

Validation of Proposed *DSM-5* Criteria for Autism Spectrum Disorder

Thomas W. Frazier, Ph.D., Eric A. Youngstrom, Ph.D., Leslie Speer, Ph.D.,
Rebecca Embacher, B.S., Paul Law, M.D., M.P.H., John Constantino, M.D.,
Robert L. Findling, M.D., Antonio Y. Hardan, M.D., Charis Eng, M.D., Ph.D.

Objective: The primary aim of the present study was to evaluate the validity of proposed *DSM-5* criteria for autism spectrum disorder (ASD). **Method:** We analyzed symptoms from 14,744 siblings (8,911 ASD and 5,863 non-ASD) included in a national registry, the Interactive Autism Network. Youth 2 through 18 years of age were included if at least one child in the family was diagnosed with ASD. Caregivers reported symptoms using the Social Responsiveness Scale and the Social Communication Questionnaire. The structure of autism symptoms was examined using latent variable models that included categories, dimensions, or hybrid models specifying categories and subdimensions. Diagnostic efficiency statistics evaluated the proposed *DSM-5* algorithm in identifying ASD. **Results:** A hybrid model that included both a category (ASD versus non-ASD) and two symptom dimensions (social communication/interaction and restricted/repetitive behaviors) was more parsimonious than all other models and replicated across measures and subsamples. Empirical classifications from this hybrid model closely mirrored clinical ASD diagnoses (90% overlap), implying a broad ASD category distinct from non-ASD. *DSM-5* criteria had superior specificity relative to *DSM-IV-TR* criteria (0.97 versus 0.86); however sensitivity was lower (0.81 versus 0.95). Relaxing *DSM-5* criteria by requiring one less symptom criterion increased sensitivity (0.93 versus 0.81), with minimal reduction in specificity (0.95 versus 0.97). **Conclusions:** Results supported the validity of proposed *DSM-5* criteria for ASD as provided in Phase I Field Trials criteria. Increased specificity of *DSM-5* relative to *DSM-IV-TR* may reduce false positive diagnoses, a particularly relevant consideration for low base rate clinical settings. Phase II testing of *DSM-5* should consider a relaxed algorithm, without which as many as 12% of ASD-affected individuals, particularly females, will be missed. Relaxed *DSM-5* criteria may improve identification of ASD, decreasing societal costs through appropriate early diagnosis and maximizing intervention resources. *J. Am. Acad. Child Adolesc. Psychiatry*, 2012;51(1):28–40. **Key Words:** autism spectrum disorder, diagnosis, factor analysis, latent class, factor mixture

Autisms are a highly heterogeneous set of disorders with wide variations in symptom severity, intellectual level, and functional disability.¹ The importance of accurately identifying individuals with autism has never been greater, particularly given the growing prevalence,² considerable family and societal costs,³ and recognized importance of early diagnosis and intervention. *DSM-5* field trials have

recently begun evaluating new diagnostic criteria that contain several important modifications relative to *DSM-IV-TR*.^{4,5} The most remarkable change was combining specific *DSM-IV-TR* diagnoses into a single broad autism spectrum disorder (ASD). This proposed change has generated considerable apprehension from patients and their families, who are concerned that individuals diagnosed with Asperger's disorder will be orphaned or will receive inappropriate service provision.⁶ Yet, to date, there is little evidence that Asperger's disorder is qualitatively distinct from other autism diagnoses at the symptom level, or that the new criteria will underidentify high-functioning ASD.⁷ The present study eval-



This article is discussed in an editorial by Dr. Bennett L. Leventhal on page 6.



Supplemental material cited in this article is available online.

uates an important aspect of this controversy—namely, symptom continuity between individuals with Asperger's disorder and other ASD cases—by explicitly testing conflicting views of the nature of autism symptom structure.

The first viewpoint proposes that autism symptoms are best represented dimensionally, with differences between typical and ASD symptom levels being a matter of degree (i.e., no distinct ASD category is present).⁸ This perspective is supported by the observations from population and family studies of a broad distribution of *observed* autism symptoms in the population,⁹ elevated levels of autism traits in siblings and other family members of affected cases,^{10,11} and shifting toward greater autism symptoms in children whose parents both show subthreshold autism traits.¹² Also compelling are possible genetic heterogeneity of symptom domains^{13,14} and the substantial variability of autism symptoms among identical twins.^{15,16} This evidence points toward the need to include a dimensional conceptualization of autism symptoms in its diagnosis.

The second viewpoint is that autism symptoms represent a category, with qualitative differences in symptom levels between ASD-affected and unaffected individuals.¹⁷⁻¹⁹ Two recent studies supported the categorical conceptualization of autism, identifying a *latent* symptom category parsing ASD and non-ASD cases.^{20,21} Longitudinal studies of early (age 2 years) and later (ages 3-9 years) diagnoses also support the notion of a single, distinct ASD category.^{22,23} These studies found strong stability of the broad ASD distinction but substantial shifting within specific *DSM-IV* diagnoses. Similarly, most studies of ASD symptoms and cognitive processes have identified only quantitative distinctions among *DSM-IV* disorders.^{7,24,25}

It is possible that both viewpoints are correct, and that categorical and dimensional aspects of autism symptoms should be considered in the conceptualization of ASD. In fact, in addition to an ASD category, *DSM-5* also identifies two symptom dimensions—Social Communication and Interaction (SCI) and Restricted, Repetitive Behavior (RRB)—that collapse the three *DSM-IV-TR* domains. Thus, *DSM-5* presents a complex model with a single ASD category superimposed on two primary symptom dimensions. The present study investigated the plausibility of the *DSM-5* conceptualization of autism symptoms by comparing categorical, dimensional, and hybrid

latent variable models (Aim 1).²⁶⁻²⁸ Based on the *DSM-5* model of ASD, we hypothesized that autism symptoms would be most parsimoniously represented by hybrid models specifying a categorical distinction (ASD versus non-ASD) and two dimensions (SCI and RRB).

The distinction between dimensional, categorical, and hybrid models has relevance beyond diagnostic conceptualization, affecting the clinical assessment approach. If autism symptoms are best represented by a latent category, the prevalence of ASD, sibling recurrence rates, and the most useful diagnostic criteria are defined. In this case, an evidence-based medicine approach to clinical assessment that includes the generation of post-test probabilities of diagnosis is indicated.²⁹ In this scenario, instruments should be used to optimize classification rather than only grading autism symptom severity. In contrast, if only a latent continuum is needed to describe autism symptoms, additional research is needed to link the continuum to clinically relevant outcomes, such as functional deficits, before setting a diagnostic threshold. If hybrid models are supported, integration of the above approaches will be needed to most accurately characterize the presence versus absence of ASD and to grade symptom severity.

Identifying the latent structure of autism also facilitates the design and analysis of future research.^{30,31} If the latent structure of autism is categorical, then traditional group comparison research is optimized if individuals are accurately sorted into ASD and non-ASD groups. Alternatively, if a continuum is identified, research designs will be more efficient if autism symptoms are measured as quantitative traits and regression or latent factor approaches are implemented rather than group comparisons. A latent ASD category might also facilitate future neurobiological and genomic research. For example, a latent category may support developmental models postulating distinct divergence in brain structure and function.^{32,33} An ASD category could also favor molecular models emphasizing the primacy of strong individual genomic effects,³⁴⁻³⁷ such as private mutations or copy number variations involving genes crucial for early brain maturation,³⁸ although this is less certain because of the potential for genetic threshold effects. In contrast, strictly dimensional representations of autism would suggest graded alterations in brain development, overlapping

typical brain trajectories, and a primary role for polygenic common variation and other additive genomic and environmental effects.^{37,39,40} Finally, hybrid models of ASD symptoms may imply a more complex pattern, such as combinations of rare and common variation as well as environmental effects contributing to the ASD phenotype.^{27,41} Comparing models of autism symptom structure will be crucial for advancing diagnostic criteria, clinical assessment, and future molecular and neurobiological research.

The present study also evaluated specific changes in the proposed *DSM-5* algorithm examined in phase I trials (Aim 2). These changes include the following: collapsing three domains into two, namely, Social Communication and Interaction (SCI) and Restricted, Repetitive Behavior (RRB); requiring all SCI criteria and at least two of four RRB criteria to be met; necessitating at least two specific symptoms present per criterion; providing a separate RRB criterion evaluating sensory sensitivities and unusual sensory interests; and adding subtle behavioral manifestations of high-functioning ASD. Thus, as a secondary aim, we explored the relative classification accuracy of *DSM-IV-TR* and proposed *DSM-5* criteria to inform future diagnostic revisions in Phase II trials and to provide guidance to evaluating clinicians.

