or are specific to the areas covered by the different methods. The ADI-R provides a developmental history, a detailed description of the individual's functioning in a variety of social contexts, and the opportunity to take into account caregivers' perceptions of the severity of different behaviors. The ADOS provides a summary of an experienced clinician's standardized observations of the individual's current behavioral strengths and limitations. Studies with other instruments that deliberately vary these factors could further help specify the critical components of a valid and reliable diagnostic judgment.

Expanding the case definition of autism to ADI-R and ADOS (ASD) criteria rather than using the more restrictive autism criteria on the ADOS allowed identification of more cases with clinical diagnoses of autism or ASD without overdiagnosis of many cases with nonspectrum disorders. However, the more conservative ADI-R and ADOS (AUT) criteria may be preferred for genetics research, where false classification as an affected case may be more problematic than false classification as unaffected or unknown. In contrast, because most false positives for autism consisted of individuals without autism but with ASD, the ADI-R and ADOS (ASD) criteria may be more useful in other research.

Unexpectedly, both instruments performed well in a sample of children younger than 36 months of age, perhaps because at this age, ADI-R algorithm scores are based on current behaviors rather than parents' recollection of the 4- to 5-year period used with older children. Although case criteria were effective in identifying autism in both the older and younger samples, the instruments were consistently over inclusive with a low functioning sample. Because both instruments performed relatively well with children younger than 3 years of age, it is unlikely that the instruments were affected by mental age alone, but rather by the combined effects of low mental age, higher chronological age, and general level of impairment. The stability of autism and ASD diagnoses was not directly analyzed in this study, but is addressed by Lord et al. (2006) in a sample of children diagnosed before 36 months of age.

It was disappointing but not unexpected that standard case criteria for nonautism ASD were less straightforward than case criteria for autism. The combined ADI-R/ADOS case criteria generally failed to improve performance over the single instruments. The and-combinations missed many true cases and the or-combinations were too inclusive. Furthermore, perusal of the ADI-Rand ADOS domain scores for the nonspectrum participants in this study (shown in Table 1) indicated that many nonspectrum participants easily fulfilled two examples of social deficits, one of a communication abnormality and one of a repetitive behavior or interest, thus meeting the total of six symptoms needed for a *DSM-IV* diagnosis of autism. These findings suggest that current *DSM-IV* criteria cannot be interpreted literally for diagnoses for autism and atypical autism or PDD-NOS. The clinicians, as well as the instrument algorithms, required more than the minimum pattern of positive or negative features specified in *DSM-IV* to determine diagnosis.

Limitations

In an ideal design, different examiners would administer the ADI-R and ADOS in random order,

and examiners blind to ADI-R and ADOS scores would make the BE diagnoses. In study 1, however, the ADOS was almost always administered after the ADI-R, and the examiner participated in the consensus diagnosis. Circumstances were closer to the ideal in study 2. Nevertheless, like study 1, study 2 also showed stronger specificity of diagnoses using two-instrument combinations and showed the trade off in specificity and sensitivity that occurs in choosing between the more stringent (ADI-R+ADOS [AUT]) and slightly less stringent criteria (ADI-R+ADOS [ASD]).

In study 1, one of the examiners making the consensus diagnosis was present for both the ADI-R and the ADOS, and the other clinician was provided with all of the information from the instruments including the algorithm scores (except for the first third of the sample when the ADOS algorithms were not yet developed). The finding that even when the instruments agreed the clinicians selected a different consensus diagnosis in >15% of the cases, suggests that the clinicians may be using or weighting information beyond the scores on the diagnostic instruments. When the instruments disagreed, the consensus diagnosis in study 1 was more likely to agree with the ADOS classification, which may reflect an emphasis placed by clinicians present during the ADOS on their own observations, particularly for children with more broadly defined autism. This is in contrast to study 2, in which the clinicians who made the diagnosis were not present during the testing, and when the instruments disagreed, the BE diagnosis was more likely to agree with the ADI-R. This could have occurred because the ADI-R provides broader contexts, historical information, or more complete descriptions of the behavior domains that define autism.

In past epidemiological studies with similar prevalence rates for ASD, rates for autism versus rates of PDD-NOS or Asperger syndrome have fluctuated dramatically (Fombonne, 2005). Our results indicate that this variability in distinguishing within ASD may still occur even when similar instruments and identical cutoffs are used.These biases may be a product of how samples were ascertained and for what purposes, providing a reminder that the original designs underlying secondary data sets may influence results in ways that are not obvious in the final data. Other biases, including demographic characteristics, are also important to consider; Hispanic subjects were not represented in the data for this study.Recognizing these issues may be particularly important as researchers are encouraged to collaborate in establishing large public data sets.

Clinical Implications

The ADI-R and ADOS provide unique and critical overlapping information that informs clinical judgments in making an ASD diagnosis. When results from these instruments are consistent and correspond with clinical impression, diagnostic decision-making is straightforward. Additional testing and alternative hypotheses must be considered when there is inconsistency in results between the instruments and when the instrument results deviate from clinical impression. Clinicians are frequently unable to administer both instruments because of lack of time or expertise. Other instruments that capture the same types of information but are less costly in terms of time to administer and that require less training are needed.

