Examining the phenotypic heterogeneity of early Autism Spectrum Disorder: subtypes and short-term outcomes

So Hyun Kim,[1] Suzanne Macari,[1] Judah Koller,[2] and Katarzyna Chawarska[1]

[1]Child Study Center, Yale University School of Medicine, New Haven, CT, USA; [2]School of Education, Hebrew University of Jerusalem, Jerusalem, Israel

Background: Phenotypic heterogeneity among toddlers presenting with ASD symptoms complicates diagnostic considerations and limits our ability to predict long-term outcomes. To address this concern, we sought to identify more homogeneous subgroups within ASD based on toddlers' clinical profiles in the second year of life, evaluating diagnostic stability and clinical outcomes within the subgroups 1–2 years later. **Methods:** One hundred toddlers referred for suspected ASD underwent comprehensive assessments at 22 months ($SD = 3$) and 37 months ($SD = 4$). At 22 months, they were clustered based on symptom severity, developmental skills, and adaptive functioning. Diagnostic stability and clinical outcomes were evaluated within the clusters. **Results:** Four clusters characterized by distinct clinical profiles at the time of the first diagnosis were identified. Diagnostic stability was excellent in 3 out of 4 clusters (93%–100%) and was lowest in the initially least affected cluster (85%). Autism symptom severity was stable, except for one group where it increased over time (16% of the sample). A large proportion of toddlers showed significant improvements in verbal and communication skills. Only a small group (17%) exhibited very low levels of functioning and limited gains over time. **Conclusions:** Diagnostic stability and developmental progression from the second to third year of life in toddlers with ASD vary depending on their initial early profiles of relative strengths and deficits. Although a small minority of toddlers with more complex clinical presentations may not retain their diagnoses by the age of three, most children continue to exhibit symptoms of autism. Despite limited improvements in symptom severity, many children show significant gains in verbal functioning. Only a small proportion of children (17%) exhibit very limited gains despite intensive intervention. These findings support continued efforts to examine determinants of developmental trajectories including factors mediating and moderating response to treatment. **Keywords:** ASD; toddlers; diagnosis; clinical outcomes.

Introduction

Autism Spectrum Disorder (ASD) is an early emerging and highly heterogeneous neurodevelopmental disorder (APA, 2013). Symptoms of ASD begin to manifest by the second year of life and include deficits in verbal and nonverbal communication and social reciprocity as well as repetitive behaviors, atypical sensory interests (Chawarska, Klin, Paul, Macari, & Volkmar, 2009; Guthrie, Swineford, Nottke, & Wetherby, 2013; Kim & Lord, 2010), and impairments in adaptive functioning (Ray-Subramanian, Huai, & Weismer, 2011). One of the paradigmatic features of ASD is its genotypic and phenotypic heterogeneity. On the phenotypic level, heterogeneity is expressed with regard to symptom severity (Georgiades et al., 2013; Wiggins, Robins, Adamson, Bakeman, & Henrich, 2011), verbal and nonverbal IQ (Munson et al., 2008), social attention (Campbell, Shic, Macari, & Chawarska, 2014), and response to treatment (Sherer & Schreibman, 2005). Such heterogeneity hinders attempts to predict clinical outcomes, develop individualized treatment targets and strategies, and identify etiological factors associated with ASD.

One strategy for improving the understanding of heterogeneity of early syndrome expression involves investigating the presence of more homogeneous subtypes among affected toddlers. Only a few studies have focused on identifying phenotypically less variable subgroups in toddlers with ASD. Wiggins et al. (2011) clustered a group of 2-year olds ($N = 186$) with ASD based on items from the Childhood Autism Rating Scale (Schopler & Reichler, 1980) and found three clusters varying in overall symptom levels (i.e., 'mild,' 'moderate,' and 'severe' impairment). Georgiades et al. (2013) also found three clusters in 3-year olds ($N = 391$) varying in levels of impairments across symptom domains based on parent report (Autism Diagnostic Interview Revised; Rutter, Le Couteur, & Lord, 2003). These studies focused their clustering analyses on autism symptoms only. To our best knowledge there have been no studies examining subgroups of toddlers with ASD based on multiple features. However, a study by Stevens et al. (2000) followed 138 preschoolers with ASD to school age and identified clusters based on measures of language, nonverbal intelligence, and normal and abnormal social behavior. In this study, high- and low-functioning subgroups were identified who continued to show distinct social, adaptive, and developmental outcomes at school age. Therefore,

---

Conflicts of interest statement: No conflicts declared.

considering that levels of social, adaptive, and developmental functioning are often interdependent, another highly informative approach to examining heterogeneity in ASD is to identify subtypes based on a broader constellation of behavioral characteristics.

The stability of an ASD diagnosis is typically high, ranging from 84% to 100% in 2–3-year olds (e.g., Chawarska, Klin, Paul, & Volkmar, 2007; Chawarska et al., 2009; Guthrie et al., 2013) and from 85% to 89% in children aged 5 years or older (see Woolfenden, Sarkozy, Ridley, & Williams, 2012 for a review). There is evidence that, for some children with ASD, symptoms lessen over time such that the children may cease to meet diagnostic criteria for ASD by school-age (Fein et al., 2013). Given recent advances in early detection and improvements in early intervention strategies, it is not clear whether the proportion of children who no longer meet ASD criteria later during school-age has increased in recent years. Assuming that clinically meaningful subtypes exist among toddlers with ASD, it is possible that diagnostic stability may vary depending on the clinical presentation of affected children in the second year of life. However, there have been no direct investigations into the diagnostic stability among different subtypes within ASD.

