A Multi-Site Study of the Clinical Diagnosis of Different Autism Spectrum Disorders

Catherine Lord, Ph.D.[1], **Eva Petkova, Ph.D.**[2,3], **Vanessa Hus, M.Sc.**[4], **Weijin Gan, M.S., M.D.**[2], **Feihan Lu, M.A.**[2], **Donna M. Martin, M.D., Ph.D.**[5], **Opal Ousley, Ph.D.**[6], **Lisa Guy, Ph.D.**[7], **Raphael Bernier, Ph.D.**[8], **Jennifer Gerdts, M.A.**[9], **Molly Algermissen, Ph.D.**[10], **Agnes Whitaker, M.D.**[10], **James S. Sutcliffe, Ph.D.**[11], **Zachary Warren, Ph.D.**[12], **Ami Klin, Ph.D.**[6,13], **Celine Saulnier, Ph.D.**[6], **Ellen Hanson, Ph.D.**[14], **Rachel Hundley, Ph.D.**[14], **Judith Piggot, M.D., Ph.D.**[15], **Eric Fombonne, M.D.**[16], **Mandy Steiman, Ph.D.**[16], **Judith Miles, M.D., Ph.D.**[17], **Stephen M. Kanne, Ph.D.**[18], **Robin P. Goin-Kochel, Ph.D.**[18], **Sarika U. Peters, Ph.D.**[12], **Edwin H. Cook, M.D.**[19], **Stephen Guter, M.A.**[19], **Jennifer Tjernagel, M.S.**[20], **Lee Anne Green-Snyder, Ph.D.**[4], **Somer Bishop, Ph.D.**[21], **Amy Esler, Ph.D.**[22], **Katherine Gotham, Ph.D.**[4], **Rhiannon Luyster, Ph.D.**[14], **Fiona Miller, Ph.D.**[4], **Jennifer Olson, Ph.D.**[4], **Jennifer Richler, Ph.D.**[23], and **Susan Risi, Ph.D.**[4]

[1]Weill Cornell Medical College; White Plains, New York

[2]Department of Child and Adolescent Psychiatry, NYU; New York City, New York

[3]Nathan Klein Institute; Orangeburg, New York

[4]University of Michigan Autism and Communication Disorders Center; Ann Arbor, Michigan

[5]Departments of Pediatrics and Human Genetics, University of Michigan; Ann Arbor, Michigan

[6]Emory University School of Medicine; Atlanta, Georgia

[7]Center for Autism Research, Children's Hospital of Philadelphia; Philadelphia, Pennsylvania

[8]Department of Psychiatry, University of Washington; Seattle, Washington

[9]Department of Psychology, University of Washington; Seattle, Washington

[10]Division of Child and Adolescent Psychiatry, Columbia University Medical Center; New York, New York

[11]Departments of Molecular Physiology & Biophysics and Psychiatry, Vanderbilt Kennedy Center; Nashville, Tennessee

[12]Departments of Pediatrics and Psychiatry, Vanderbilt University Medical Center; Nashville, Tennessee

[13]Marcus Autism Center, Children's Healthcare of Atlanta

[14]Division of Developmental Medicine, Children's Hospital Boston, Harvard Medical School; Boston, Massachusetts

[15]Centre for Autism Research and Treatment and Department of Psychiatry, Semel Institute of Neuroscience, University of California Los Angeles; Los Angeles, California

[16]Montreal Children's Hospital Department of Psychiatry; Montreal, Québec

[17]Thompson Center for Autism and Neurodevelopmental Disorders, University of Missouri; Columbia, Missouri

[18]Department of Pediatrics, Baylor College of Medicine; Houston, Texas

[19]Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago; Chicago, Illinois

[20]Simons Foundation; New York City, New York

[21]Cincinnati Children's Hospital Medical Center; Cincinnati, Ohio

[22]University of Minnesota; Minneapolis, Minnesota

[23]Department of Psychological and Brain Sciences, Indiana University; Bloomington, Indiana

Abstract

Context—Clinical best estimate diagnoses of specific autism spectrum disorders (autistic disorder, pervasive developmental disorder-not otherwise specified, Asperger's disorder) have been used as the diagnostic gold standard, even when information from standardized instruments is available.

