

Published in final edited form as:

J Am Acad Child Adolesc Psychiatry. 2012 January ; 51(1): 28–40.e3. doi:10.1016/j.jaac.2011.09.021.

Validation of Proposed DSM-5 Criteria for Autism Spectrum Disorder

Thomas W. Frazier, PhD, Eric A. Youngstrom, PhD, Leslie Speer, PhD, Rebecca Embacher, BS, Paul Law, MD, MPH, John Constantino, MD, Robert L. Findling, MD, Antonio Y. Hardan, MD, and Charis Eng, MD, PhD

Drs. Frazier, Speer is 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

Abstract

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; 5,863 non-ASD) included in a national registry, the Interactive Autism Network. Youth aged 2–18 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 sub-dimensions. Diagnostic efficiency statistics evaluated the proposed DSM-5 algorithm in identifying ASD.

Results—A hybrid model that included both a category (ASD vs. 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 sub-samples. 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 (.97 vs. .86), however sensitivity was lower (.81 vs. .95). Relaxing DSM-5 criteria by requiring one less symptom criterion increased sensitivity (.93 vs. .81), with minimal reduction in specificity (.95 vs. .97).

© 2011 American Academy of Child & Adolescent Psychiatry. Published by Elsevier Inc. All rights reserved.

Correspondence to Thomas W. Frazier, Ph.D., Center for Autism, Cleveland Clinic, 2801 Martin Luther King Jr. Drive CRS10, Cleveland, OH 44104, USA, Fax: 216-448-6445. fraziet2@ccf.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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

Disclosure: Drs. Law, Speer, Eng, and Hardan, and Mrs. Embacher report no biomedical financial interests or potential conflicts of interest.

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.

Keywords

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 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 under-identify high functioning ASD.⁷ The present study evaluates an important aspect of this controversy - 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 (ie. 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 sub-threshold 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) and later (ages 3-9) diagnoses also support the notion of a single, distinct ASD category.^{22, 23} These studies found strong stability of the broad ASD distinction, while substantial shifting occurred 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 vs. non-ASD) and two dimensions (SCI and RRB).

The distinction between dimensional, categorical, and hybrid models has relevance beyond diagnostic conceptualization - impacting 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 Under 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, prior to setting a diagnostic threshold. If hybrid models are supported, integration of the above approaches will be needed to most accurately characterize the presence vs. 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 would 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: 1) collapsing three domains into two - Social Communication and Interaction (SCI) and Restricted, Repetitive Behavior (RRB), 2) requiring all SCI criteria and at least 2 of 4 RRB criteria to be met, 3) necessitating at least 2 specific symptoms present per criterion, 4) providing a separate RRB criterion evaluating sensory sensitivities and unusual sensory interests, and 5) 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 US, internet-based registry for families with one or more ASD-affected children (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 US, 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, 98.7% scored in the ASD-affected range for one or both instruments. A recent study of youth from the IAN registry provided additional support for the validity of clinical ASD diagnoses.⁴² In this study, randomly selected verbal youth with a score >12 on the SCQ were evaluated using the Autism Diagnostic Interview-Revised and by an expert clinician's observation. All but a handful 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 sub-sample 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 non-trivial 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/guardians for all participants prior to 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, ordinal-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 other psychiatric conditions.^{44, 47–49} SRS sub-scales provide a quantitative assessment of ASD symptoms and decrease the likelihood of falsely identifying a latent category due to 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 to 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 vs. non-ASD) and two sub-dimensions (SCI and RRB) was evaluated by estimating a series of latent class (LCA; 2–6 classes) and exploratory factor analyses (EFA; 1–3 factors) (Aim 1). Latent class analysis assumes the existence of discrete categories, while 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 *and* 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 (ex. ASD vs. non-ASD *OR* Autism vs. Asperger's vs. 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 since it is unlikely three-factor FM models would improve fit when both factors and classes are included.

