ORIGINAL ARTICLE

ONLINE FIRST

A Multisite Study of the Clinical Diagnosis of Different Autism Spectrum Disorders

Catherine Lord, PhD; Eva Petkova, PhD; Vanessa Hus, MSc; Weijin Gan, MS, MD; Feihan Lu, MA; Donna M. Martin, MD, PhD; Opal Ousley, PhD; Lisa Guy, PhD; Raphael Bernier, PhD; Jennifer Gerdts, MA; Molly Algermissen, PhD; Agnes Whitaker, MD; James S. Sutcliffe, PhD; Zachary Warren, PhD; Ami Klin, PhD; Celine Saulnier, PhD; Ellen Hanson, PhD; Rachel Hundley, PhD; Judith Piggot, MD, PhD; Eric Fombonne, MD; Mandy Steiman, PhD; Judith Miles, MD, PhD; Stephen M. Kanne, PhD; Robin P. Goin-Kochel, PhD; Sarika U. Peters, PhD; Edwin H. Cook, MD; Stephen Guter, MA; Jennifer Tjernagel, MS; Lee Anne Green-Snyder, PhD; Somer Bishop, PhD; Amy Esler, PhD; Katherine Gotham, PhD; Rhiannon Luyster, PhD; Fiona Miller, PhD; Jennifer Olson, PhD; Jennifer Richler, PhD; Susan Risi, PhD

Context: Best-estimate clinical diagnoses of specific autism spectrum disorders (autistic disorder, pervasive developmental disorder–not otherwise specified, and Asperger syndrome) have been used as the diagnostic gold standard, even when information from standardized instruments is available.

Objective: To determine whether the relationships between behavioral phenotypes and clinical diagnoses of different autism spectrum disorders vary across 12 university-based sites.

Design: Multisite observational study collecting clinical phenotype data (diagnostic, developmental, and demographic) for genetic research. Classification trees were used to identify characteristics that predicted diagnosis across and within sites.

Setting: Participants were recruited through 12 university-based autism service providers into a genetic study of autism.

Participants: A total of 2102 probands (1814 male probands) between 4 and 18 years of age (mean [SD] age, 8.93 [3.5] years) who met autism spectrum criteria on the Autism Diagnostic Interview–Revised and the Autism Diagnostic Observation Schedule and who had a clinical diagnosis of an autism spectrum disorder.

Main Outcome Measure: Best-estimate clinical diagnoses predicted by standardized scores from diagnostic, cognitive, and behavioral measures.

Results: Although distributions of scores on standardized measures were similar across sites, significant site differences emerged in best-estimate clinical diagnoses of specific autism spectrum disorders. Relationships between clinical diagnoses and standardized scores, particularly verbal IQ, language level, and core diagnostic features, varied across sites in weighting of information and cutoffs.

Conclusions: Clinical distinctions among categorical diagnostic subtypes of autism spectrum disorders were not reliable even across sites with well-documented fidelity using standardized diagnostic instruments. Results support the move from existing subgroupings of autism spectrum disorders to dimensional descriptions of core features of social affect and fixated, repetitive behaviors, together with characteristics such as language level and cognitive function.

Arch Gen Psychiatry. 2012;69(3):306-313.
Published online November 7, 2011.
doi:10.1001/archgenpsychiatry.2011.148

Author Affiliations are listed at the end of this article.

IN THE FIELD OF AUTISM SPECTRUM disorders (ASDs), diagnostic instruments have been helpful in defining populations,[1] merging samples,[2] and comparing results across studies.[3,4] Nevertheless, best-estimate clinical (BEC) diagnoses have long been the gold standard.[5-7] In single-site studies, BEC diagnoses have added information to standardized instruments to predict later diagnoses[8,9] and classify children according to developmental trajectories of adaptive and language functioning.[10,11] However, researchers have recently expressed skepticism about the scientific and clinical value of categorical ASD groupings in the *DSM-IV-TR*[12] and the *International Statistical Classification of Diseases, 10th Revision*[13] (ie, autistic disorder, pervasive developmental disorder–not otherwise specified [PDD-NOS], Asperger syndrome), on which BEC diagnoses are based.[5,14,15]

The Simons Simplex Collection is a multisite project that aims to study de novo genetic variations in families that have 1

child with ASD and 1 or more unaffected siblings. Diagnostic parameters for probands were intentionally set to include common forms of ASD: autistic disorder, PDD-NOS, and Asperger syndrome. Stringent requirements for training and maintenance to ensure reliability in the selection, administration, and scoring of standardized instruments and cognitive tests were set. However, there was a deliberate decision to provide no specific training in diagnosis; rather, senior clinicians were asked to consider all available information to make BEC diagnoses (autistic disorder, PDD-NOS, and Asperger syndrome) using *DSM-IV-TR* criteria as they normally would in their practices, thereby allowing examination of relationships between BEC diagnoses of different ASDs, demographics, and standardized developmental and behavioral phenotype measures across sites. This design allows us to assess whether there are differences in BEC diagnoses of children with ASD across sites that are not associated with differences in the characteristics of the children but are instead associated with site- and clinician-based differences in how information is used to make diagnoses.

METHODS

PARTICIPANTS

A total of 2102 probands from 4 to 18 years of age were evaluated at 12 university-based centers. To prioritize children more likely to have de novo copy number variations, inclusion criteria for probands were as follows: (1) meeting criteria for ASD on the Autism Diagnostic Observation Schedule (ADOS)[16]; (2) meeting Collaborative Programs for Excellence in Autism–defined ASD criteria on the Autism Diagnostic Interview–Revised (ADI-R),[17] which has less stringent cutoffs for social and communication domains than "autism" criteria and no requirement for repetitive behaviors or age of onset[3,18]; (3) having a nonverbal mental age of at least 18 months; and (4) receiving a BEC diagnosis of autistic disorder, PDD-NOS, or Asperger syndrome (http://sfari.org).[15] Families were excluded if the proband had significant hearing, vision, or motor problems likely to affect interpretation of behavioral data and, because of the focus on de novo variations, if any known relative (third degree or less) had ASD. Families were also excluded if a sibling had substantial language or psychological problems related to ASD or if the proband had fragile X syndrome, tuberous sclerosis, Down syndrome, or a significant early medical history (eg, very low birth weight). The different sites contributed between 97 and 229 families.