In sum, several questions regarding heterogeneity in diagnostic and clinical outcomes in toddlers with ASD remain to be investigated. First, considering that studies have mainly focused on identifying subgroups based on a single domain (i.e., autism symptoms), we expand our investigation by considering not only symptom severity but also levels of cognitive and adaptive functioning. Moreover, toddlers in the most recent studies examining diagnostic and clinical outcomes were either born between 1997 and 2006 (Guthrie et al., 2013) or referred between 2001 and 2006 (Chawarska et al., 2009). Here, we focus our inquiry on the cohort of toddlers diagnosed with ASD more recently, between 2006 and 2012, a period that has been marked by significant increases in the awareness of early diagnosis, as well as major improvements in early intervention policies in our capture area (Connecticut, New York, New Jersey, and Massachusetts). By focusing on this particular cohort, we address the question of phenotypic variability and developmental dynamics in toddlers who were diagnosed with ASD in the very early stages of the disorder and who had access to early intensive treatment. Toward these ends, we examine patterns of phenotypic variability among 100 toddlers presenting with ASD symptoms between 14 and 27 months who were followed prospectively to the age of 3 years for a confirmatory diagnostic assessment. The main objectives of the study are to: (1) evaluate diagnostic stability in the whole sample ($N = 100$); (2) identify more homogeneous subgroups within ASD among a subset of toddlers who received an initial diagnosis of ASD ($n = 95$) based on a constellation of key clinical features in the second year of life; (3) examine cluster characteristics with regard to demographic characteristics and intervention history; and (4) examine stability of the diagnostic and clinical presentation in the identified clusters.

Methods

This study was approved by the Human Investigations Committee of Yale University School of Medicine, New Haven, Connecticut; written informed consent was obtained from all parents. Assessments were conducted at the Toddler Developmental Disabilities Clinic at the Yale Child Study Center.

Participants

One hundred toddlers (84 males) were referred by parents or professionals to a university-based clinic or research study in an urban setting between 2006 and 2012, due to concerns regarding their cognitive, language, or social development. All toddlers were first evaluated between 14 and 27 months (Time 1), and were re-evaluated between 30 and 49 months (Time 2). There was notable heterogeneity in early clinical presentations in our sample (Table 1). To be included in the present study, a toddler had to receive an ASD diagnosis either at Time 1 or 2, or at both times. Participants were primarily drawn from the northeastern states of the United States: Connecticut, New Jersey, New York, and Massachusetts. In 79% of cases, parents reported Caucasian racial background, 5% Asian, 5% African American, and 7% mixed racial heritage; 4% did not provide race information. Children of Hispanic origin constituted 6% of the sample. Mothers were, on average, 36 years of age ($SD = 5.6$) and fathers 38 years ($SD = 6.1$). Around 80% and 71% of mothers and fathers, respectively, had a 4-year college degree.

Procedures

All children underwent a comprehensive assessment conducted by a multidisciplinary team specializing in autism and other developmental disorders, consisting of measures of autism symptoms along with language, cognitive, and adaptive skills.

Table 1 Sample characterization at Time 1 and Time 2

	Time 1	Time 2
N (male:female)	100 (84:16)	100 (84:16)
Age	21.72 (2.83)	37.14 (3.72)
ADOS Comparison Scores	6.92 (2.14)	7.20 (2.09)
Verbal DQ	51.25 (27.21)	77.56 (36.09)
% with VDQ > 70	39%	63%
Nonverbal DQ	80.04 (18.71)	80.27 (23.22)
% with NVDQ > 70	74%	68%
Vineland: Communication	76.09 (14.17)	85.71 (18.18)
Vineland: Daily Living	80.62 (11.94)	82.38 (14.07)
Vineland: Socialization	78.63 (7.58)	79.08 (11.56)

	N	Range	*M* (*SD*)
Intensity of Total Intervention (hr/week)	99	0.9–44	17.44 (9.22)
Educational	97	0.3–40	14.08 (8.40)
Speech/Language	95	0.3–12	2.07 (1.50)
Occupational/Physical	92	0.3–15	1.79 (1.73)

ADOS, Autism Diagnostic Observation Schedule.

Assessment procedures. Autism symptom severity was evaluated with the ADOS-G (Lord et al., 2000), using Module 1 ($N = 100$) at Time 1 and either Module 1 ($n = 52$) or Module 2 ($n = 48$) at Time 2, depending on language level. To facilitate comparisons of symptom severity across modules, comparison scores (CS) for both algorithm totals (Gotham, Pickles, & Lord, 2009) and domain scores under the algorithm (Social Affect [SA] and Restricted and Repetitive Behaviors [RRBs]; Esler et al., 2015; Hus, Gotham, & Lord, 2014) were computed based on ADOS algorithm scores. Verbal and nonverbal developmental skills were assessed using the Mullen Scales of Early Learning (MSEL; Mullen, 1995), which captured Fine Motor (FM), Visual Reception (VR), Receptive Language (RL), and Expressive Language (EL) skills and age equivalents (AE). Developmental quotient (DQ) scores were computed based on age equivalents from FM and VR scales for nonverbal DQ and from RL and EL scales for verbal DQ ([average AE scores from the two scales/chronological age]*100). Levels of adaptive functioning in the areas of Socialization, Communication, and Daily Living were quantified with the Vineland Adaptive Behavioral Scales, 2nd edition (VABS-II; Sparrow, Cicchetti, & Balla, 2005) using standard scores ($M = 100$, $SD = 15$).