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

Design—Multi-site observational study collecting clinical phenotype data (diagnostic, developmental and demographic) for genetic research. Classification trees were employed 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—2102 probands (1814 males) between 4 and 18 years of age (M age=8.93, SD=3.5 years) who met autism spectrum criteria on the Autism Diagnostic Interview–Revised and Autism Diagnostic Observation Schedule and had a clinical diagnosis of an autism spectrum disorder.

Main Outcome Measures—Best estimate clinical diagnoses predicted by standardized scores from diagnostic, cognitive, and behavioral measures.

Results—Though 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 cut-offs.

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 sub-groupings 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.

Introduction

In the field of autism spectrum disorders (ASD), diagnostic instruments have been helpful in defining populations,[1] merging samples,[2] and comparing results across studies,[3,4] nevertheless, best estimate clinical diagnoses (BEC) have long been the gold standard.[5,6,7] In single-site studies, BEC diagnoses 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 DSM-IV-TR[12] and ICD-10[13] (i.e., autistic disorder (AUT), pervasive developmental disorder-not otherwise specified (PDD-NOS), Asperger's disorder (ASP)), upon which BEC diagnoses are based.[5,14,15]

The Simons Simplex Collection (SSC) is a multi-site project, aiming to study *de novo* genetic variations in families that have one child with ASD and one or more unaffected siblings. Diagnostic parameters for probands were intentionally set to include common forms of ASD: AUT, PDD-NOS and ASP. Stringent requirements for training and maintenance of 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 (AUT, PDD-NOS, ASP) 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 characteristics of the children, but rather that are associated with site- and clinician-based differences in how information is used to make diagnoses.

Methods

Participants

2102 probands from 4 to 18 years were evaluated at 12 university-based centers. To prioritize children more likely to have *de novo* copy number variations, inclusion criteria for probands were: a) meeting criteria for ASD on the Autism Diagnostic Observation Schedule (ADOS[16]), b) meeting Collaborative Programs for Excellence in Autism (CPEA) ASD criteria on the Autism Diagnostic Interview-Revised (ADI-R[17]), which has less stringent cut-offs for social and communication domains than "autism" criteria and no requirement for repetitive behaviors or age of onset,[3,18] c) having a nonverbal mental age of at least 18 months and d) a BEC diagnosis of AUT, PDD-NOS or ASP (see www.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; a sibling had substantial language or psychological problems related to ASD or the proband had Fragile X, Tuberous Sclerosis, Down Syndrome or a significant early medical history (e.g., very low birthweight). Sites contributed between 97 and 229 families.

Procedures

Each proband was administered the ADOS and a hierarchy of cognitive tests was implemented across sites, with 88% receiving the Differential Ability Scales, Second Edition (DAS-II[19]), 7% the Mullen Scales of Early Learning (Mullen[20]) and 2-3% each receiving the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV[21]), Wechsler Abbreviated Intelligence Scale (WASI[22]), or other scales. Parents were interviewed using the ADI-R and Vineland Adaptive Behavior Scales, Second Edition (Vineland-II[23]), and completed questionnaires, including the Aberrant Behavior Checklist (ABC[24]). Parents provided informed consent and children provided assent, approved by Institutional Review Boards at each university.

Examiners attended standard research trainings and maintained research reliability with project consultants through semi-annual workshops and video scoring (details in eMethods). Following review of all information and observing the proband in person or on video, the senior clinician (47 psychologists, 6 physicians -- psychiatrists, pediatricians, a clinical geneticist -- and 3 master's level clinicians) specified a BEC diagnosis of AUT, PDD-NOS or ASP according to DSM-IV-TR criteria. Clinicians' years of experience in ASD ranged from less than 5 to more than 20 (see Table 1). Because one goal 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: ADI-Social, ADI-R Verbal Communication (ADI-VC), ADI-R Nonverbal Communication (ADI-NVC), Restricted and Repetitive Patterns of Behavior (ADI-RRB); ADOS domain scores: Social + Communication (ADOS-S+C) and Restricted Repetitive Behavior (ADOS-RRB) totals from Modules 1-4, Social Affect (ADOS-SA) from revised algorithms[25] for Modules 1-3; Calibrated Severity Scores (ADOS-CSS) from Modules 1-3[26]] or demographic, developmental and behavioral [age, gender, race, ethnicity, maternal education, site, verbal IQ (VIQ), performance IQ (NVIQ), Vineland Adaptive Behavior Composite (Vineland-Composite), and irritability and hyperactivity scores from the ABC]. We also considered *diagnosticians' characteristics* (type of degree; 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. Intra-class correlation coefficients; i.e., the ratios of between-site 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 employing permutation tests.[29,30] Categorical measures were described through ranges of proportions across sites; sites differences were assessed with χ^2 tests for independence.

