Trajectories of Autism Severity in Children Using Standardized ADOS Scores

WHAT'S KNOWN ON THIS SUBJECT: Autism spectrum disorders are characterized by heterogeneous severity. Previous latent variable analyses of longitudinal data have focused on trajectories of related features such as IQ, and not on changes over time in standardized, observational measures of core autism symptoms.

WHAT THIS STUDY ADDS: Autism Diagnostic Observation Schedule—calibrated severity scores allow comparisons of observational data from toddlerhood to adolescence. This first report of latent autism severity trajectory classes indicates that most children show stability in core symptom severity over many years; small groups improved or worsened.

AUTHORS: Katherine Gotham, PhD,^a Andrew Pickles, PhD,^b and Catherine Lord, PhD^c

^aUniversity of Michigan Autism and Communication Disorders Center, Ann Arbor, Michigan; ^bDepartment of Biostatistics, Institute of Psychiatry, King's College London, London, England; and ^cCenter for Autism and the Developing Brain, Weill-Cornell Medical College, White Plains, New York

KEY WORDS

autism spectrum disorders, severity, trajectory, Autism Diagnostic Observation Schedule Comparison Score

ABBREVIATIONS

ABA—applied behavior analysis
ADI-R—Autism Diagnostic Interview-Revised
ADOS—Autism Diagnostic Observation Schedule
ASD—autism spectrum disorder
BIC—Bayesian Information Criteria
CSS—calibrated severity score
MPST—mentored, parent-implemented structured teaching
NVIQ—nonverbal IQ
PDD-NOS—pervasive developmental disorder-not otherwise specified
VIQ—verbal IQ

Dr Gotham's current affiliation is Vanderbilt Kennedy Center, Nashville, Tennessee.

www.pediatrics.org/cgi/doi/10.1542/peds.2011-3668

doi:10.1542/peds.2011-3668

Accepted for publication Jul 24, 2012

Address correspondence to Katherine Gotham, PhD, PMB74, 230 Appleton Place, Nashville, TN 37203. E-mail: katherine.gotham@vanderbilt.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2012 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: Dr Lord receives royalties for the ADOS; profits related to this study were donated to charity; Dr Gotham will receive royalties from the ADOS-2, the second edition of the measure described here and plans to donate all proceeds from research use to charity, and Dr Pickles has indicated that he has no financial relationships relevant to this article to disclose.

FUNDING: This study was funded by the National Institute of Mental Health (NIMH R01 MH57167, MH066469, and T32-MH18921), the National Institute of Child Health and Human Development (HD 35482-01 and P30HD15052), and an Autism Speaks Pre-doctoral Training Fellowship. Funded by the National Institutes of Health (NIH).

abstract

OBJECTIVES: To plot longitudinal trajectories of autism spectrum disorder (ASD) severity from early childhood to early adolescence. In line with reported trajectories in toddlers, we hypothesize that a substantial minority of children will show marked changes in ASD severity over time, with “Improvers” demonstrating the highest mean baseline and rate of growth in verbal IQ (VIQ).

METHODS: Patients included 345 clinic referrals and research participants with best-estimate clinical diagnoses of ASD at 1 or more time points, and repeated Autism Diagnostic Observation Schedule (ADOS), VIQ, and nonverbal IQ scores. Standardized ADOS severity scores were applied to 1026 assessments collected longitudinally between the ages of 2 and 15 (VIQ at most recent assessment: mean = 58, SD = 35). Scores were fitted for latent severity trajectory classes with and without covariates. Adaptive behavior and VIQ trajectories over time were modeled within each of the best-fit latent classes.

RESULTS: A 4-class model best represented the observed data. Over 80% of participants were assigned to persistent (stable) high or moderately severe classes; 2 small classes respectively increased or decreased in severity over time. Age, gender, race, and nonverbal IQ did not predict class membership; VIQ was a significant predictor. Baseline VIQ was highest in the improving and worsening classes; it increased at the greatest rate in the improving class. Adaptive behavior declined in all but the improving class, with consistent impairment in all classes.