METHOD

Sample

Data were obtained from the Interactive Autism Network (IAN; www.ianproject.org), a U.S., Internet-based registry for families with one or more ASD-affected child (IAN Data Export ID: IAN_DATA_2010-07-06). Families were eligible for enrollment in IAN if the parent or legal guardian who provided information was English speaking, the family lived in the United States, and their child was diagnosed with an ASD by a professional. To be included in the present study, caregivers must have reported ASD symptom data for at least one ASD-affected child. Caregivers also reported whether each registered child was clinically diagnosed with ASD and, if ASD affected, the specific diagnosis obtained from a prior clinical evaluation. The vast majority of clinical ASD diagnoses (93%) were provided by a doctoral level professional or team. A substantial proportion of youth (74.9%) were diagnosed using the Autism Diagnostic Interview—Revised, the Autism Diagnostic Observation Schedule, or both. Of these youth, 98.7% scored in the ASD-affected range for one or both instruments. A recent study of youth from the IAN registry provided addi-

tional support for the validity of clinical ASD diagnoses.⁴² In this study, randomly selected verbal youth with a score higher than 12 on the SCQ were evaluated using the Autism Diagnostic Interview—Revised and by an expert clinician's observation. The great majority of youth (98%) were confirmed to have a *DSM-IV-TR* clinical diagnosis of ASD.

Autism symptoms, ages of developmental milestones, and other developmental concerns were also reported for siblings without a clinical ASD diagnosis. This non-ASD sibling subsample includes substantial proportions of youth who have caregiver-reported diagnoses of ADHD (11.8%), anxiety disorder (5.4%), mood disorder (4.2%), intellectual disability (0.5%), motor delay (6.6%), and/or language delay (20.2%). A nontrivial fraction of these siblings also show characteristics of the broad autism phenotype.⁴³ For these reasons, we refer to these youth as parent-designated non-ASD siblings. The heterogeneity of these siblings provides a reasonable comparison for examining both the latent structure and the relative efficiency of diagnostic criteria.

Informed consent was obtained from parents or guardians for all participants before entry into the IAN data collection. The procedures of IAN were reviewed and approved by the institutional review board of Kennedy Krieger Institute. The procedures of the present study were reviewed and approved by the institutional review board of the Cleveland Clinic.

Measures

Autism symptom data were provided using the Social Responsiveness Scale (SRS)⁴⁴ and/or the Social Communication Questionnaire (SCQ).⁴⁵ The SRS is 65-item, ordinaly scaled (1 = "not true" to 4 = "almost always true") questionnaire that provides a quantitative assessment of the severity of autism traits. The SCQ is a dichotomously keyed (yes/no) rating scale that consists of 40 questions, many of which tap *DSM-IV-TR* symptom domains. Lifetime ratings referenced the child's behavior throughout their developmental history, increasing diagnostic validity.⁴⁶ Both the SRS and SCQ have been extensively validated and distinguish youth with autism from those with other psychiatric conditions.^{44,47-49} SRS subscales provide a quantitative assessment of ASD symptoms and decrease the likelihood of falsely identifying a latent category because of limited scaling. Because the SRS includes only one scale evaluating RRB behaviors, we randomly divided the 12 items on the autism mannerisms scale into four packets to improve identification of the RRB domain in latent variable models. Packets or parcels have important psychometric advantages compared with directly analyzing items scores, including increasing the reliability and variance of the scales, and improving model stability and interpretability.⁵⁰

Data Analysis

The hypothesis that autism symptoms would be most parsimoniously explained by a category (ASD versus non-ASD) and two subdimensions (SCI and RRB) was evaluated by estimating a series of latent class (LCA; two to six classes) and exploratory factor analyses (EFA; one to three factors) (Aim 1). Latent class analysis assumes the existence of discrete categories, whereas exploratory factor analysis assumes latent dimensions or continua of symptoms. Recent developments in latent variable modeling now permit the simultaneous integration of categories and dimensions.^{26,27} These hybrid factor mixture (FM) models integrate both EFA and LCA, with latent factors modeled across classes. Comparing the fit of EFA, LCA, and FM models addresses the question of whether categories (classes), dimensions (factors), or both are needed to represent autism symptoms. When applied to autism symptom data, hybrid models examine whether autism symptoms measure one or more dimensions as well as whether individual differences in those dimensions result from one or more groups of individuals. If a single group provides optimal fit, then dimensions capture graded increases in autism symptom levels. If two or more groups provide superior fit, then symptom dimensions result from two or more distinct groups or distributions (e.g., ASD versus non-ASD, or autism versus Asperger's versus non-ASD). Based on empirical work, the results of LCA and EFA models guided the selection of plausible FM models.^{26,51} For example, if a two-factor EFA fit better than three-factor models, FM models emphasized one- and two-factor solutions, as it is unlikely that three-factor FM models would improve fit when both factors and classes were included.

Models were fit in the total available samples (including ASD and non-ASD siblings) and the non-ASD sibling subsamples. Sibling comparators were examined separately because the full sample may be biased toward identification of a pseudo-category because of the polarizing effects of social comparison processes or rater contrast effects (i.e., rating presumed non-ASD siblings as very healthy and ASD siblings as very severely affected). Estimating models in non-ASD siblings circumvents this potential bias. We expected to observe a low base rate class with significant ASD symptoms in the sibling comparison subsample. This expectation was based on previous analyses,²⁰ the observation of high SCQ and SRS scores in a nontrivial minority of caregiver-designated non-ASD siblings in the IAN registry (>3% show SCQ total score ≥ 15 and/or SRS total t score ≥ 70), a similar observation of high scoring youth with autism or neurobehavioral disorders in population samples,⁴⁷ and recent findings of language delay with autistic qualities of speech in a minority of these siblings.⁴³ Primary analyses focused on the SRS because this measure evaluates autism symptoms as a quantitative trait and is less biased

toward identifying categorical latent structure. Sensitivity analyses were also conducted with the SCQ. These analyses were conducted to internally replicate latent variable model findings using this very different type of measure and to determine the sensitivity of model findings to variations in sampling (see Table S1, available online, which presents the sample/sub-sample, item/scale set, and purpose of each latent variable model analysis).

All models were estimated accounting for the clustered nature of the sibling data. LCA models assumed conditional independence of variables within classes. FM models were estimated specifying strict or strong measurement invariance across classes.⁵² The pattern of model comparisons was similar, and therefore results were presented with strong measurement invariance. The best-fitting models were determined by comparing the Bayesian Information Criteria (BIC),^{26,28} in which lower values indicate better model fit. Results were highly similar for the Akaike and sample-adjusted Bayesian criteria. Class proportions were also evaluated for FM models to ensure that better fitting models did not overfit the observed score distribution. Empirically derived class assignments (ASD versus non-ASD) were computed using the best-fitting model.⁵² These assignments served as an additional proxy of ASD diagnosis,⁵³ complementing clinical ASD diagnoses. The terms "ASD class" and "non-ASD class" described these empirical assignments even when they were derived from the caregiver-designated non-ASD sibling subsample. Independent-sample t tests and χ^2 tests evaluated differences in developmental milestones across empirical classifications in the non-ASD siblings. Differences in the ages of developmental milestones were anticipated to be small in magnitude across ASD and non-ASD empirical classifications. Thus, they were not expected to be clinically meaningful, but, rather, to provide theoretical validation of the notion of a small ASD class in these caregiver-designated non-ASD siblings.

To explore the relative classification accuracy of DSM-IV-TR and DSM-5, autism symptoms from the SCQ and SRS were mapped onto DSM-IV-TR and proposed DSM-5 diagnostic criteria (see Table S2, available online). SCQ items were mapped first with a limited number of SRS items (eight total) chosen only to fill gaps where SCQ items did not evaluate a DSM-5 symptom. Because SCQ items are based on DSM-IV-TR criteria, only six SRS items were included to provide a balance of the total number of items mapped to DSM-IV-TR and DSM-5 criteria. SCQ items were predominant to ensure that subsequent analyses were favored DSM-IV-TR rather than DSM-5 criteria. This yields a stronger test of the relative efficiency of DSM-5 criteria rather than predominantly SRS items, which were not developed in relation to diagnostic criteria. Mapped DSM-IV-TR criteria were considered met if any of the algorithms for autistic disorder, PDD

NOS, or Asperger's disorder were endorsed. Algorithms were followed as specified in the Phase I Field Trials criteria. Iterations of this criteria were also examined, including the most recent iteration as of this writing that requires only one specific symptom present per criterion.⁵ SCQ and SRS item mapping was done in a subsample with complete data on both measures. Because most individuals enrolling in IAN receive the SCQ, this subsample has highly similar composition to the SRS sample (see Results for comparison of SCQ and SRS samples).