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

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: CART.1 included only diagnostic

scales and clinician characteristics; CART.2 included diagnostic scales, clinician characteristics and site; CART.3 included diagnostic scales, clinician characteristics and site, as well as proband demographic, developmental and behavior characteristics. Finally, separate CART models were fit for each site.

Tree models were first fully 'grown' and then 'pruned' (see eMethods). All analyses were performed with R[32] using the recursive partitioning library *rpart*. Due to space constraints, the main text focuses on the CART.2 model, with a brief discussion of CART.1 and CART. 3 (see eResults for details).

Results from parametric models regarding site differences are also presented using classic inferential procedures. After diagnostic scales associated with BEC as outcome were identified in CART.1, we fit logistic regression models for AUT vs. PDD-NOS or ASP and for ASP vs. AUT 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 to the second two 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 three ASD diagnostic categories (AUT, ASP and PDD-NOS) using BEC diagnoses, $\chi^2(22)=358$, p<0.001. Two out of 12 sites gave fewer than half of the probands AUT diagnoses, while one site gave AUT diagnoses to all probands (see Table 1). Two sites gave PDD-NOS diagnoses to more than 40% of probands. Sites also showed significant differences in the proportion of probands receiving diagnoses of ASP, 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 related to differences in the children, either in specific diagnostic, 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 (see Table 1). In part, this was a function of the CPEA-defined ASD diagnostic criteria[see 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.

Though 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 (See Table 2), with only one site falling outside a 99.5% tolerance band (compared to 11 other sites) on the ADI-Social and ADI-Communication domains and none on the ADI-RRB score. Site density distributions and permutation tolerance bands of ADI-Social, ADOS-RRB domains and NVIQ are shown online in eFigure 1 as examples (additional figures available upon 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 (see eResults for further explanation). Thus, the large

site differences in BEC diagnoses were not accompanied by equivalent differences in standardized diagnostic scores.

Demographic/behavioral characteristics—Mean chronological age was 8.93 years (SD 3.5), with similar distributions of age across sites (see Table 2), with 1814 males and 288 females; differences in sites' proportions of males: females ranged from 5:1 to 9:1 but were not statistically significant. Maternal education was high and homogeneous. Participants from all but three sites were 70% to 90% Non-Hispanic Caucasian, with 4% indicating Asian-American and 4% African-American ancestry and 8% more than one race. Mean IQs were relatively high; Vineland-Composite scores were lower, with less variation within and across sites.

What determines BEC diagnosis?

Following the sequential model fitting strategy outlined earlier, classification trees were grown for BEC diagnosis using different sets of predictors. Details of CART.1 are presented in eResults and eFigure 2, with a brief description here. 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 AUT; diagnoses for the remaining 39% of children with milder social-communication deficits, including most of the children with BEC PDD-NOS and ASP diagnoses and about one-third of the children with BEC AUT, 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. More experienced diagnosticians gave a higher proportion of AUT diagnoses; Ph.D.-level clinicians used PDD-NOS as a diagnosis more often than M.D.s or master's level clinicians. This model reduced the misclassification error from 0.30 (with random assignment based on prevalence) to 0.24, a 20% percent reduction in misclassification rate (explained error, which corresponds to percent-explained variation in a linear regression).

Did sites use the diagnostic scales differently 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 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, ADOS-S+C 12 (left branch) was the only information used in 9 of 12 sites; of the children who had moderate to high observed social-communication deficits at these 9 sites, 91% were given BEC diagnosis of AUT. In the 3 remaining sites, additional information was associated with differentiation of PDD-NOS, ASP, and AUT..

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, <12, on the ADOS), five sites in the left sub-branch (acfgi) made proportionately more AUT diagnoses than the other seven sites, with one site (g) giving only AUT. Four of these five sites further differentiated children using ADOS-CSS (which takes into account age, language level and RRBs). Children with less severe ADOS-CSS scores were split by site again, with two sites

(af) further split by abnormalities in parent reports of children's verbal communication (ADI-VC).