Clinical best estimate diagnosis. At both time points, a clinical best estimate diagnosis was assigned by at least two expert clinicians based on direct clinical assessment of autism symptoms (ADOS-G), verbal and nonverbal skills (MSEL), and adaptive functioning (VABS-II), as well as detailed information regarding medical and developmental history. Clinicians made independent diagnostic judgments, and a full consensus among all clinicians was necessary for the diagnostic assignment (Chawarska et al., 2009). All examiners had previously established reliability with the ADOS training site and with each other.

Intervention. At Time 1, all toddlers who received a provisional diagnosis of ASD or other developmental disorders were referred to their local early intervention agencies for treatment. Information regarding the type (i.e., educational, speech and language, occupational and physical therapies) and duration of intervention received between Times 1 and 2 were gathered at the follow-up visit via face-to-face parent interview by clinical social workers. Speech and language as well as educational therapies were typically delivered using either behavioral (e.g., Applied Behavioral Analysis) or developmental (e.g., Floor Time) approaches or both. The hours of each intervention type were then divided by the weeks elapsed between Times 1 and 2 to obtain the average intensity of intervention (hr/week).

Statistical analysis

To identify subgroups among the 95 toddlers who received an ASD diagnosis at Time 1, we applied Hierarchical Clustering (HC) analysis to the Euclidean distances between each pair of subjects with Ward's method as linkage criterion to the following set of variables from Time 1: ADOS Social Affect and Restricted and Repetitive domain scores, Mullen VR, FM, RL, and EL DQ scores, and Vineland Socialization, Communication, and Daily Living standard scores. All scores were standardized into Z-scores. The optimal number of clusters was chosen by the location of an 'elbow' in the cluster tree of height differences on the y-axis versus cluster number on the x-axis (Tibshirani, Walther, & Hastie, 2001). Cluster characteristics with regard to clinical presentations and demographic information at Time 1 and intervention intensity between Times 1 and 2 were evaluated using a series of ANOVAs with Bonferroni correction for multiple comparisons. Subsequently, we evaluated the stability of diagnosis over time within the identified clusters, as well as changes in symptom severity,

verbal and nonverbal functioning, and adaptive functioning using Generalized Linear Mixed Models with an unstructured covariance matrix, while controlling for the intensity of intervention. Cohen's ds were calculated to examine the magnitudes of changes in these domains using the raw means and standard deviations. Data analysis was implemented in SPSS 19.

Results
Preliminary analyses

At Time 1, 95 toddlers referred for a differential diagnosis were diagnosed with ASD. An additional three toddlers were diagnosed with global developmental delay, one with language delay, and one with reactive attachment disorder (Table 2); all five toddlers were diagnosed with ASD at Time 2. At Time 2, 6 (6.9%) of 95 toddlers initially diagnosed with ASD no longer met diagnostic criteria for ASD. Five of these six toddlers received another diagnosis (e.g., language delays or global developmental delays), and one did not meet criteria for any disorder, although clinicians noted continued subthreshold vulnerabilities in social communication. The overall stability of the ASD diagnosis was 93%.

Cluster analysis

The HC analysis identified four clusters among the 95 toddlers who received the ASD diagnosis at Time 1 (Fig. 1, Table 3). There were no significant differences in the distribution of gender (χ^2 (3) = .398, p = .94), multiplex status (χ^2 (3) = 3.197, p = .36), gestational age ($F_{(3,89)}$ = 2.00, p = .92), maternal ($F_{(3,91)}$ = .953, p = .42) or paternal age ($F_{(3,89)}$ = 1.65, p = .18), or age at the time of first diagnosis ($F_{(3,91)}$ = .179, p = .91) (See Table S1).

As shown in Table 3, Clusters 1 (36% of the sample) and 2 (16%) were similar in terms of verbal and nonverbal skills. Their nonverbal skills were largely in the average range, but their verbal skills were significantly delayed at Time 1. Their levels of restrictive and repetitive behaviors (RRB) were comparable. However, Cluster 2 had less severe Social Affective (SA) symptoms and lower adaptive skills than Cluster 1. Clusters 3 (31%) and 4 (17%) had SA symptoms more severe than both Clusters 1 and 2, but the differences in the RRB levels were less

Table 2 Stability of the provisional best estimate clinical diagnosis from Time 1 to Time 2

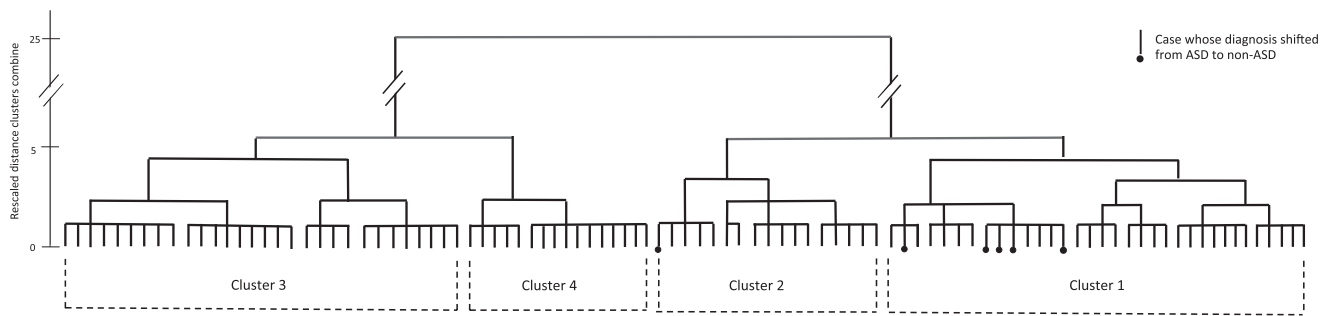| Provisional diagnosis at Time 1	Confirmatory diagnosis at Time 2		
	ASD	Non-ASD	Total
ASD	89	6	95
Non-ASD	5	0	5
Total	94	6	100