Analyses compared the sensitivity and specificity of *DSM-IV-TR*, proposed *DSM-5*, and alterations of *DSM-5* to proxy ASD diagnoses. Greater weight was given to specificity to limit false-positive ASD diagnoses. Analyses of the *DSM-5* algorithm are exploratory because criteria were not perfectly mapped, the non-ASD sibling sample is not equivalent to a clinical comparison sample, clinical observation was not included, and proxies for ASD diagnosis are only estimates of true ASD. However, because of the important role of parent report, particularly for symptom history in older individuals and for behaviors that are difficult to observe in an office visit,⁵⁴ these analyses provide a reasonable evaluation of the relative efficiency of different versions of diagnostic criteria.

Latent variable models were estimated using MPlus v5.2. All other analyses used SPSS version 18 (SPSS Inc., Chicago, IL). Effect sizes are presented as Cohen's *d* with conventions of very small ($d < 0.10$), small ($d = 0.20$), medium (0.50), and large ($d = 0.80$).⁵⁵

RESULTS

Sample Characteristics

The total IAN sample included symptom data from 14,774 youth aged 2 to 18 years (8,911 ASD; 5,863 caregiver-designated non-ASD siblings) provided by 10,038 caregivers. There was a predominance of males with clinical ASD diagnoses (82.3%) and females in the sibling comparison group (53.9%), consistent with the sex ratio of autism. SRS data were available from 6,949 youth (4,248 ASD and 2,701 non-ASD siblings), and SCQ data were available from 14,200 youth (8,606 ASD and 5,594 non-ASD siblings), respectively. The smaller SRS sample had highly similar demographic and clinical characteristics to the larger SCQ sample with a three notable exceptions. First, the SRS sample was slightly older (SRS mean = 8.40, standard deviation [SD] = 3.99, SCQ mean = 7.53, SD = 4.37; $t(14772) = 12.69$, $p < .001$, $d = 0.21$) because the instrument is completed only for individuals aged 4 years and older, whereas the SCQ is completed for individuals aged 2 and older. In addition, the

total SRS sample included fewer nonverbal individuals even when only youth ages 4 and older are considered (SRS 8.4%, SCQ 12.8%; $\chi^2(1) = 61.10$, $p < .001$; $d = 0.14$) and a slightly higher proportion of white/non-Hispanic individuals (SRS 86.3%, SCQ 85.0%; $\chi^2(1) = 5.42$, $p = .020$; $d = 0.04$). The above differences, although statistically significant, ranged from very small to small in magnitude. The SRS and SCQ samples had similar distributions across sex and specific *DSM-IV-TR* ASD diagnoses, and showed equivalent proportions of individuals receiving the ADOS or ADI-R as part of their clinical diagnostic evaluation (largest $\chi^2(1) = 3.65$, $p > .05$). Additional characteristics of the IAN sample have been described elsewhere.^{20,43}

Model Fit (Aim 1)

Table 1 displays comparisons for models using SRS subscales. LCA models showed steady improvements in fit from two- to six-class solutions, but no minimum was reached for any information criterion, and fit was generally weaker than EFA models. The two-factor EFA model divided SRS social and autism mannerisms variables consistent with proposed *DSM-5* SCI and RRB domains. The three-factor EFA model fit better, but had only one minor loading on the third factor. Therefore, FM models were estimated with two factors separating social and autism mannerism variables.²⁸

The two-factor/three-class FM model fit slightly better than all other models. However, the third class appears to overfit the symptom distribution by splitting ASD-affected youth according to extreme and less extreme groups across all SRS scales. Despite the large sample size, there were no significant differences in the proportion of individuals diagnosed with autistic disorder, PDD NOS, or Asperger's disorder between these classes [$\chi^2(1) = 2.45$, $p = .118$], indicating that no additional information was added by retaining the third class. Furthermore, the third class tended to fit more poorly and was not replicated in other analyses. Thus, the two-factor/two-class FM model represented the most parsimonious solution (Figure 1).

Supporting our prediction, similar results were obtained in the non-ASD sibling subsample. EFA and LCA models fit more poorly (lowest BIC = 42420) than FM models. The two-factor/three-class FM model fit best (BIC = 41480) but the third class (5%) overfit extreme

TABLE 1 Model Comparisons, Fit Statistics, and Class Proportions Based on Autism Symptoms

Model	Factor/Class No.	LL	Par.	BIC	Comparison	Class Proportions
LCA	Two-class	-134017	25	268256		0.57 , 0.43
	Three-class	-127148	34	254598		0.31 , 0.32 , 0.37
	Four-class	-124544	43	249468		0.17 , 0.20 , 0.30 , 0.33
	Five-class	-123150	52	246759		0.10 , 0.13 , 0.21 , 0.25 , 0.31
	Six-class	-122499	61	245537	Best fitting LCA	0.09 , 0.11 , 0.19 , 0.19 , 0.12, 0.30
EFA	One-factor	-123767	24	247746		
	Two-factor	-122803	31	245881	Parsimonious EFA	
	Three-factor	-122575	37	245478	Best-fitting EFA	
FM	Two-factor one-class	-122994	25	246208		
	Two-factor two-class	-122055	35	244420	Parsimonious model	0.63 , 0.37
	Two-factor three-class	-118305	47	237027	Best-fitting overall	0.14 , 0.56 , 0.30

Note: Lower values of the Bayesian Information Criterion (BIC) indicate better fit. Akaike Information Criterion and sample-adjusted BIC showed an identical pattern. Factor mixture models specified strong measurement invariance. Boldface type designates classes with predominantly autism spectrum disorder-affected youth. EFA = exploratory factor analysis; FM = factor mixture model analysis; LCA = latent class analysis; LL = log likelihood; No. = number; Par. = number of estimated parameters.

scores with no qualitative differences in symptom pattern. The two-factor/two-class FM model was most parsimonious (BIC = 41981).

Figure 1 displays results of the two-factor/two-class hybrid model. Base rates of the ASD and non-ASD empirically derived classes are presented for both the total sample and non-ASD sibling subsample. In the total sample, the ASD class had a high base rate (63%) that was similar to the base rate of caregiver-reported clinical ASD diagnosis (60%). In the non-ASD sibling subsample, a low base rate of the ASD class (9%) was observed. This was consistent with the prediction of only a small proportion of non-ASD siblings having ASD and should be viewed in light of suggestions from the empirical literature that small base rates such as this may be overestimated.⁵²

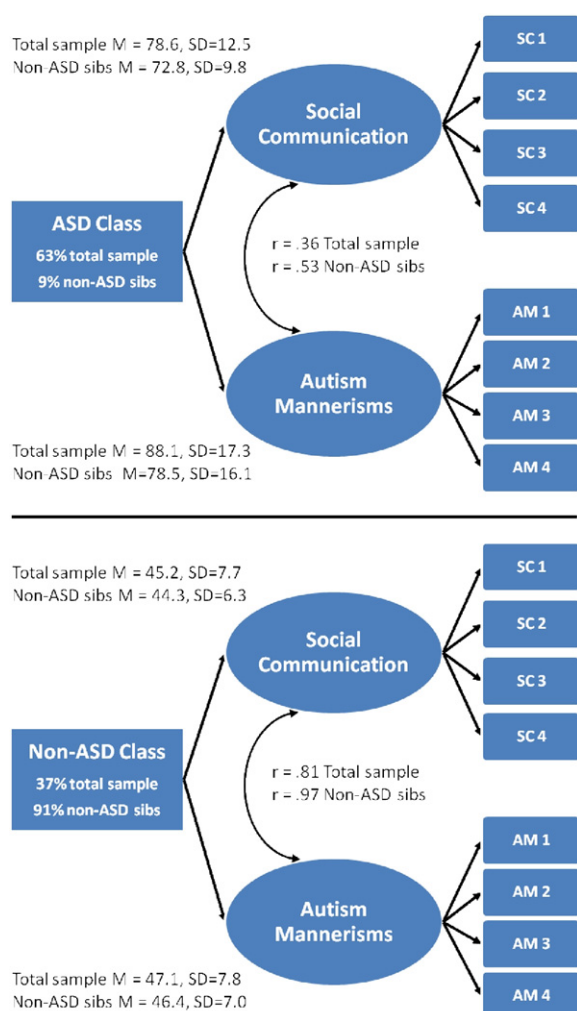
Figure 1 also presents ASD and non-ASD class means and standard deviations for SRS social and autism mannerism T scores, separately for the total and non-ASD sibling subsamples. The means and standard deviations show strong separation of scores across ASD and non-ASD classes in both samples. Figure 2 plots SRS social and autism mannerism T-scores for youth in ASD and non-ASD classes, separately for the total sample (left panel) and the non-ASD sibling subsample (right panel). Inspection of the total sample (left panel) reveals clustering of class values into ASD (orange) and non-ASD classes (blue). However, broad dimensions of symptom severity were also present within each cluster, with higher variance and a weaker correlation