The seven sites in the rightmost sub-branch (bdehjkl) predominantly gave children with milder social-communication impairments PDD-NOS BEC diagnoses. The ADOS-CSS was again taken into account, with children scoring <6 (milder severity) receiving mostly PDD-NOS diagnoses and those   6 given any of the three ASD diagnostic classifications depending on parent-reported historical accounts (ADI-Social and ADI-RRB), as well as diagnosticians' years of experience. When differentiation by site was included, misclassification error rate improved from 0.24 to 0.21 (29% reduction of the total misclassification rate, which constitutes 9% improvement over CART.1).

The importance of site in making BEC diagnosis was also 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) to CART.2 (which also included site and site-by-scale interactions). All p-values comparing the respective nested models were highly significant ($p<1e^{-10}$). From the models where site was treated as a random factor, the variances of the random effects for site and site-by-covariate were quite large - the coefficient of variation (CV=SD of the random effect/mean effect of the covariate) ranged from 0.33 to 4.95, with the largest CV 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.

Did demographic/developmental/behavioral variables matter in making BEC diagnoses?—In CART.3 (see eFigure 3), demographic, developmental and specific behavioral characteristics were added. The primary difference from previous CART models was that, among children with moderate to severe social-communication deficits, the most important factor for BEC diagnosis became VIQ. When children had ADOS-S+C >12 and VIQs <85, 93% received AUT diagnoses across all sites.

In contrast, BEC diagnoses of children with ADOS-S+C >12 and VIQ >85or children with milder ADOS-S+C (<12; right branch), were affected by site differences and many different interactions with each of the diagnostic variables at different stages, as shown in CART.3. Splits were also made on VIQ and NVIQ at a number of places in the tree with cut-offs in IQ ranging from 85 to 122, depending on site. There were no effects of gender, 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 (43% reduction of the total misclassification rate, which is an improvement of 23% compared to CART.1 and 14% compared to CART.2).

Replicability of models

How was BEC diagnosis made at each site?—Individual trees were generated for each site using diagnostic, developmental and demographic variables as predictors. The numbers, although smaller compared to the number of participants used for CART.1-3, are sufficient to have relative confidence in the results (n from 97 to 229). In order 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); see eMethods. Presented online in eFigure 4 are models for 11 out of 12 sites, omitting the site where all probands had AUT.

Several findings for the 11 individual-site CART were striking. As shown in eTable 1, VIQ was the single feature most related to BEC diagnoses in five sites and the second or third strongest predictor in five others (see eFigure 4). However, there were striking site differences in VIQ cut-points and whether IQ was associated with differentiating AUT from PDD-NOS/ASP or AUT/PDD-NOS from ASP. The next most frequent predictors across sites were ADOS social-communication or repetitive behaviors, emerging first in four and two sites respectively. For 9 sites, one of these three measures predicted an entire "node" of diagnosis, in most cases, AUT; but in one case, ASP. Six sites had age effects, primarily such that ASP diagnoses were given to older children, though the age cut-points varied from 5.25 to 12 years. Cut-points for AUT vs. PDD-NOS/ASP for the ADOS-S+C domain varied from 8 to 16. Only one site had an effect of gender and also of maternal education.

Did individual diagnosticians' characteristics affect 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, though 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.

Discussion

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 best estimate clinical (BEC) diagnoses within the autism spectrum (i.e., autistic disorder, PDD-NOS, Asperger's disorder) 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, while 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, parent-reported (ADI-R, ABC) 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), they took into account different factors in deciding which diagnosis to apply to children within the spectrum. Though predictors overlapped across sites, they also differed markedly in "cut-points" (e.g., individual site VIQ cut-points between AUT/PDD-NOS and ASP ranged from 62 to 127) and the order in which information was used.

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

An important concern is the stability of findings based on CART models, which is a tool for discovery, rather than hypothesis testing and inference. To assess this, we evaluated how well the models developed on the 1,169 most recent participants compared to those

developed on the first 933 subjects in the data collection. The misclassification rates for both models were very similar (see eMethods).

Another potential concern for the results presented in eFigure 4 is the relatively small sample size for the individual sites. Again, results for models developed on the most recent 1,169 participants were nearly identical to those from the first 933 participants, except that maternal education played less of a role in the larger sample and one site had a gender effect.