Figure 1 Dendrogram for clustering analysis

Table 3 Descriptive statistics for the four clusters identified at Time 1

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
N (%)	34 (36%)	15 (16%)	30 (31%)	16 (17%)
ADOS:SA	12.41 (3.48)[a]	9.6 (2.5)[b]	16.97 (1.8)[c]	18.31 (2.20)[c]
ADOS:RRB	3.68 (16.43)[a]	3.80 (1.74)[a]	5.27 (1.55)[b]	4.88 (1.78)[a]
Mullen: VRDQ	89.11 (16.43)[a]	88.97 (26.17)[a]	72.69 (12.60)[b]	49.64 (9.0)[c]
Mullen: FMDQ	90.98 (11.84)[a]	95.95 (22.05)[a]	81.60 (10.81)[b]	53.76 (8.18)[c]
Mullen: RLDQ	73.15 (29.82)[a]	70.99 (31.90)[a]	27.49 (13.93)[b]	11.87 (10.47)[c]
Mullen: ELDQ	72.12 (22.89)[a]	60.26 (21.86)[a]	38.73 (11.41)[b]	27.75 (10.61)[b]
Vineland: Communication	88.38 (10.34)[a]	77.67 (8.92)[b]	66.13 (11.41)[c]	62.75 (4.31)[c]
Vineland: Daily Living	91.44 (10.31)[a]	76.33 (4.14)[b]	77.50 (7.47)[b]	66.44 (7.55)[c]
Vineland: Socialization	85.41 (6.68)[a]	74.27 (4.79)[b]	76.17 (5.1)[b]	72.19 (4.56)[b]
Intensity of Intervention (hrs/week)				
Educational	12.23 (7.78)[a]	14.00 (7.41)[ab]	14.06 (7.28)[ab]	19.94 (10.37)[b]
Speech/Language	1.85 (1.19)	2.26 (1.06)	1.73 (0.98)	2.93 (2.74)
Occupational/Physical	2.01 (2.64)	1.95 (1.31)	1.35 (0.85)	1.98 (1.08)

SA, social affect, RRBs, restricted and repetitive behaviors; VR, visual reception; FM, fine motor; RL, receptive language; EL, expressive language, Means with different superscripts within each row differ significantly; All comparisons with Bonferroni correction for multiple comparisons.

consistent. Clusters 3 and 4 were also lower functioning with regard to verbal and nonverbal skills than either Clusters 1 or 2. In comparison to Cluster 3, Cluster 4 had lower nonverbal skills and very limited ability to understand and respond to language. Cluster 4 also had very delayed adaptive daily living and socialization skills. Thus, Clusters 1 and 4 appeared to represent the classic extremes of the 'autism spectrum,' whereas Clusters 2 and 3 represented children with more intermediate skill levels, distinguished largely by their levels of social impairment and verbal and nonverbal skills.

Clusters also differed significantly in the intensity of the educational services ($F_{(3,88)} = 3.310$, $p < .05$), such that Cluster 4 received significantly more hours of educational services than Cluster 1 ($p < .05$) (Table 3). No significant differences in intervention intensity were observed for speech and language therapy ($F_{(3,90)} = 1.753$, $p = .16$) or occupational and physical therapy ($F_{(3,83)} = .831$, $p = .48$).

Stability of ASD diagnosis within clusters

The stability of ASD diagnosis differed by cluster (See Table S1). In the more severely affected clusters (Cluster 3 and 4), the stability was 100%. In Cluster 2, the stability was 93%; one of the six children who

did not retain the ASD diagnosis fell into this cluster. In Cluster 1, the stability was the lowest at 85%; five of the six children who did not retain the ASD diagnosis fell into Cluster 1.

Stability of clinical presentation within clusters

Social impairment and socialization skills. The GLMM analysis on symptom severity (ADOS Comparison Scores for Algorithm Totals) indicated a significant effect of a cluster × time interaction ($p < .001$) (Fig. 2, Table 4). Post hoc comparisons indicated that symptom severity was stable in Clusters 1, 3, and 4, but in Cluster 2, symptoms increased over time. Analyses examining the ADOS domain comparison scores also revealed that symptoms in both Social Affect and RRB domains increased over time for Cluster 2 (from $M = 6$, $SD = 2$ to $M = 7$, $SD = 2$ for Social Affect; from $M = 7$, $SD = 2$ to $M = 8$, $SD = 1$ for RRBs; $ps < .05$). Similarly, there were significant effects of cluster ($p < .001$) and a cluster x time interaction ($p < .001$) for adaptive social skills. Cluster 1 showed the highest scores and Cluster 4 the lowest scores at both times. Clusters 2 and 3 did not differ from one another. Cluster 2 showed a significant increase in the adaptive social skills over time, whereas Cluster 4
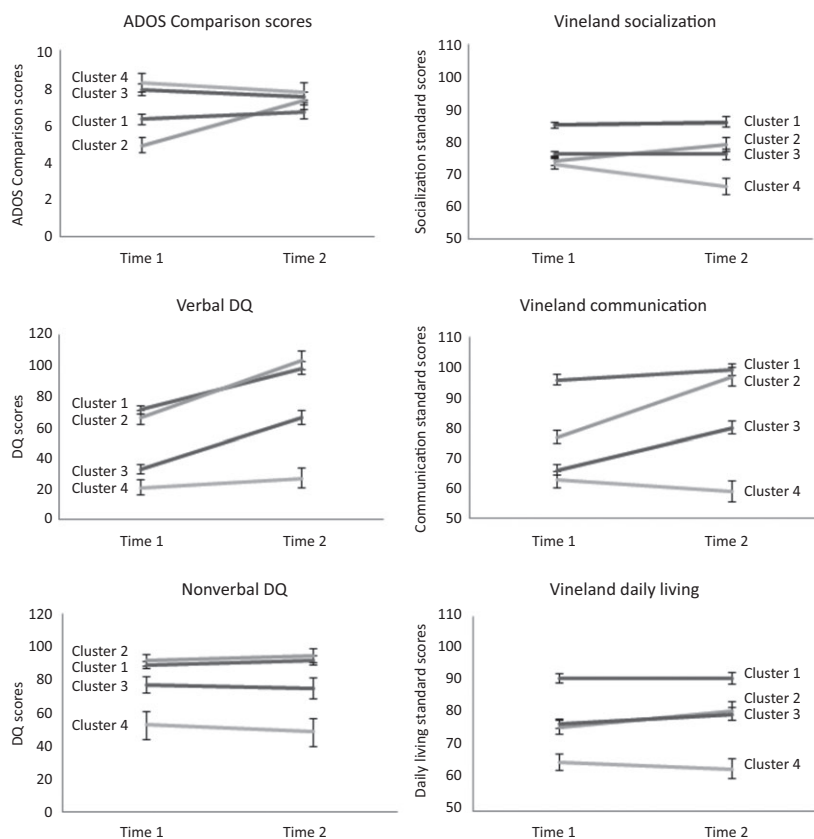
Figure 2 Least square means (±2 Standard Errors) for ADOS Comparison Scores, Mullen Verbal and Nonverbal DQ, and Vineland Socialization, Communication, and Daily Living Scores at Time 1 and Time 2

showed a significant decrease in these skills. The scores for the other clusters remained stable.

Verbal and communication skills. For VDQ and adaptive communication, there were significant effects of cluster ($p < .001$), time ($p < .001$), and a cluster × time interaction ($p < .01$ and $p < .001$ for VDQ and adaptive communication respectively) (Fig. 2, Table 4). Clusters 1 and 2 had higher verbal and adaptive communication scores than Cluster 3, which in turn had higher verbal scores than Cluster 4. All clusters but one (Cluster 4) showed a significant improvement in these domains.

Nonverbal and daily living skills. With regard to both NVDQ and adaptive daily living skills, there was a significant effect of cluster ($p < .001$). Clusters 1 and 2 had the highest NVDQ scores, followed by Clusters 3 and 4. Vineland Daily Living scores were significantly higher for Cluster 1, followed by Clusters 2 and 3, which showed similar levels of skills; Cluster 4 showed the lowest scores ($p < .001$). The scores for all clusters remained stable over time (Fig. 2, Table 4).

Magnitude of change over time

Cluster 2 showed the most notable improvements, with moderate to large effect sizes in most areas of

skills and Cohen's ds ranging from $d = 0.9$ (adaptive socialization) to $d = 2.5$ (adaptive communication), except for NVDQ ($d = 0.1$) (Table 5). However, Cluster 4 showed stable or worsening scores in most domains, with effect sizes ranging from $d = 0$ (symptom severity) to $d = -1.8$ (adaptive socialization). Clusters 1 and 3 showed improvements in VDQ and adaptive communication, with moderate to large effect sizes ranging from $d = 0.8$ to $d = 1.6$, while maintaining stability in NVDQ and adaptive daily living and socialization skills, with small effect sizes (ds = 0–0.3).

Discussion

This study aimed to identify phenotypically homogeneous subgroups among 14–27 month old toddlers presenting with ASD symptoms and to evaluate their outcomes within 1–2 years. Four clusters were identified based on autism symptom severity, nonverbal and verbal skills, and adaptive functioning. Similar to other studies (e.g., Georgiades et al., 2013; Wiggins et al., 2011), we found both high- (Cluster 1) and low-functioning (Cluster 4) groups of toddlers with ASD. These groups constituted 53% of the sample. The remaining children constituted two intermediate subtypes (Clusters 2 and 3) characterized by specific combinations of relative strengths and deficits.

Table 4 Stability and change in ASD symptoms, nonverbal and verbal skills and adaptive functioning from Time 1 to Time 2

	F	Sig	Post hoc comparisons
ADOS Comparison Scores			
Cluster	1.894	.132	
Time	3.642	.058	
Cluster*Time	6.805	<.001	T1 < T2 for C2
Intervention	0.353	.563	
Cluster*Intervention	0.282	.838	
Vineland: Social			
Cluster	6.257	<.001	C1 > C2 = C3 > C4
Time	0.050	.824	
Cluster*Time	4.986	<.01	T1 < T2 for C2, T1 > T2 for C4
Intervention	3.200	.075	
Cluster*Intervention	0.395	.757	
VDQ			
Cluster	14.180	<.001	C1 = C2 > C3 > C4
Time	97.561	<.001	T1 < T2
Cluster*Time	5.975	<.01	T1 < T2 for C1, C2, C3
Intervention	0.018	.894	
Cluster*Intervention	1.418	.239	
Vineland: Communication			
Cluster	12.233	<.001	C1 = C2 > C3 > C4
Time	51.571	<.001	T1 < T2
Cluster*Time	11.152	<.001	T1 < T2 for C1, C2, C3
Intervention	2.876	.092	
Cluster*Intervention	0.331	.803	
NVDQ			
Cluster	8.020	<.001	C1 = C2 > C3 > C4
Time	0.018	.894	
Cluster*Time	1.221	.304	
Intervention	1.844	.176	
Cluster*Intervention	0.077	.972	
Vineland: Daily Living			
Cluster	8.524	<.001	C1 > C2 = C3 > C4
Time	1.955	.164	
Cluster*Time	1.568	.199	
Intervention	0.514	.474	
Cluster*Intervention	0.924	.430	

Table 5 Mean (SD) and Cohen's ds for social, developmental, and adaptive functioning at Time 1 and Time 2

	Cluster 1			Cluster 2			Cluster 3			Cluster 4		
	T1	T2	d	T1	T2	d	T1	T2	d	T1	T2	d
ADOS Comparison Scores	6 (1)	6 (2)	0	5 (1)	7 (2)	1.3	8 (1)	7 (1)	−1	7 (1)	7 (1)	0
Mullen NVDQ	90 (11)	93 (16)	0.2	92 (23)	95 (22)	0.1	77 (8)	75 (17)	−0.2	51 (6)	48 (10)	−0.4
Mullen VDQ	72 (22)	99 (24)	1.2	65 (19)	102 (20)	1.9	33 (9)	65 (27)	1.6	19 (8)	25 (15)	0.5
Vineland: Communication	88 (10)	96 (10)	0.8	77 (8)	97 (8)	2.5	66 (11)	80 (16)	1.0	62 (4)	58 (9)	−0.6
Vineland: Daily Living	91 (10)	91 (12)	0	76 (4)	81 (6)	1	77 (7)	80 (11)	0.3	66 (7)	64 (8)	−0.3
Vineland: Socialization	85 (6)	86 (9)	0.1	74 (4)	80 (9)	0.9	76 (5)	76 (10)	0	72 (4)	65 (4)	−1.8

Despite early detection and access to intensive community-based interventions, the stability of ASD diagnosis from age two to three was very high at 93%. This is consistent with reports based on comparable clinic-referred samples (Chawarska et al., 2009; Guthrie et al., 2013). Importantly, however, the stability of ASD diagnosis varied by cluster, i.e., was dependent on the clinical presentation of the affected children in the second year of life. All children who initially presented with severe impairments in social communication, marked delays in language understanding and use, and severe to moderate delays in nonverbal functioning retained the ASD diagnosis (Clusters 3 and 4). Similarly, stability of diagnosis was also high (93%) in toddlers with somewhat less severe verbal and nonverbal delays and social affective deficits (Cluster 2). However, for children who initially presented with the most intact levels of nonverbal and adaptive skills (Cluster 1), the stability was somewhat lower (85%).

As many as 36% of toddlers in our sample fell into the highest functioning group (Cluster 1). Despite moderate impairments in social communication and

the presence of repetitive behaviors these toddlers made significant progress from age 2 to 3 years in the areas of formal language and adaptive social communication. This high-functioning group included most of the toddlers who did not retain an ASD diagnosis at follow-up (n = 5 out of 6), which raises important questions: did these five toddlers respond particularly favorably to early intervention, or did their presentations in the second year reflect complex combinations of developmental challenges that resembled autism but were driven by other underlying factors? The current study was not designed to answer these questions in a definitive manner. We do note, however, that despite intense treatment, the proportion of children in the entire sample who did not retain an ASD diagnosis was modest (about 6%), consistent with previous studies reporting that very few individuals with ASD show marked improvements in autism symptoms over time (Anderson, Oti, Lord, & Welch, 2009; Gotham, Pickles, & Lord, 2012). Furthermore, most of these six children met criteria for other diagnoses (e.g., language delays or global developmental delays), suggesting that vulnerabilities associated with early deficits in social communication or developmental delays persisted over time. Similarly, school-age children who do not retain earlier ASD diagnoses have been found to present with persistent difficulties with pragmatic and semantic language (Kelley, Fein, & Naigles, 2006), as well as attention deficits and poor executive functions (Kelley, Naigles, & Fein, 2010), despite having overall levels of cognitive and adaptive functioning within normal limits (Fein et al., 2013). Therefore, even for toddlers who do not retain a diagnosis of ASD from early years, continued monitoring and provision of appropriate services will be crucial to ameliorate potential residual difficulties and to prevent other secondary or comorbid symptoms.

In contrast to Cluster 1, toddlers in Cluster 4 had severe autism symptoms paired with significant delays in all areas of functioning at both assessment points. Interestingly, this very low-functioning group constituted only about 17% of our sample, suggesting that, due to an increasing awareness of early symptoms of ASD across the spectrum, children referred for a differential diagnosis in the second year of life represent a broad spectrum of symptom severity and that only a small minority of these toddlers present with the highest degree of impairment. Although children in this most affected group received about 25 hr of intervention services per week, equivalent to the amount of intervention recommended by the National Research Council (2001), their gains in the targeted areas were limited. Consistent with other findings (Flanagan et al., 2015) these findings suggest that early cognitive and adaptive skill levels play an important role in predicting later outcomes among toddlers with ASD.

A more in-depth examination of the mediators and moderators of treatment effects, especially for children who make the least progress in response to early intervention, is necessary to address the differential needs within the autism spectrum (Sherer & Schreibman, 2005). These results also highlight the importance of identifying more effective treatment approaches tailored to unique developmental profiles and dynamics. In addition, Cluster 4 also needs to be contrasted with Cluster 3, which had comparable levels of social affective symptoms at 2 years, but was higher functioning with respect to verbal, nonverbal, and adaptive skills. Unlike toddlers in Cluster 4, those in Cluster 3 showed a significant increase in verbal and communication skills.

Finally, approximately 16% of children (Cluster 2) were also relatively high-functioning with regard to cognitive and adaptive skills, but their initially mild autism symptoms worsened over time. This pattern has been also reported in other prospectively ascertained samples of multiplex cases (Ozonoff et al., 2015). Considering that for these toddlers the full clinical picture did not emerge until the third year, continued monitoring and repeated diagnostic screening even in children with initially mild symptoms is essential.

Taken together, as many as 83% of toddlers (Clusters 1, 2, & 3) showed either stable or worsening autism symptoms but large gains in verbal skills and adaptive communication, suggesting a potential dissociation between the developmental dynamics of autism symptoms and of language and communication skills. Such limited 'yoking' of the developmental trajectories of autism symptom severity and adaptive functioning has also been found in older preschool children with ASD followed from 3 to 6 years of age (Szatmari et al., 2015). This may indicate that the factors and mechanisms facilitating the development of cognitive and adaptive functioning are not exactly the same as those mediating how autism symptomatology unfolds over time. In contrast, Cluster 4 showed a stronger 'yoking' between autism symptom severity and other domains of functioning as severe delays and abnormalities in these domains were either relatively stable over time or showed slight worsening. These results have important clinical implications; the variability of associations between the trajectories of different domains of functioning we observed is not well-captured under the current diagnostic framework. Although the current DSM-5 now allows clinicians and researchers to indicate severity levels for autism symptoms in order to determine the level of support needed for each individual, these levels are merely determined by autism symptomatology. Therefore, the severity level scales also based on verbal, cognitive, and adaptive functioning will be useful to determine the degrees of support needed in these areas of functioning.

In three of four clusters (~80% of the sample), gains with regard to language development and adaptive communication over the course of 1–2 years were large, with effect sizes ranging from 0.8 to 2.5. The magnitude of these effects is comparable to those reported in toddlers with ASD receiving intensive, manualized, and systematic intervention programs delivered in randomized control studies (e.g., Dawson et al., 2010; Kasari, Paparella, Freeman, & Jahromi, 2008; Wetherby et al., 2014). Given the design of our study, it is not clear what factors may have contributed to such large gains in the verbal and adaptive communication domains. Specifically, it is not clear what proportion of variance might be accounted for by the child's inherent characteristics and maturation, environmental factors (e.g., family environment), or treatment effects. It is plausible that verbal functioning in young children is more malleable than other (e.g., social or nonverbal) skills or that the existing intervention strategies may be most effective in fostering skills in these particular areas of functioning. These not necessarily exclusive hypotheses will have to be tested directly through future studies.

Limitations and future directions

Although we were able to collect detailed information about the types and intensity of intervention, the quality of services was not captured by our metrics. Given the generally eclectic approaches to early intervention, a rigorous comparison between different treatment approaches was not possible. Longer term follow-up with this cohort will be necessary to evaluate developmental trajectories with regard to core and comorbid symptoms. More importantly, given better prognosis for toddlers whose cognitive and adaptive skills were intact early on despite pronounced symptom severity, examining the clinical features that may mediate response to treatment will strengthen the development of more effective treatment approaches. A replication and extension of this study with more diverse samples with regard to socioeconomic and racial characteristics is warranted.

Conclusions

Subtyping based on autism symptom severity, verbal and nonverbal skills, and adaptive functioning identified four groups of toddlers characterized by different levels of impairments across domains and rates of progress over time. Despite very early identification and implementation of intensive early intervention, a diagnosis of ASD remained highly stable from the second into third year of life in 95 clinic-referred toddlers. Relatedly, severity of autism symptoms was in general stable within the identified subgroups, except for 16% of toddlers who experienced worsening of the symptoms between the ages of 2 to 3 years. Although highly variable across the subgroups, the levels of nonverbal skills and adaptive daily living skills were also very stable over time. Greatest gains over time were noted in the areas of verbal skills and adaptive communication, with a magnitude commensurate with that observed in controlled treatment studies. Only a small minority (17%) of children exhibited very limited gains despite intensive intervention. These findings support continued efforts to examine determinants of developmental trajectories in young children with ASD including factors mediating and moderating their response to treatment.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1. Demographic characteristics among clusters at Time 1.

Correspondence

So Hyun Kim, Yale University School of Medicine, 40 Temple St. Suite 7-D, New Haven, CT 06510, USA; Email: so-hyun.kim@yale.edu

Key points

- Phenotypic heterogeneity among toddlers presenting ASD symptoms complicates predictions regarding their later diagnostic and clinical outcomes, pointing to the importance of identifying more homogeneous subgroups in ASD.
- Based on early clinical presentations at 2 years of age, we identified four subgroups with varying levels of autism symptoms and cognitive and adaptive skills as well as intervention utilization.
- Stability of the diagnosis was high at 93% from the second to the third year of life, but stability varied depending on the early clinical presentations, with the least affected children showing lower stability at 85%.
- Many children were able to maintain or improve their initial average or below average developmental and adaptive skills with intervention, but a small group of children showed very limited progress in all areas of development.
- Continued efforts to identify subgroups within very young children with ASD will be crucial in order to capture heterogeneity in developmental trajectories, response to treatment, and possible etiological markers.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM–5®)* (5th edn). Washington, DC: Author.

Anderson, D.K., Oti, R.S., Lord, C., & Welch, K. (2009). Patterns of growth in adaptive social abilities among children with autism spectrum disorders. *Journal of Abnormal Child Psychology*, 37, 1019–1034.

Campbell, D.J., Shic, F., Macari, S., & Chawarska, K. (2014). Gaze response to dyadic bids at 2 years related to outcomes at 3 years in autism spectrum disorders: A subtyping analysis. *Journal of Autism and Developmental Disorders*, 44, 431–442.

Chawarska, K., Klin, A., Paul, R., Macari, S., & Volkmar, F. (2009). A prospective study of toddlers with ASD: Short-term diagnostic and cognitive outcomes. *Journal of Child Psychology and Psychiatry*, 50, 1235–1245.

Chawarska, K., Klin, A., Paul, R., & Volkmar, F. (2007). Autism spectrum disorder in the second year: Stability and change in syndrome expression. *Journal of Child Psychology and Psychiatry*, 48, 128–138.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., . . . & Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics*, 125, e17–e23.

Esler, A.N., Bal, V.H., Guthrie, W., Wetherby, A., Weismer, S.E., & Lord, C. (2015). The autism diagnostic observation schedule, toddler module: Standardized Severity Scores. *Journal of Autism and Developmental Disorders*, 1–17. Advanced online publication. doi: 10.1007/s10803-015-2432-7.

Fein, D., Barton, M., Eigsti, I.M., Kelley, E., Naigles, L., Schultz, R.T., . . . & Tyson, K. (2013). Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry*, 54, 195–205.

Flanagan, H.E., Smith, I.M., Vaillancourt, T., Duku, E., Szatmari, P., Bryson, S., . . . & Georgiades, S. (2015). Stability and change in the cognitive and adaptive behaviour scores of preschoolers with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 1–13. Advanced online publication. doi: 10.1007/s10803-015-2433-6.

Georgiades, S., Szatmari, P., Boyle, M., Hanna, S., Duku, E., Zwaigenbaum, L., . . . & Thompson, A. (2013). Investigating phenotypic heterogeneity in children with autism spectrum disorder: A factor mixture modeling approach. *Journal of Child Psychology and Psychiatry*, 54, 206–215.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39, 693–705.

Gotham, K., Pickles, A., & Lord, C. (2012). Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics*, 130, 1278–1284.

Guthrie, W., Swineford, L.B., Nottke, C., & Wetherby, A.M. (2013). Early diagnosis of autism spectrum disorder: Stability and change in clinical diagnosis and symptom presentation. *Journal of Child Psychology and Psychiatry*, 54, 582–590.

Hus, V., Gotham, K., & Lord, C. (2014). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders*, 44, 2400–2412.

Kasari, C., Paparella, T., Freeman, S., & Jahromi, L.B. (2008). Language outcome in autism: Randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology*, 76, 125.

Kelley, E., Fein, D., & Naigles, L. (2006). Residual language deficits in optimal outcome children with a history of autism. *Journal of Autism and Developmental Disorders*, 36, 807–828.

Kelley, E., Naigles, L., & Fein, D. (2010). An in-depth examination of optimal outcome children with a history of Autism Spectrum Disorders. *Research in Autism Spectrum Disorders*, 4, 526–538.

Kim, S.H., & Lord, C. (2010). Restricted and repetitive behaviors in toddlers and preschoolers with autism spectrum disorders based on the autism diagnostic observation schedule (ADOS). *Autism Research*, 3, 162–173.

Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Jr, Leventhal, B.L., DiLavore, P.C., . . . & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30, 205–223.

Mullen, E.M. (1995). *Mullen Scales of early learning*. San Antonio, TX: Pearson.

Munson, J., Dawson, G., Sterling, L., Beauchaine, T., Zhou, A., Koehler, E., . . . & Abbott, R. (2008). Evidence for latent classes of IQ in young children with autism spectrum disorder. *Journal Information*, 113, 439–452.

National Research Council. (2001). Goals for educational services. In C. Lord, & J.P. McGee (Eds.), *Educating children with autism*. Washington, DC: National Academy Press.

Ozonoff, S., Young, G.S., Landa, R.J., Brian, J., Bryson, S., Charman, T., . . . & Iosif, A.M. (2015). Diagnostic stability in

young children at risk for autism spectrum disorder: A baby siblings research consortium study. *Journal of Child Psychology and Psychiatry*. Advanced online publication. doi: 10.1111/jcpp.12421.

Ray-Subramanian, C.E., Huai, N., & Weismer, S.E. (2011). Brief report: Adaptive behavior and cognitive skills for toddlers on the autism spectrum. *Journal of Autism and Developmental Disorders*, *41*, 679–684.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism diagnostic interview-revised*. Los Angeles, CA: Western Psychological Services.

Schopler, E., & Reichler, J. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders*, *10*, 91–103.

Sherer, M.R., & Schreibman, L. (2005). Individual behavioral profiles and predictors of treatment effectiveness for children with autism. *Journal of Consulting and Clinical Psychology*, *73*, 525–538.

Sparrow, S., Cicchetti, D., & Balla, D. (2005). *Vineland-II. Vineland Adaptive Behavior Scales*. Survey forms manual. Minneapolis, MN: NCS Pearson Inc.

Stevens, M.C., Fein, D.A., Dunn, M., Allen, D., Waterhouse, L.H., Feinstein, C., & Rapin, I. (2000). Subgroups of children with autism by cluster analysis: A longitudinal examination. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*, 346–352.

Szatmari, P., Georgiades, S., Duku, E., Bennett, T.A., Bryson, S., Fombonne, E., ... & Thompson, A. (2015). Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. *JAMA Psychiatry*, *72*, 276–283.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*, 411–423.

Wetherby, A.M., Guthrie, W., Woods, J., Schatschneider, C., Holland, R.D., Morgan, L., & Lord, C. (2014). Parent-implemented social intervention for toddlers with autism: An RCT. *Pediatrics*, *134*, 1084–1093.

Wiggins, L.D., Robins, D.L., Adamson, L.B., Bakeman, R., & Henrich, C. (2011). Support for a dimensional view of autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders*, *42*, 191–200.

Woolfenden, S., Sarkozy, V., Ridley, G., & Williams, K. (2012). A systematic review of the diagnostic stability of autism spectrum disorder. *Research in Autism Spectrum Disorders*, *6*, 345–354.