between social and autism mannerisms scales in the ASD class (correlations are presented in Figure 1). For the total sample (left panel), social and autism mannerism T scores in the ASD class (orange) ranged from 45 to more than 100. Social and autism mannerism T scores were lower in the non-ASD class, ranging from 35 to 65. A similar pattern was observed for the non-ASD sibling subsample (right panel). The only exception was that the ASD class derived from the non-ASD sibling subsample is considerably smaller and has few very high scores relative to the ASD class from the total sample. Importantly, and despite overlapping score ranges, ASD and non-ASD classes showed very large ($d > 2.0$), highly significant separation on SRS social and autism mannerism t scores [smallest $t(2699) = 57.3$, $p < .001$, $d = 2.21$]. Large separation on these scores suggested that low and high scores might have clinical utility for separating ASD and non-ASD classifications. To examine this possibility, we computed likelihood ratios for low ($T \leq 50$) and high score ($T \geq 65$) ranges on SRS scales in the total sample. Low scores on both the SRS social and autism mannerism scales were highly useful for decreasing the probability of ASD (likelihood ratio = 0.007). High scores on both scales greatly increased the probability of ASD (likelihood ratio = 14.3).

Sensitivity Analyses

To examine the sensitivity of model comparisons to the measure used (SRS versus SCQ) and sam-

FIGURE 1 Two-factor/two-class model using the Social Responsiveness Scale in the total available sample ($N = 6,949$) and sibling subsample without autism spectrum disorder (ASD) ($N = 2,701$). Note: The ASD class had substantially higher symptom levels (T scores) than the non-ASD class in both the total sample and non-ASD sibling subsample ($p < .001$). Non-ASD siblings designate model results for the non-ASD sibling subsample. SC 1-4 = social communication/interaction scales from the Social Responsiveness Scale. AM 1-4 = autism mannerisms item packets from the Social Responsiveness Scale. Horizontal black line separates results for the empirically derived ASD and non-ASD classes. Curved double-headed arrows represent factor correlations in the ASD and non-ASD classes.



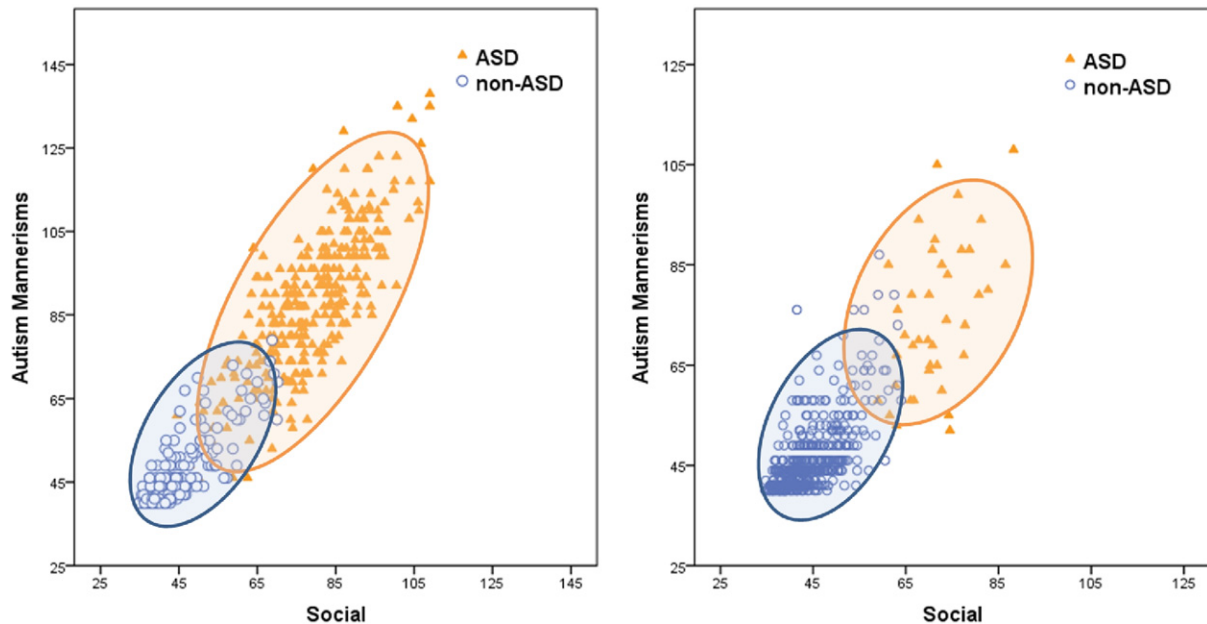
pling characteristics, the same set of models were estimated using SCQ items across the total available sample and several demographic subsamples (non-ASD siblings, verbal youth, multiple-incidence families, females, age <7

years, and age ≥ 7 years). The pattern of results for each set of model comparisons was highly similar to the primary models using SRS indicators. EFA supported the presence of two primary factors. LCA tended to require a large number of latent classes to approximate FM model fit. The two-factor/two-class FM model was typically the best fitting and always the most parsimonious model in the SCQ total and subsamples. The SCQ total sample closely matches the complete IAN registry,⁵⁶ ensuring that results were not a function of a biased subset of IAN. Results also generalized across changes in FM model specifications.^{26,28,52}

Validity of Empirically Defined ASD

Total sample FM classifications closely mirrored clinical ASD diagnosis (90% of youth classified identically; $\kappa = 0.78$). Empirical classifications from the non-ASD subsample included a higher proportion of females identified as belonging to the ASD class (50.9%) relative to empirical classifications from the total sample (20.8%), reinforcing the notion that female siblings may be underidentified by caregivers as having ASD.^{10,57} Figure 3 presents the ages at which developmental milestones were achieved in caregiver-designated non-ASD siblings, separately for empirically identified ASD and non-ASD classes. Individuals in the ASD class were reported to have later ages of first words (ASD class: mean = 12.6 months, $SD = 9.0$; non-ASD class: mean = 10.9 months, $SD = 4.9$), walking (ASD class: mean = 12.8 months, $SD = 3.2$; non-ASD class: mean = 12.1 months, $SD = 2.5$), meaningful phrase speech (ASD class: mean = 22.6 months, $SD = 11.4$; non-ASD class mean = 18.7 months, $SD = 5.9$), and toilet training (ASD class mean = 35.6 months, $SD = 12.7$; non-ASD class: mean = 31.0 months, $SD = 9.3$) relative to youth in the non-ASD class. These differences were all highly significant (all $p < .001$) and remained highly significant when excluding youth with intellectual disability. Thus, caregiver-designated non-ASD siblings empirically classified into an ASD category showed a reliable pattern of developmental delay across a range of early milestones. However, these delays tended to be subtle and highly variable across cases. As a result the above differences tended to be small (largest $d = 0.33$) and are not large enough to be clinically useful for distinguishing ASD from non-ASD cases.

FIGURE 2 Social Responsiveness Scale (SRS) social and autism mannerism T scores cluster within autism spectrum disorder (ASD) and non-ASD classes in the total available sample (N = 6,949; left panel) and non-ASD sibling subsample (N = 2,701; right panel). Note: SRS social T scores are the average of the four social subscales. Higher scores indicate greater impairment. Each figure is based upon a random subsample (n = 500).