PROCEDURES

Each proband was administered the ADOS, and a hierarchy of cognitive tests were implemented across sites, with 88% of probands receiving the Differential Ability Scales–Second Edition,[19] 7% receiving the Mullen Scales of Early Learning,[20] and 2% to 3% receiving the Wechsler Intelligence Scale for Children–Fourth Edition,[21] the Wechsler Abbreviated Intelligence Scale,[22] or some other scale. Parents were interviewed using the ADI-R and the second edition of the Vineland Adaptive Behavior Scales,[23] and they completed questionnaires, including the Aberrant Behavior Checklist.[24] Parents provided informed consent, children provided assent, and the institutional review board at each university provided approval.

Examiners attended standard research training sessions and maintained research reliability in conjunction with project consultants during semiannual workshops and video scoring (eAppendix [http://www.archgenpsychiatry.com]). After reviewing all the information and observing the proband in person or on video, the senior clinician specified a BEC diagnosis of autistic disorder, PDD-NOS, or Asperger syndrome according to *DSM-IV-TR* criteria. We enlisted the help of 47 psychologists, 6 physicians (3 psychiatrists, 2 pediatricians, and 1 clinical geneticist), and 3 master's level clinicians. The clinicians' years of experience in ASD ranged from less than 5 to more than 20 years (**Table 1**). Because one of our goals was to examine the contribution of BEC diagnoses in a protocol that asked experienced clinicians to consider information as they would in other research or their own practice, no training was provided in clinical diagnoses of ASD.

ANALYSIS

Relevant proband characteristics were classified as diagnostic (ADI-R standard algorithm domain totals: social, verbal communication, nonverbal communication, and restricted and repetitive patterns of behavior; ADOS domain scores: social + communication [ADOS-S + C] and restricted repetitive behavior [ADOS-RRB] total scores from modules 1-4, social affect from revised algorithms[25] for modules 1-3, and calibrated severity scores [ADOS-CSSs] from modules 1-3[26]) or as demographic, developmental, and behavioral (age, sex, race/ethnicity, maternal education, site, verbal IQ, performance IQ, composite Vineland Adaptive Behavior Scale scores, and irritability and hyperactivity scores from the Aberrant Behavior Checklist). We also considered diagnosticians' characteristics (type of degree and years of experience).

Differences between sites were assessed as follows. Continuous variables were described through minimum and maximum values, means, and standard deviations within each site. Distributions of continuous characteristics were approximated using kernel density estimation[27]; site densities were overlaid for visualization. Variance was partitioned into within- and between-site variances using mixed-effects models[28] for means that included random site effects. Intraclass correlation coefficients (ie, the ratios of between-site variance to total variance) are reported. Sites significantly deviating from the rest with respect to mean values were identified based on tolerance bands under the assumption of no differences between sites using permutation tests.[29,30] Categorical measures were described through ranges of proportions across sites; differences between sites were assessed with χ^2 tests for independence.

To investigate how the BEC diagnosis was associated with behavioral domains from diagnostic measures of ASD and to determine whether there were differences between sites with regard to the use of demographic, developmental, and behavioral measures in making BEC diagnoses, we used a classification and regression tree (CART).[31] This recursive partitioning technique is a statistical technique for discovering relationships between variables. It is different from the more familiar linear and generalized linear models, which evaluate and test for significant relationships of known forms. A CART is particularly well suited here because we do not know how various specific diagnostic features influence a clinician's decisions about distinctions among ASDs, whether scores on standardized instruments are linearly related to BEC diagnoses, or whether the same relationship between one scale and a diagnosis exists for all levels of other scales (eg, interactions between scales). In such situations, a CART can reveal relationships between variables that might go unnoticed using other analytic techniques and can generate empirically derived cut points within continuous variables. It is important to note, however, that a CART is not a probabilistic model, which means that formal inferences regarding the significance of predictors cannot be made (see details in eAppendix).

Table 1. Summary of Variability of Factors Characterizing 2102 Probands and 56 Diagnosticians Across 12 University-Based Sites[a]

Factor	Overall Proportion	Range Across Sites	SD Proportion Across Sites
Sex, % of probands			
Male	86.3	82.5-89.6	2.6
Female	13.7	10.4-17.5	2.6
Race[b]			
White	78.5	47.2-90.3	12.9
African American	4.0	0.5-12.7	3.3
Asian	4.0	1.4-8.0	2.2
Native American	0.1	0.0-0.7	0.3
Native Hawaiian	0.1	0.0-0.5	0.1
Other	4.3	0.0-18.2	5.3
More than 1 race	8.0	2.1-19.8	5.4
Not specified	1.0	0.0-5.7	1.7
Highest level of education obtained,[b] % of probands' mothers			
Graduate	25.3	17.5-40.6	7.3
Baccalaureate	36.0	30.3-43.7	3.9
Associate or some college	29.1	20.1-36.0	5.8
High school, GED, or some high school	9.4	6.3-11.8	1.7
<9th grade	0.1	0.0-1.1	0.4
ADOS diagnostic classification,[b] % of probands			
Autism	87.8	79.4-95.9	5.1
Autism spectrum	12.2	4.1-20.6	5.1
ADOS module,[b] % of probands			
1	16.9	8.0-26.6	5.1
2	23.0	15.9-31.4	5.1
3	57.1	48.3-67.9	6.7
4	3.0	0.0-6.5	2.1
ADI-R diagnostic classification (CPEA), % of probands			
Autism	90.3	87.3-93.5	2.1
Autism spectrum	9.8	6.6-12.7	2.1
Clinician's best-estimate diagnosis,[b] % of probands			
Autistic disorder	69.9	46.7-100.0	16.3
PDD-NOS	20.7	0.0-45.3	14.1
Asperger syndrome	9.4	0.0-20.7	7.1
Highest degree,[b] % of diagnosticians			
Doctor of medicine	12.2	0.0-84.8	26.0
Doctor of philosophy	87.5	15.1-100.0	25.9
Master of arts	0.3	0.0-2.1	0.7
Years of experience,[b] % of diagnosticians			
<5	8.2	0.0-41.3	13.9
5-10	30.7	0.0-100.0	36.1
≥10	61.1	0.0-100.0	39.5