Previous research[2,9] has shown that within a site, clinicians' diagnoses can add information to standardized scores. With consistent application of BEC decision rules, or if standard training had been offered, BEC diagnoses might have been an important source of information in this 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 DSM-V and ICD-11. Recurrent evidence of the importance of information external to a psychiatric diagnosis, particularly verbal IQ and current language level (e.g., ADOS module), supports the need for cognitive function and language level to be considered as essential to BEC diagnoses of ASD. Diagnostic classifications based upon 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-report measures of social communication and repetitive behaviors clearly contributed to clinical diagnoses. Within these 12 sites with experienced and well-trained staff, distributions of dimensional measures of standardized instruments were much more consistent than categorical BEC diagnoses. More precise diagnostic criteria might have improved them, 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 (e.g., 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 clinical categorical diagnoses of autistic disorder, PDD-NOS or Asperger's disorder.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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. [PubMed: 11569587]

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. [PubMed: 16754843]

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. [PubMed: 16926617]

4. Gotham K, Risi S, Dawson G, Tager-Flusberg H, Joseph R, Carter A, Hepburn S, McMahon W, Rodier P, Hyman S, Sigman M, Rogers S, Landa R, Spence A, 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. [PubMed: 18434924]

5. Chakrabarti S, Fombonne E. Pervasive Developmental Disorders in preschool children: Confirmation of high prevalence. Am J Psychiatry. 2005; 162(6):1133–1141. [PubMed: 15930062]

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. [PubMed: 11944873]

7. Volkmar F, Chawarska K, Klin A. Autism in infancy and early childhood. Annu Rev Psychol. 2005; 56(1):315–336. [PubMed: 15709938]

8. Lord C. Follow-up of two-year-olds referred for possible autism. J Child Psychol Psychiatry. 1995; 36(8):1365–1382. [PubMed: 8988272]

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. [PubMed: 10188704]

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. [PubMed: 19521762]

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. [PubMed: 17663613]

12. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition, Text Revision: DSM-IV-TR. American Psychiatric Association; Washington, DC: 2000.

13. World Health Organization. International Statistical Classification of Diseases and Related Health Problems. 10th revision. World Health Organization; Geneva, Switzerland: 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. [PubMed: 1787139]

15. Klin A, Volkmar FR. Asperger syndrome. Child Adolesc Psychiatr Clin N Am. 2003; 12(1):xiii–xvi. [PubMed: 12512394]

16. Lord, C.; Rutter, M.; DiLavore, PS.; Risi, S. Autism Diagnostic Observation Schedule (ADOS). Western Psychological Services; Los Angeles, CA: 1999.

17. Rutter, M.; Le Couteur, A.; Lord, C. The Autism Diagnostic Interview - Revised (ADI-R). Western Psychological Services; Los Angeles, CA: 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. 2006; 140A(21):2257–2274. [PubMed: 17022081]

19. Elliott, CD. Differential Ability Scales—Second Edition (DAS-II). The Psychological Corporation; San Antonio, TX: 2006.

20. Mullen, E. The Mullen Scales of Early Learning. American Guidance Service, Inc.; Circle Pines, MN: 1995.

21. Wechsler, D. Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV). The Psychological Corporation; San Antonio, TX: 2003.

22. Wechsler, D. Wechsler Abbreviated Scale of Intelligence. The Psychological Corporation; San Antonio, TX: 1999.

23. Sparrow, SS.; Cicchetti, DV.; Balla, DA. Vineland Adaptive Behavior Scales. Second Edition. AGS Publishing; Circle Pines, MN: 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. [PubMed: 3993694]

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. [PubMed: 17180459]

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. [PubMed: 19082876]

27. Scott, DW. Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley-Interscience; Hoboken, NJ: 1992.

28. Diggle, P. Analysis of Longitudinal Data. Oxford University Press; United Kingdom: 2002.

29. Fan J, Lin S-K. Test of significance when data are curves. J Am Stat Assoc. 1998; 93(443):1007–1021.

30. Westfall, PH.; Young, SS. Resampling-based multiple testing: Examples and methods for p-value adjustment. John Wiley and Sons; Hoboken NJ: 1993.

31. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C.; Steinberg, D.; Colla, P. CART: Classification and Regression Trees. Wadsworth; Belmont, CA: 1983.

32. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2010. Available at: http://www.R-project.org

33. Wing, L. Asperger Syndrome. The Guilford Press; New York, NY: 2000. Past and future of research on Asperger syndrome; p. 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. [PubMed: 12042174]

35. Alarcón M, Cantor RM, Liu J, Gilliam TC, Geschwind DH. Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. Am J Hum Genet. 2002; 70(1):60–71. [PubMed: 11741194]

36. National Research Council. Committee on Educational Interventions for Children with Autism. Educating Children with Autism. The National Academies Press; Washington, D.C.: 2001.
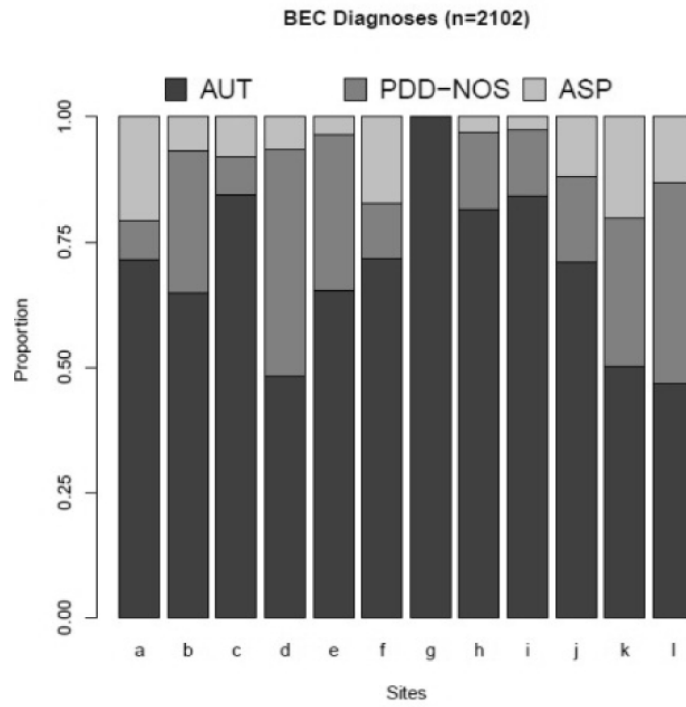
Figure 1.
BEC diagnoses across sites.
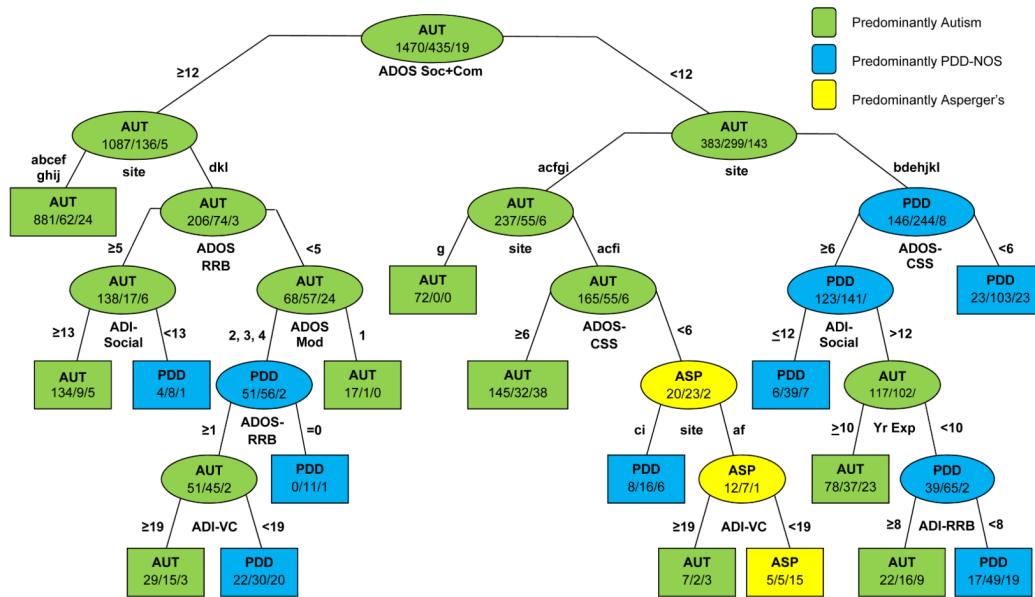
Figure 2. CART.2 for BEC diagnoses with diagnostic scales, site, and diagnostician characteristics as predictors
Numbers denote Ns for each diagnostic group: AUT/PDD-NOS/ASP. ADOS-Soc +Com=ADOS Social + Communication Domain Total; ADOS-CSS=ADOS Calibrated Severity Score; ADOS Mod=ADOS Module; ADI-Social=ADI-R Social Total; ADI-VC=ADI-R Verbal Communication Total; ADI-NVC=ADI-R Nonverbal Communication Total; ADI-RRB=ADI-R Restricted and Repetitive Behaviors Total; Yr Exp=Senior Diagnostician's Number of Years of Experience. .