DSM-IV-TR Versus DSM-5 (Aim 2)

Table 2 presents sensitivity and specificity of DSM-IV-TR, proposed DSM-5, and modified DSM-5 criteria to proxy ASD diagnoses. DSM-IV-TR and DSM-5 criteria showed substantial, but far from complete, overlap ($\kappa = 0.74$, agreement = 87.1%). DSM-5 criteria had superior specificity, but lower sensitivity relative to

DSM-IV criteria ($p < .001$). The increase in specificity is remarkable in that it could reduce false positives by more than four times the estimated DSM-IV-TR rate (DSM-5 = 3% versus DSM-IV-TR = 14%). Interestingly, relaxing the DSM-5 algorithm to require one less symptom from either SCI or RRB provided the best balance of sensitivity and specificity. With the relaxed algorithm an additional 11% to 12% of ASD cases would be captured (Table 2, comparing sensitivity values in row 2 versus row 3). The superior sensitivity of a relaxed DSM-5 algorithm was at least partly due to decreased endorsement of two SCI criteria and one RRB criterion in older youth. Specifically, we observed a weak, but highly significant, negative relationship between age of the child and endorsement of the DSM-5 non-verbal communication, relationships, and repetitive behavior criteria (age with SCI criterion A2: nonverbal communication $r = -0.10$; age with SCI criterion A3: relationships, $r = -0.09$; age with RRB criterion B1: repetitive behaviors, $r = -0.09$; all $p < .001$).

Including symptoms tapping subtler behavioral manifestations of high-functioning ASD improved sensitivity (0.81 versus 0.64). Symptoms

FIGURE 3 Age of reported developmental milestones (mean \pm 95% confidence interval) for siblings without autism spectrum disorder (ASD) empirically classified into the ASD and non-ASD classes.

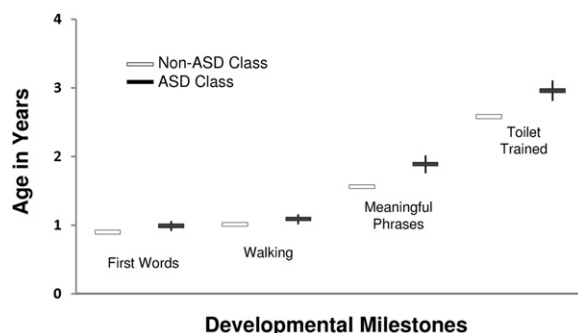


TABLE 2 Sensitivity and Specificity of *DSM-IV-TR*, Proposed *DSM-5*, and Modified Criteria for Clinical Diagnosis and Empirically Derived Classifications

	Clinical ASD Diagnosis		Empirically Derived Classifications	
	Sensitivity	Specificity	Sensitivity	Specificity
<i>DSM-IV-TR</i>	0.95	0.86	0.92	0.83
<i>DSM-5</i> : Field Trial Phase I	0.81	0.97	0.78	0.95
<i>DSM-5</i> : Relaxed Algorithm (one less criterion)	0.93	0.95	0.89	0.92
<i>DSM-5</i> : Without High-Functioning Symptoms	0.64	0.98	0.61	0.97
<i>DSM-5</i> : Without Sensory Sensitivity/Interests	0.78	0.97	0.75	0.96
<i>DSM-5</i> : One symptom per criterion	0.96	0.90	0.92	0.87

Note: Comparison of *DSM-IV-TR* and proposed *DSM-5* criteria was done using a subsample with complete data on the Social Communication Questionnaire (SCQ) and Social Responsiveness Scale (SRS) ($N = 6,426$). This sample had highly similar demographic and clinical characteristics to the total available SRS sample. All 95% confidence intervals were ± 0.01 . Symptoms tapping high-functioning manifestations of autism spectrum disorder (ASD) were determined from field trials criteria and included the following: literal or pedantic use of language (SRS 10), unusual prosody (SRS 53), body orientation or social distance (SRS 55), impaired social understanding or awareness (SCQ 4), and difficulties in adjusting behavior to various contexts (SRS 52). *DSM-5*: Field Trials Phase I refers to Phase I Field Trials criteria that required two symptoms per criterion. *DSM-5*: Relaxed Algorithm (one less criterion) refers to relaxed Phase I Field Trials criteria with the modification that one less criterion from either social communication/interaction or restricted, repetitive behavior is required for diagnosis. *DSM-5*: One Symptom Per Criterion refers to the currently proposed *DSM-5* criteria that required only one symptom per criterion.

evaluating social difficulties in high-functioning ASD included: impaired social understanding or awareness (SRS 4), literal or pedantic use of language (SRS 10), difficulties in adjusting behavior to various contexts (SRS 52), unusual prosody (SRS 53), and problems with body orientation or social distance (SRS 55). The sizable improvement in sensitivity was attributable to a higher proportion of individuals with Asperger's disorder being correctly identified when these symptoms were included (78.3% versus 58.7%). Detection of Asperger's disorder was further enhanced using the relaxed algorithm described above (90.9%). Adding sensory sensitivities and unusual interests as a RRB criterion also improved sensitivity (0.81 versus 0.78) without substantially altering specificity. Requiring two specific symptom endorsements per *DSM-5* criterion enhanced specificity (0.97 versus 0.90), although it is important to note that sensitivity is highest (0.96) when only one symptom is required per criterion. This version of criteria may be preferred in high base rate settings where specificity is less crucial. The same pattern of findings was observed across both ASD diagnostic proxies. Results were also consistent across clinically relevant groups (age <7 years, age ≥ 7 years, females, verbal youth, and multiple-incidence families), with the notable exception that sensitivity and specificity tended to be slightly weaker in

youth age less than 7 years of age and in multiple-incidence families (see Table S3, available online).

DISCUSSION

In this investigation, a hybrid model, closely matching proposed *DSM-5* criteria for ASD, yielded the most parsimonious representation of autism symptoms. The model included both a categorical distinction between youth with and without ASD and dimensional representations of social communication and interaction difficulties and restricted, repetitive behavior. Sensitivity analyses supported this hybrid conceptualization of ASD with generalization across measurement scales and demographic subsamples. Furthermore, results were consistent with taxometric analyses of autism symptom data from this sample.²⁰ If the present results are replicated, large-sample neurobiological studies will be useful for determining whether biological markers of the categorical ASD distinction can be identified⁵⁸ or whether the *DSM-5* ASD diagnosis can be empirically parsed into biologically meaningful subphenotypes.

The hybrid model merges two previously competing views of ASD symptoms—dimensional and categorical perspectives.^{9,20} Figure 2 provides a

visualization of how an ASD diagnostic category can be superimposed on quantitative symptom distributions, clarifying the underlying nature of autism symptoms. Merging categorical and dimensional models presents the opportunity to enhance clinical assessment through a two-pronged approach. In this approach, assessment tools will provide both quantitative evaluation of symptom severity and generate post-test probabilities of ASD diagnosis following evidence-based medicine procedures.²⁹ A key next step will be to identify or adapt gold-standard measures that map *DSM-5* criteria and can be used in an evidence-based medicine fashion to generate post-test probabilities of ASD.⁵⁹ Dissemination of these measures to clinicians is essential to meet the growing demand for early diagnosis of ASD.

The presence of an ASD versus non-ASD distinction coheres with data identifying a divergent trajectory of brain development in ASD.^{32,33} The hybrid model of ASD also meshes with an emerging understanding of the complexity of autism genetics. Clearly, the present findings do not specify a genomic mechanism. However, they do cohere with recent findings from genomic studies identifying roles for highly penetrant mono- or oligogenic effects and modifying effects from common variants and epigenetic processes.^{41,60-62} It is important to note that the presence of a categorical distinction (ASD versus no ASD) does not rule out a role for quantitative biological variation. These constitute different levels of analysis: For example, combinations of quantitative trait loci can produce threshold effects that might manifest as categorical distributions of symptoms. With this important caveat in mind, future genomic and neurobiological studies will benefit from designs and analytic methods that incorporate both the categorical and dimensional nature of ASD symptoms. Doing so will facilitate capturing the full nature and intricacy of ASD.

Findings from exploratory analyses supported the diagnostic efficiency of the proposed *DSM-5* Phase I Field Trials algorithm in identifying ASD. The only exception was possible relaxing of the algorithm by requiring one less SCI or RRB criterion, which may increase capture of ASD-affected individuals by as much as 11% to 12%. Relaxing the number of criteria required may be particularly important for improving sensitivity

to individuals previously classified as having Asperger's disorder or other underidentified subgroups of ASD-affected youth^{57,63} and for enhancing detection in settings in which reports of early-life symptoms are not readily obtained (e.g., adopted children or youth in foster or juvenile justice settings). Accounting for these individuals would be consistent with the *DSM* task force efforts to improve the applicability of the diagnostic system across subgroups. Additional studies of the influence of age on *DSM-5* ASD criteria will also be vital, particularly given the superior validity of retrospective reports of earlier symptoms and the primary role of caregiver report in evaluations of older individuals.⁴⁶ The present results suggest that caregiver report may be less helpful in distinguishing younger children (age <7 years) with ASD and affected and unaffected youth from multiple-incidence families. Careful clinical observation may be crucial for accurate diagnosis in these groups.

The superior specificity of *DSM-5*—especially criteria used in the Phase I Field Trials—relative to *DSM-IV-TR* criteria was striking. High specificity is a desirable trait for the new *DSM-5* ASD diagnosis, reducing the likelihood that individuals without ASD are inappropriately given the diagnosis. However, the improved specificity of *DSM-5* criteria to ASD raises the possibility that, if adopted, some individuals meeting criteria for *DSM-IV-TR* may not meet *DSM-5* criteria. Studies examining youth with autism symptoms who do not meet full *DSM-5* Phase I Field Trials and currently proposed *DSM-5* criteria are needed to determine whether the newly planned Social Communication Disorder captures all of these youth or whether some youth with high-functioning ASD might be misclassified. Revision and phase II testing of *DSM-5* criteria should continue to include sensory sensitivities/unusual sensory interests and subtler SCI symptoms to maximize sensitivity to high-functioning ASD. The possibility of requiring two or more specific symptoms per criterion (as specified in Phase I Field Trials criteria) should also be re-examined, as this appeared to improve specificity substantially over the currently proposed criteria.

The primary limitations of this study are as follows: use of a self-selected Internet registry, comparison with caregiver-designated non-ASD siblings, mapping of caregiver-reports of autism symptoms to diagnostic criteria, and

reliance on clinical ASD diagnoses. The IAN Internet registry is not representative of all families affected by ASD. However, the national scope and large size of this registry make it preferable to smaller community samples that limit generalizability and preclude examining hybrid FM models. Replication in population samples will be important for estimating the base rate and demographic characteristics of ASD. Very large population samples ($N > 10,000$) will be needed to ensure a sufficient number of ASD cases are captured and to limit the influence of ascertainment biases on the detection of low-base-rate latent classes.

Non-ASD siblings are not equivalent to a clinical comparison sample. For this reason, the specificity values obtained may overestimate values seen in typical clinical assessment situations in which many cases mimicking ASD may present. However, the sibling comparison group may provide a reasonable approximation to the heterogeneous presentations often seen in clinical practice. Specifically, this group included a substantial proportion of individuals with other developmental concerns, including language delay, characteristics of the broad autism phenotype,⁴³ ADHD, anxiety, and intellectual disability. Furthermore, analyses of this subsample identified a small proportion of caregiver-designated non-ASD siblings as falling in the ASD category. Thus, it is possible that diagnostic efficiency may actually be underestimated. Regardless, the aim of the present study was to compare the relative efficiency of different iterations of diagnostic criteria rather than absolute values that would mirror those obtained in clinical settings. Phase II Field Trials and additional studies using clinical comparison samples will be needed to provide clinically realistic absolute values.

Clinical ASD diagnoses are not as reliable and valid as diagnoses based on gold-standard semi-structured research interviews or observational instruments. However, these diagnoses were often based on gold-standard measures, and available data suggest that clinical diagnoses in IAN are quite accurate.⁴² To circumvent the limitations of this diagnostic proxy, we also included FM-based empirical classifications. The consistency of findings across these surrogates is reassuring and implies that many of the present findings may also apply to gold-standard ASD diagnoses.

Mapping caregiver reports of symptoms on the SRS and SCQ to diagnostic criteria is also a limitation. The SCQ and SRS items may not have been mapped optimally, and/or the items themselves may not adequately capture the targeted diagnostic criteria. It is possible that inclusion of clinical observation and judgment or other clinical information such as adaptive impairment would have improved sensitivity to ASD without the need to alter the proposed *DSM-5* algorithm. However, caregiver report is often a crucial aspect of a thorough diagnostic evaluation and it is sometimes difficult to elicit accurate historical information for early-life ASD symptoms.⁵⁴ Thus, assessment approaches that lean heavily on caregiver-report in determining ASD may benefit from relaxing of *DSM-5* Phase I Field Trials criteria. Relaxed criteria should also be considered in conjunction with adaptive impairment to determine the simplest, most efficient criteria. Finally, it should also be noted that the present findings do not *prima facie* imply that *DSM-IV-TR* diagnoses should be lumped. It is possible that retaining a separate diagnosis or a specific diagnostic modifier for particular cases (such as individuals previously diagnosed with Asperger's disorder) might be clinically useful if these cases were found to have a different course, prognosis, or treatment response.

Despite these limitations, the present findings provided a comprehensive validation of *DSM-5* criteria and generated several promising leads for future examinations of the *DSM-5* algorithm. Adopting relaxed *DSM-5* Phase I Field Trials criteria may substantially decrease false positive diagnoses and simultaneously maintain or improve identification of ASD cases. The importance of enhanced detection is magnified by increasing recognition of the prevalence of ASD. Many ASD cases will have their first, and possibly only, point of contact in non-ASD specialty settings.⁶⁴ In these settings, clinicians without expertise in ASD are and will be called upon on an increasing basis to make an appraisal of the diagnosis, to initiate additional referral for specialty evaluation, and to provide preliminary treatment recommendations while they await diagnostic confirmation. Thus, improvements in ASD detection will be meaningful for all consumers and providers, facilitating earlier and more appropriate intervention, diminishing inappropriate resource allocation, and reducing family and societal costs. *©*

Accepted September 29, 2011.

Drs. Frazier and Speer are with the Center for Pediatric Behavioral Health and Center for Autism, Cleveland Clinic. Dr. Youngstrom is with the University of North Carolina at Chapel Hill. Ms. Embacher is with the Center for Autism, Cleveland Clinic. Dr. Law is with the Kennedy Krieger Institute. Dr. Constantino is with Washington University in St. Louis. Dr. Findling is with Case Western Reserve University. Dr. Hardan is with Stanford University, Stanford, CA. Dr. Eng is with Genomic Medicine Institute, Taussig Cancer Institute, and Stanley Shalom Zielony Institute of Nursing Excellence, Cleveland Clinic.

This publication was made possible by the Case Western Reserve University/Cleveland Clinic Clinical and Translational Grant Number UL1 RR024989 from the National Center for Research Resources. Autism Speaks provided support for the Interactive Autism Network project and Dr. Law. Funding from the National Institute of Child Health and Human Development Grant Number HD42541 supported Dr. Constantino's involvement. Dr. Eng is a Doris Duke Distinguished Clinical Scientist and holds the Sondra J. and Stephen R. Hardis Chair of Cancer Genomic Medicine at the Cleveland Clinic.

Dr. Law designed and maintains the Interactive Autism Network registry. Drs. Frazier, Youngstrom, and Eng designed the present study. Dr. Frazier obtained funding to support analyses. Drs. Frazier, Youngstrom, Speer, Constantino, Findling, and Hardan, and Ms. Embacher supervised data interpretation of the study. Drs. Frazier and Ms. Embacher conducted data management and data analyses. All authors contributed to writing and revision.

The authors gratefully acknowledge the efforts of families contributing to the Interactive Autism Network. Drs. Frazier and Youngstrom served as the statistical experts.

Disclosure: Dr. Frazier has received federal funding or research support from, acted as a consultant to, or received travel support from Shire Development, Inc., Bristol-Myers Squibb, Integragen, the National Institutes of Health (NIH), and the Brain and Behavior Research Foundation. Dr. Youngstrom has received travel support from Bristol-Myers Squibb. Dr. Findling receives or has received research support, acted as a consultant to and/or served on a speakers' bureau for Abbott, Addrenex, AstraZeneca, Biovail, Bristol-Myers Squibb, Forest, GlaxoSmithKline, Johnson and Johnson, KemPharm, Eli Lilly and Co., Lundbeck, Merck, Neuropharm, Novartis, Noven, Organon, Otsuka, Pfizer, Rhodes Pharmaceuticals, Sanofi-Aventis, Schering-Plough, Seaside Therapeutics, Sepracore, Shire, Solvay, Sunovion, Supernus Pharmaceuticals, Transcept, Validus, and Wyeth. Dr. Constantino receives grant or research support from NIH, National Institute of Child Health and Human Development, the National Institute of Mental Health, the Department of Health and Human Services, Autism Speaks, and the Centers for Disease Control and Prevention. He receives royalties from Western Psychological Services for the commercial distribution of the Social Responsiveness Scale, one of the metrics used in this study; no royalties were generated by any of the assessments performed in the present research. Drs. Law, Speer, Eng, and Hardan, and Ms. Embacher report no biomedical financial interests or potential conflicts of interest.