As reported in this article, a valid grouping resulted when a slightly broader classification of autism was operationalized using the ADI-R and ADOS (ASD). This reflects our emerging but limited understanding of ASDs as a classification. Moving out from autism to further extend and clarify the boundaries of ASDs, atypical autism, and PDD-NOS is a much less straightforward question that will require serious consideration and more empirical data (Buitelaar et al., 1999; Walker et al., 2004). More information is needed as to which social deficits are specific to ASDs. The development of metrics for the severity of autism and for broader ASDs may improve our ability to represent what appears to be a continuum (or continua) rather than a discrete classification, despite the fact that categorization is often necessary for clinical work and research.

Disclosure: Drs. Risi, Lord, Corsello, and Pickles receive royalties for the ADI-R and/or ADOS; profits accrued from this study were donated to charity. The other authors have no financial relationships to disclose.

REFERENCES

American Psychiatric Association (2000), *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR)* Washington, DC: American Psychiatric Association

Buitelaar JK, Van der Gaag R, Klin A, Volkmar F (1999), 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* 29:33–43

Constantino JN, Hudziak JJ, Todd RD (2003), Deficits in reciprocal social

behavior in male twins: evidence for a genetically independent domain of psychopathology. *J Am Acad Child Adolesc Psychiatry* 42:458–467
Dunn G (2000), *Statistics in Psychiatry* London: Arnold
Elliot CD (1990), *Differential Abilities Scale* San Antonio, TX: The Psychological Corporation
Fombonne E (2005), Epidemiology of autistic disorder and other pervasive developmental disorders. *J Clin Psychiatry* 66:3–8
International Molecular Genetic Study of Autism Consortium (2001), Agenomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *Am J Hum Genet* 69:570–581
Kraemer HC (1992), *Evaluating Medical Tests: Objective and Quantitative Guidelines* Newbury Park, CA: Sage Publications
Le Couteur A, Bailey A, Goode S et al. (1996), A broader phenotype of autism: the clinical spectrum in twins. *J Child Psychol Psychiatry* 37:785–801
Levine M (1986), *Leiter International Performance Scale: A Handbook* Los Angeles: Western Psychological Services
Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Pickles A (2006), Autism from 2 to 9 years of age. *Arch Gen Psychiatry* 63:694–701
Lord C, Rutter ML, Goode S et al. (1989), Autism Diagnostic Observation Schedule: a standardized observation of communicative and social behavior. *J Autism Dev Disord* 19:185–212
Lord C, Rutter M, Le Couteur A (1994), Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 24:659–685
Lord C, Rutter M, DiLavore P, Risi S (1999), *Autism Diagnostic Observation Schedule Manual* Los Angeles: Western Psychological Services
Mahoney WJ, Szatmari P, MacLean JE et al. (1998), Reliability and accuracy of differentiating pervasive developmental disorder subtypes. *J Am Acad Child Adolesc Psychiatry* 37:278–285
Mullen E (1995), *Mullen Scales of Early Learning* Circle Pines, MN: American Guidance Service
Rutter M, Le Couteur A, Lord C (2003), *Autism Diagnostic Interview-Revised* Los Angeles: Western Psychological Services
Seed PT, Tobias A (2001), Summary statistics for diagnostic tests *Stata Technical Bull* 59:9–12
Sparrow SS, Balla DA, Cicchetti DV (1984), *Vineland Adaptive Behavior Scales* Circle Pines, MN: American Guidance Service
Sung Y, Dawson G, Munson J, Estes A, Schellenberg GD, Wijsman EM (2005), Genetic investigation of quantitative traits related to autism: use of multivariate polygenic models with ascertainment adjustment. *Am J Hum Genet* 76:68–81
Szatmari P, Merette C, Bryson SE et al. (2002), Quantifying dimensions in autism: a factor-analytic study. *J Am Acad Child Adolesc Psychiatry* 41:467–474
Walker DR, Thompson A, Zwaigenbaum L et al. (2004), Specifying PDD-NOS: a comparison of PDD-NOS, Asperger syndrome, and autism. *J Am Acad Child Adolesc Psychiatry* 43:172–180

Should Parents Accompany Critically Ill Children During Inter-hospital Transport? J. Davies, S.M. Tibby, I.A. Murdoch

Background: Parental accompaniment during inter-hospital transportation (retrieval) of critically ill children is not commonplace in the United Kingdom. *Methods:* A three-month pilot of parental accompaniment was undertaken in 2002 (143 retrievals), after which time the policy was adopted as standard practice. A follow-up audit was performed in 2004 (136 retrievals). *Results:* Findings were remarkably consistent between the two periods. Staff perceived little or no added stress during the majority of transfers (96% in 2002, 98% in 2004), and felt able to perform medical interventions without hindrance (98% in 2002, 100% in 2004). There was good agreement between medical and nursing staff regarding perception of stress and ability to perform interventions (phi statistic 0.57 to 1.00). Adverse events occurred during 11 (3.9%) retrievals; six of these involved a parent exclusively. Stress tended to be associated with adverse events or parental behaviour rather than disease acuity. Staff vetoed the offer of accompaniment on 11 occasions, for a variety of reasons. The majority of parents found the experience safe, beneficial, and perceived a reduction in stress as a result. These data may inform other retrieval services who are considering adopting a similar policy. **Archives of Disease in Childhood** 2005;90:1270–1273.