Table 1

Summary of variability of factors characterizing probands and diagnosticians across sites

	Overall proportion (%)	Range across sites (%)	SD proportion across sites (%)
Gender			
Female	13.7	10.36~17.47	2.56
Male	86.3	82.53~89.64	2.56
Race ***			
White	78.54	47.16~90.34	12.90
African American	4.04	0.49~12.68	3.25
Asian	3.95	1.38~7.95	2.20
Native American	0.14	0~0.69	0.27
Native Hawaiian	0.05	0~0.49	0.14
Other	4.28	0~18.18	5.34
More than one race	7.99	2.06~19.81	5.35
Not specified	1.00	0~5.68	1.65
Maternal Education (highest level obtained) ***			
Graduate	25.30	17.46~40.58	7.34
Baccalaureate	36.01	30.29~43.66	3.88
Associate/Some College	29.12	20.1~35.98	5.83
High school/GED/Some HS	9.42	6.29~11.77	1.72
Less than 9th grade	0.14	0~1.05	0.35
ADOS Diagnostic Classification ***			
Autism	87.82	79.41~95.88	5.09
Autism Spectrum	12.18	4.12~20.59	5.09
ADOS Module ***			
Module 1	16.94	7.98~26.57	5.14
Module 2	22.98	15.86~31.43	5.05
Module 3	57.09	48.31~67.88	6.69
Module 4	3.00	0~6.52	2.05
ADI-R Diagnostic Classification (CPEA)			
Autism	90.25	87.32~93.45	2.07
Autism Spectrum	9.75	6.55~12.68	2.07
Clinician's Best Estimate Diagnosis ***			
AUT	69.93	46.74~100	16.25
PDD-NOS	20.69	0~45.29	14.12
ASP	9.37	0~20.71	7.08
Most senior diagnostician: Highest degree ***			
a-MD	12.18	0~84.83	25.95

	Overall proportion (%)	Range across sites (%)	SD proportion across sites (%)
b-PhD	87.54	15.17~100	25.92
c-MA	0.29	0~2.06	0.67
Most senior diagnostician: Years of experience ***			
10 or more years	61.13	0~100	39.46
5-10 years	30.69	0~100	36.12
Less than 5 years	8.18	0~41.30	13.91

X^2 test for independence between site and a factor

Degrees of freedom =(# sites–1)×(# levels–1) = 11×(# levels–1)

Total N= 2102, Site range from N=97 to N=229

***
p .001

Table 2

Summary of variation between sites with respect to diagnostic scales and continuous demographic and behavior characteristics

	Range across Site				Variance				ICC
	Min	Max	Mean	SD	Overall Mean	Overall SD	Within	Between	Between/ Total
Chronological Age	48~51	209~216	101.8~117.3	38.5~45	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 VMA							435.9	5.5	0.013
Nonverbal IQ									
no control	9~30	133~161	79.7~89.8	22~27.7	86.1	25.3	635.1	4.7	0.007
control for NVMA							375.6	3.7	0.010
Autism Diagnostic Interview-Revised									
Social Interaction	8~9	30~30	18.5~22.6	5.1~6.1	20.1	5.7	31.8	1.4	0.043
Communication - Verbal	6~8	24~26	15.6~18.4	3.5~4.8	16.4	4.2	17.5	0.7	0.036
Communication - Nonverbal	0~3	14~14	8.2~10.3	3.1~3.6	9.1	3.4	11.5	0.5	0.038
Restricted/Repetitive Behavior	0~2	12~12	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	10~10	6.8~8.1	1.6~1.8	7.4	1.7	2.8	0.1	0.041
Social+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	8~8	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	8.9~11	16.5	10.5	107.3	2.4	0.022
Sample Size	97	229	175	38					