Correspondence to Thomas W. Frazier, Ph.D., Center for Autism, Cleveland Clinic, 2801 Martin Luther King Jr. Drive CRS10, Cleveland, OH 44104; e-mail: frazier2@ccf.org

0890-8567/\$36.00/©2012 American Academy of Child and Adolescent Psychiatry

DOI: 10.1016/j.jaac.2011.09.021

REFERENCES

1. Geschwind DH. Advances in autism. *Annu Rev Med*. 2009;60:367-380.
2. Fombonne E. Epidemiology of pervasive developmental disorders. *Pediatr Res*. 2009;65:591-598.
3. Ganz ML. The lifetime distribution of incremental societal costs of autism. *Arch Pediatr Adolesc Med*. 2007;161:343-349.
4. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Fourth Edition, Text Revision. Washington, DC: American Psychiatric Association; 2000.
5. DSM-5 Development. 299.00 Autistic Disorder. 2011. Available at: http://www.dsm5.org/ProposedRevisions/Pages/proposed_revision.aspx?rid=94. Accessed February 1, 2011.
6. Tanner L. Proposed autism diagnosis changes anger "Aspies". ABC News / Health. 2010. Accessed February 11, 2010.
7. Miller JN, Ozonoff S. The external validity of Asperger disorder: lack of evidence from the domain of neuropsychology. *J Abnorm Psychol*. 2000;109:227-238.
8. Constantino JN. How continua converge in nature: cognition, social competence, and autistic syndromes. *J Am Acad Child Adolesc Psychiatry*. 2009;48:97-98.
9. Constantino JN, Todd RD. Autistic traits in the general population: a twin study. *Arch Gen Psychiatry*. 2003;60:524-530.
10. Constantino JN, Lajonchere C, Lutz M, et al. Autistic social impairment in the siblings of children with pervasive developmental disorders. *Am J Psychiatry*. 2006;163:294-296.
11. Piven J, Palmer P, Jacubbi D, Childress D, Arndt S. Broad autism phenotype: evidence from a family history study of multiple-incidence autism families. *Am J Psychiatry*. 1997;154:185-190.
12. Constantino JN, Todd RD. Intergenerational transmission of subthreshold autistic traits in the general population. *Biol Psychiatry*. 2005;57:655-660.
13. Ronald A, Happe F, Bolton P, et al. Genetic heterogeneity between the three components of the autism spectrum: a twin study. *J Am Acad Child Adolesc Psychiatry*. 2006;45:691-699.
14. Ronald A, Simonoff E, Kuntsi J, Asherton P, Plomin R. Evidence for overlapping genetic influences on autistic and ADHD behaviours in a community twin sample. *J Child Psychol Psychiatry*. 2008;49:535-542.
15. Rosenberg RE, Law JK, Yenokyan G, McGready J, Kaufmann WE, Law PA. Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Arch Pediatr Adolesc Med*. 2009;163:907-914.
16. Ho A, Todd RD, Constantino JN. Brief report: autistic traits in twins vs. non-twins—a preliminary study. *J Autism Dev Disord*. 2005;35:129-133.
17. Kanner L, Rodriguez A, Ashenden B. How far can autistic children go in matters of social adaptation? *J Autism Child Schizophr*. 1972;2:9-33.
18. Rutter M. The autistic child. *R Inst Public Health Hyg J*. 1967;30:130-132.
19. Wing JK, O'Connor N, Lotter V. Autistic conditions in early childhood: a survey in middlesex. *Br Med J*. 1967;3:389-392.
20. Frazier TW, Youngstrom EA, Sinclair L, et al. Autism spectrum disorders as a qualitatively distinct category from typical behavior in a large, clinically ascertained sample. *Assessment*. 2010;17:308-320.
21. Ingram DG, Takahashi TN, Miles JH. Defining autism subgroups: a taxometric solution. *J Autism Dev Disord*. 2008;38:950-960.
22. Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Pickles A. Autism from 2 to 9 years of age. *Arch Gen Psychiatry*. 2008;63:694-701.
23. Van Daalen E, Kemner C, Dietz C, Swinkels SHN, Buitelaar J, Van Engeland H. Inter-rater reliability and stability of diagnoses of autism spectrum disorder in children identified through screening at a very young age. *Eur Child Adolesc Psychiatry*. 2009;18:663-674.
24. Mayes SD, Calhoun SL. Influence of IQ and age in childhood autism: lack of support for DSM-IV Asperger's disorder. *J Dev Phys Disabilities*. 2004;16:257-272.
25. Mahoney W, Szatmari P, Maclean J, et al. Reliability and accuracy of differentiating pervasive developmental disorder subtypes. *J Am Acad Child Adolesc Psychiatry*. 1998;37:278-285.
26. Lubke GH, Muthen B. Investigating population heterogeneity with factor mixture models. *Psychol Methods*. 2005;10:21-39.
27. Muthen B, Asparouhov T, Rebollo I. Advances in behavioral genetics modeling using Mplus: applications of factor mixture modeling to twin data. *Twin Res Hum Genet*. 2006;9:313-324.

28. Lubke GH, Muthen B, Moilanen IK, *et al.* Subtypes versus severity differences in attention-deficit/hyperactivity disorder in the Northern Finnish Birth Cohort. *J Am Acad Child Adolesc Psychiatry.* 2007;46:1584-1593.
29. Kraemer HC. *Evaluating Medical Tests: Objective and Quantitative Guidelines.* Newbury Park, CA: Sage Publications; 1992.
30. Ruscio J, Ruscio AM. A structure-based approach to psychological assessment: matching measurement models to latent structure. *Assessment.* 2002;9:4-16.
31. Ruscio J, Ruscio AM. Clarifying boundary issues in psychopathology: the role of taxometrics in a comprehensive program of structural research. *J Abnorm Psychol.* 2004;113:24-38.
32. Amaral DG, Schumann CM, Nordahl CW. Neuroanatomy of autism. *Trends Neurosci.* 2008;31:137-145.
33. Pardo CA, Eberhart CG. The neurobiology of autism. *Brain Pathol.* 2007;17:434-447.
34. Lykken DT, McGue M, Tellegen A, Bouchard TJ. Emergence: genetic traits that may not run in families. *Am Psychol.* 1992;47:1565-1577.
35. Haslam N. Evidence that male sexual orientation is a matter of degree. *J Pers Soc Psychol.* 1997;73:862-870.
36. Lenzenweger MF, McLachlan G, Rubin DB. Resolving the latent structure of schizophrenia endophenotypes using expectation-maximization-based finite mixture modeling. *J Abnorm Psychol.* 2007;116:16-29.
37. Haslam N. Categorical versus dimensional models of mental disorder: the taxometric evidence. *Aust N Z J Psychiatry.* 2003;37:696-704.
38. Cook EH, Jr., Scherer SW. Copy-number variations associated with neuropsychiatric conditions. *Nature.* 2008;455:919-923.
39. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry.* 2003;160:636-645.
40. Cannon TD, Keller MC. Endophenotypes in the genetic analyses of mental disorders. *Annu Rev Clin Psychol.* 2006;2:267-290.
41. Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet.* 2008;9:341-355.
42. Lee H, Marvin AR, Watson T, *et al.* Accuracy of phenotyping of autistic children based on Internet implemented parent report. *Am J Med Genet B Neuropsychiatr Genet.* 2010;153B:1119-1126.
43. Constantino JN, Zhang Y, Frazier TW, Abbacchi AM, Law P. Sibling recurrence and the genetic epidemiology of autism. *Am J Psychiatry.* 2010;167:1349-1356.
44. Constantino JN, Gruber CP. *Social Responsiveness Scale: Manual.* Los Angeles, CA: Western Psychological Services; 2005.
45. Rutter M, Bailey A, Lord C. *The Social Communication Questionnaire Manual.* Los Angeles, CA: Western Psychological Services; 2003.
46. Lord C, Pickles A, McLennan J, *et al.* Diagnosing autism: analyses of data from the Autism Diagnostic Interview. *J Autism Dev Disord.* 1997;27:501-517.
47. Chandler S, Charman T, Baird G, *et al.* Validation of the social communication questionnaire in a population cohort of children with autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry.* 2007;46:1324-1332.
48. Virkud YV, Todd RD, Abbacchi AM, Zhang Y, Constantino JN. Familial aggregation of quantitative autistic traits in multiplex versus simplex autism. *Am J Med Genet B Neuropsychiatr Genet.* 2009;150B:328-334.
49. Corsello C, Hus V, Pickles A, *et al.* Between a ROC and a hard place: decision making and making decisions about using the SCQ. *J Child Psychol Psychiatry.* 2007;48:932-940.
50. Nasser F, Wisenbaker J. A Monte Carlo study investigating the impact of item parceling on measures of fit in confirmatory factor analysis. *Educ Psychol Measure.* 2003;63:729-757.
51. Lubke G, Neale M. Distinguishing between latent classes and continuous factors with categorical outcomes: class invariance of parameters of factor mixture models. *Multivar Behav Res.* 2006;41:499-532.
52. Lubke G, Tueller S. Latent class detection and class assignment: a comparison of the MAXEIG taxometric procedure and factor mixture modeling approaches. *Struct Equat Model.* 2010;17:605-628.
53. Faraone SV, Tsuang MT. Measuring diagnostic accuracy in the absence of a "gold standard". *Am J Psychiatry.* 1994;151:650-657.
54. Mildenerberger K, Sitter S, Noterdaeme M, Amorosa H. The use of the ADI-R as a diagnostic tool in the differential diagnosis of children with infantile autism and children with a receptive language disorder. *Eur Child Adolesc Psychiatry.* 2001;10:248-255.
55. Cohen J. *Statistical power analysis for the behavioral sciences.* 2nd edition ed. Hillsdale, NJ: Erlbaum; 1987.
56. Kalb LG, Law JK, Landa R, Law PA. Onset patterns prior to 36 months in autism spectrum disorders. *J Autism Dev Disord.* 2010;40:1389-1402.
57. Wing L, Gould J, Gillberg C. Autism spectrum disorders in the DSM-V: better or worse than the DSM-IV? *Res Dev Disabil.* 2011;32:768-773.
58. Dinstein I, Pierce K, Eyler L, *et al.* Disrupted neural synchronization in toddlers with autism. *Neuron.* 2011;70:1218-1225.
59. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine: How to Practice and Teach EBM.* 2nd ed. Edinburgh: Churchill Livingstone; 2000.
60. Sanders SJ, Ercan-Sencicek AG, Hus V, *et al.* Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron.* 2011;70:863-885.
61. Wang K, Zhang H, Ma D, *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature.* 2009;459:528-533.
62. Gregory SG, Connelly JJ, Towers AJ, *et al.* Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. *BMC Med.* 2009;7:62.
63. Moss J, Magiati I, Charman T, Howlin P. Stability of the Autism Diagnostic Interview—Revised from pre-school to elementary school age in children with autism spectrum disorders. *J Autism Dev Disord.* 2008;38:1081-1091.
64. Pinto-Martin JA, Young LM, Mandell DS, Poghosyan L, Giarelli E, Levy SE. Screening strategies for autism spectrum disorders in pediatric primary care. *J Dev Behav Pediatr.* 2008;29:345-350.

TABLE S1 Sample and Items/Scale Set Included, and Purpose of Primary and Secondary Latent Variable Model Analyses

Sample	Item/Scales	Purpose
Primary Analyses		
Total available (N = 6,949)	SRS – Subscales with autistic mannerisms divided into four items sets	To evaluate categorical, dimensional, and hybrid models of ASD symptoms
Non-ASD siblings (N = 2,701)	SRS subscales (same as above)	To determine whether a small ASD category is identified; rule out rater contrast effects as an explanation
Sensitivity Analyses		
Total available (N = 14,200)	SCQ items	Internal replication/extension of total sample SRS above
Non-ASD siblings (N = 5,594)	SCQ items	Internal replication/extension of sibling sample SRS above
Verbal youth (N = 12,014)	SCQ items	Determine whether latent structure was influenced by the presence of non-verbal youth in the total sample
Females only (N = 4,013)	SCQ items	Determine whether latent structure was consistent in females
Age <7 years (N = 6,821)	SCQ items	Determine whether latent structure was consistent across age
Age ≥7 years (N = 7,379)	SCQ items	Determine whether latent structure was consistent across age
Multiple incidence (N = 1,022)	SCQ items	Determine whether multiple incidence families have a different latent structure
Note: ASD = autism spectrum disorder; SCQ = Social Communication Questionnaire, SRS = Social Responsiveness Scale.		

TABLE S2 Social Communication Questionnaire (SCQ) and Social Responsiveness Scale (SRS) Item Mapping to DSM-IV-TR and Proposed DSM-5 Criteria

DSM-IV-TR	SCQ Item No.	SRS Item No.
Social Interaction		
a	26, 27, 33	16
b	19, 36, 37, 39, 40	
c	28, 29, 30	
d	9, 10, 31, 32	
Communication		
a	22, 23, 24, 25	13
b	2, 20	
c	3, 4, 5, 6	
d	21, 34, 35	
Restricted, Repetitive, Stereotyped Behavior		
a	11, 13	39
b	7, 8	24
c	15, 16	
d	12, 14	
Proposed DSM-5	SCQ Item No.	SRS Item no.
Social Communication and Interaction		
1	2, 9, 20, 28, 30, 31, 34, 37, 38, 40	33, 10, 53
2	10, 22, 23, 24, 25, 26, 27, 33	55
3	4, 19, 36, 39	52
Restricted, Repetitive Behavior		
1	3, 5, 6, 12, 15, 16	24
2	7, 8	
3	11, 13, 18	
4	14	

Note: This table depicts scale items that were used to cover the respective DSM-IV and proposed DSM-5 criterion sets, preferentially drawing from the SCQ, and incorporating SRS items as necessary to ensure complete coverage of each criterion domain. This was done to bias the item algorithms toward DSM-IV-TR and ensure that improvements in sensitivity and specificity using DSM-5 criteria were not due to item selection. Many additional SRS items relate to the constructs encompassed by DSM-IV and DSM-5. The specific symptom criteria for DSM-5 criteria are not listed but can be viewed at <http://www.dsm5.org>.

TABLE S3 Sensitivity and Specificity of DSM-IV-TR and Iterations of DSM-5 Criteria to Clinical Autism Spectrum Disorder (ASD) Diagnosis Across Relevant Demographic and Clinical Subsamples

	Clinical ASD Diagnosis	
	Sensitivity	Specificity
Youth aged <7 years		
DSM-IV-TR	0.95	0.84
DSM-5: Field Trial Phase I	0.78	0.98
DSM-5: Relaxed Algorithm (1 less criterion)	0.92	0.95
DSM-5: One Symptom Per Criterion	0.95	0.89
Youth aged ≥7 years		
DSM-IV-TR	0.96	0.87
DSM-5: Field Trial Phase I	0.85	0.97
DSM-5: Relaxed Algorithm (1 less criterion)	0.94	0.95
DSM-5: One Symptom Per Criterion	0.97	0.92
Females		
DSM-IV-TR	0.96	0.87
DSM-5: Field Trial Phase I	0.78	0.98
DSM-5: Relaxed Algorithm (1 less criterion)	0.91	0.96
DSM-5: One Symptom Per Criterion	0.95	0.92
Verbal Youth		
DSM-IV-TR	0.95	0.87
DSM-5: Field Trial Phase I	0.81	0.97
DSM-5: Relaxed Algorithm (1 less criterion)	0.93	0.95
DSM-5: One Symptom Per Criterion	0.96	0.91
Multiple-Incidence Families		
DSM-IV-TR	0.91	0.81
DSM-5: Field Trial Phase I	0.73	0.95
DSM-5: Relaxed Algorithm (1 less criterion)	0.87	0.92
DSM-5: One Symptom Per Criterion	0.92	0.83
<p>Note: DSM-5: Field Trials refers to Phase I Field Trials criteria that require two symptoms per criterion. DSM-5: Relaxed Algorithm (one less criterion) refers to relaxed Phase I Field Trials criteria with the modification that one less criterion from either social communication/interaction or restricted, repetitive behavior is required for diagnosis. DSM-5: One Symptom Per Criterion refers to the currently proposed DSM-5 criteria that require only one symptom per criterion.</p>		