CONCLUSIONS: If replicated, identified trajectory classes of ADOS severity may contribute to clinical prognosis and to subtyping samples for neurobiological and genetic research. *Pediatrics* 2012;130:e1278–e1284

Research on developmental trajectories in autism spectrum disorders (ASDs) has focused on the stability of categorical diagnoses, verbal and cognitive outcomes, and symptom domain change over time. In terms of cognitive trajectories, groups with initially higher IQs often make great gains, whereas less able groups remain relatively stable or show small improvements over time.^{1,2} Within the small but growing body of literature on the trajectory of ASD-specific symptom expression, severity has most often been quantified with scores from the Autism Diagnostic Interview-Revised³ and the Childhood Autism Rating Scale.⁴ A 2004 review⁵ indicates general improvement over the life span in all 3 core *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*⁶ symptom domains. Yet building a coherent picture of change and stability in ASD symptom expression is obscured by the variability of participant demographics, the measures used, and study design (eg, retrospective versus prospective data analysis). Uniform methods of tracking changes in ASD symptom profiles over time may lead to more detailed prognostic estimates, as well as opportunities to study the course of these disorders over the life span.

The Autism Diagnostic Observation Schedule (ADOS)⁷ has shown strong sensitivity and specificity for best-estimate diagnoses,⁸ making it a common choice among phenotyping measures. As a child develops, he/she often moves through age- and language-specific ADOS modules, which contributes to the measure's predictive validity across developmental levels but makes raw scores not directly comparable across time. Comparing ADOS data longitudinally is similarly confounded by the effects of age and language level on algorithm totals.^{8–10} Two recent updates have been made to the ADOS with the purpose of increasing

the comparability of the modules used with children and adolescents. First, revised algorithms were created with the same number of items and of similar content across modules 1 to 3,⁸ which resulted in minimal association between ADOS totals and chronological age and verbal IQ (VIQ).^{8,11} Second, the ADOS-revised algorithm raw total scores were standardized within 1807 assessments from participants with ASDs to produce a 10-point scale with more uniform distribution across age- and language-level-determined groups, and less variance accounted for by VIQ, than raw total scores.¹² It is important to note that these calibrated severity scores (CSS; known as Comparison Scores in the ADOS-2) do not measure functional impairment but rather provide a marker of ASD symptom severity on the ADOS relative to age and language level.

The current study used standardized ADOS scores to plot changes in ASD severity over time in prospective repeat-assessment data. Our primary goal was to identify latent trajectory classes, or patterns of change over time, in autism severity from early childhood to adolescence, covarying VIQ, nonverbal IQ (NVIQ), gender, and race. A further aim was to compare trajectories of VIQ and measures of adaptive functioning across the resulting classes. We also examined the effect of treatment variables on severity classes; when using a subset of the same data, Anderson and colleagues¹³ found that individuals who as young children received >20 hours of mentored, parent-implemented structured teaching (MPST) had substantially greater Vineland Adaptive Behaviors Scales¹⁴ Socialization domain scores and verbal skills at age 13 than those with less or no exposure to MPST, even in comparison with those with many hours of early applied behavior analysis (ABA). Based on a recent report in which approximate thirds of a toddler ASD

sample improved, worsened, or remained stable on ADOS algorithm totals over 2 years,¹⁵ we hypothesize that a substantial minority of children with ASDs will show marked changes in symptom expression over time. Alternatively, children may show wide variability in severity over time because of developmental influences or instability of the measure. Second, we hypothesize that children with the greatest magnitude of symptom improvement will have the highest mean baseline VIQ and greatest rate of VIQ growth despite the relative independence of ADOS severity scores and VIQ.

METHODS

Participants and Procedures

Analyses were conducted on data from 345 individuals aged 2 to 15 years who were referred for, and at least 1 time point were diagnosed with, ASD. Inclusion also required repeated ADOS administrations with best-estimate clinical diagnoses, VIQ and NVIQ scores, and gender and race data (see Table 1 for participant demographics).