Models were fit in the total available samples (including ASD and non-ASD siblings) and the non-ASD sibling sub-samples. Sibling comparators were examined separately because the full sample may be biased toward identification of a pseudo-category due to the polarizing effects of social comparison processes or rater contrast effects (ie. rating presumed non-ASD sibs as very healthy and ASD sibs 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 sub-sample. This expectation was based on previous analyses,²⁰ the observation of high SCQ and SRS scores in a non-trivial 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. Best fitting models were determined by comparing the Bayesian Information Criteria (BIC),^{26, 28} where 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 over-fit the observed score distribution. Empirically-derived class assignments (ASD vs. 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 sub-sample. Independent samples t-tests and chi-square 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 (8 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 6 SRS items were included to ensure 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 provides 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 1 specific symptom present per criterion.⁵ SCQ and SRS item mapping was done in a sub-sample with complete data on both measures. Because most individuals enrolling in IAN receive the SCQ, this sub-sample 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 v18. Effect sizes are presented as Cohen’s *d* with conventions of very small ($d < .10$), small ($d = .20$), medium ($d = .50$), and large ($d = .80$).⁵⁵

Results

Sample Characteristics

The total IAN sample included symptom data from 14,774 youth aged 2–18 (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 (53.9%) sibling comparison group, consistent with the sex ratio of autism. SRS data were available from 6,949 youth (4,248 ASD, 2,701 non-ASD siblings) and SCQ data were available from 14,200 youth (8,606 ASD, 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 $M = 8.40$, $SD = 3.99$, SCQ $M = 7.53$, $SD = 4.37$; $t(14772) = 12.69$, $p < .001$, $d = .21$) due to the fact that the instrument

is completed only for individuals ages 4 and older, while the SCQ is completed for ages 2 and older. Additionally, the total SRS sample included fewer non-verbal individuals even when only youth ages 4 and older are considered (SRS 8.4%, SCQ 12.8%; $X^2(1)=61.10$, $p<.001$; $d=.14$) and a slightly higher proportion of white/non-Hispanic individuals (SRS 86.3%, SCQ 85.0%; $X^2(1)=5.42$, $p=.020$; $d=.04$). The above differences, while 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 $X^2(1)=3.65$, $p>.05$). Additional characteristics of the IAN sample have been described elsewhere.^{20, 43}

Model Fit (Aim 1)

Table 1 displays comparisons for models using SRS sub-scales. LCA models showed steady improvements in fit from 2- to 6-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 mannerisms 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. In spite of 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 ($X^2(1)=2.45$, $p=.118$), indicating 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 sub-sample. 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 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 sub-sample. 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 sub-sample, 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 over-estimated.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 sub-samples. 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 sub-sample (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 >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 sub-sample (right panel). The only exception was that the ASD class derived from the non-ASD sibling sub-sample is considerably smaller and has few very high scores relative to the ASD class from the total sample. Importantly, and in spite of 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 very useful for decreasing the probability of ASD (Likelihood ratio = .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 vs. SCQ) and sampling characteristics, the same set of models were estimated using SCQ items across the total available sample and several demographic sub-samples (non-ASD siblings, verbal youth, multiple incidence families, females, ages <7, and ages ≥ 7). 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 sub-samples. 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; $K = .78$). Empirical classifications from the non-ASD sub-sample 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 under-identified 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 $M = 12.6$ months, $SD = 9.0$; non-ASD class $M = 10.9$ months, $SD = 4.9$), walking (ASD class $M = 12.8$ months, $SD = 3.2$; non-ASD class $M = 12.1$ months, $SD = 2.5$), meaningful phrase speech (ASD class $M = 22.6$ months, $SD = 11.4$; non-ASD class $M = 18.7$ months, $SD = 5.9$), and toilet training (ASD class $M = 35.6$ months, $SD = 12.7$; non-ASD class $M = 31.0$ 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 $d = .33$) and are not large enough to be clinically useful for distinguishing ASD from non-ASD cases.

DSM-IV-TR vs. 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 ($\text{Kappa}=.74$, $\text{agreement}=87.1\%$). 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% vs. 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 11–12% of ASD cases would be captured (Table 2 – comparing sensitivity values in row 2 vs. row 3). The superior sensitivity of a relaxed DSM-5 algorithm was at least partly due to decreased endorsement of 2 SCI criteria and 1 RRB criterion in older youth. Specifically, we observed a weak, but highly significant, negative relationship between age of the child and endorsement of the DSM5 non-verbal communication, relationships, and repetitive behavior criteria (age with SCI criterion A2:non-verbal communication $r=-.10$; age with SCI criterion A3:relationships $r=-.09$; age with RRB criterion B1:repetitive behaviors $r=-.09$; all $p<.001$).