Abbreviations: ADI-R, Autism Diagnostic Interview–Revised; ADOS, Autism Diagnostic Observation Schedule; CPEA, Collaborative Programs for Excellence in Autism; GED, General Education Development; PDD-NOS, pervasive developmental disorder–not otherwise specified.

[a] Differences between sites were assessed with χ^2 tests for independence. The different sites contributed between 97 and 229 families. Degrees of freedom = (number of sites – 1) × (number of levels – 1) = 11 × (number of levels – 1).

[b] $P \leq .001$.

In the CART analyses, we sequentially fit models, adding groups of predictors at each step. This is akin to forward variable selection in classic regression analysis. The order for inclusion of sets of predictors of BEC diagnosis was as follows: CART.1 included only diagnostic scales and clinician characteristics; CART.2 included diagnostic scales, clinician characteristics, and site; and CART.3 included diagnostic scales, clinician characteristics, and site as well as proband demographic, developmental, and behavioral characteristics. Finally, separate CART models were fit for each site.

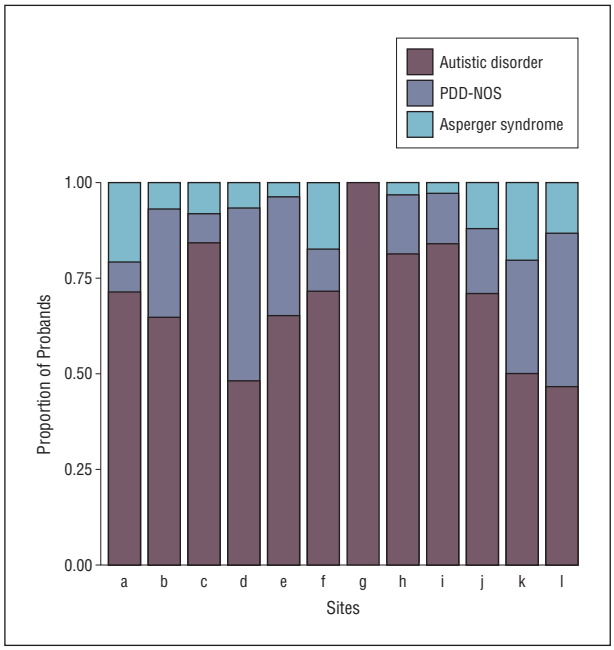

Figure 1. Best-estimate clinical diagnoses across 12 university-based sites (ie, autism service providers) for 2102 probands assigned to 3 autism spectrum disorder diagnostic categories (autistic disorder, Asperger syndrome, and pervasive developmental disorder–not otherwise specified [PDD-NOS]).

Tree models were first fully "grown" and then "pruned" (eAppendix). All analyses were performed with R using the recursive partitioning library rpart.[32] Owing to space constraints, the main text focuses on the CART.2 model, with a brief discussion of CART.1 and CART.3 (eAppendix).

Results from parametric models regarding site differences are also presented using classic inferential procedures. After diagnostic scales associated with BEC diagnosis as outcome were identified in CART.1, we fit logistic regression models for autistic disorder vs PDD-NOS or Asperger syndrome and for Asperger syndrome vs autistic disorder or PDD-NOS as functions of these scales and clinician characteristics. We then fit models that added site as either a fixed or random effect and tested interactions for site by each scale. Finally, the first model was compared with the second 2 models to assess the effect of site, using likelihood ratio tests.

RESULTS

SITE DIFFERENCES

BEC Diagnosis

As shown in **Figure 1**, statistically significant differences emerged across sites in the proportion of probands assigned to the 3 ASD diagnostic categories (autistic disorder, Asperger syndrome, and PDD-NOS) using BEC diagnoses ($\chi^2_{22}=358$, $P<.001$). Of the 12 sites, 2 gave fewer than half of the probands' conditions a diagnosis

Table 2. Summary of Variation Between Sites With Respect to Diagnostic Scales and Continuous Demographic and Behavioral Characteristics

	Ranges Across Sites				Variance			
	Minimum	Maximum	Mean	SD	Overall Mean (SD)	Within Sites	Between Sites	ICC[a]
Chronological age, mo	48-51	209-216	101.8-117.3	38.5-45.0	107.2 (42.1)	1770.0	4.6	0.003
Verbal IQ								
No control	5-13	138-167	72.4-85.4	27.5-33.4	79.3 (30.5)	918.4	13.1	0.014
Control for verbal mental age						435.9	5.5	0.013
Nonverbal IQ								
No control	9-30	133-161	79.7-89.8	22.0-27.7	86.1 (25.3)	635.1	4.7	0.007
Control for nonverbal mental age						375.6	3.7	0.010
Autism Diagnostic Interview–Revised								
Social interaction	8-9	30[b]	18.5-22.6	5.1-6.1	20.1 (5.7)	31.8	1.4	0.043
Verbal communication	6-8	24-26	15.6-18.4	3.5-4.8	16.4 (4.2)	17.5	0.7	0.036
Nonverbal communication	0-3	14[b]	8.2-10.3	3.1-3.6	9.1 (3.4)	11.5	0.5	0.038
Restricted/repetitive behavior	0-2	12[b]	5.8-7.1	2.2-2.7	6.5 (2.5)	6.1	0.2	0.032
Autism Diagnostic Observation Schedule								
Calibrated severity score	4-4	10[b]	6.8-8.1	1.6-1.8	7.4 (1.7)	2.8	0.1	0.041
Social and communication	4-7	22-24	12.1-14.7	3.8-4.6	13.3 (4.2)	17.4	0.4	0.022
Social affect	3-6	19-20	10.3-12.7	3.7-4.3	11.0 (4.0)	16.0	0.3	0.021
Restricted/repetitive behavior	0-0	8[b]	3.4-4.5	1.8-2.3	3.9 (2.0)	4.1	0.1	0.026
Vineland-II composite	27-52	95-115	68.9-75.9	9.3-14.4	73.8 (11.7)	134.0	3.5	0.026
Aberrant Behavior Checklist								
Irritability	0-0	28-42	8.4-13.1	6.8-9.2	11.3 (8.6)	72.7	1.6	0.022
Hyperactivity	0-1	37-48	13.1-18.0	8.9-11.0	16.5 (10.5)	107.3	2.4	0.022
Sample size, No.	97	229	175	38				

Abbreviations: ICC, intraclass correlation coefficient; Vineland-II, the second edition of the Vineland Adaptive Behavior Scales.[23]
[a] The ratio of between-site variance to total variance.
[b] The highest possible score (ie, ceiling) on the instrument.

of autistic disorder (sites d and l), whereas 1 site gave all of the probands' conditions a diagnosis of autistic disorder (site g) (Figure 1 and Table 1). Two sites gave more than 40% of probands' conditions a diagnosis of PDD-NOS. Sites also showed significant differences in the proportion of probands receiving diagnoses of Asperger syndrome, ranging from 0% to nearly 21%.

Because the sites were clinics known for different strengths, differences in recruitment were expected to yield site differences in behavioral phenotypes and demographics. The question is the degree to which differences in particular ASD diagnoses across sites were related to differences in the children (in either specific diagnostic features or other features) or to differences in the clinicians and their use of information about the children.

Diagnostic Variables

In contrast to differences in BEC diagnoses, sites showed no statistically significant differences in ASD diagnostic classifications yielded by standardized instruments (Table 1). In part, this was a function of the Collaborative Programs for Excellence in Autism–defined ASD diagnostic criteria,[18] which requires relatively mild social communication deficits on both the ADI-R and ADOS, but does not require the presence of any repetitive behavior.

Although there was substantial variation in measures of core features of ASD and developmental scores across individual children within sites, distributions were surprisingly similar across sites (**Table 2**), with only 1 site falling outside a 99.5% tolerance band (compared with 11 other sites) on the ADI-R social and communication domains and no sites falling outside a 99.5% tolerance band on the ADI-R restricted and repetitive patterns of behavior score. Site density distributions and permutation tolerance bands of ADI-R social and ADOS-RRB domains and performance IQ are shown in eFigure 1 as examples (additional figures available on request). All but 14 participants met ADI-R criteria for onset of symptoms before 3 years.[17]

Patterns of across-site variability for ADOS domain scores were similar. No site-related intraclass correlation exceeded 0.07 (eAppendix). Thus, the large site differences in BEC diagnoses were not accompanied by equivalent differences in standardized diagnostic scores.

Demographic and Behavioral Characteristics

The mean (SD) chronological age was 8.93 (3.5) years, with similar distributions of age across sites (Table 2). Of the 2102 probands, 1814 were men, and 288 were women; differences in sites' proportions of males to females ranged from 5:1 to 9:1 but were not statistically significant. Maternal education was high and homogeneous. With regard to race/ethnicity, participants from all but 3 sites were 70% to 90% non-Hispanic white, 4% Asian American, 4% African American, and 8% more than 1 race. Mean IQs were relatively high; composite Vineland Adaptive Behavior Scale scores were lower, with less variation within and across sites.

