CrossMark

ORIGINAL PAPER

Classifying Autism Spectrum Disorders by ADI-R: Subtypes or Severity Gradient?

Hannah Cholemkery[1] · Juliane Medda[1] · Thomas Lempp[1] · Christine M. Freitag[1]

Abstract To reduce phenotypic heterogeneity of Autism spectrum disorders (ASD) and add to the current diagnostic discussion this study aimed at identifying clinically meaningful ASD subgroups. Cluster analyses were used to describe empirically derived groups based on the Autism Diagnostic Interview-revised (ADI-R) in a large sample of n = 463 individuals with ASD aged 3–21. Three clusters were observed. Most severely affected individuals regarding all core symptoms were allocated to cluster 2. Cluster 3 comprised moderate symptom severity of social communication impairments (SCI) and less stereotyped repetitive behavior (RRB). Minor SCI and relatively more RRB characterized cluster 1. This study offers support for both, a symptom profile, and a gradient model of ASD within the spectrum due to the sample included.

Keywords Autism Diagnostic Interview-revised · Psychometric assessment · Diagnosis · Autism spectrum disorder · Cluster analysis · DSM-5

Introduction

Autism spectrum disorder (ASD) comprises a phenotypically and genetically heterogeneous group of disorders. Phenotypic aspects indicating heterogeneity are, among others, age of onset, severity and combination of symptoms, as well as language and cognitive development. The heterogeneity has often been offered as explanation for the difficulties in replicating specific genetic risk factors involved in autism (Hus et al. 2007; Hu and Steinberg 2009; Hu and Lai 2013). Consequently, numerous attempts to stratify children into distinct subgroups have been done. Several researchers already aimed at eliciting distinct underlying genetic etiologies by subgrouping the ASD phenotype (Veatch et al. 2014, Hu and Steinberg 2009; Hu and Lai 2013). Yet it has not become clear whether a categorical or a dimensional view of ASD is more appropriate to delineate etiology, diagnostic, and treatment options for children and adolescents with ASD (Wiggins et al. 2011). It is commonly expected, that the delineation of phenotypically distinct ASD subgroups will allow a more comprehensive understanding of the disorder and the individual's needs with regard to diagnostic and treatment options (Brennan et al. 2014).

Empirical, unbiased approaches for exploring underlying substructures of disorders include factor analyses and cluster analyses. Results derived from factor analyses can be interpreted as an association between the observed variables and underlying latent variables (Snow et al. 2008). Recent factor analytic results supported the novel DSM-5 two domain ASD model with the social communication domain on the one hand, and stereotyped repetitive behavior on the other one in a sample of 1861 individuals aged 4–18 years with a mean estimated IQ of 99.4 (SD 19.3) provided by the Autism Genetic Research Exchange program (Snow et al. 2008). In contrast, cluster analytic approaches aim at grouping individuals according to their observed phenotypic similarities and differences. One advantage of both methods is the un-biased analysis allowing researchers to reduce data in an objective manner based on the descriptive, atheoretical, multivariate

✉ Hannah Cholemkery
Hannah.Cholemkery@kgu.de

1 Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, JW Goethe University Hospital, Deutschordenstraße 50, 60528 Frankfurt Am Main, Germany

⚛ Springer

techniques (Hair and Black 2000). Past research using the cluster analytic approach on different ASD symptom measures identified a range of two to four ASD subgroups (Eaves et al. 1994, Sevin et al. 1995, Fein et al. 1999). Some studies described distinct diagnostic categories (Siegel et al. 1986), others a low- and a high functioning group (Fein et al. 1999, Stevens et al. 2000). Limitations of some studies were the relatively small sample sizes. Also, differing ASD populations were studied. Some studies concentrated on high functioning children (Verté et al. 2006; Ring et al. 2008), others included toddlers (Wiggins et al. 2011, Georgiades et al. 2013, Jansen et al. 2013) or pre-pubertal boys only (Obafemi-Ajayi et al. 2014). A broad range of measurement variables were also used to identify distinct cluster groups. While most of the studies included data of one primary instrument (e.g. the Autism Diagnostic Interview-Revised (ADI-R, Rutter et al. 2003), the childhood autism rating scale (CARS, Schopler et al. 1988), or the Social Communication Questionnaire (SCQ, Rutter et al. 2001), some analyzed data from a combination of phenotypic measures (Brennan et al. 2014, Veatch et al. 2014). These methodological differences may have resulted in the heterogeneous findings.

The ADI-R is a comprehensive assessment instrument, which contains items on age of onset, language development, social interaction and communication abilities as well as stereotyped and repetitive behaviors. Thus, it is an ideal measure for studies on the autism phenotype (Snow et al. 2008), and cluster analyses on the ADI-R will likely result in clinically meaningful subgroups (Wiggins et al. 2012).

In a cluster analysis of ADI-R algorithm subscale scores, verbal-nonverbal status, and IQ scores, a three cluster solution was observed in autism affected sib-pairs from n = 171 multiplex families excluding children with severe mental retardation (Spiker et al. 2002). Clusters were characterized by symptom severity (mild, moderate, severe) in all three DSM-IV-TR domains. Limitations of the study are the strict inclusion criteria, presumably not including individuals on the whole spectrum (Asperger Syndrome, pervasive developmental disorders not otherwise specified (PDD-nos). A three-cluster solution was also identified by Verté et al. (2011) in a different sample, including children within the high functioning range (IQ > 80) (Asperger's Syndrome, high functioning autism, PDD-nos) aged 6–13 years old (N = 135). This cluster analysis was based on an agglomerative hierarchical approach on z-standardized ADI-R subscale scores. The cluster groups were defined by one cluster including children with high functioning Autism (HFA), the second cluster combined HFA and Asperger's Syndrome, and the third cluster group mostly comprised individuals with PDD-nos. The authors concluded that these findings

support the severity concept; still cluster groups were described by distinct DSM-IV-TR diagnostic categories. It remains unclear if these results also indicate differing subtypes apart from a severity gradient, or if the diagnostic categories might be based on an underlying severity gradient labeled as categories.

Four phenotypic clusters based on similarity of symptom severity across the ADI-R algorithm items of nearly 2000 autistic individuals from the Autism Genetic Research Exchange (AGRE) were identified by Hu and Steinberg (2009). The sample included a broad range of ages from 1.85 to 47.68 year olds with an estimated nonverbal IQ of 73–140. One cluster was characterized by severe language deficits, another exhibited intermediate severity, the third one included milder symptoms across the domains, and a last one showed a high frequency of savant skills. This study offers support for both, a symptomatic profile of distinct subtypes, and a severity gradient.

With regard to the severity concept proposed in DSM-5 and current efforts to develop ICD-11 ASD criteria, the present study builds on previous attempts by exploring homogenous subgroups within the ASD spectrum. Cluster groups were derived from a large sample including all three ICD-10/DSM-IV-TR diagnoses with a broad age range and an estimated IQ in line with previous studies (e.g. Snow et al. 2008; Hu and Steinberg 2009) in terms of the ADI-R algorithm scores. The study's specific objective was to identify homogeneous and clinically meaningful subgroups. We first performed cluster analysis to describe empirically derived subgroups from the ADI-R algorithm scores. Second, subgroups were carefully characterized by additional phenotypic measures, covering ASD and global psychopathological symptoms.

Methods

Participants

Participants were consecutive referrals receiving a comprehensive diagnostic workup between March 1992 and January 2014 at a Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy in Germany. Individuals with a clinical diagnosis of autistic disorder (F84.0), Asperger's syndrome (F84.5), or atypical autism (F84.1) according to ICD-10 research criteria (WHO 1992) were included into the study. All of the participants were aged ≥3–≤21 years-old. Diagnoses were established by a team of independent and experienced clinicians (clinical psychologists, psychiatrists) all trained to research standards by certified ADOS/ADI-R trainers as clinical best estimate based on information of the Autism Diagnostic Interview-revised (ADI-R; Rühl et al. 2004),

the Autism Diagnostic Observation Schedule (ADOS, Bölte et al. 2006), the Social Responsiveness Scale (Bölte and Poustka 2008), and/or the Social Communication Questionnaire (Bölte and Poustka 2006). Children who scored on the ADI-R as nonverbal, individuals with IQ <35, and those ones who could not be tested were excluded from the study. The final sample with complete ADI-R data consisted of n = 463 individuals, including n = 194 with autistic disorder (41.9 %), n = 113 (24.4 %) with atypical autism (24.4 %), and n = 156 (33.7 %) with Asperger's syndrome.

Measures

The *Autism Diagnostic Interview-revised* is a semi-structured parent/primary caregiver interview on DSM-IV-TR/ICD-10 autism criteria. The ADI-R provides empirically derived diagnostic algorithms for three subdomains of qualitative impairments. The *social interaction* domain includes questions about emotional sharing, offering and seeking comfort, social smiling, and social responding comprising four subscores. The second domain (*communication*) assesses stereotyped expressions, pronoun reversal, and the social usage of language (B1–B4). In the third domain *repetitive and stereotyped behavior* as hand- and finger mannerisms, unusual sensory interest or activities are investigated by four subscales. Furthermore, a score for *abnormality of development evident at/before 36 months* (D) can be calculated from five items. For a description of the specific subscales please see Table 2. The interviewer determines a rating score for each item based on her/his evaluation of the caregiver's response to the semi-structured questions. Scores are defined from 0 ("Behavior of the type specified in the coding is not present"), 1 ("Behavior of the type specified is present in an abnormal form, but not sufficiently severe or frequent to meet the criteria for a 2"), 2 ("Definite abnormal behavior"), 3 ("Extreme severity of the specified behavior"), 7 ("Definite abnormality in the general area of the coding, but not of the type specified"), 8 ("Not applicable"), 9 ("Not known"). For calculating subdomain scores, 3 is recoded to 2, the scores of 7, 8, 9 are recoded to 0. A total score is then calculated for each of the three subdomains. In this study, data of the ever/4–5 diagnostic algorithm were used for children aged ≥4, the current/ever algorithm was used for the sample of children aged 3,0 to 3,11.

The *Autism Diagnostic Observation Schedule* is a semi-structured, cross sectional observation tool, designed for the assessment of ASD children, adolescents, and adults with varying developmental age and language abilities by four different modules. It measures *reciprocal interaction* and *communication impairment*, as well as *repetitive, stereotyped behaviors*. Here, calibrated severity scores for

social affect (SA), restrictive, repetitive behavior (RRB) (Hus et al. 2014; Hus and Lord 2014), and the total severity score (Gotham et al. 2009) were calculated for modules 1–4. The separately calibrated severity scores allow a comparison of ASD severity across the different modules.

Both, ADI-R and ADOS are well-established, and excellently validated diagnostic tools for children and adolescents with ASD (Risi et al. 2006). In this study, both instruments were administered by independent clinical experts (psychologists, psychiatrists) who were trained to research standards.

The *Social Responsiveness Scale* is a 65-item rating scale that measures symptoms indicative of ASD over the previous 6 months in 4- to 18-year-olds. It is a parent/teacher questionnaire. Each item is scaled from 0 (never true) to 3 (almost always true), generating a total score ranging from 0 to 195. Scores can also be generated for five subdomains: *social awareness, social cognition, social communication, social motivation*, and *autistic mannerisms*. For comparability, the usage of SRS raw scores is recommended for research (Bölte and Poustka 2008), and raw scores of the German version were applied in this study. Reliability and validity findings for this version were similar to the data of the US original sample (Bölte et al. 2008).

The German version of the *Social Communication Questionnaire* (Bölte and Poustka 2006) as a further parent-report screening questionnaire for autism was obtained in this study. It is widely available and has good psychometric properties (Bölte et al. 2000). The 40-item SCQ (composed of "yes" or "no"- questions) is based on ADI-R items. Here, the total score was analysed.

General psychopathology was assessed by the German version of the *Child Behavior Checklist* (CBCL 4-18) (Achenbach 1998; Döpfner et al. 1994). The CBCL comprises 113 items. A total score, second order scores for *internalizing* and *externalizing problems*, and first order syndrome scales for behavioral and emotional problems can be derived: *withdrawn, somatic complaints, anxious/depressed, social problems, thought problems, attention problems, delinquent behavior and aggressive behavior*. Responses are coded on a Likert Scale from 0 (not true), 1 (sometimes true) to 2 (often true). The CBCL is an internationally validated and widely used parent report form (Achenbach 1998).

Intelligence quotient (IQ) was measured in n = 277 individuals by the age appropriate German version of the Wechsler Intelligence Scales for children (*Hamburg-Wechsler-Intelligence Test for children*, HAWIK-IV, HAWIVA) (Petermann and Petermann 2010), and adults (Wechsler Intelligence Test for adults, WIE) (Aster et al. 2006), or by the current version of the revised *Culture Fair Intelligence Test* (CFT 20-R) (Weiss 2006). Additional

measurements of cognitive estimation comprised the *Kaufman Assessment Battery for Children* (K-ABC) (Melchers and Preuß 2009), *Peabody Picture Vocabulary Test* (PPVT) (Dunn and Dunn, 2007), *Standard Progressive Matrices* (SPM) (Horn 2009), or *Coloured Progressive Matrices* (CPM) (Bulheller and Häcker 2002) were assessed in n = 105 participants.

Statistical Analysis

IBM SPSS statistics version 22 was used for all statistical analyses. For descriptive statistics mean scores, and standard deviations (SD) were calculated. The first order ADI-R domains were tested for internal consistencies by Cronbach's α. To achieve comparability among subscales, the ADI-R subscales A1–A4, B1–B4, C1–C4, and D raw scores were transformed into standardized z-scores (mean = 0, SD = 1). By converting raw into z-scores they can be accurately compared to each other. A positive z-score means that the score is above the group mean, a negative below, and a value of 0 is equal to the group mean. Multiple cluster analyses were employed to classify symptom patterns on the basis of similarity derived from the scores of the described ADI-R algorithm subscales. In the first step, a single-linkage analysis was carried out to identify outliers with atypical profiles for elimination. In the second step, an agglomerative hierarchical clustering method was applied, using the squared Euclidean distance as proximity measure. Fusions were made by Ward's method. The number of clusters was statistically determined by dendrogram, implementing the elbow criterion, and the development of error sum of squares. Subsequently, the K-means procedure was implemented for finally optimizing the cluster solution. K-means is a partition based clustering method to minimize the sum of squared error over all clusters. Subsequently, the same cluster analytic procedure was done for the summarized social communication, and stereotyped behavior domain, as postulated by DSM-5. For controlling the quality of the clustering solution, a MANOVA with cluster affiliation as independent variable and ADI-R algorithm subscale scores as dependent variables was administered. Discriminant analysis was used to validate the final cluster results. The clustering solution is considered as satisfying, when 90–95 % of the cases are correctly classified (Steinhausen and Langer 1977). After obtaining the clusters, comparisons of the cluster means were done by (M)ANOVA (or appropriate χ^2-test) for age, IQ, gender, ICD-10 diagnoses, as well as ADOS scores for social affect (SA), and restricted repetitive behaviors (RRB), and the severity score, SCQ, SRS, and CBCL scores with post hoc Bonferroni correction. Effect sizes were calculated by Cohen's φ, or partial η^2, as appropriate.

Results

Descriptive Data of the Sample

The entire sample consists of n = 405 (87.5 %) male, and n = 58 (12.5 %) female individuals, in total n = 463 individuals. The age ranged from ≥3 to ≤21 years with a mean score of 10.41 years (SD = 4.15). IQ test results were available for n = 382 individuals ranging from an IQ of 41–147 (mean score = 94.60, SD = 20.64).

ADI-R subscale scores were available for all n = 463 participants with scores for social interaction (mean = 15.64, SD = 5.78), communication (mean = 12.31, SD = 4.63), stereotyped behavior (mean = 4.48, SD = 2.41), and abnormal development (mean = 2.27, SD = 1.48). *ADOS* (modules 1, 2, 3, 4) was available for n = 415 individuals. Severity scores were calculated above all of the four modules (mean = 6.70, SD = 2.06). Raw scores for ADOS module 1 were available for n = 28 children (SA: mean raw score = 13.43, SD = 3.93; RRB: mean raw score = 3.00, SD = 1.66), n = 97 for module 2 (SA: mean raw score = 10.38, SD = 3.75; RRB: mean raw score = 2.54, SD = 1.88), n = 220 for module 3 (SA: mean raw score = 9.97, SD = 3.80; RRB: mean raw score = 1.64, SD = 1.32), n = 78 for module 4 (SA: mean raw score = 11.72, SD = 5.04; RRB: mean raw score = 2.49, SD = 1.75). The calibrated severity scores for both domains and each module are presented for each cluster in Table 3. Scores on the *Social Responsiveness Scale* (SRS) were available for n = 334 (mean = 97.08, SD = 28.64). N = 358 raw scores were obtained for the *Social Communication Questionnaire* (SCQ) (mean = 17.92, SD = 6.90). Scores on the *Child Behavior Checklist* (CBCL) were available for n = 347 children and adolescents with a mean total T-score of 70.20 (SD = 7.96).

Cluster Analysis: Cluster Solution

Analysis of the three ADI-R algorithm domain scores resulted in fair reliability for all domains (Cronbach's α = .80 for social interaction, α = .77 for communication, α = .78 for stereotyped behaviors), except for the subscale abnormal development (Cronbach's α = .36).

The first single linkage procedure indicated one outlier by using the dendrogram elbow criteria. Examination of the dendrogram identified one spike presenting an atypical profile of the indicated participant compared with the rest of the sample. The outlier was excluded from further analysis. By using the hierarchical clustering analysis the dendrogram and inspection of reduction of the error sums of squares indicated a three-cluster solution. The subsequent adjusted cluster solution showed the following

characteristics: cluster 1 comprised n = 168 (36.68 %) individuals, cluster 2 n = 137 (29.91 %), and cluster 3 n = 153 (33.41 %). A MANOVA with the ADI-R subscale scores as dependent and the three clusters as independent variables showed a significant main effect [F(26) = 54.47, $p < .000$, $\eta^2 = .615$]. Similar results were found for all subscale scores ($F_{A1}(2) = 115.23$; $F_{A2}(2) = 70.26$; $F_{A3}(2) = 137.97$; $F_{A4}(2) = 156.17$; $F_{B1}(2) = 77.12$; $F_{B2}(2) = 85.13$; $F_{B3}(2) = 119.21$; $F_{B4}(2) = 97.48$; $F_{C1}(2) = 24.63$; $F_{C2}(2) = 20.71$; $F_{C3}(2) = 47.07$; $F_{C4}(2) = 42.31$; $F_D(2) = 34.47$; $p_{all} < .000$). For details, see Table 1 presenting the z-standardized subscale scores with means and standard deviation. Please also see Table 3 for more details on the comparison of the ADI-R raw scores with same results for a significant main effect [F(26) = 54.47, $p < .000$, $\eta^2 = .615$] and significant differences for all of the subscale scores.

The most severely affected individuals were allocated to cluster 2. All scores besides A1 ("Failure to use nonverbal behaviors to regulate social interaction") showed the highest impairments in any of the subscores related to *reciprocal social interaction, communication, stereotyped behaviors*, and *abnormal development*.

Cluster 3 comprised individuals with a moderate symptom severity in the domains of *reciprocal social interaction, communication,* and *abnormal development*. For *stereotyped, repetitive or idiosyncratic speech and behaviors* (B3, C1-C4) the individuals of cluster 3 showed minor symptom severity in comparison with the distribution of social affect (SA) symptoms within the cluster 3.

Overall, the least affected individuals in the *social interaction, communication*, and *abnormal development* domains were included in cluster 1. In contrast to the social affect domains, the *restricted and stereotyped behavior* domain was moderately pronounced in cluster 1, indicating a trend for relatively higher restricted and repetitive behaviors (RBB) than social affect symptoms in cluster 1.

Compared to the other cluster groups there were only main effects for the social affect domains (cluster 2 > cluster 3 > cluster 1), while the RBB domains showed comparable severity between cluster 3 and cluster 1 (cluster 2 > cluster 1 = cluster 3) with a descriptive trend

Table 1 The z-standardized ADI-R algorithm subscales due to a three cluster solution

	Cluster 1 N = 168 M (SD)	Cluster 2 N = 137 M (SD)	Cluster 3 N = 153 M (SD)	Statistical group differences F-value (df)	p value	Post hoc test*
Impairments in reciprocal social interaction						
A1 Failure to use nonverbal behaviors to regulate social interaction	−.82 (.77)	.50 (.80)	.50 (.77)	115.23 (2)	<.000	2 > 1, 3 > 1
A2 Failure to develop peer relationships	−.56 (.93)	.58 (.72)	.17 (.89)	70.26 (2)	<.000	2 > 3>1
A3 Lack of shared enjoyment	−.74 (.67)	.73 (.86)	.21(.83)	137.97 (2)	<.000	2 > 3>1
A4 Lack of socio-emotional reciprocity	−.70 (.72)	.86 (.82)	.05 (.77)	156.17 (2)	<.000	2 > 3>1
Impairments in communication						
B1 Lack/delay of spoken language and failure to compensate through gesture	−.61 (.81)	.58 (.92)	.19 (.87)	77.12 (2)	<.000	2 > 3>1
B2 Failure to initiate/sustain conversational interchange	−.60 (.90)	.66 (.76)	.11 (.87)	85.13 (2)	<.000	2 > 3>1
B3 Stereotyped, repetitive or idiosyncratic speech	−.22 (.82)	.87 (.93)	-.53 (.66)	119.21 (2)	<.000	2 > 1>3
B4 Lack of spontaneous make-believe or social imitative play	−.64 (.90)	.64 (.69)	.21 (.83)	97.48 (2)	<.000	2 > 3>1
Repetitive behaviors and stereotyped pattern						
C1 Encompassing preoccupation or circumscribed pattern of interest	−.17 (.89)	.48 (1.05)	−.23 (.90)	24.63 (2)	<.000	2 > 1; 2 > 3
C2 Apparently compulsive adherences to non-functional routines or rituals	−.18 (.93)	.45 (1.08)	−.18 (.88)	20.71 (2)	<.000	2 > 1; 2 > 3
C3 Stereotyped and repetitive motor mannerisms	−.19 (.92)	.62 (.95)	−.35 (.86)	47.07 (2)	<.000	2 > 1; 2 > 3
C4 Preoccupations with part-objects or non-functional elements of material	−.23 (.92)	.61 (.94)	−.29 (.90)	42.31 (2)	<.000	2 > 1; 2 > 3
D Abnormality of development evident at/before 36 months	−.29 (.96)	.56 (92)	−.14 (.91)	34.47 (2)	<.000	2 > 3; 2 > 1

M mean, *SD* standard deviation

* Bonferroni correction > means significant differences

for more severely affected individuals in cluster 1. To provide a better overview, the average behavioral profile of each cluster is displayed in Fig. 1.

For greater clarity, comparisons were also done for the z-standardized overall scores. Instead of summarizing the subscales into the original ADI-R algorithm domains, subscales were reorganized according to current DSM-5 criteria. Impairments in reciprocal social interaction and communication were combined to a social affect score. The subscale B3 on stereotyped speech was reorganized under the overall score of stereotyped, repetitive behavior as recommended by DSM-5. Significant main differences between the cluster groups were found for both, social affect (cluster 2 > cluster 3 > cluster 1), and RRB (cluster 2 > cluster 1 > cluster 3) (see Table 2).

Comparisons within the single cluster groups showed, that the confidence interval of the mean scores for social affect and repetitive behavior in cluster 1 did not overlap, indicating a significant difference between both domains with less deficits in the SA domain than in the RRB domain (SA < RRB). In contrast, confidence intervals of both mean scores overlapped in cluster 2, indicating a similar severity of difficulties in SA, and RRB (SA = RRB). Results for cluster 3 were similar to cluster 1: confidence intervals of mean SA and RRB scores did not overlap, indicating more problems in the SA domain than in the RRB domain (SA > RRB). Results are visualized in Fig. 2.

As last step, a discriminant analysis was computed to confirm the results of the cluster solution. In total, 96.7 % of the cases were correctly classified by the presented three cluster solution. In detail, 98.2 % of the participants were correctly classified in cluster 1, 94.9 % were correctly allocated to cluster 2, and 96.7 % to cluster 3 (see Fig. 3).

Cluster Comparisons

Demographic and Diagnostic Variables

No differences between clusters were observed for demographic data (Table 3), namely age [$F(2) = 1.49$, $p = .228$, $\eta^2 = .008$] and gender [$\chi^2(2) = 2.49$, $p = .293$, $\varphi = .073$]. IQ [$F(2) = .40$, $p = .670$, $\eta^2 = .006$] did not differ between the cluster groups. Cluster groups were also compared with regard to ICD-10 diagnostic classification, for which again no differences were observed between the three cluster groups [$\chi^2(4) = 2.74$, $p = .603$, $\varphi = .077$].

Autistic Symptoms

The *Social Responsiveness Scale* total score did not differed between clusters [$F(2) = 1.69$, $p = .187$, $\eta^2 = .010$]. No significant group comparisons were found with regard to the SRS subscales, but trends were observed with the most severe social impairments in cluster 2 and elevated

Fig. 1 The average behavioral profile for each of the three cluster groups. *Note*: A1–A4 = impairments in social interaction, B1–B4 = impairments in communication and language, C1–C4 = restricted, repetitive and stereotyped behaviors/interests, D = abnormality of development evident at/before 36 months; Cluster 2: high impairments in all subscales; Cluster 3: more impairments in the social affect domain (except for B3: stereotyped speech) than for stereotyped behaviors; Cluster 1: less impairments in all subscales with more severe pronounced stereotyped behaviors than social affect deficits
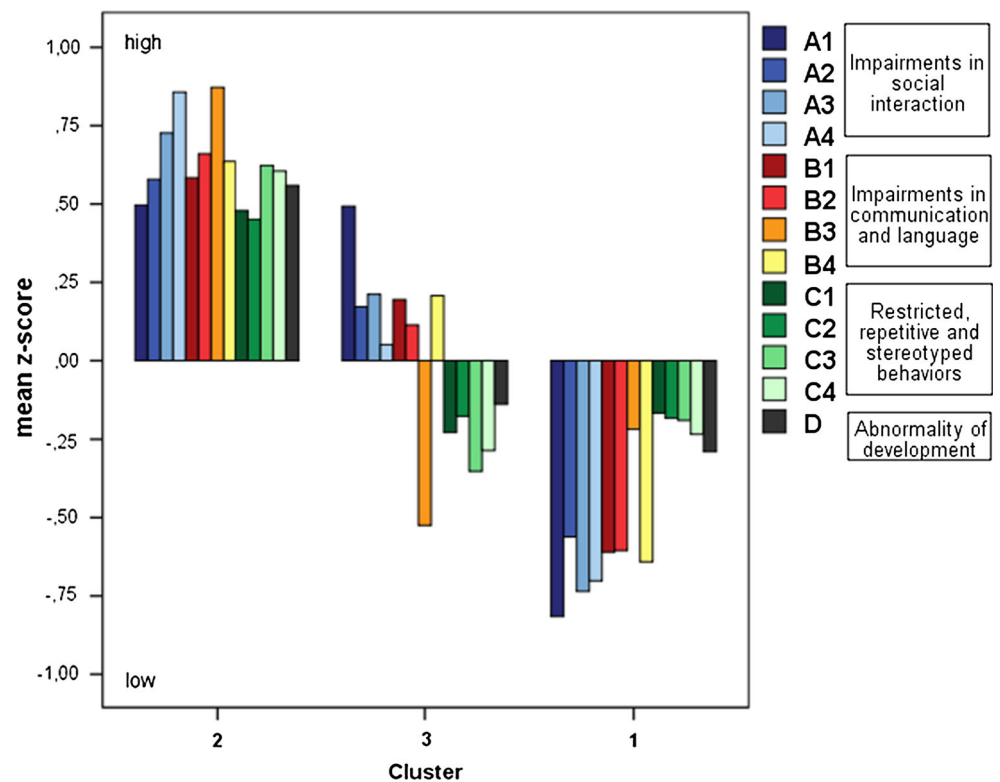
Table 2 Comparisons of the z-standardized overall scores including confidence intervals

	Cluster 1 N = 168 M (SD) CI	Cluster 2 N = 137 M (SD) CI	Cluster 3 N = 153 M (SD) CI	Statistical group differences		Post hoc test*
				F-value (df)	p value (η^2)	
Social affect: impairments in reciprocal social interaction and communication (A1, A2, A3, A4, B1, B2, B4)	−.99 (.45) −1.07 to −.93	.96 (.68) .85 to 1.07	.30 (.47) .23 to .38	540.90 (2)	.000 (.704)	2 > 3>1
Repetitive behaviors and stereotyped pattern (B3, C1, C2, C3, C4)	−.33 (.80) −.45 to −.20	1.01 (.79) .88 to 1.15	−.55 (.63) −.65 to −.45	184.21 (2)	.000 (.447)	2 > 1>3
Abnormality of development evident at/before 36 months (D)	−.29 (.96) −.44 to −.13	.56 (.91) .41 to .71	−.14 (.91) −.28 to .00	34.47 (2)	.000 (.132)	2 > 3,1

M mean, *SD* standard deviation, *CI* confidence interval, *η2* partial eta-squared
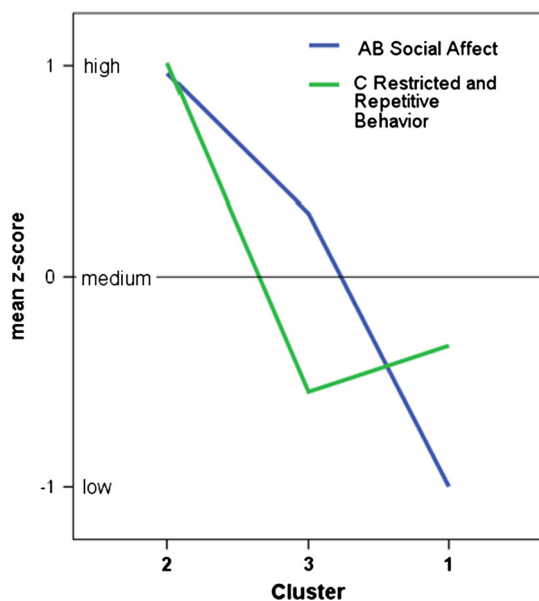
* Bonferroni correction > means significant differences

Fig. 2 Overall scores for Social Affect (domain AB) and Restricted, Repetitive Behavior (domain C) according to their cluster affiliation

Fig. 3 Discriminant analysis of correct cluster classification

autistic mannerisms in cluster 1 similar to the cluster groups.

Statistically significant differences were found for the *Social Communication Questionnaire* [F(2) = 3.20, $p = .042$, $\eta^2 = .018$] with highest scores in cluster 2, followed by cluster 1, and cluster 3.

To assess directly observed autistic symptoms *ADOS* modules 1, 2, 3 and 4 were applied. The ADOS severity scores [F(2) = .76 (2), $p = .469$, $\eta^2 = .004$] did not differ according to the cluster affiliation. None of the calibrated severity scores for SA or RRB for module 1, 2, 3, and 4 differed between the three cluster groups [module 1/SA: F(2) = 1.13, $p = .341$, $\eta^2 = .086$; RRB: F(2) = .1.91,
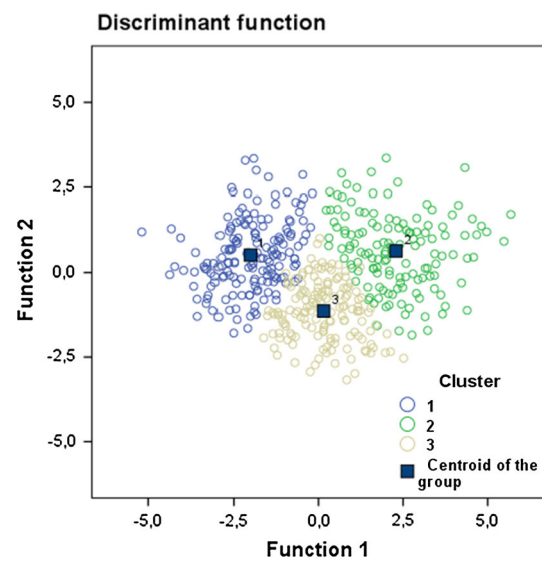
$p = .169$, $\eta^2 = .138$; module 2/SA: F(2) = .75, $p = .476$, $\eta^2 = .016$; RRB: F(2) = .58, $p = .560$, $\eta^2 = .012$; module 3/SA F(2) = .16, $p = .851$, $\eta^2 = .002$; RRB: F(2) = .19, $p = .825$, $\eta^2 = .002$; module 4/SA F(2) = .70, $p = .502$, $\eta^2 = .019$; RRB: F(2) = .16, $p = .854$, $\eta^2 = .004$]. For details please see Table 3.

Comorbid Psychopathology

Comparing the *Child Behavior Checklist* (CBCL) total score between clusters no differences could be observed [F(2) = .67, $p = .574$, $\eta^2 = .309$]. Similar results were found for all CBCL subscale scores (see Table 3). Descriptively, cluster 2 showed the highest comorbid psychopathology.

Table 3 Cluster comparison due to sample characteristics, autism specific, and comorbid variables

	Cluster 1 M (SD)	Cluster 2 M (SD)	Cluster 3 M (SD)	Statistical group differences	
				F/χ^2value (df)	p value
Age	N = 168	N = 137	N = 153		
N = 458	10.50 (4.39)	10.10 (3.85)	10.61 (4.16)	1.49 (2)	.228
Gender	N = 168	N = 137	N = 153		
N = 458; ♀/♂	26/142	13/124	19/134	2.46 (2)	.293
Intelligence	N = 133	N = 121	N = 124		
N = 378	95.39 (20.95)	94.80 (19.62)	93.14 (21.47)	.40 (2)	.670
ICD-10 diagnoses					
N = 458	N = 168	N = 137	N = 153		
Autistic Disorder	74 (44.0 %)	60 (43.8 %)	58 (37.9 %)		
Asperger's syndrome	51 (30.4 %)	45 (32.8 %)	59 (38.6 %)		
Atypical Autism	43 (25.6 %)	32 (23.4 %)	36 (23.5 %)	2.74 (4)	.603
ADI-R raw scores	N = 168	N = 137	N = 153		
Social interaction					
A1 nonverbal behavior	1.93 (1.39)	4.32 (1.46)	4.31 (1.34)	155.23 (2)	.000
A2 peer relationships	4.11 (1.84)	6.38 (1.43)	5.57 (1.78)	70.26 (2)	.000
A3 shared enjoyment	1.55 (1.24)	4.23 (1.58)	3.29 (1.52)	137.97 (2)	.000
A4 reciprocity	2.61 (1.55)	5.96 (1.76)	4.23 (1.64)	156.17 (2)	.000
Communication					
B1 delay of language	2.15 (2.02)	5.12 (2.30)	4.16 (2.15)	77.12 (2)	.000
B2 conversation	1.72 (1.15)	3.33 (.96)	2.63 (1.11)	85.13 (2)	.000
B3 Stereotyped speech	1.83 (1.39)	3.68 (1.57)	1.31 (1.11)	119.21 (2)	.000
B4 imitative play	2.93 (1.42)	4.94 (1.09)	4.27 (1.30)	97.48 (2)	.000
Repetitive behaviors					
C1 circumscribed interest	1.30 (.99)	2.01 (1.16)	1.23 (1.00)	24.63 (2)	.000
C2 nonfunctional routines	1.08 (1.11)	1.83 (1.28)	1.08 (1.05)	20.71 (2)	.000
C3 motor mannerisms	.62 (.80)	1.33 (.83)	.48 (.75)	47.07 (2)	.000
C4 Preoccupations with part-objects	.74 (.68)	1.36 (.70)	.71 (.67)	42.31 (2)	.000
D Abnormality of development	1.85 (1.40)	3.09 (1.34)	2.07 (1.34)	34.47 (2)	.000
ADOS (all modules)					
N = 410	N = 154	N = 117	N = 139		
Total severity score	6.65 (2.15)	6.58 (2.09)	6.88 (1.91)	.76 (2)	.469
Module 1	N = 10	N = 7	N = 10		
Severity score SA	7.40 (1.65)	6.86 (1.68)	6.20 (1.99)	1.13 (2)	.341
Severity score RRB	6.10 (1.10)	5.00 (2.83)	6.70 (1.34)	1.91 (2)	.169
Module 2	N = 36	N = 31	N = 29		
Severity score SA	6.56 (1.87)	6.39 (2.03)	6.97 (1.72)	.75 (2)	.476
Severity score RRB	5.86 (2.33)	5.45 (2.23)	6.03 (1.84)	.58 (2)	.560
Module 3	N = 74	N = 63	N = 77		
Severity score SA	7.19 (2.16)	7.17 (2.06)	7.35 (2.03)	.16 (2)	.851
Severity score RRB	5.36 (2.72)	5.33 (2.77)	5.58 (2.52)	.19 (2)	.825
Module 4	N = 34	N = 17	N = 22		
Severity score SA	7.21 (1.90)	7.88 (2.03)	7.59 (2.13)	.70 (2)	.502
Severity score RRB	6.28 (1.94)	5.94 (2.29)	6.13 (2.05)	.16(2)	.854
SRS N = 329	N = 117	N = 96	N = 116		
Total score	97 (29.47)	101.14 (29.22)	93.93 (26.79)	1.69 (2)	.187
Social awareness	11.81 (4.00)	12.42 (3.94)	11.83 (3.60)	.82 (2)	.441
Social cognition	17.31 (6.86)	18.87 (6.60)	16.98 (6.15)	2.45 (2)	.088

Table 3 continued

	Cluster 1 M (SD)	Cluster 2 M (SD)	Cluster 3 M (SD)	Statistical group differences	
				F/χ^2value (df)	p value
Social communication	33.38 (10.63)	34.16 (10.68)	32.64 (10.05)	.56 (2)	.574
Social motivation	17.20 (5.85)	17.86 (5.86)	16.91 (5.26)	.78 (2)	.459
Autistic mannerisms	17.22 (7.16)	17.82 (7.24)	15.58 (7.20)	2.85 (2)	.060
SCQ N = 353	N = 133	N = 98	N = 122		
Total score	18.16 (6.93)	19.08 (7.37)	16.77 (6.35)	3.20 (2)	.042*
CBCL N = 331	N = 117	N = 95	N = 119		
Total score	70.11 (8.06)	70.36 (7.95)	70.01 (7.78)	.67 (2)	.574
Internalizing	66.64 (10.04)	68.18 (8.34)	67.49 (9.04)	.74 (2)	.478
Externalizing	61.37 (10.31)	62.63(9.45)	63.16 (9.51)	1.04 (2)	.356
Withdrawn	70.32 (11.60)	71.44 (10.34)	70.73 (10.52)	.28 (2)	.756
Somatic complaints	59.08 (11.14)	59.58 (9.09)	59.35 (9.35)	.07 (2)	.935
Anxious/depressed	62.96 (10.68)	65.21 (9.84)	64.80 (9.25)	1.61 (2)	.201
Social problems	72.91(11.86)	73.08 (9.83)	74.20 (10.33)	.49 (2)	.614
Thought problems	70.85 (13.10)	71.36 (10.90)	68.56 (11.05)	1.78 (2)	.170
Attentional problems	71.76 (11.42)	70.65 (11.12)	71.62 (9.48)	.33 (2)	.722
Delinquent behavior	59.32 (9.02)	60.43 (7.25)	61.24 (7.59)	1.71 (2)	.183
Aggressive behavior	62.80 (11.26)	64.24 (10.34)	64.53 (10.48)	.86 (2)	.424

M mean, *SD* standard deviation, *ADOS* Autism Diagnostic Observation Schedule, *SRS* Social Responsiveness Scale, *SCQ* Social Communication Questionnaire, *CBCL* Child Behavior Checklist, *SA* social affect domain, *RRB* restricted and repetitive behaviors domain

* Post hoc Bonferroni correction $p = .04$ (2 > 3)

Discussion

Research that attempts to stratify children and adolescents with ASD into distinct subgroups assumes on an improving comprehension of the underlying phenotype structure to better delineate etiology, diagnostic, and treatment options. The major aim of the present study thus was to identify homogenous and clinically meaningful ASD subgroups based on the ADI-R algorithm subscale scores. This is a highly topical question with regard to the previous changes from DSM-IV-TR to DSM-5 and current efforts to develop ICD-11.

Our approach to reduce the heterogeneity of ASD samples differs from many other studies, which have attempted to analyze the factor structure of the ADI-R (Snow et al. 2009; Duku et al. 2013), or utilized the method of clustering in very circumscribed samples (e.g. Brannan et al. 2014). By using the hierarchical cluster analytic approach in a large sample with a broad range of IQ and age over the full spectrum of ASD symptoms, our study identified three empirically derived homogenous subgroups. Discriminant analysis revealed an excellent rate of correctly classified participants, confirming the presented cluster solution. The empirically derived number of clusters is in line with previously reported results on differing ASD populations from toddlers to high functioning adolescents (Verté et al. 2006; Wiggins et al. 2011; Brannan et al. 2014).

The three derived clusters indicate a different pattern of social interaction and communication problems versus stereotype behaviors. The characteristics of social interaction and communication problems fit with the assumption of a severity gradient across all three cluster groups with cluster 2 comprising the most affected individuals, followed by cluster 3, and least affected ASD individuals in cluster 1. These results support the novel DSM-5 concept of a severity gradient with regard to social communication, former comprising social interaction and communication problems. This result is in line with replicated findings that a differentiation between both DSM-IV-TR/ICD-10 domains seems to be arbitrary (Lord and Bishop 2015). Based on our results, a severity gradient can also be described for the stereotyped, repetitive behavior domain; however, the pattern of severity considerably differed from the first one. Even though the strongest impairments again were found for cluster 2, the less affected individuals in cluster 1 and 3 did not show a distinct symptom pattern for the stereotyped behavior domain as found for the social communication domains. Interestingly, the subscale for stereotyped speech (B3) (former included in the communication domain in DSM-IV-TR) showed the same cluster pattern as demonstrated for the repetitive and stereotyped behavior domain. These findings again support the shift of idiosyncratic and repetitive speech under the umbrella of the stereotyped behavior domain in DSM-5. Due to the

differing symptom distribution, our findings also strongly support the reorganization from the triadic to the dyadic domain architecture, as suggested by the DSM-5 concept.

As a consequence, clusters were subsequently compared with regard to the DSM-5 based dyadic structure. For both domains, social communication (SA) impairments and stereotyped, repetitive behavior (RRB), differences between clusters were observed. The significant difference, especially for the RRB domain, might be developed by combining the subscale scores and shifting the stereotyped speech under the RRB domain. Non-overlapping confidence intervals between the social communication and stereotyped behavior domains for cluster 1 and 3 revealed different symptom patterns than demonstrated for cluster 2. Cluster 1 included ASD individuals exhibiting significant more stereotyped behaviors than social communication impairments. ASD individuals with significantly less stereotyped behaviors in contrast to moderate impairments in the social communication domain were allocated to cluster 3. The most severely affected individuals with regard to both (not differing) domains, who also showed the earliest age of onset, were allocated to cluster 2. Taken together, a severity gradient as well as differing symptom pattern can be found within ASD. While the most affected ASD individuals were allocated to one cluster group (cluster 2), the less affected individuals were grouped into two clusters with a differing symptom profile. In consequence our study offers support for both, a gradient model, and a distinct phenotypic profile model. Nevertheless the minor number of items in the RRB domain and the size of the standard deviation need to be considered by interpreting the non-overlapping confidence intervals.

Thus, no cluster group characterized by social-communication problems only was observed, which supports the concept of a phenotypic distinction of *social communication disorder* (SCD), and ASD. Still, ASD children in cluster 3 showed the least stereotyped behavior, and probably included some children who may also meet SCD criteria according to DSM-5. Groups mainly exhibiting social communication problems without many restrictive and repetitive behaviors were also described by cluster analyses of ASD samples before (Greaves-Lord et al. 2012; Brennan et al. 2014). Similar to the study of Greaves-Lord et al. (2012), almost 30 % of our sample showed the described symptom pattern. A study of Mandy et al. (2011) indicated that the majority of the former diagnoses PDD-NOS presented a combination of social interaction and communication impairments with less or without repetitive and stereotyped behaviors. To date, it is widely discussed, but still unclear if indeed a substantial proportion of children and adolescents with ASD will be shifted into the SCD category. Due to the fact that we did not include any diagnostic assessment of SCD, we cannot provide any empirical data on this important question. Further studies including a broader range of communication and language impaired children with and without ASD need to clarify this issue.

After deriving the cluster groups, we aimed at further characterizing them by additional phenotypic measures. No influence of demographic variables or cognitive abilities was observed, which replicates previous results (Verté et al. 2006; Wiggins et al. 2012; Kim et al. 2013) providing some preliminary evidence for the cluster groups being independent of IQ, age and gender. Further studies found that the ADI-R algorithm totals are influenced by the developmental level, particularly the expressive language of the children, in linear regression models (Hus and Lord 2013). This influence was stronger for the ADI-R current than the ADI-R diagnostic algorithm. Future studies, also including non-verbal participants using the current algorithm, are needed to analyse if there might also be any similar effects of language-level for the derived cluster groups.

Evidence that DSM-IV-TR defined subtypes cannot be distinguished by the previous characteristics has led to the proposal of merging them into a single dimensional ASD group (Lord et al. 2012). In our study, there were also no differences for the distinct ICD-10 diagnoses autistic disorder, Asperger's syndrome, and PDD-NOS between the three empirically derived cluster groups. This result is also in line with previous studies (e.g. Brennan et al. 2014). These findings again support the revision of previously existing sub-groupings of ASD, as implemented by the DSM-5 concept, and underscore the need for a more empirically based conceptualization in the prospective eleventh revision of the ICD criteria.

Cluster comparisons with additional phenotypic measures, covering ASD symptoms, showed significant differences for the Social Communication Questionnaire including the most affected individuals in cluster 2 similar to the presented ADI-R cluster distribution, but less affected individuals in cluster 3 instead of cluster 1. Even though the results for the Social Responsiveness Scale did not differ significantly, descriptive statistics also indicated a trend for highest mean scores in cluster 2. Trends for less affected individuals due to autistic mannerisms measured by the SRS were found for cluster 3 in line with the cluster analytic results. We did not find any significant cluster differences on the ADOS severity score. Similar results were found by Georgiades et al. (2013). Several measurement issues may have influenced this result. One explanation may be that the ADOS severity metric used in our analyses combines scores of social-communication symptoms, and stereotyped, repetitive behaviors, and may be confounded by summarizing all symptoms into one combined score. For overcoming this effect, we also compared the ADOS social affect and restricted, repetitive behavior scores between the three cluster groups. But again, no

significant differences between cluster groups were found. The lack of statistically significant cluster differences on the ADOS scores might also reflect the differences between a retrospective parent interview, and a current direct observation of ASD symptoms. One further explanation could be that the cluster groups derived by retrospective report of symptoms at ages 4–5 or ever were compared to current ASD symptoms. Past studies comparing agreement of the ADI-R and ADOS algorithms similarly observed that both instruments allow independent, additive contributions on the basis of rather low correlations (Kim and Lord 2012). Another reason might also be the comparative small sample sizes for modules 1, 2 and 4 for each cluster group. In sum, the lack of significant differences between cluster groups for most of the ASD questionnaires might be attributed to the differing measurement issues, probably also age confounded.

Since individuals with ASD vary not only with regard to level and type of core autistic symptoms, but also in terms of co-occurring psychopathological symptoms, clusters were also compared with regard to the Child Behavior Checklist syndrome scales. No differences between clusters were observed, contrasting findings of previous studies of higher rates of psychiatric comorbidity in more severely affected ASD individuals (Greaves-Lord et al. 2013). But previous findings in this area are very heterogeneous. While some data showed a tendency of more co-occurring behavioral problems in individuals with more severe ASD traits (e.g. Pearson et al. 2006), others also report high rates of comorbidities among less affected children with PDD-NOS (de Bruin et al. 2006), and further studies revealed few differences between the ASD subtypes (e.g. Snow and Lecavalier 2011) in line with this study. As a limitation, the cluster comparisons were based on a subset of children assessed with the CBCL. Thus, the number of missing data needs to be taken into account.

Our study has methodological strengths that enhance the validity of the results compared to previous studies. First, we included a large sample with a broad range of ages, above the broad spectrum of ASD symptoms from the less severe end (e.g. atypical autism, Asperger's syndrome) to the severe affected end (autistic disorder), enhancing the generalization of the study results. Diagnoses were carefully implemented by using the gold standard diagnostic system, underlining the internal validity of our study. The cognitive level was available for most of the participants, and mean IQ scores are remarkably similar across clusters. The homogenous distribution enables the core symptom domains to emerge within the cluster analyses. The empirically derived cluster groups were additionally compared by various child psychiatric variables, as cognitive level, comorbidities, and autism specific questionnaires, which were not part of the former cluster analyses. Second,

by using the ADI-R as one of the most widely used and comprehensive assessment for the broad range of behavioral and functional ASD symptoms (Snow et al. 2008), an ideal measure to derive clinically relevant and homogenous subgroups, was chosen for the cluster analytic approach (Hu and Steinberg 2009). Besides the limitation already noted, we have to emphasize that we just used data assessed by an interview with the primary caregiver for the cluster analyses. Including data of one primary measurement instead of analyzing data from multiple instruments allowed us post hoc comparisons of the clusters by variables, which were not part of the cluster analytic process before. When interpreting the results of the current study, it has to be noted, that the ADI-R measurement was also included, besides other instruments, into the diagnostic process. So ICD-10 diagnoses were not totally independent of the ADI-R algorithm scores used in the cluster analyses. Non-verbal children were excluded for reducing missing data, and enhancing the homogeneity of the sample, what also limits the generalization of our results, although Snow et al. (2008) did not find any differences between verbal and non-verbal children due to a factor analysis of the ADI-R algorithm items. At last, two aspects due to the measurement and the statistic procedure have to be addressed. First, the ADI-R is developed as categorical diagnostic instrument. That might limit the identification of a severity model and affect the cluster pattern. This aspect has to be taken into account even if previous studies already showed evidence for an underlying severity gradient using the ADI-R (e.g. Spiker et al. 2002). A study of Constantino et al. (2004) also found similar results for a continuously distributed underlying factor for both, the ADI-R, and the SRS as a dimensional measurement. Second, next to the strengths of creating z-scores by eliminating inherent subscale differences, the reduction of variance can also be noted as a shortcoming while generating relative instead of absolute differences. This is of special interest as the social communication differences across cluster groups might be more meaningful with regard to the number of items included. Finally, even if the sample size and the sample characteristics strengthen the representativeness of the findings, the results have to be replicated in further samples.

Despite these limitations, results from this study could have important clinical and research implications. First, the empirically derived cluster groups mainly differed from the diagnostic single subtypes as classified by ICD-10. Asperger's syndrome, PDD-NOS, and autistic disorder were nearly equally distributed across the cluster groups, indicating that the underlying ASD phenotype visibly differs from the previously proposed diagnostic criteria of the subtypes. In line with numerous other studies, this leads to a need for a more empirically based conceptualization in

the prospective eleventh revision of the ICD concept. Second, the ASD phenotype according to the distribution of social communication deficits confirms the novel dyadic concept as well as the proposed severity gradient by DSM-5. This finding should also be considered by the revision of the ICD-10 diagnostic concept. A closer look on the data even showed a somewhat differing pattern for the stereotyped behavior domain, also indicating a symptom profile for the less affected groups. This is of special interest for diagnostic concepts, but needs to be replicated in further studies especially by considering the minor item number in the RRB domain.

All in all, our results essentially offer support for both, the severity gradient model suiting for the severe affected individuals, and specific symptom profiles due to the less affected participants. Due to the differing symptom profiles at the less affected end we want to alert to the risk given for treatment approaches when distinct subtypes will no longer be distinguished. Different children with different symptoms might need differing intervention approaches. At last it has to be taken into account that the presented cluster solution is dependent upon the variables and the specific sample included into these analyses. Our results have to be replicated in independent samples first, especially with cluster variables not used in the diagnostic process before, to verify the external validity of these findings.

Compliance with Ethical Standards

References

Achenbach, T. M. (1998). *Elternfragebogen über das Verhalten von Kindern und Jugendlichen; Deutsche Bearbeitung der Child Behavior Checklist (CBCL/4-18)*. Köln: Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik.

Aster, M., Neubauer, A., & Horn, R. (2006). *Wechsler Intelligenztest für Erwachsene (WIE)*. Frankfurt a. Main: Harcourt Test Services.

Bölte, S., Crecelius, K., & Poustka, F. (2000). Der Fragebogen über Verhalten und soziale Kommunikation: psychometrische Eigenschaften eines Autismus-Screening-Instruments für Forschung und Praxis. *Diagnostica, 46*, 149–155.

Bölte, S., & Poustka, F. (2006). *Fragebogen zur Sozialen Kommunikation—Autismus Screening (FSK). Deutsche Fassung des Social Communication Questionnaire (SCQ)*. Bern: Hans Huber, Hogrefe AG.

Bölte, S., & Poustka, F. (2008). *Skala zur Erfassung Sozialer Reaktivität (SRS)*. Bern: Hans Huber, Hogrefe AG.

Bölte, S., Poustka, F., & Constantino, J. N. (2008). Assessing autistic traits: Cross-cultural validation of the social responsiveness scale (SRS). *Autism Research, 1*, 354–363.

Bölte, S., Rühl, D., Schmötzer, G., & Poustka, F. (2006). Diagnostisches Interview für Autismus—Revidiert. Deutsche Fassung des Autism Diagnostic Interview—Revised von Michael Rutter, Ann LeCouteur und Catherine Lord; 1. Auflage. Bern: Hans Huber, Hogrefe AG.

Brennan, L., Barton, M., Chen, C.-M., Green, J., & Fein, D. (2014). Detecting subgroups in children with pervasive developmental disorders—not otherwise specified. *Journal of Autism Developmental Disorders, 45*(5), 1329–1344.

Bulheller, S., & Häcker, H. O. (Eds.). (2002). *Coloured progressive matrices (CPM). Deutsche Bearbeitung und Normierung nach J. C. Raven*. Frankfurt: Pearson Assessment.

Constantino, J. N., Gruber, C. P., Davis, S., Hayes, S., Passanante, N., & Przybeck, T. (2004). The factor structure of autistic traits. *Journal of Child Psychology and Psychiatry, 45*(4), 719–726.

Döpfner, M., