Journal of Autism and Developmental Disorders, Vol. 27, No. 5, 1997

Diagnosing Autism: Analyses of Data from the Autism Diagnostic Interview[1]

Catherine Lord[2]
Department of Psychiatry, University of Chicago

Andrew Pickles
MRC Child Psychiatry, London

John McLennan
University of Pittsburgh

Michael Rutter
MRC Child Psychiatry, London

Joel Bregman
Emory University

Susan Folstein
New England Medical Center

Eric Fombonne
INSERM Centre Alfred Binet, Paris

Marion Leboyer
INSERM U155-Unite de Recherche on Genetique Epidernogique, Paris

Nancy Minshew
University of Pittsburgh

[1]This research was supported by the National Institute of Mental Health K05 MH01196, MH19726, and a grant from the John D. and Catherine T. MacArthur Foundation in association with the DSM-IV Field Trials to the first author.
[2]Address all correspondence to Catherine Lord, Department of Psychiatry, University of Chicago, 5841 South Maryland Avenue, MC 3077, Chicago, Illinois 60637.

Results from ROC curves of items from two scales, the Autism Diagnostic Interview (ADI) and Autism Diagnostic Interview-Revised (ADI-R), operationalizing DSM-IV criteria for autism are presented for 319 autistic and 113 other subjects from 8 international autism centers. Analyses indicate that multiple items were necessary to attain adequate sensitivity and specificity if samples with varying levels of language were considered separately. Although considering only current behavior was generally sufficient when a combination cutoff and additive model was employed, predictive power was highest when history was taken into account. A single set of criteria, as operationalized by individually structured questions in the ADI/ADI-R, was effective in differentiating autism from mental handicap and language impairment in subjects with a range of chronological ages and developmental levels.

In 1987, the American Psychiatric Association released a revised version of their third *Diagnostic and Statistical Manual* (American Psychiatric Association [APA], 1980), the *Diagnostic and Statistical Manual III-Revised* (DSM-III-R; APA, 1987). Almost immediately, work was begun on the next version to be published, DSM-IV (APA, 1994). The intention was to make the process of revision logical in its clinical impact and as empirically based as possible. Four goals were set in revising the diagnostic criteria for autism and pervasive developmental disorders in the DSM-III-R (APA 1987): (a) to make the diagnostic criteria less complex, (b) to address concerns that DSM-III-R criteria for autism were overly inclusive (c) to maintain comparability between proposed DSM-IV criteria (APA, 1994) and those suggested in the 10th draft of the International Classification of Diseases (ICD-10 draft; World Health Organization, 1987), and (d) to assess the role of age of onset as a factor in the diagnosis of autism (Volkmar et al., 1994). The multicenter analysis presented here uses data from a standard parent interview employed in eight different sites to address these issues.

One concern about DSM-III-R was that the diagnostic algorithm was overly complicated, with the result that some clinicians may have elected to employ the residual category of Pervasive Developmental Disorders Not Otherwise Specified (PDDNOS), rather than systematically assessing whether or not individuals met full criteria for autism. A paper using signal detection theory to assess the predictive power of individual DSM-III-R items in a relatively small sample (Siegel, Vukicevic, Elliott, & Kramer, 1989) proposed that diagnosis could be accurately made using clinician's ratings of only one or two items: *marked lack of awareness of others* and preoccupations with parts of objects. The present data analysis builds on these findings in asking whether criteria less cumbersome than those employed in DSM-III-R could be equally powerful.

First, specificity and sensitivity of classification of autism are compared for three methods of computation across DSM-IV draft items: (a) addition of scores for all items across the three areas, (b) achievement of cutoffs in each of the three areas (social interaction, communications, and restricted, repetitive behaviors) that have been used to define autism, and (c) combination of the two approaches (which was the "Chinese menu" approach employed in DSM-III-R).

Second, we asked if separate criteria are necessary for high-functioning, verbal autistic individuals; severely mentally handicapped, nonverbal autistic individuals, and other persons with autism who fall in between the two extremes of functioning. Third, the sensitivity (i.e., the proportion of true cases identified) and specificity (i.e., the proportion of true noncases correctly identified) of single items from the DSM-IV draft is compared to the predictive power yielded by the combination of DSM-IV domains. A final question of interest is the usefulness of historical versus current information. To address this issue, the specificity and sensitivity of diagnosis made on the basis of current behavior only or from descriptions of behaviors or deficits that "ever" occurred are examined.

Signal detection is a statistical method that has now gained widespread use in psychiatric research. Receiver Operating Curves (ROC) are used to derive the optimally sensitive (i.e., able to diagnose "positive" cases of autism) and specific (i.e., able to differentiate "negative" cases of not-autism) combination of scoring criteria (Kraemer, 1988). Through serial testing of various possible cutoffs, a scoring rule can be derived that optimizes the ability to predict a diagnosis.

METHOD

Sites

Eight sites contributed data on 432 children and adults for whom satisfactory scores were available from either the Autism Diagnostic Interview (ADI: Le Couteur et al., 1989) or the Autism Diagnostic Interview-Revised (ADI-R: Lord, Rutter, & Le Couteur, 1994) together with data on clinical diagnosis, chronological age, and IQ. The sample from the Institute of Psychiatry, University of London [89 autistic (AUT), 8 nonautistic (NONAUT)], included autistic children and adolescents from family, twin, and follow-up studies of autistic and language-impaired children. Subjects from the Greensboro-High Point TEACH Center (65 AUT, 28 NONAUT) represented consecutive clinic referrals for possible autism over a period of about 18 months and subjects from a study of sex differences in autism.

Subjects from Johns Hopkins University (60 AUT, 11 NONAUT) included children with Down syndrome and autistic children from a family study of autism in the division of psychiatric genetics. Subjects from Glenrose Hospital in Edmonton, Alberta (48 AUT, 11 NONAUT) had participated in a study of preschool autistic and mentally handicapped children, a treatment program for autistic adults or were consecutive referrals to a clinic for children with autism and/or communication disorders. One French sample (20 AUT, 26 NONAUT) was reunited in an ongoing multicenter longitudinal study of low-functioning autistic and nonautistic adolescents conducted by an INSERM research team (Fombonne, 1992, 1995). The other French sample (25 AUT, 27 NONAUT) was a series of clinic referrals to a university-based psychiatric service for autistic individuals and a mentally handicapped control group assessed as part of a validity study of a French translation of the ADI. Subjects from the University of Pittsburgh Clinic for Social Dysfunction were part of a project on the neurological bases of autism without mental handicap (12 AUT, 0 NONAUT). Subjects from Emory University (0 AUT, 12 NONAUT) were part of a project on fragile X syndrome originally begun at Yale University. It is important to note that in many of the research samples, diagnostically borderline cases had been excluded. Thus, the total sample for the study was not representative of a typical population seen in clinics for autism or pervasive developmental disorders.

Subjects

Diagnosis of autism [212 verbal (V) subjects, 107 nonverbal (NV) subjects] was determined by the clinical judgment of principal investigators/senior research associates (see below) using Rutter (1978) criteria for autism. These diagnoses were based on earlier contact or on measures employed at the time of administration of the interview. DSM-III-R criteria (APA, 1987) were used for cases of PDDNOS (8V, INV) and language delay (4V, 5NV). Mental retardation was defined as nonverbal IQ below 70 and no other psychiatric disorder (73V, 19NV). Three verbal subjects had non-PDD psychiatric diagnoses in addition to language impairment (i.e., conduct disorder, schizophrenia), with no mental handicap. Because the focus of the present analyses is on autism versus other disorders, the nine subjects with PDDNOS were grouped with other nonautistic subjects. Chronological age of the sample ranged from 3 to 43 years.

Nonverbal IQs were obtained on the basis of availability, using the following ranking of tests: WISC-R (Wechsler, 1974) or WAIS-R (Wechsler, 1981). Raven's Progressive Matrices (Raven, 1956), Leiter International Performance Scale (Arthur, 1952), and Merrill-Palmer Scale of Mental Tests (Stutsman, 1931). For some profoundly handicapped nonverbal adults (30 AUT, 35 NONAUT), the standard score from the Daily Living Scale from the Vineland Adaptive Behavior Scale (VABS; Sparrow, Balla, & Cicchetti, 1984) was used as an estimate of some aspects of developmental level when nonverbal IQ was not available. This score was selected because it is less confounded with the defining feature of autism (i.e., communication, social skills) than the remainder of the VABS and provides a general index of functioning in some capacities.

For the purpose of comparing autism to other diagnoses, three developmental groupings of subjects were defined, each of which had subjects with autism and with nonautism classifications. These groups were formed to provide clinical comparisons. They are not intended to be considered as "matched." Comparison groups were created such that the autistic subjects were never more severely retarded or lower in mental age than the nonautistic subjects in the same group. This strategy was employed to avoid having the autistic subjects characterized by more severe mental retardation than the nonautistic subjects.