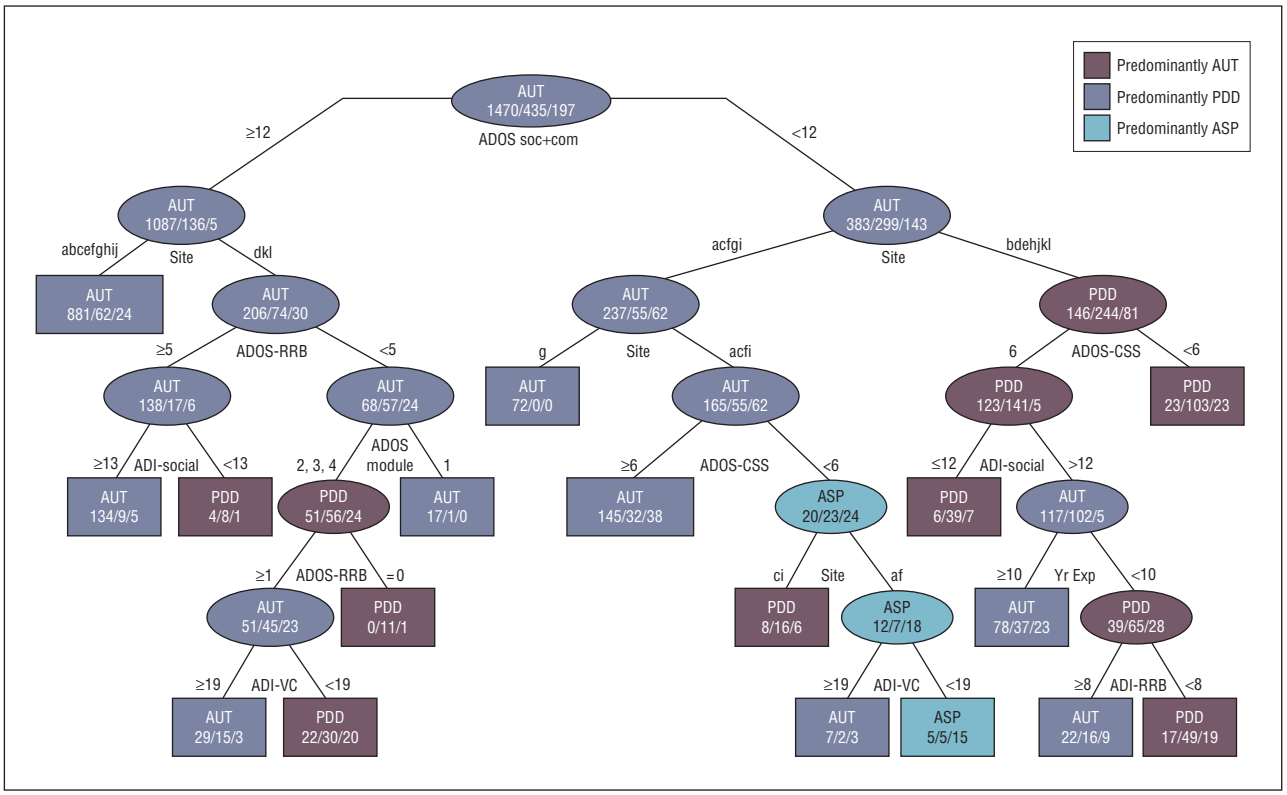

Figure 2. Classification and regression tree (CART.2 model) for best-estimate clinical diagnoses with diagnostic scales, site, and diagnostician characteristics as predictors. The sets of numbers separated by slashes denote the numbers for each diagnostic group: autistic disorder (AUT)/pervasive developmental disorder–not otherwise specified (PDD)/Asperger syndrome (ASP). ADI-RRB indicates Autism Diagnostic Interview–Revised restricted and repetitive behaviors total score; ADI-social, ADI-R social total score; ADI-VC, ADI-R verbal communication total score; ADOS-CSS, Autism Diagnostic Observation Schedule calibrated severity score; ADOS-RRB, ADOS restricted and repetitive behaviors total score; ADOS-soc+com, ADOS social and communication domain total score; Yr Exp, senior diagnostician's number of years of experience.

DETERMINING BEC DIAGNOSIS

Following the sequential model-fitting strategy already outlined, classification trees were "grown" for BEC diagnosis using different sets of predictors. Details of CART.1 are presented in the eAppendix and eFigure 2, with a brief description herein. The most powerful predictor selected was ADOS-S + C, a standard measure of clinician-observed social communication available for all participants. The 61% of children with moderate to severe social communication deficits were primarily classified as having autistic disorder; diagnoses for the remaining 39% of children with milder social communication deficits, including most of the children who received a BEC diagnosis of PDD-NOS or ASP and about one-third of the children who received a BEC diagnosis of autistic disorder, showed interactions with a series of predictors, including ADOS module and calibrated severity scores, each of the ADI-R domains, and clinicians' years of experience and type of degree. Even the smallest nodes were heterogeneous across different ASDs. The more experienced diagnosticians gave a higher proportion of diagnoses of autistic disorder; clinicians with a doctor of philosophy degree used PDD-NOS as a diagnosis more often than did clinicians with either a doctor of medicine degree or a master's degree. This model reduced the misclassification error from 0.30 (with random assignment based on prevalence) to 0.24, a 20% reduction in the misclassification rate (explained error, which corresponds to percent-explained variation in a linear regression).

Differential Use of Diagnostic Scales by the Sites to Make BEC Diagnoses

The CART.2 model (**Figure 2**) added site as a predictor. The first branching was identical to CART.1. However, in CART.2, the second step in both right and left branches was site, indicating that site differences accounted for more variance in BEC diagnoses than did any other factor after ADOS-S + C. When site was included in the model, most effects of clinician characteristics disappeared.

In general, similar biases affected several sites at a time. For example, an ADOS-S + C score of 12 or greater (left branch) was the only information used in 9 of 12 sites; of the children who had moderate to high social communication deficits at these 9 sites, 91% received a BEC diagnosis of autistic disorder. In the 3 remaining sites, additional information was associated with the differentiation of PDD-NOS, Asperger syndrome, and autistic disorder.

Site differences also appeared at several steps in the right branch, indicating interactions between site and diagnostic scales for children with less severe social communication deficits. "Walking through" the first few steps of the right branch of CART.2 (which includes the 825

children with relatively mild social communication scores [ie, <12] on the ADOS), we found that 5 sites in the left subbranch (a, c, f, g, and i) made proportionately more autistic disorder diagnoses than did the other 7 sites, with 1 site (g) giving only autistic disorder diagnoses. Four of these 5 sites further differentiated children using the ADOS-CSS (which takes into account age, language level, and restricted and repetitive patterns of behaviors). Children with less severe ADOS-CSSs were split by site again, with 2 sites (a and f) further split by abnormalities in parent reports of children's verbal communication (ADI-R verbal communication).

The 7 sites in the rightmost subbranch (b, d, e, h, j, k, and l) predominantly gave children with milder social communication impairments BEC diagnoses of PDD-NOS. The ADOS-CSS was again taken into account, with children who had a score of less than 6 (milder severity) receiving mostly PDD-NOS diagnoses and those who had a score of 6 or greater receiving 1 of the 3 ASD diagnostic classifications, depending on the parent-reported historical account (ADI-R social and ADI-R restricted and repetitive patterns of behavior) and the diagnostician's years of experience. When differentiation by site was included, the misclassification error rate improved from 0.24 to 0.21 (29% reduction in the total misclassification rate, which constitutes a 9% improvement over CART.1).