Including symptoms tapping subtler behavioral manifestations of high functioning ASD improved sensitivity (.81 vs. .64). Symptoms 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% vs. 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 (.81 vs. .78) without substantially altering specificity. Requiring two specific symptom endorsements per DSM-5 criterion enhanced specificity (.97 vs. .90), although it is important to note that sensitivity is highest (.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, age≥7, females, verbal youth, and multiple incidence families) with the notable exception that sensitivity and specificity tended to be slightly weaker in youth age<7 and 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 sub-samples. 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 if the DSM-5 ASD diagnosis can be empirically parsed into biologically meaningful sub-phenotypes.

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 identifying or adapting gold-standard measures which 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 vs. 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 vs. 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 11–12%. Relaxing the number of criteria required may be particularly important for improving sensitivity to individuals previously classified as Asperger's disorder or other under-identified sub-groups of ASD-affected youth^{57, 63} and for enhancing detection in settings where 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 sub-groups. 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) 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: 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 over-estimate values seen in typical clinical assessment situations where 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 sub-sample 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 under-estimated. 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 suggests 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 comforting 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.

In spite of 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.

Acknowledgments

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. Charis 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.

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

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, the National Institute 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 speaker's 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 royalties from Western Psychological Services for the commercial distribution of one of the metrics used in this study (Social Responsiveness Scale); no royalties were generated by any of the assessments performed in the present research.

References

1. Geschwind DH. Advances in autism. *Annu. Rev. Med.* 2009; 60:367–380. [PubMed: 19630577]
2. Fombonne E. Epidemiology of pervasive developmental disorders. *Pediatr. Res.* Jun. 2009; 65(6): 591–598.
3. Ganz ML. The lifetime distribution of incremental societal costs of autism. *Arch. Pediatr. Adolesc. Med.* 2007; 161:343–349. [PubMed: 17404130]
4. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Fourth Edition. Washington, DC: American Psychiatric Association; 2000. Text Revision
5. [Accessed February 01, 2011] DSM-5 Development. 299.00 Autistic Disorder. 2011. <http://www.dsm5.org/ProposedRevisions/Pages/proposedrevision.aspx?rid=94>
6. Tanner L. Proposed autism diagnosis changes anger "Aspies". ABC news / health. 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. [PubMed: 10895561]
8. Constantino JN. How continua converge in nature: cognition, social competence, and autistic syndromes. *J. Am. Acad. Child Adolesc. Psychiatry.* 2009 Feb; 48(2):97–98. [PubMed: 20040821]