The first group consisted of verbal subjects with estimated nonverbal/performance mental ages (MAs) from 3 years up to, but not including, 12 years. The 152 autistic subjects in this group had a mean chronological age (CA) of 14.5 years (SD = 7.2; range = 3–37) and a mean nonverbal IQ of 66.2 (SD = 24.1; range = 35–91). The 82 nonautistic subjects had corresponding means of 13.6 years for CA (SD = 5.9; range = 3–31) and nonverbal IQ of 56.0 (SD = 17.9; range = 39–84). All subjects had received codes of 0 on the ADI item of overall level of language, indicating that they had regular, spontaneous communicative use of at least three-word phrases, some of which contained a verb. Nonautistic subjects were almost all mildly to moderately mentally handicapped with some children with fragile X and some with dual diagnoses of language impairment and conduct disorder.

The second and highest functioning group was made up of 60 verbal autistic adolescents and adults with estimated nonverbal MAs of 12 years or above with a mean CA of 21.4 years (SD = 6.9; range = 12–40) and mean nonverbal IQ of 94.8 (SD = 14.3; range = 80–144). To obtain sufficient numbers, the MA required for nonautistic subjects for entry to this group was less stringently set at 7 years and above. This group included some nonautistic subjects (i.e., 30) who were also in the nonautistic group of verbal subjects with MAs between 7 and 12 years. Inclusion of younger, overlapping groups of nonautistic subjects seemed reasonable because, if anything, it would have made it more difficult to distinguish autistic from slightly less developmentally advanced nonautistic subjects. Diagnoses of nonautistic subjects included language impairment and attention deficit or conduct disorder and adolescent onset of schizophrenia. Means for the nonautistic subjects were 17.5 years (*SD* = 5.2; range = 7–38) for CA and 63.8 (*SD* = 16.7; range = 55–117) for nonverbal IQ. All subjects received 0 scores for ADI overall level of language.

The last group was composed entirely of low-functioning individuals. All were nonverbal. The 107 autistic subjects within this group had a mean CA of 11.9 years (*SD* = 8.0; range = 3–43) and mean nonverbal IQ of 46.1 (*SD* = 23.9; range = 20–69). The 25 nonautistic subjects had corresponding means of 9.9 years (*SD* = 6.8; range = 2–43) and a mean nonverbal IQ of 50.3 (*SD* = 26.0; range = 18–58). Scores for VABS were used for 46 profoundly mentally handicapped subjects (20 AUT, 26 NONAUT).

There were 326 males and 106 females, with similar distributions of gender across developmental level in autistic samples and nonautistic samples. This occurred in part because several sites provided only male subjects and because samples recruited from research projects had often been matched for gender. Of the 194 subjects for whom ethnic background was provided, 133 were Caucasian, 43 were African American or of West Indian descent, 6 were Asian, 4 were from other ethnic backgrounds, and the remainder unknown.

Procedures

The ADI (Le Couteur et al., 1989) and ADI-R (Lord et al., 1994) are standardized investigator-based interviews intended for use in the differential diagnosis of pervasive developmental disorders. At each site, the interview was administered by a trained clinician who had established reliability on training tapes and with other interviewers in the same group and, except for the French contributors, with researchers in at least one other site. at each site, reliability on administration of the ADI/ADI-R was initially obtained by the senior researchers' participation in a standard training course offered by Lord or Rutter. Criteria for reliability was set at a minimum 90% agreement on individual items for scoring of three consecutive interviews of children with autism or a related disorder. One of these interviews was conducted by the trainee, one was scored and administered live by a trainer, and one was scored from a standard set of videotapes. Research teams were then charged with training their own staff to meet the same standards (typically three consecutive interviews with 90% item by item agreement). About 30% of the interviews available were from research cases conducted by an interviewer blind to the diagnosis of the subject; about 30% were research cases in which the interviewer was aware of the source of recruitment and possible diagnosis. The remainder were part of multidisciplinary clinical assessments in which the ADI or ADI-R was usually administered as part of the first contact with the family. In these cases, the interviewer was typically not aware of the diagnosis but was aware of the reasons for referral.

The ADI and ADI-R contain items in three content areas: communication, social interaction, and restricted, repetitive behavior. Generally, items are coded 0 (*no evidence of abnormality*), 1 (*some evidence of abnormality*), and 2 (*evidence of marked abnormality*). A small number of items from the ADI and ADI-R are used in an algorithm to operationalized ICD-10 and DSM-IV criteria. Scores on these algorithm items are summed to create area scores. Diagnostic algorithm cutoffs used in this paper had been identified for the ADI based on an earlier data set: 10 out of 28 points for the *social* area, 6 out of 8 in *communication* for nonverbal subjects or 8 out of 24 in *communication* for verbal subjects, and 4 out of 12 for *restricted, repetitive behaviors* (Le Couteur et al., 1989). To be classified as having autism, a subject must have exceeded cutoff scores in each of the three areas: social, communication (nonverbal or verbal); and restricted, repetitive behaviors. The ADI-R contains more items than the earlier ADI (Lord et al., 1994); it includes almost all of the items from the earlier ADI. Consequently, the algorithm from the original ADI (Le Couteur et al., 1989) was used for both the ADI and ADI-R. When an identical item was not available, the comparable score on the most similar question from the ADI-R was used.