The sample included 1026 “assessments,” defined as contemporaneous ADOS data and clinical diagnosis; 258 individuals had 2 or 3 assessments, and 87 had between 4 and 8 assessments. Ninety-seven percent of assessments came from participants with final (ie, most recent) diagnoses of ASD (see Table 1 for details). Ten children (contributing 3% of assessments) ultimately received nonspectrum diagnoses ($n = 5$ language disorders, $n = 2$ intellectual disability, $n = 1$ Tourette's disorder, $n = 1$ mood disorder, $n = 1$ oppositional defiant disorder with attention-deficit/hyperactivity disorder).

Within this sample, 159 individuals were consecutive referrals to the Treatment and Education of Autistic and Communication Handicapped Children Centers at the University of North Carolina, Chapel Hill, or the University of Chicago Developmental Disorders Clinic, who

TABLE 1 Sample Description

Participant gender, <i>n</i> (%)	
Female	63 (18)
Male	282 (82)
Race/ethnicity, <i>n</i> (%)	
Black	62 (18)
Asian American	7 (2)
White	272 (79)
Multiracial/other	4 (1)
Maternal education, <i>n</i> (%)	
Graduate/professional	86 (25)
Bachelors/some college	186 (54)
High school or less	73 (21)
Most recent diagnosis, <i>n</i> (%)	
Autism	231 (67)
PDD-NOS	104 (30)
Nonspectrum	10 (3)
ADOS module at most recent assessment, <i>n</i> (%)	
Module 1	139 (40)
Module 2	103 (30)
Module 3	103 (30)
Chronological age, <i>y</i>	
Initial assessment	
Mean (SD)	3.3 (1.4)
Quartiles (25th/50th/75th)	2.4/2.8/3.8
Most recent assessment	
Mean (SD)	7.3 (2.7)
Quartiles (25th/50th/75th)	4.7/7.6/9.5
VIQ	
Initial assessment	
Mean (SD)	48 (27)
Quartiles (25th/50th/75th)	25/43/67
Most recent assessment	
Mean (SD)	58 (35)
Quartiles (25th/50th/75th)	29/53/87
NVIQ	
Initial Assessment	
Mean (SD)	73 (24)
Quartiles (25th/50th/75th)	55/72/89
Most Recent Assessment	
Mean (SD)	73 (31)
Quartiles (25th/50th/75th)	47/74/97

participated in a longitudinal study of the “Early Diagnosis of ASD.” These participants (hereafter, “inception cohort”) were referred for possible autism before 36 months of age and were evaluated again around ages 5 and 9 by examiners blind to their previous diagnosis and scores.^{16,17} This subsample maintained a high level of participation over time (80.4% follow-up rate at age 9),¹⁶ with attrition unrelated to initial diagnosis, language level, IQ, adaptive functioning, or gender.¹⁶ The remainder (*n* = 186 individuals) received repeated diagnostic

evaluations as clinic patients or research participants at 2 university-based autism clinics.

A standard research protocol was used across sites and projects. The Autism Diagnostic Interview-Revised (ADI-R),³ a standardized, semistructured parent/caregiver interview yielding developmental history specific to ASD features, was followed by the Vineland Adaptive Behavior Scales, first or second edition,^{14,18} a standardized parent/caregiver interview of adaptive functioning across social, communication, daily living, and motor skills domains. Child assessments included psychometric testing and the ADOS. Reevaluations often did not include the ADI-R. For all assessments, a clinical diagnosis was made by a psychologist and/or psychiatrist after review of all data.

Study Design and Analyses

ADOS calibrated severity scores¹² were analyzed for patterns of stability or change using Generalized Linear Latent and Mixed Models (gllamm)¹⁹ in Stata version 10.²⁰ Latent class growth curve models with 3 to 6 trajectory classes and random intercept linear and quadratic age terms were compared for goodness of fit,²¹ and the most parsimonious model was chosen. Local optima problems were addressed by using both standard starting points and those derived from a gateaux derivative search. The fixed part coefficients, representing linear and quadratic relationships of age with ADOS severity scores for the whole sample, were tested for significance using an overall likelihood ratio χ^2 test to assess a common trend. Multinomial logistic regression was used to examine the association of the model posterior assigned class membership to the baseline covariates VIQ, NVIQ, gender, and race.