9. Constantino JN, Todd RD. Autistic traits in the general population: A twin study. *Arch. Gen. Psychiatry*. 2003; 60:524–530. [PubMed: 12742874]
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. [PubMed: 16449484]
11. Piven J, Palmer P, Jacohbi 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. [PubMed: 9016266]
12. Constantino JN, Todd RD. Intergenerational transmission of subthreshold autistic traits in the general population. *Biol. Psychiatry*. 2005; 57:655–660. [PubMed: 15780853]
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(6):691–699. [PubMed: 16721319]
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(5):535–542. [PubMed: 18221348]
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 Oct; 163(10):907–914. [PubMed: 19805709]
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(1):129–133. [PubMed: 15796128]
17. Kanner L, Rodriguez A, Ashenden B. How far can autistic children go in matters of social adaptation? *J. Autism Child. Schizophr.* 1972 Jan–Mar; 2(1):9–33. [PubMed: 5067404]
18. Rutter M. The autistic child. *R. Inst. Public Health Hyg. J.* 1967 Jul–Dec; 30(4):130–132. [PubMed: 5588915]
19. Wing JK, O'Connor N, Lotter V. Autistic conditions in early childhood: a survey in middlesex. *Br. Med. J.* 1967 Aug 12; 3(5562):389–392. [PubMed: 20791298]
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(3):308–320. [PubMed: 20040725]
21. Ingram DG, Takahashi TN, Miles JH. Defining autism subgroups: a taxometric solution. *J. Autism Dev. Disord.* 2008 May; 38(5):950–960. [PubMed: 17985224]
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. [PubMed: 16754843]
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(11):663–674. [PubMed: 19421728]
24. Mayes SD, Calhoun SL. Influence of IQ and age in childhood autism: Lack of support for DSM-IV Asperger's disorder. *Journal of Developmental and Physical 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. [PubMed: 9519632]
26. Lubke GH, Muthen B. Investigating population heterogeneity with factor mixture models. *Psychol Methods*. 2005 Mar; 10(1):21–39. [PubMed: 15810867]
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 Jun; 9(3):313–324. [PubMed: 16790142]
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 Dec; 46(12):1584–1593. [PubMed: 18030080]
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. [PubMed: 11911233]
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(1):24–38. [PubMed: 14992654]
32. Amaral DG, Schumann CM, Nordahl CW. Neuroanatomy of autism. *Trends Neurosci.* 2008; 31(3):137–145. [PubMed: 18258309]
33. Pardo CA, Eberhart CG. The neurobiology of autism. *Brain Pathol.* 2007 Oct; 17(4):434–447. [PubMed: 17919129]
34. Lykken DT, McGue M, Tellegen A, Bouchard TJ. Emergenesis: Genetic traits that may not run in families. *Am. Psychol.* 1992; 47(12):1565–1577. [PubMed: 1476327]
35. Haslam N. Evidence that male sexual orientation is a matter of degree. *J. Pers. Soc. Psychol.* 1997; 73:862–870. [PubMed: 9325597]
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. [PubMed: 17324013]
37. Haslam N. Categorical versus dimensional models of mental disorder: The taxometric evidence. *Aust. N. Z. J. Psychiatry.* 2003; 37:696–704. [PubMed: 14636384]
38. Cook EH Jr, Scherer SW. Copy-number variations associated with neuropsychiatric conditions. *Nature*. 2008 Oct 16; 455(7215):919–923. [PubMed: 18923514]
39. Gottesman II, Gould TD. The endophenotype concept in psychiatry: Etymology and strategic intentions. *Am. J. Psychiatry.* 2003; 160(4):636–645. [PubMed: 12668349]
40. Cannon TD, Keller MC. Endophenotypes in the genetic analyses of mental disorders. *Annu Rev Clin Psychol.* 2006; 2:267–290. [PubMed: 17716071]
41. Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet.* 2008 May; 9(5):341–355. [PubMed: 18414403]
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 Sep; 153B(6):1119–1126. [PubMed: 20552678]
43. Constantino JN, Zhang Y, Frazier TW, Abbacchi AM, Law P. Sibling recurrence and the genetic epidemiology of autism. *Am. J. Psychiatry.* 2010 Nov; 167(11):1349–1356. [PubMed: 20889652]
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 Oct; 27(5):501–517. [PubMed: 9403369]
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 Oct; 46(10):1324–1332. [PubMed: 17885574]
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 Apr 5; 150B(3):328–334. [PubMed: 18618672]
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 Sep; 48(9):932–940. [PubMed: 17714378]
50. Nasser F, Wisenbaker J. A Monte Carlo study investigating the impact of item parceling on measures of fit in confirmatory factor analysis. *Educational & Psychological Measurement.* 2003; 63(5):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. *Multivariate Behav Res.* 2006 Feb; 41(4):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. *Structural Equation Modeling: A Multidisciplinary Journal*. 2010; 17(4):605–628.
53. Faraone SV, Tsuang MT. Measuring diagnostic accuracy in the absence of a "gold standard". *Am. J. Psychiatry*. 1994 May; 151(5):650–657. [PubMed: 8166304]
54. Mildenberger 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 Dec; 10(4):248–255. [PubMed: 11794550]
55. Cohen, J. *Statistical power analysis for the behavioral sciences*. 2nd edition. 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 Nov; 40(11):1389–1402. [PubMed: 20361243]
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 Jan 3; 32(2):768–773. [PubMed: 21208775]
58. Dinstei I, Pierce K, Eyer L, et al. Disrupted neural synchronization in toddlers with autism. *Neuron*. 2011 Jun 23; 70(6):1218–1225. [PubMed: 21689606]
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 Jun 9; 70(5):863–885. [PubMed: 21658581]
61. Wang K, Zhang H, Ma D, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*. 2009 May 28; 459(7246):528–533. [PubMed: 19404256]
62. Gregory SG, Connelly JJ, Towers AJ, et al. Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. *BMC Med*. 2009; 7:62. [PubMed: 19845972]
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 Jul; 38(6):1081–1091. [PubMed: 18058215]
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 Oct; 29(5): 345–350. [PubMed: 18852608]