Two different versions of the algorithm were computed: (a) one based on judgments on items describing current behaviors only, and (b) one based on judgments of whether abnormalities had ever occurred (with "ever" always including "current") and, for items describing the development, or lack of development, of normal social and communicative behaviors, judgments based on whether behaviors had not developed by age 4 to 5 years. In most cases, identical items were available describing current and earlier behavior; but when they were not, similar items were used. Whenever identical items were not used, it is indicated on the tables.

A fourth criterion proposed for DSM-IV consisted of evidence of abnormality in communication, play, or social development prior to age 36 months. This factor was not analyzed because there was no variability within the autistic sample (i.e., all autistic subjects in these samples showed such abnormalities by age 36 months).

Clinical diagnoses were made at each site on the basis of observation and access to all available information. Consensus diagnoses were reached between two experienced clinicians, generally the principal investigator and a senior researcher. Clinicians were aware of information from the ADI or ADI-R but not informed of specific codes or algorithm scores. Clinicians made "best estimate" diagnoses (Volkmar et al., 1994) based on Rutter (1978) criteria for autism and on DSM-III-R criteria for other disorders.

Specificities, sensitivities, and total predictive powers of items correctly classifying autism were calculated using macros written in GLIM (Baker & Nelder, 1978/1982). The ratios of subjects receiving clinical diagnosis of autism to those with other diagnoses varied from about 1.75:1 (for the comparisons of two verbal groups) to over 4:1 for the low-functioning group. Thus, numbers of subjects with and without autism in each of the three groups were not representative of the normal population or of typical distributions of autistic versus nonautistic cases in clinic referrals (Lord, Storoschuk, Rutter, & Pickles, 1993). Total predictive power is reported in the text, though not in the tables, to allow comparisons with other studies. However, its interpretation is limited by the relatively high proportion of autistic subjects in each sample, particularly the nonverbal, low-functioning group.

Discrimination was examined over the range of scores (i.e., for most items, one analysis of 0 scores vs. grouping 1 and 2 and then a second analysis grouping 0 and 1 vs. 2). Results are reported for the grouping with the highest total predictive power. Item performance is most easily compared using a consistent set of subjects. We therefore used only subjects with complete data within any one domain. This resulted in varying numbers of subjects in different analyses. These numbers are reported in the tables. Because items for "current" behaviors and those that "ever" occurred were sometimes not identical, it was possible for sensitivity to be greater for a current score than for an ever score (which could not have occurred if the items were the same).

RESULTS

Results are reported according to the major issues addressed by the data analysis. Questions concerning issues in computation of diagnostic algorithms are presented first, followed by questions concerning type of approach and subgroups within autism. Total predictive power and negative predictive powers were computed, but are reported below only when they offer relevant information.

Table I. Sensitivity and Specificity of Different Methods of Generating ICD-10 Draft Algorithm Cutoff Scores of "Ever" Items[a]

	Verbal with MA = 3–12 (n = 234)		High functioning (n = 96)		All nonverbal (n = 132)	
Variable[b]	SE	SP	SE	SP	SE	SP
Original ADI cutoffs						
C8/6[c], S10, R4	.86	.91	.86	.93	.90	.96
Best cutoffs						
C7/5[c], S7, R3	.96	.85	.94	.90	.97	.79
Simple total: 21	.99	.84	1.00	.85	.98	.75
Total + cutoffs						
22/21[c] + C5/4[c], S5, R3	.99	.88	1.00	.90	.98	.83

[a]SE = sensitivity, SP = specificity.
[b]C = communication, S = social reciprocity, R = restricted and repetitive behavior. These represent cutoff scores for these three domains.
[c]The lower score is for nonverbal subjects.

Using Area Scores Versus Sums Across Items

As shown in Table I, the first sensitivities and specificities are for subjects who met originally specified cutoffs for autism in each of the three domains in the ADI/ADI-R (Le Couteur et al., 1989) and who received overall clinical judgments of autism versus any other disorder. With these populations, specificities were generally quite high (.90–.91). ROC resulted in adjusting the cutoffs in each area downward by 1–3 points from the original ADI algorithm. These "best" cutoffs, shown in Table I, made classification more inclusive and resulted in higher sensitivities and lower specificities. Total predictive power ranged from .92 to .94. Negative predictive power (i.e., proportion of those identified as negative who were true noncases) was also high, .77 to .85, except for nonverbal subjects, .52.