Importance of Site in Making BEC Diagnosis: Formally Assessed via Logistic Regression

Site was a very important factor, both as a main effect and in interaction with diagnostic scales, based on comparisons of CART.1 (using only the diagnostic scales and clinician characteristics) with CART.2 (which also included site and site-by-scale interactions). All *P* values comparing the respective nested models were highly significant ($P < 1 \times 10^{-10}$). From the models in which site was treated as a random factor, the variances of the random effects for site and site × covariate interaction were quite large; the coefficient of variation (ie, the standard deviation of the random effect/mean effect of the covariate) ranged from 0.33 to 4.95, with the largest coefficient of variation corresponding to the interaction between site and ADOS-CSS, indicating variability between sites in interpreting observed overall severity of autism symptoms in the context of children's ages and language levels.

Use of Demographic, Developmental, and Behavioral Variables in Making BEC Diagnoses

In CART.3 (eFigure 3), demographic, developmental, and specific behavioral characteristics were added. The primary difference between it and previous CART models was that, among children with moderate to severe social communication deficits, the most important factor for BEC diagnosis became verbal IQ. Across all sites, 93% of children who had an ADOS-S + C score of 12 or greater and a verbal IQ of 85 or less received an autistic disorder diagnosis.

In contrast, BEC diagnoses of children with an ADOS-S + C score of 12 or greater and a verbal IQ of greater than 85 or children with an ADOS-S + C score of less than 12 (right branch) were affected by the site differences and by the many different interactions with each of the diagnostic variables at different stages, as shown in CART.3. Splits were also made on verbal IQ and performance IQ at a number of places in the tree, with cutoffs in IQ ranging from 85 to 122, depending on site. There were no effects of sex, ethnicity/race, or maternal education, but there were effects of chronological age, adaptive behavior, and hyperactivity. When demographic, developmental, and behavioral measures were included, the misclassification rate decreased to 0.17 (a 43% reduction in the total misclassification rate, which is an improvement of 23% compared with CART.1 and 14% compared with CART.2).

REPLICABILITY OF MODELS

The BEC Diagnosis Made at Each Site

Individual trees were generated for each site using diagnostic, developmental, and demographic variables as predictors. The numbers of participants, although smaller than those used for CART.1 and CART.3, are sufficient enough for us to have relative confidence in the results (n=97-229). To test the stability of models, results for CART.2 generated from the first half of the sample (n=933) were applied to the second half of the sample (n=1169). Misclassification rates were nearly identical (0.23 vs 0.25) (eAppendix). Models for 11 of the 12 sites, omitting the site where all probands had autistic disorder, are presented in eFigure 4.

Several findings for the 11 individual-site CARTs were striking. As shown in the eTable, verbal IQ was the single feature most related to BEC diagnoses in 5 sites and the second or third strongest predictor in 5 other sites (eFigure 4). However, there were striking site differences in verbal IQ cut points and in whether or not IQ was associated with differentiating autistic disorder from PDD-NOS/Asperger syndrome or differentiating autistic disorder/PDD-NOS from Asperger syndrome. The next most frequent predictors across sites were ADOS social communication or repetitive behaviors, emerging first in 4 and 2 sites, respectively. For 9 sites, 1 of these 3 measures predicted an entire "node" of diagnosis; in most cases, that diagnosis was autistic disorder, but in one case, it was Asperger syndrome. Six sites had age effects, primarily such that Asperger syndrome diagnoses were given to older children, although the age cut points varied from 5.25 to 12 years. Cut points for autistic disorder vs PDD-NOS/Asperger syndrome for the ADOS-S + C domain varied from 8 to 16. Only 1 site had an effect of sex and also of maternal education.

Diagnosticians' Characteristics Affecting BEC Diagnosis

Findings of differences in BEC diagnoses related to the training or level of experience of senior diagnosticians appeared to be accounted for by site differences in almost all cases, although the direction of effect (whether senior clinicians influenced others in their sites) cannot be determined. Within sites, clinician differences did not have significant effects on BEC diagnoses.

Several conclusions are inescapable. In these 12 university-based sites, with research clinicians selected for their expertise in ASD and trained in using standardized diagnostic instruments, there was great variation in how BEC diagnoses within the autism spectrum (ie, autistic disorder, PDD-NOS, and Asperger syndrome) were assigned to individual children. Clinical diagnoses were not random. It is not surprising that clinicians often feel strongly that their distinctions among the various ASD diagnoses mean something. However, although patterns within and across the sites were clearly discernible, they were idiosyncratic and complex.

Despite the fact that the sample was somewhat restricted in age and skewed in IQ, and that children were required to meet minimal ASD criteria on the ADI-R and ADOS, we anticipated recruitment differences associated with different referral populations. Had these restrictions not been in place, even greater site differences might have been expected. Nevertheless, in contrast to differences in BEC diagnoses, differences in distributions among children's scores on standardized diagnostic measures across sites were almost never significant. Observational (ADOS) summary scores and verbal IQ, as well as children's ages and parent-reported (using the ADI-R and Aberrant Behavior Checklist) information about repetitive behaviors, communication abnormalities, and hyperactivity, influenced diagnoses in many sites. However, careful examination suggested that patterns within sites varied considerably in how and when (along a decision tree) clinicians took into account different factors in deciding which diagnosis to apply to children within the spectrum. Although predictors overlapped across sites, they also differed markedly in "cut points" (eg, individual-site verbal IQ cut points between autistic disorder/PDD-NOS and Asperger syndrome ranged from 62 to 127) and in the order in which information was used.

Differences in BEC diagnosis could reflect regional variation. For example, in some regions, children with diagnoses of autistic disorder receive different services than do children with other ASD diagnoses; elsewhere, autistic disorder diagnoses may be avoided as more stigmatizing than diagnoses of PDD-NOS or Asperger syndrome.

An important concern is the stability of findings based on CART models, which are tools for discovery rather than hypothesis testing and inference. To assess this, we evaluated how well the models developed for the 1169 most recent participants compared with the models developed for the first 933 participants in the data collection. The misclassification rates for both sets of models were very similar (eAppendix).