Classes were assessed for differences in rates of parent-reported regression

in communicative or other skills, as measured by scores of 1 or 2 on items 11 or 20 of the ADI-R. We also used an overall likelihood ratio χ^2 to examine trajectory class differences in treatment variables available in the inception cohort subsample, comparing children with >20 hours of MPST or >1667 hours of early ABA with those children with less or no exposure to either type of therapy.¹³

To examine the concurrent development of VIQ and the Vineland Daily Living Skills V-scale scores, we plotted smoothed (fractional polynomial) mean scores by age for each trajectory class. Wald tests from generalized estimating equations multivariate regression models with an exchangeable working correlation matrix (equivalent to repeated measures analysis of variance without requiring complete data and with a robust parameter covariance matrix estimator that does not assume constant error variance) were used to test for class differences in the intercept (centered at age 6 to allow intercepts to provide estimates of class means at this point), linear, and quadratic age trends.

RESULTS

Latent Classes by ADOS Severity Score Trajectory

A linear model of 5 latent trajectory classes had the most parsimonious fit to longitudinal ADOS CSS data, as suggested by the lowest Bayesian Information Criteria (BIC) in comparison with other models (see Table 2). Greater numbers of dimensions or classes led to models with higher BIC.

The linear fixed part coefficients of the 5-class model showed no evidence of a significant relationship between ADOS severity and chronological age (0.2[2] = 0.33, *P* = .8), suggesting no significant overall age trend masked by the grouping into latent classes.

One of the 5 classes in this best-fit model included only 6 participants (final

TABLE 2 Latent Severity Class Model Comparison

Dimensions	Classes		
	4	5	6
Intercept	2148.5 (4355.4)	2148.8 (4367.1)	—
Intercept	2134.2 (4344.4)	2122.8 ^a (4339.1) ^a	2115.2 (4341.4)
Linear slope	2133.6 (4360.6)	2120.9 (4358.7)	2110.7 (4361.7)
Quadratic slope	—	—	—

Log-likelihoods shown, with BIC below them in parentheses.

^a Lowest BIC = most parsimonious fitting model.

diagnoses: $n = 1$ autism, $n = 2$ pervasive developmental disorder-not otherwise specified (PDD-NOS), $n = 3$ intellectual disability/language delay) with 22 assessments. These children had stable mild-severity scores from 1 to 3 over time, with 1 outlying score of 6. Because of the small size of this class, these participants were dropped from further analyses. The 4 remaining latent trajectory classes are shown in Fig 1. Participant chronological age was restricted to a maximum of 10 years for graphical representation, because data for the 11 to 15 age span were more sparse.

The 4 classes included a persistent high-severity class (class 1: persistent high; 46% of 339 remaining participants), a moderately severe class (class 2: persistent moderate; 38%), a class that tended to increase in ASD severity over time (class 3: worsening; 9%), and a class that decreased in ASD severity over time (class 4: improving; 7%). Trajectory class was not associated with the number of assessments per individual ($F[3,335] = 1.3$, $P = .27$). The average probability with which children were assigned to their best class was high for classes 1, 3, and 4 ($P = .82$, $.79$, and $.81$ respectively) but lower ($P = .68$) for class 2 (persistent moderate). The average probability that children assigned to this class might have belonged to class 3 (worsening) was not small ($P = .21$). The worsening class was marked by score variability: 70% of children assigned to this class had most recent severity scores higher than previous scores, whereas the remaining 30% showed wide variability across time, some of them “ending” on an improved score. In contrast, all children assigned to the improving group

had most recent scores milder than previous scores.

Table 3 describes initial and final diagnostic measures and demographic variables in the 4 latent classes. ADI-R domain totals are reported as sums of “current” scores of only those algorithm items comparable across age groups at both initial and final assessment, to compare stability or change over time by latent class. Trends in raw scores were observed to fall (ie, improve) slightly over time in Current Social-Communication on the ADI-R and Social Affect scores on the ADOS, and to rise (ie, worsen) slightly over time on ADOS Restricted Repetitive Behavior scores across the first 3 classes. The worsening class was the only group to exhibit greater severity over time in any ADI-R Current domain mean score (Verbal Communication and Restricted Repetitive Behavior).