Using a total score of 21 points (without domain cutoffs) resulted in even higher sensitivities and only slightly lower specificities for verbal subjects, but yielded lower specificity for nonverbal subjects. Total predictive power was increased to .93 to .94% for all groups, with good negative power, except for nonverbal subjects (.43). Using a combination approach (as in DSM-III-R), in which subjects had to meet individual criteria in each area *and* have a total of 22 points (with nonverbal subjects "given" an extra point for having no language and thus not being codable on a number of items), resulted in equally high sensitivities and higher specificities, with total predictive powers of .95 to .96, and negative predictive powers ranging from .82–.86 for verbal subjects. However, negative predictive power for nonverbal subjects continued to be low, at .51.

Issues Related to the Very High-Functioning Group and Nonverbal Low-Functioning Group

As shown in Table I, by giving nonverbal subjects 1 point in the area of communication (and not scoring verbal items), the same cutoffs correctly classified a high proportion of low-functioning nonverbal, verbal high-functioning, and mildly to moderately handicapped verbal autistic individuals and nonautistic individuals of equivalent chronological ages and intellectual levels. If only a total score was used, criteria were overinclusive and negative predictive power poor (< .50) for low-functioning, nonverbal subjects. Using only domain cutoffs also resulted in incorrect classification of about one fifth of the nonautistic, severely mentally handicapped individuals. Furthermore, specificity was consistently lowest for the nonverbal group across the three content areas. Eliminating subjects with MAs below 2 years improved specificities markedly, but because the nonverbal sample then became quite small, this result requires cautious interpretation.

Classificatory Power of Individual Items Versus Domain Scores

Sensitivity and specificity were computed for items that most closely paralleled those used by Siegel et al. (1989) in their signal detection analysis of data from the DSM-III-R field trials. The item, *social reciprocity*, scored on the basis of current behavior only, was found to have a sensitivity of .93 and specificity of .54 for the entire study population. Sensitivity of *unusual sensory behavior*, also scored on the basis of current behavior only, was .52 and specificity was .90. When a positive score on either item was treated as indicative of autism, sensitivity was .87, but specificity was reduced to .48. As seen in Tables II–IV, when the autistic groups were divided by general developmental level and compared to individuals at similar levels, sensitivity of current social reciprocity remained high (>.90), but specificity was quite low (.47–.68). Nearly the reverse pattern held true for current unusual sensory behaviors. Sensitivity was low for all three groups (<.75) and specificity high for the two comparisons of verbal groups (.94), but relatively low for nonverbal subjects (.71).

Current Versus Ever Contexts

Another question examined was whether judgments should be made on the basis of an individual's current behavior or on the presence of the behavior at any point in the person's development. The issue is of particular concern in autism because, while impairment associated with the disorder

Table II. Sensitivity and Specificity for ADI Items for Verbal Subjects with Mental Ages of 3–12 Years

ADI items	Current SE	Current SP	Ever SE	Ever SP
Social interaction (n = 134 autistic, 77 other)				
Social responsiveness	.93	.78	.96	.70
Social reciprocity	.95	.68	.95	.68
Greeting	.70	.88	.86	.83
Come for comfort	.70	.75	.82	.73
Affection			.83	.79
Separation anxiety			.75	.57
Friendship	.84	.83	.89	.75
Cooperative play	.94	.52		
Turn taking/imitation[a]	.82	.64	.73	.89
Share pleasure	.80	.84	.81	.83
Inappropriate facial expression	.82	.88	.91	.86
Quality of social overtures			.89	.79
Communication (n = 110 autistic, 59 other)				
Instrumental gesture/point[a]	.92	.85	.85	.76
Expressive, inactive gesture	.92	.80	.95	.49
Attention to instructions	.60	.89	.90	.66
Range of facial expression	.78	.78	.78	.78
Nonverbal intentionality			.91	.76
Intonation/rhythm/rate	.54	.84	.55	.83
Vocal expression[a]	.76	.52	.76	.53
Stereotypic speech	.75	.92	.82	.88
Pronoun reverslas	.51	.88	.70	.64
Neologisms, idiosyncratic speech	.35	.92	.44	.89
Initiation of activity/play[a]	.88	.66	.93	.92
Conversation[a]	.85	.71	.68	.80
Amount of social language[a]	.68	.80	.66	.93
Restricted, repetitive behaviors (n = 142 autistic, 78 other)				
Resistance to change	.70	.68	.84	.65
Compulsions and rituals	.79	.68	.84	.76
Verbal rituals	.68	.87	.74	.85
Unusual sensory behavior	.66	.94	.74	.91
Hand and finger mannerisms	.67	.78	.82	.73
Curiosity/play[a]	.70	.74	.82	.86
Share others' activities	.89	.74		
Sensitivity to noise			.60	.54
Self-injury	.55	.86	.70	.77
Unusual preoccupations	.89	.74	.92	.68
Unusual attachments	.43	.91	.60	.86

[a]Items for current and ever scoring involved slightly different probes and/or coding.

persists across the life-span, deficits and abnormalities change as a child grows older. Patterns of deficit also differ across individuals according to developmental level (Rutter, 1978). As shown in Tables II–IV, on the

Table III. Sensitivity and Specificity for ADI Items for High-Functioning Autistic (MA > 12 Years) and Highest Functioning (MA > 7 Years) Nonautistic Subjects

	Current		Ever	
	SE	SP	SE	SP
Social interaction (*n* = 41 autistic, 32 other)				
Social responsiveness	.73	.75	.93	.75
Social reciprocity	.90	.50	.91	.50
Greeting	.54	.88	.95	.88
Come for comfort	.76	.81	.78	.72
Affection			.98	.88
Separation anxiety			.78	.84
Friendship	.96	.61	.96	.57
Cooperative play	.76	.41		
Turn taking/imitation[a]	.20	.72	.80	.84
Share pleasure	.56	.84	.56	.81
Inappropriate facial expression	.70	.85	.70	.86
Quality of social overtures			.80	.78
Communication (*n* = 36 autistic, 23 other)				
Instrumental gesture/point[a]	.81	.91	.78	.87
Expressive, inactive gesture	.85	.83	.97	.78
Attention to instructions	.25	.91	.94	.78
Range of facial expression	.81	.78	.80	.78
Nonverbal intentionality			.97	.87
Intonation/rhythm/rate	.83	.26	.83	.26
Vocal expression[a]	.45	.52	.53	.52
Stereotypic speech	.58	.91	.86	.87
Pronoun reverslas	.22	.91	.42	.91
Neologisms, idiosyncratic speech	.28	.91	.42	.91
Initiation of activity/play[a]	.78	.61	.91	.90
Conversation[a]	.86	.74	.64	.83
Amount of social language[a]	.81	.48	.53	.96
Restricted, repetitive behaviors (*n* = 142 autistic, 78 other)				
Resistance to change	.63	.69	.82	.69
Compulsions and rituals	.80	.69	.90	.66
Verbal rituals	.49	.94	.59	.91
Unusual sensory behavior	.45	.94	.74	.91
Hand and finger mannerisms	.53	.80	.69	.80
Curiosity/play[a]	.63	.63	.82	.77
Share others' activities	.80	.74		
Sensitivity to noise			.61	.71
Self-injury	.20	.91	.47	.89
Unusual preoccupations	.84	.77	.96	.77
Unusual attachments	.42	.94	.58	.89

[a]Items for current and ever scoring involved slightly different probes and/or coding.

whole, there were only relatively minor differences in the sensitivity and specificity of current and ever items. Items coded on the basis of ever occurring (i.e., ever occurring for a minimum of 3 consecutive months when

Table IV. Sensitivity and Specificity for ADI Items for Nonverbal Subjects

	Current		Ever	
	SE	SP	SE	SP
Social interaction (n = 98 autistic, 23 other)				
Social responsiveness	.95	.52	.97	.39
Social reciprocity	.95	.47	.95	.48
Greeting	.79	.65	.83	.70
Come for comfort	.76	.81	.84	.70
Affection			.73	.83
Separation anxiety			.66	.52
Friendship	.96	.61	.96	.57
Cooperative play	.94	.30		
Turn taking/imitation[a]	.94	.65	.96	.35
Share pleasure	.91	.52	.91	.48
Inappropriate facial expression	.82	.82	.82	.82
Quality of social overtures			.96	.57
Communication (n = 100 autistic, 19 other)				
Instrumental/point[a]	.96	.32	.91	.47
Expressive, inactive gesture	.93	.56	1.00	.26
Attention to instructions	.71	.58	.75	.68
Range of facial expression	.81	.68	.81	.63
Nonverbal intentionality			.95	.42
Initiation of activity/play[a]	.92	.42	.99	.67
Restricted, repetitive behaviors (n = 91 autistic, 21 other)				
Resistance to change	.49	.86	.60	.86
Compulsions and rituals	.59	.76	.65	.71
Unusual sensory behavior	.75	.71	.84	.71
Hand and finger mannerisms	.83	.81	.89	.76
Curiosity/play[a]	.75	.71	.99	.57
Share others' activities	.91	.81		
Sensitivity to noise			.67	.67
Self-injury	.54	.86	.67	.67
Unusual preoccupations	.69	.81	..79	.76
Unusual attachments	.45	.95	.60	.95