Another potential concern for the results presented in eFigure 4 is the relatively small sample sizes of the individual sites. Again, results from the models developed for the most recent 1169 participants were nearly identical to the results from the models developed for the first 933 participants, except that maternal education played less of a role in the larger sample and that 1 site had a sex effect.

Previous research[2,9] has shown that, within a site, clinicians' diagnoses can add information to standardized scores. If clearer BEC decision rules had been specified and applied consistently or if standard training had been offered, BEC diagnoses might have been an important source of information in our study. However, given the evidence that there is little standard meaning of BEC diagnoses across sites, their utility in research is questionable.

These results have implications for revisions of current diagnostic frameworks such as the *DSM-V* and the *International Statistical Classification of Diseases, 11th Revision*. Recurrent evidence of the importance of information external to a psychiatric diagnosis, particularly verbal IQ and current language level (eg, ADOS module), supports the need for cognitive function and language level to be considered as essential to BEC diagnoses of ASD. Diagnostic classifications based on retrospectively recalled information from the ADI-R were not as useful as expected in discriminating groupings within the autism spectrum in this selected population, perhaps because there was so little variability. However, dimensional, observational, and parent-reported measures of social communication and repetitive behaviors clearly contributed to clinical diagnoses. Within these 12 sites with experienced and well-trained staff, the distributions of dimensional measures of standardized instruments were much more consistent than were the distributions of categorical BEC diagnoses. More precise diagnostic criteria might have made these distributions more consistent, but how to do this succinctly and address the range of developmental and individual variability in ASD is not clear. As others have suggested,[33] the conceptualization and measurement of ASD as a behavioral diagnosis, based on different dimensions (eg, social communication and repetitive behaviors) that are strongly influenced by intelligence and language skills, may be more useful in providing links to brain function,[34] genetics,[35] and services[36] than are clinical categorical diagnoses of autistic disorder, PDD-NOS, or Asperger syndrome.

Submitted for Publication: June 11, 2011; final revision received August 18, 2011; accepted August 30, 2011.
Published Online: November 7, 2011. doi:10.1001/archgenpsychiatry.2011.148
Author Affiliations: Institute for Brain Development, Weill Cornell Medical College, White Plains (Dr Lord), Nathan Klein Institute for Psychiatric Research, Orangeburg (Dr Petkova), and Department of Child and Adolescent Psychiatry, New York University (Drs Petkova and Gan and Ms Lu), Division of Child and Adolescent Psychiatry, Columbia University Medical Center (Drs Algermissen and Whitaker), and Simons Foundation (Ms Tjernagel), New York, New York; Autism and Communication Disorders Center (Drs Green-Snyder, Gotham, Miller, Olson, and Risi and Ms Hus) and Departments of Pediatrics and Human Genetics (Dr Martin), University of Michigan; Ann Arbor; Emory University School of Medicine (Drs Ousley, Klin, and Saulnier), and Marcus Autism Center, Children's Healthcare of Atlanta (Dr Klin), Georgia; Center for Autism Research, Children's Hospital of Philadelphia, Pennsylvania (Dr Guy); Departments of Psychiatry (Dr Bernier) and Psychology (Dr Gerdts), University of Washington, Seattle; Departments of Molecular Physiology and Biophysics and Psy-

chiatry, Vanderbilt Kennedy Center (Dr Sutcliffe), and Departments of Pediatrics (Drs Warren and Peters) and Psychiatry (Dr Warren), Vanderbilt University Medical Center, Nashville, Tennessee; Division of Developmental Medicine, Children's Hospital Boston, Harvard Medical School, Massachusetts (Drs Hanson, Hundley, and Luyster); Center for Autism Research and Treatment and Department of Psychiatry, Semel Institute of Neuroscience, University of California Los Angeles (Dr Piggot); Department of Psychiatry, Montreal Children's Hospital, Québec, Canada (Drs Fombonne and Steiman); Thompson Center for Autism and Neurodevelopmental Disorders, University of Missouri, Columbia (Dr Miles); Department of Pediatrics, Baylor College of Medicine, Houston, Texas (Drs Kanne and Goin-Kochel); Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago (Dr Cook and Mr Guter); Cincinnati Children's Hospital Medical Center, Ohio (Dr Bishop); Department of Pediatrics, University of Minnesota, Minneapolis (Dr Esler); and Department of Psychological and Brain Sciences, Indiana University, Bloomington (Dr Richler).

Correspondence: Dr Catherine Lord, PhD, Institute for Brain Development, Weill Cornell Medical College, Bard House, 21 Bloomingdale Rd, White Plains, NY 10605-1504 (cal2028@med.cornell.edu).

Financial Disclosure: Dr Lord receives royalties from the manufacturer of the diagnostic instruments described in this article. She gives all the profits generated by the University of Michigan Autism and Communication Disorders Center and by all other projects at the center, including the Simons Simplex Collection, to charity.

Funding/Support: This research was funded by the Simons Foundation and the National Institute of Mental Health (grant R01 MH081873-01A1 to Dr Lord).

Role of the Sponsors: The Simons Foundation played a role in the design and conduct of the study; this role included independently funding the data management group so that they would be able to store data.

Online-Only Material: The eAppendix, eTable, and eFigures are available at http://www.archgenpsychiatry.com.

Additional Information: Phenotypic data were accessed through the SFARI base, a central database of clinical and genetic information about families affected by autism; this database is provided as part of the Simons Foundation Autism Research Initiative (SFARI). SFARI-approved researchers can obtain the Simons Simplex Collection population data set described in this study by applying at https://base.sfari.org.

Additional Contributions: We thank all of the families at the participating SFARI Simplex Collection sites.