Covariates as Predictors of Latent Class Membership

As shown in Table 3, gender, race, and initial NVIQ did not significantly predict latent class membership. Higher initial VIQ significantly predicted membership in the improving, worsening, and moderate classes in comparison with the persistent high reference class. Relative risk ratios were generated from multinomial logistic regression; race and gender were entered as binary predictors (0 = white or male; 1 = other race or female), and VIQ and NVIQ scores were standardized before being entered into the model. Relative risk ratios indicate the multiple of odds for specific class membership (eg, improving) in a particular group (eg, females) in comparison with membership in the persistent high class (eg, 1 SD difference in VIQ increased the odds of being in the improving class, relative to the persistent high class, by 383%).

Note. Y-axis depicts ADOS Calibrated Severity Score.

FIGURE 1

ADOS severity score latent trajectory classes. y axis depicts ADOS CSS.

TABLE 3 Latent Severity Classes: Descriptives and Predictors

		Class 1		Class 2		Class 3		Class 4	
		Persistent High		Persistent Moderate		Worsening		Improving	
		First	Last	First	Last	First	Last	First	Last
Age, y	M (SD)	3.4 (1.6)	7.5 (2.7)	3.3 (1.3)	7.0 (2.7)	2.8 (0.6)	7.3 (3.0)	3.4 (1.7)	7.5 (2.7)
ADI-R (C) Social	M (SD)	13.5 (4.6)	12.3 (5.0)	11.3 (4.9)	10.2 (5.2)	8.3 (4.0)	5.6 (3.6)	9.4 (5.3)	6.7 (5.4)
ADI-R (C) Comm Verbal	M (SD)	9.9 (4.1)	9.8 (4.4)	8.5 (4.1)	8.3 (4.1)	5.4 (2.5)	5.8 (3.2)	5.3 (4.4)	5.3 (4.5)
ADI-R (C) Comm Nonverbal	M (SD)	6.1 (2.2)	5.3 (2.5)	5.5 (2.5)	4.2 (2.7)	3.6 (2.6)	1.6 (1.7)	4.3 (2.7)	2.8 (2.8)
ADI-R (C) RRB	M (SD)	4.7 (2.3)	4.6 (2.2)	3.8 (2.4)	3.7 (2.5)	2.9 (2.4)	3.4 (2.6)	3.3 (1.9)	3.1 (2.3)
ADOS SA	M (SD)	16.9 (3.2)	15.4 (3.3)	13.8 (4.2)	11.6 (4.2)	7.4 (3.7)	8.3 (4.6)	12.2 (4.2)	3.0 (1.8)
ADOS RRB	M (SD)	4.8 (1.8)	5.2 (2.1)	3.3 (1.8)	4.0 (2.1)	2.1 (1.5)	2.6 (2.0)	3.3 (1.9)	1.2 (1)
ADOS CSS	M (SD)	8.8 (1.7)	8.6 (1.3)	6.6 (1.3)	6.9 (1.4)	3.5 (1.7)	5.7 (2.5)	7.2 (1.9)	2.3 (1.1)
VDLS	M (SD)	56.5 (14.8)	43.2 (22.4)	60.4 (12.8)	49.7 (24.6)	65.6 (14.6)	59.5 (22.6)	72.2 (10.7)	77.7 (15.0)
VIQ	M (SD)	41.0 (25.9)	45.6 (29.2)	48.8 (27.3)	61.0 (35.4)	55.8 (20.9)	70.7 (28.8)	74.2 (24.0)	100.8 (19.0)
	RRR (<i>P</i>)	—		1.63 (0.01*) ^a		3.09 (0.001*) ^a		4.83 (0.001*) ^a	
NVIQ	M (SD)	67.6 (23.9)	64.9 (29.4)	74.6 (23.7)	73.4 (31.1)	75.1 (23.6)	81.2 (32.0)	88.1 (15.7)	100.3 (18.4)
	RRR (<i>P</i>)	—		0.99 (0.94)a		0.70 (0.27)a		0.91 (0.80)a	
Gender	Female	17.3%		23.8%		10.0%		8.7%	
	RRR (<i>P</i>)	—		1.52 (0.17) ^a		0.51 (0.31) ^a		0.45 (0.33) ^a	
Race	Minority	20.5%		23.8%		26.7%		8.7%	
	RRR (<i>P</i>)	—		1.53 (0.16) ^a		2.14 (0.12) ^a		0.83 (0.82) ^a	
Final diagnosis (% observed in each class)	Autism (<i>n</i> = 230)	137 (88%)		83 (64%)		9 (30%)		1 (4%)	
	PDD-NOS (<i>n</i> = 102)	19 (12%)		47 (36%)		18 (60%)		18 (78%)	
	Non-ASD (<i>n</i> = 7)	—		—		3 (10%)		4 (18%)	
	<i>n</i> by class	156		130		30		23	

M, mean; First, data at initial assessment; Last, data at final assessment; ADI-R (C), current ADI-R algorithm scores on items comparable across ages 2 through 15, summed within ADI-R domains; ADOS SA, ADOS Social Affect domain raw total; ADOS RRB, ADOS Restricted Repetitive Behavior domain raw total; ADOS CSS, ADOS calibrated severity score; VDLS, Vineland Daily Living Skills standard score; RRR, relative risk ratio; *P* = *P* value with 16 degrees of freedom; **P* < .01. —, indicate groups for which no relevant data exists.

^a Results of multinomial logistic regression, with reference group = class 1, persistent high.

Diagnosis, Regression Status, and Treatment Variables by Latent Severity Class

Almost all children with a final diagnosis of autism were assigned to persistent high (60%) or moderate (36%). Participants with PDD-NOS most commonly were assigned to persistent moderate (45%), worsening, and improving classes (17.3% each). Three children in the worsening severity class ultimately received nonspectrum diagnoses (*n* = 1 language disorder, *n* = 1 disruptive behavior disorder, *n* = 1 intellectual disability), as did 4 children in the improving class (*n* = 1 Tourette's syndrome, *n* = 1 mood disorder, *n* = 2 language disorders).

Language regression scores did not differ significantly across the 4 classes, $F(3,439) = 2.3$, $P = .08$. We found no significant class difference between participants with the highest levels of either MPST or ABA therapy hours in comparison with children who had

received less or none of either type of intervention, $\chi^2(4) = 4.3$, $P = .36$ for MPST; $\chi^2(4) = 3.5$, $P = .48$ for ABA.

IQ and Adaptive Behavior Trajectories Within ADOS Latent Severity Classes

The pattern of change in sample mean VIQ and Vineland Daily Living V-scale (standard scores) in each trajectory group is shown in Figs 2 and 3.

All classes showed an increasing (improving) trend over time in VIQ but with marked between-class differences (generalized estimating equations Wald test over intercept, linear and quadratic terms; $0.2[9] = 219.60$, $P < .001$). The improving class means exhibited a much steeper curve indicating progress that occurred earlier and was greater overall than in the other 3 classes. At baseline, the improving class had significantly higher VIQ than the persistent high and moderate classes ($F(3,322) = 18.21$, $P < .001$), although it

did not differ significantly from the worsening class ($P = .642$). Tests at age 6, when IQ appeared to stabilize, indicated the improving class had significantly higher mean VIQ than the remaining 3 classes ($P < .001$); persistent moderate and worsening were similar ($P = .164$), although both were above the persistent high class ($P < .001$).

For Vineland Daily Living Skills (including such skills as toileting, dressing, and chores), the classes showed similar and relatively unimpaired scores at age 2 but diverged thereafter ($0.2[9] = 103.16$, $P < .001$). Modest gains were made by the improving class, with marked declines noted in the other groups. By age 6, the improving class was significantly better than the other classes (at $P = .006$ or smaller), with no significant differences among those 3 ($P = .243$ or greater).

DISCUSSION

Based on standardized scores with the use of a single instrument relatively

FIGURE 2
VIQ trajectories by latent severity class.

independent of age and IQ, the majority of children remained surprisingly stable in terms of ASD severity scores over 8 to 12 years, with >80% assigned to 2 stable latent trajectory groups. This overall stability is more remarkable because the data came from an observational measure administered by different examiners over time, most of whom were blind to any previous information; thus, results cannot be attributed to informant bias (eg, parent report). Only 15% of the sample was assigned to improving or worsening classes, suggesting that, contrary to our hypothesis, severity changes in childhood are observed in a relatively small proportion

of this population. Note that the majority of participants were identified with ASDs at early ages during the 1990s, and, therefore, the sample is skewed toward lower verbal ability and higher average severity than we would expect to see in a more recent young population cohort.

Severity classes were not predicted by gender, race, or NVIQ. VIQ was a significant predictor of latent class membership, with higher scores predicting assignment to all classes over persistent high. Contrary to our hypothesis, baseline VIQ was not significantly highest in the improving class, although improvers tended to make the earliest and largest

gains in VIQ over time. In general, VIQ was maintained or increased over time in all groups. Adaptive behavior worsened in all groups, with the exception of the improving class, in which it remained stable, although in the impaired range. These findings highlight that ADOS severity scores functioned as intended, that is, to measure severity of ASD symptoms and *not* functional impairment. However, the fact that persistent autism severity continued to be associated with lower mean VIQ and adaptive behavior indicates that autism characteristics and cognitive and adaptive functioning are not entirely independent features.

Regression did not appear to “follow” children over time in terms of increasing autism severity, although it has been reported to be associated with ongoing functional impairment in terms of verbal and/or intellectual ability.^{22,23} We did not find class differences in parent-mediated or ABA intervention hours by age 5 in the inception cohort subsample, but treatment data were based on parent report, and children were not randomly assigned to type or amount of intervention. Future examinations in carefully controlled intervention data are needed.

Limitations

This sample comprised approximately one-third of the participants on whose data the ADOS CSS was based. Although this introduces the potential for some degree of circularity of findings, severity scores were calibrated within narrow age-by-language cells¹² without regard for a participant's previous or future scores; thus, his/her set of severity scores were free to vary or remain stable across assessments. We also might expect caregivers of clinic patients to self-refer for repeated evaluations more often in the case of persistently severe autism characteristics; however, trajectory class was not associated with number of repeat assessments.

FIGURE 3
Vineland Adaptive Behavior Scales “Daily Living” V scores by latent class.

Again, the inception cohort was identified at early ages during a historical period of limited public awareness of ASDs. Thus, the persistently mild class consisting of 6 participants (2% of the sample that was dropped from further analyses) might be more prevalent in studies using samples diagnosed at age 2 in more recent years (see Lord et al¹⁵ for trajectories within a recent sample of toddlers with overall greater verbal ability). Replication of these findings is needed in other large, preferably epidemiological, datasets.

CONCLUSIONS

These findings underscore the stability of core ASD features in children across

diverse ages and levels of functioning. Overall stability in autism severity is more striking given that data were based on a relatively brief standardized observation by an experienced clinician who often was blind to previous diagnoses or scores. The current findings can aid clinical prognostic estimates by allowing professionals to provide families with “benchmark” statistics on stability and change in this population. They also provide an initial model of the direction, magnitude, and age periods associated with observed ASD severity changes in a small proportion of children, with which more recent outcome data could be compared as they become available. Future directions

include exploration of other risk and protective factors to class membership, with particular emphasis on treatment effects. Latent classes can be tested for association with distal outcomes (eg, academic placement, peer relationships) and genetic or neurobiological profiles.^{1,24–26}

ACKNOWLEDGMENTS

We gratefully acknowledge the help of Drs Brady West, Lingling Zhang, Al Cain, Mohammed Ghaziuddin, and Israel Liberzon; the staff of the University of Michigan Autism and Communication Disorders Center; and the families that participated in this research.

REFERENCES

- Anderson DK, Lord C, Risi S, et al. Patterns of growth in verbal abilities among children with autism spectrum disorder. *J Consult Clin Psychol*. 2007;75(4):594–604
- Gabriels RL, Hill DE, Pierce RA, Rogers SJ, Wehner B. Predictors of treatment outcome in young children with autism: a retrospective study. *Autism*. 2001;5(4):407–429
- Rutter M, Le Couteur A, Lord C. *Autism Diagnostic Interview-Revised*. Torrance, CA: Western Psychological Services; 2003
- Schopler E, Reichler R, Renner B. *The Childhood Autism Rating Scale (CARS) for Diagnostic Screening and Classification of Autism*. Diagnosis and Teaching Curricula for Autism and Developmental Disabilities Series. New York, NY: Irvington; 1986
- Seltzer MM, Shattuck P, Abbeduto L, Greenberg JS. Trajectory of development in adolescents and adults with autism. *Ment Retard Dev Disabil Res Rev*. 2004;10(4):234–247
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: APA Press; 1994
- Lord C, Risi S, Lambrecht L, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord*. 2000;30(3):205–223
- Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. *J Autism Dev Disord*. 2007;37(4):613–627
- Joseph RM, Tager-Flusberg H, Lord C. Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. *J Child Psychol Psychiatry*. 2002;43(6):807–821
- de Bildt A, Sytema S, Ketelaars C, et al. Interrelationship between Autism Diagnostic Observation Schedule-Generic (ADOS-G), Autism Diagnostic Interview-Revised (ADI-R), and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. *J Autism Dev Disord*. 2004;34(2):129–137
- Gotham K, Risi S, Dawson G, et al. A replication of the Autism Diagnostic Observation Schedule (ADOS) revised algorithms. *J Am Acad Child Adolesc Psychiatry*. 2008;47(6):642–651
- Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *J Autism Dev Disord*. 2009;39(5):693–705
- Anderson DK, Oti RS, Lord C, Welch K. Patterns of growth in adaptive social abilities among children with autism spectrum disorders. *J Abnorm Child Psychol*. 2009;37(7):1019–1034
- Sparrow S, Balla D, Cicchetti D. *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service; 1984
- Lord C, Luyster R, Guthrie W, Pickles A. Patterns of developmental trajectories in toddlers with autism spectrum disorders. *J Consult Clin Psych*. 2012;80(3):477–489
- Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Pickles A. Autism from 2 to 9 years of age. *Arch Gen Psychiatry*. 2006;63(6):694–701
- Lord C. Follow-up of two-year-olds referred for possible autism. *J Child Psychol Psychiatry*. 1995;36(8):1365–1382
- Sparrow S, Cicchetti DV, Balla D. *Vineland Adaptive Behavior Scales*. 2nd ed. Circle Pines, MN: American Guidance Service Inc.; 2005
- Rabe-Hesketh S, Skrondal A, Pickles A. Generalized multi-level structural equation modeling. *Psychometrika*. 2004;69(2):167
- StataCorp. *Stata Statistical Software*. Release 10.0. College Station, TX: Stata Corporation; 2007
- Pickles A, Croudace T. Latent mixture models for multivariate and longitudinal outcomes. *Stat Methods Med Res*. 2010;19(3):271–289
- Lord C, Shulman C, DiLavore P. Regression and word loss in autistic spectrum disorders. *J Child Psychol Psychiatry*. 2004;45(5):936–955
- Richler J, Luyster R, Risi S, et al. Is there a ‘regressive phenotype’ of Autism Spectrum Disorder associated with the measles-mumps-rubella vaccine? A CPEA Study. *J Autism Dev Disord*. 2006;36(3):299–316
- Morrow EM, Yoo SY, Flavell SW, et al. Identifying autism loci and genes by tracing recent shared ancestry. *Science*. 2008;321(5886):218–223
- DeLong GR. Autism: new data suggest a new hypothesis. *Neurology*. 1999;52(5):911–916
- Pelphrey K, Adolphs R, Morris JP. Neuroanatomical substrates of social cognition dysfunction in autism. *Ment Retard Dev Disabil Res Rev*. 2004;10(4):259–271