[a]Items for current and ever scoring involved slightly different probes and/or coding.

the subject was beyond 18 months mental age), were slightly more sensitive than current codes, particularly those items involving abnormal language characteristics in the verbal groups of subjects.

DISCUSSION

Overall, given the international and heterogeneous nature of the sample, with subjects ranging in age from 3 to 43 years and ranging in intellectual level from profound mentally handicap to superior functioning, the strength of the proposed diagnostic criteria was impressive. Although there was some variation related to differences in verbal skills and IQ, similarities across autistic samples were far more striking than were differences.

Current Versus Ever Contexts

In general, there were few differences in predictive power for these samples when behaviors were scored only current or ever, except that sensitivity was stronger for high-functioning individuals when ever codes were used. Greater difference might have emerged had the sample been restricted to adolescents or adults, for whom discrepancies between current and ever scores would be expected to be more frequent. On the basis of these findings, it was recommended that the scope of DSM-IV criteria be defined as behaviors that ever occurred consistently over any 3-month period during an individual's development, excluding infancy, and including the present. However, if historical information is not available, relying solely on current behaviors should exclude only a very small number of subjects, who will be primarily higher functioning adolescents and adults.

Computing an Algorithm

Various methods of computing a diagnostic algorithm were compared. Original domain cutoffs for the ADI were slightly underinclusive for all except very low-functioning individuals, compared to those generated by ROC for this larger sample. Using a single total score rather than cutoffs within each domain resulted in high sensitivity, but specificities of .75 to .85, which were judged to be inadequate. A combination of cutoffs in each domain and a total number of items, as in DSM-III-R, was found to yield the best sensitivity and specificity. Effective allowance for the inability of nonverbal subjects to receive scores on questions regarding language was made by allocating them 1 point for having insufficient language to score those items.

Use of Individual Items in Diagnosis

The individual items, *social reciprocity* and *unusual sensory behavior*, suggested by Siegel et al. (1989) were two of the strongest items in overall predictive power of any of the ADI/ADI-R questions, When the full subject population was considered as a single group, specificity for social reciprocity alone was very similar (i.e., >.88) to the specificity yielded by the strategy selected for DSM-IV; that is, use of a total score plus cutoffs for each domain. Sensitivity using only social reciprocity was lower than for the total plus cutoff procedure (.87 vs. .99). Tests of statistical significance for these differences are not available, but clinical differences seem likely. Including unusual sensory behaviors as an alternative did not increase sensitivity and lowered specificity. Taking the two items together resulted in lower sensitivity with no change in specificity.

The strength of the current item, social reciprocity, as an indication of autism initially seemed impressive. However, closer perusal of the data for the groups separated by level of functioning revealed that this effect was not consistent when autistic subjects were compared to other individuals with comparable language and developmental levels. Sensitivity remained high but specificities were relatively low (i.e., < .70), suggesting that the item in itself is not sufficient to make a diagnosis that differentiates within developmental level. On the other hand, at least on the basis of these and Siegel et al.'s (1989) data, this single item has the potential to serve as a powerful screening question, to be followed by more detailed evaluation.

Issues Related to Very High Functioning and to Nonverbal, Lower Functioning Individuals

In general, items showed lower specificity for low-functioning individuals and lower sensitivity for high-functioning subjects. However, with the use of domain totals and an algorithm based on ever occurrence, it was possible to predict correctly the classification of a very high proportion of the sample. Negative predictive power was sometimes lower than optimal, in part due to relatively small samples of the nonautistic groups (Fombonne 1992).

Separate criteria for autism associated with various developmental levels are not necessary. However, because of the difficulty in distinguishing autism-specific deficits from the general absence of directed behaviors in individuals with developmental levels under 2 years of age, the question arises whether lower limits for mental age should be specified. This issue has also been raised for use of the ADI/ADI-R in general (Lord et al., 1993). When individuals functioning at an infant to toddler level were excluded from these analyses, specificity rates increased markedly for social items, total domain, and algorithm scores. At this point, criteria for autism must be considered of limited value when applied without further information to developmentally delayed individuals functioning at the level of infants or toddlers.