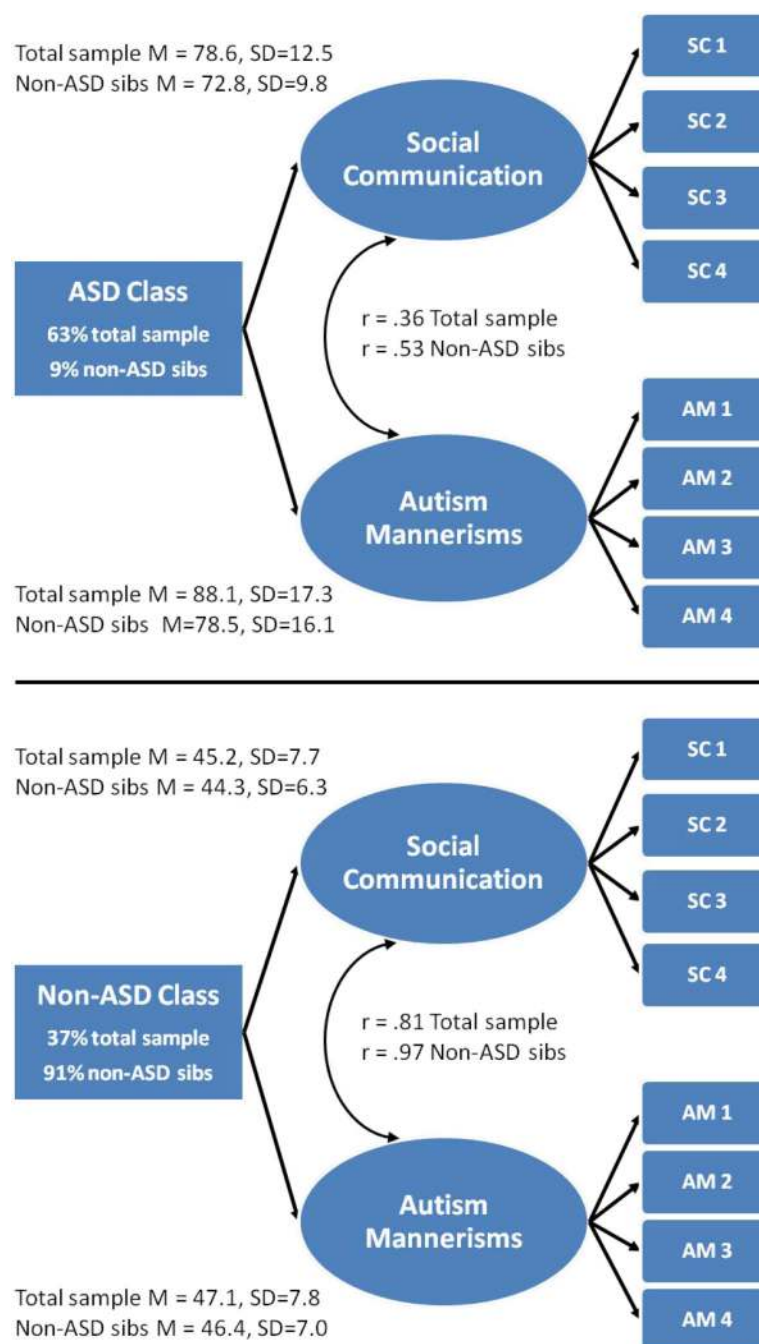


Figure 1.

Two-factor/two-class model using the Social Responsiveness Scale in the total available sample ($N=6,949$) and non-autism spectrum disorder (ASD) sibling sub-sample ($N=2,701$). 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 sub-sample ($p<.001$). Non-ASD sibs designate model results for the non-ASD sibling sub-sample. 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.