REFERENCES

1. Beglinger LJ, Smith TH. A review of subtyping in autism and proposed dimensional classification model. *J Autism Dev Disord*. 2001;31(4):411-422.
2. Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Pickles A. Autism from 2 to 9 years of age. *Arch Gen Psychiatry*. 2006;63(6):694-701.
3. Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, Cook EH Jr, Leventhal BL, Pickles A. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry*. 2006; 45(9):1094-1103.
4. Gotham K, Risi S, Dawson G, Tager-Flusberg H, Joseph R, Carter A, Hepburn S, McMahon W, Rodier P, Hyman SL, Sigman M, Rogers S, Landa R, Spence MA, Osann K, Flodman P, Volkmar F, Hollander E, Buxbaum J, Pickles A, Lord C. A replication of the Autism Diagnostic Observation Schedule (ADOS) revised algorithms. *J Am Acad Child Adolesc Psychiatry*. 2008;47(6):642-651.
5. Chakrabarti S, Fombonne E. Pervasive developmental disorders in preschool children: confirmation of high prevalence. *Am J Psychiatry*. 2005;162(6):1133-1141.
6. Charman T, Baird G. Practitioner review: Diagnosis of autism spectrum disorder in 2- and 3-year-old children. *J Child Psychol Psychiatry*. 2002;43(3):289-305.
7. Volkmar F, Chawarska K, Klin A. Autism in infancy and early childhood. *Annu Rev Psychol*. 2005;56(1):315-336.
8. Lord C. Follow-up of two-year-olds referred for possible autism. *J Child Psychol Psychiatry*. 1995;36(8):1365-1382.
9. 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(2):219-226.
10. Anderson DK, Oti RS, Lord C, Welch K. Patterns of growth in adaptive social abilities among children with autism spectrum disorders. *J Abnorm Child Psychol*. 2009;37(7):1019-1034.
11. Anderson DK, Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Welch K, Pickles A. Patterns of growth in verbal abilities among children with autism spectrum disorder. *J Consult Clin Psychol*. 2007;75(4):594-604.
12. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition, Text Revision): DSM-IV-TR*. Washington, DC: American Psychiatric Association; 2000.
13. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems, 10th Revision*. Geneva, Switzerland: World Health Organization; 1992.
14. Ozonoff S, Rogers SJ, Pennington BF. Asperger's syndrome: evidence of an empirical distinction from high-functioning autism. *J Child Psychol Psychiatry*. 1991; 32(7):1107-1122.
15. Klin A, Volkmar FR. Asperger syndrome. *Child Adolesc Psychiatr Clin N Am*. 2003; 12(1):xiii-xvi.
16. Lord C, Rutter M, DiLavore PS, Risi S. *Autism Diagnostic Observation Schedule (ADOS)*. Los Angeles, CA: Western Psychological Services; 1999.
17. Rutter M, Le Couteur A, Lord C. *The Autism Diagnostic Interview–Revised (ADI-R)*. Los Angeles, CA: Western Psychological Services; 2003.
18. Lainhart JE, Bigler ED, Bocian M, Coon H, Dinh E, Dawson G, Deutsch CK, Dunn M, Estes A, Tager-Flusberg H, Folstein S, Hepburn S, Hyman S, McMahon W, Minshew N, Munson J, Osann K, Ozonoff S, Rodier P, Rogers S, Sigman M, Spence MA, Stodgell CJ, Volkmar F. Head circumference and height in autism: a study by the Collaborative Program of Excellence in Autism. *Am J Med Genet A*. 2006; 140(21):2257-2274.
19. Elliott CD. *Differential Ability Scales–Second Edition (DAS-II)*. San Antonio, TX: The Psychological Corporation; 2006.
20. Mullen E. *The Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service, Inc; 1995.
21. Wechsler D. *Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV)*. San Antonio, TX: The Psychological Corporation; 2003.
22. Wechsler D. *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation; 1999.
23. Sparrow SS, Cicchetti DV, Balla DA. *Vineland Adaptive Behavior Scales*. 2nd ed. Circle Pines, MN: AGS Publishing; 2005.
24. Aman MG, Singh NN, Stewart AW, Field CJ. The aberrant behavior checklist: a behavior rating scale for the assessment of treatment effects. *Am J Ment Defic*. 1985;89(5):485-491.
25. Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. *J Autism Dev Disord*. 2007;37(4):613-627.
26. Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *J Autism Dev Disord*. 2009;39(5):693-705.
27. Scott DW. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Hoboken, NJ: Wiley-Interscience; 1992.
28. Diggle P. *Analysis of Longitudinal Data*. London, England: Oxford University Press; 2002.
29. Fan J, Lin SK. Test of significance when data are curves. *J Am Stat Assoc*. 1998; 93(443):1007-1021 http://www.jstor.org/stable/2669845. Accessed June 10, 2011.
30. Westfall PH, Young SS. *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment*. Hoboken, NJ: John Wiley and Sons; 1993.
31. Breiman L, Friedman J, Olshen R, Stone C, Steinberg D, Colla P. *CART: Classification and Regression Trees*. Belmont, CA: Wadsworth; 1983.
32. The R Project for Statistical Computing. R Foundation for Statistical Computing Web site. http://www.R-project.org. Accessed June 10, 2011.
33. Wing L. Past and future of research on Asperger syndrome. In: Klin A, Sparrow SS, Volkmar FR, eds. *Asperger Syndrome*. New York, NY: The Guilford Press; 2000:418-432.
34. Klin A, Jones W, Schultz R, Volkmar F, Cohen D. Defining and quantifying the social phenotype in autism. *Am J Psychiatry*. 2002;159(6):895-908.
35. Alarcón M, Cantor RM, Liu J, Gilliam TC, Geschwind DH; Autism Genetic Research Exchange Consortium. Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. *Am J Hum Genet*. 2002;70 (1):60-71.
36. National Research Council; Committee on Educational Interventions for Children with Autism. *Educating Children With Autism*. Washington, DC: The National Academies Press; 2001.