Final Note

Overall, these data provide strong support for the ability of a single set of general criteria to differentiate autism from comparison groups of similar nonverbal mental age across a range of chronological ages and developmental levels. The methods of data collection here, that is, use of scores from a standard parent/caregiver interview, differ substantially from those of the DSM-IV field trials (Volkmar et al., 1994) and thus provide independent support for some of the features of DSM-IV. These data were based on a carefully circumscribed set of descriptions of behavior solicited from parents; they were collected by trained interviewers who followed more explicit rules for judgments on specific items than are part of DSM-IV or ICD-10 criteria. The sample comprised several data sets, some of which had been carefully screened to eliminate borderline cases. This factor likely contributed to the good predictive power that was found.

On the other hand, these data are more limited, and thus the task of discrimination more demanding, than procedures typically used in clinical diagnosis. In most clinical situations, information from parent interviews is combined with direct observation and information from other sources and weighted according to the clinicians' judgments. Consequently, the robust findings here are particularly encouraging. Finally, although the population studied was more selective than a sample of consecutive referrals, comparison groups were cases likely to be included in differential diagnoses of autism, such as individuals with mental retardation and severe language impairments, as opposed to the full range of psychiatric disorders in children and adolescents.

The results of these analyses are most important in providing exploratory data concerning how best to formulate a diagnostic framework for autism. The implications for methods of devising a diagnostic algorithm are also important. The conceptualization arising out of DSM-III-R of autism as a disorder involving deficits in three areas: communication, social development, and restricted, repetitive behaviors, is well-supported, as is the need for a range of items and criteria. The diagnostic system for autism represented in DSM-IV can be operationalized to be applied effectively across groups with wide variation in age, intellectual level, and language skills.

REFERENCES

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders (DSM-III-R)* (3rd ed., Rev.). Washington, DC: Author

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Arthur, G. (1952). *The Arthur adaptation of the Leiter International Performance Scale*. Chicago: The Psychological Service Center Press.

Baker, R. J. & Nelder, J. A. (1978/1982). In A. L. Baldwin, R. E. Cole, & C. P. Baldwin (Eds.), *The GLIM system manual, release 3*. Oxford, U.K.: Numerical Algorithms Group.

Fombonne, E. (1992). Diagnostic assessment in a sample of autistic and developmentally delayed adolescents. *Journal of Autism and Developmental Disorders, 22*, 563-582.

Fombonne, E. (1995). Troubles sévères du dévelopement. In *Le bilan a l'adolescence* (p. 262). Paris: éditions du CTNERHI.

Kraemer, H. C. (1988). Assessment of 2 × 2 associations: Generalization of signal detection methodology. *American Statistician, 42*, 37-49.

Le Couteur, A., Rutter, M., Lord, C., Rios, P. Robertson, S., Holdgrafer, M., & McLennan, J. D. (1989). Autism Diagnostic Interview: A semistructured interview for parents and caregivers of autistic persons. *Journal of Autism and Developmental Disorders, 19*, 363-387.

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. *Journal of Autism and Developmental Disorders, 24*, 659-685.

Lord, C., Storoschuk, S., Rutter, M., & Pickles, A. (1993). Using the ADI-R to diagnose autism in preschool children. *Infant Mental Health Journal, 14*, 234-252.

Raven, J. C. (1956). Guide to using the Coloured Progressive Matrices. London: H. K. Lewis.

Rutter, M. (1978). Diagnosis and definition. In M. Rutter & E. Schopler (Eds.), *Autism: A reappraisal of concepts and treatment* (pp. 1-26). New York: Plenum Press.

Siegel, B., Vukicevic, J., Elliott, G., & Kraemer, H. (1989). The use of signal detection theory to assess DSM-III-R criteria for autistic disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*, 542-548.

Sparrow, S., Balla, D., & Cicchetti, D. (1984). *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service.

Stutsman, R. (1931). Guide for administering the Merrill-Palmer Scale of Mental Tests. In L. M. Terman (Ed.), *Mental measurement of preschool children* (pp. 139-262). New York: Harcourt, Brace & World.

Volkmar, F. R., Klin, A., Siegal, B., Szatmari, P., Lord, C., Campbell, M., Freeman, B. J., Cicchetti, D. V., & 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. (1994). Field trial for autistic disorder in DSM-IV. *American Journal of Psychiatry, 151*, 1361-1367.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. San Antonio, TX: Psychological Corp.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corp.

World Health Organization. (1987). *ICD-10 1986 draft of chapter 5 categories F00-F99. Mental, behavioural and developmental disorders*. Geneva: Author.