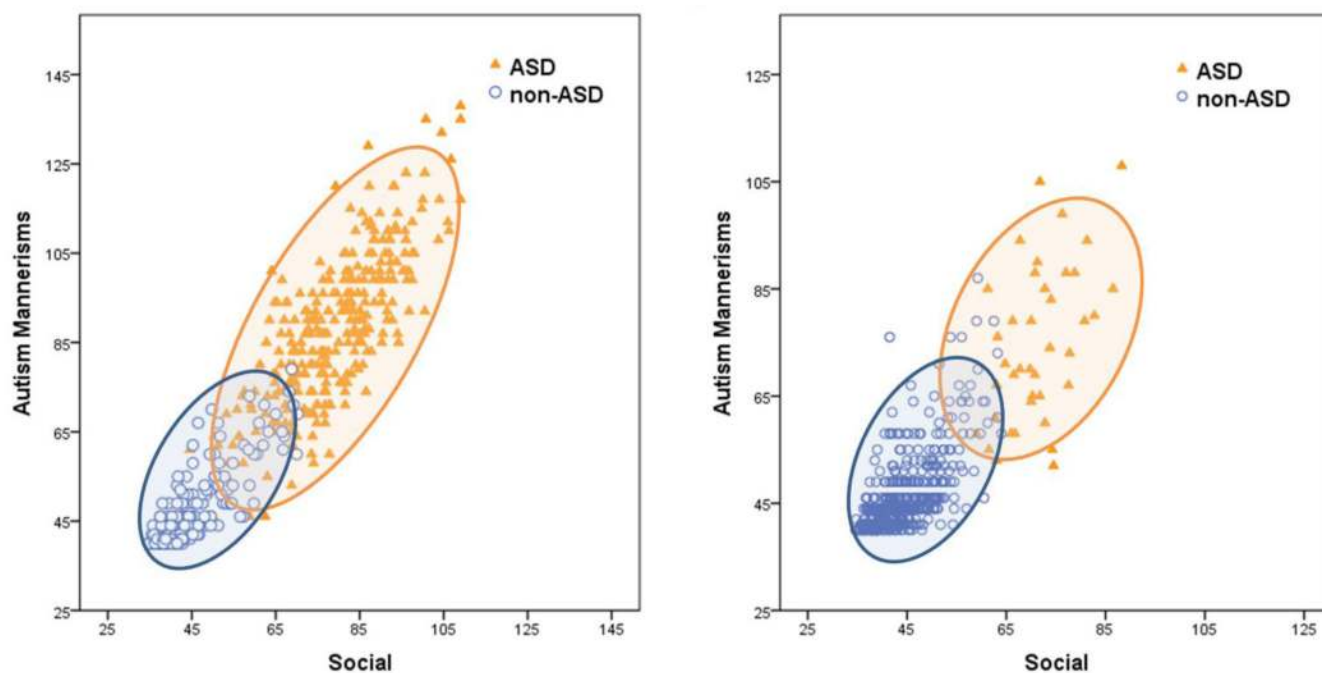


Figure 2.

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 sub-sample (N=2,701; right panel). Note: SRS social T-scores are the average of the four social sub-scales. Higher scores indicate greater impairment. Each figure is based upon a random sub-sample (n=500).

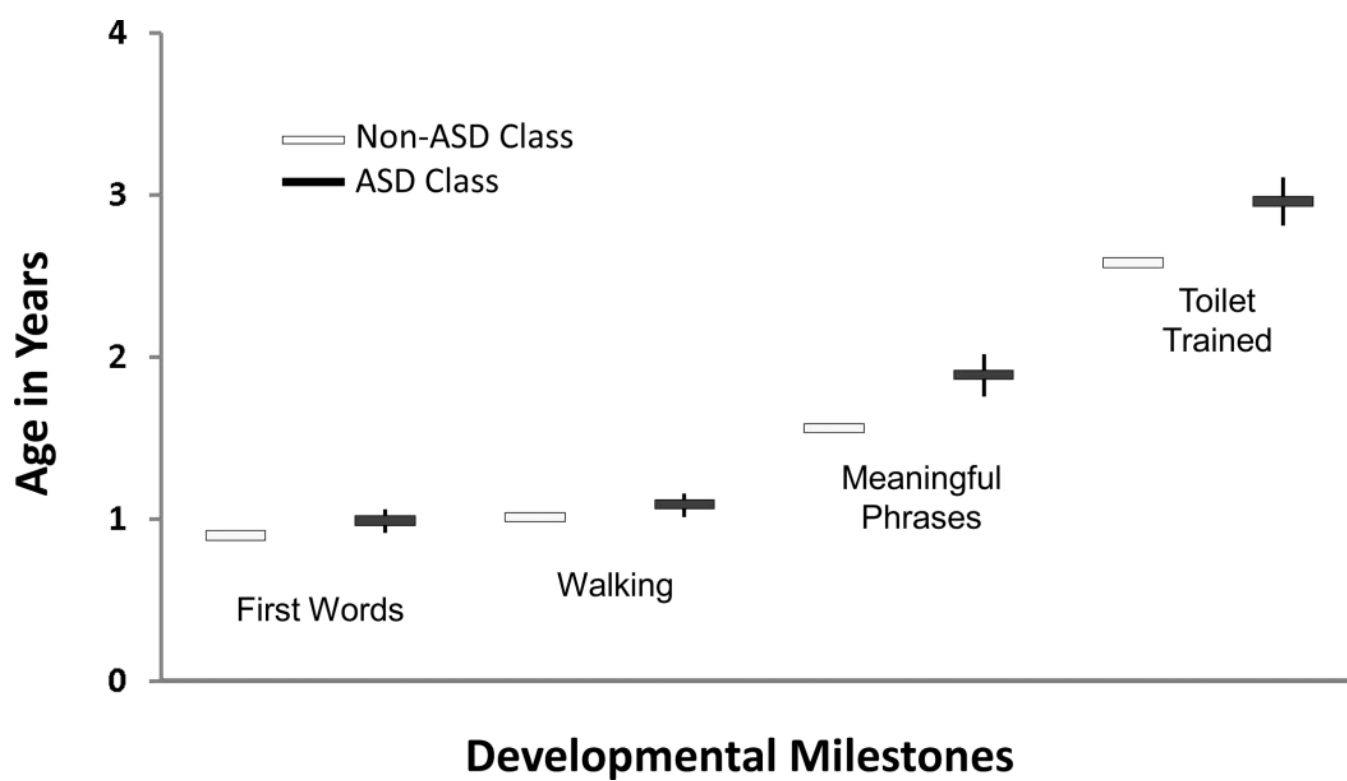


Figure 3.

Age of reported developmental milestones (M \pm 95% CI) for non-autism spectrum disorder (ASD) siblings empirically classified into the autism spectrum disorder and non-autism spectrum disorder classes.

Table 1

Model comparisons, fit statistics, and class proportions based on autism symptoms.

Model	Factor/ Class #	LL	Par	BIC	Comparison	Class Proportions
LCA	2 class	-134017	25	268256		.57 , .43
	3 class	-127148	34	254598		.31 , .32 , .37
	4 class	-124544	43	249468		.17 , .20 , .30 , .33
	5 class	-123150	52	246759		.10 , .13 , .21 , .25 , .31
	6 class	-122499	61	245537	Best fitting LCA	.09 , .11 , .19 , .19 , .12, .30
EFA	1 factor	-123767	24	247746		
	2 factor	-122803	31	245881	Parsimonious EFA	
	3 factor	-122575	37	245478	Best fitting EFA	
FM	2 factor 1 class	-122994	25	246208		
	2 factor 2 class	-122055	35	244420	Parsimonious model	.63 , .37
	2 factor 3 class	-118305	47	237027	Best overall fit	.14 , .56 , .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. **Bold** 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, Par = No. of estimated parameters.

Table 2

Sensitivity and specificity of DSM-IV-TR, proposed DSM-5, and modified criteria to clinical diagnosis and empirically-derived classifications.

	Clinical ASD Diagnosis		Empirically-Derived Classifications	
	Sensitivity	Specificity	Sensitivity	Specificity
DSM-IV-TR	.95	.86	.92	.83
DSM-5: Field Trial Phase I	.81	.97	.78	.95
DSM-5: Relaxed Algorithm (1 less criterion)	.93	.95	.89	.92
DSM-5: Without High Functioning Symptoms	.64	.98	.61	.97
DSM-5: Without Sensory Sensitivity / Interests	.78	.97	.75	.96
DSM-5: One Symptom Per Criterion	.96	.90	.92	.87

Note: Comparison of DSM-IV-TR and proposed DSM-5 criteria was done using a sub-sample 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 $\pm .01$. Symptoms tapping high functioning manifestations of Autism Spectrum Disorder (ASD) were determined from field trials criteria and included: 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 (1 less criterion) refers to relaxed phase I field trials criteria with the modification that 1 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.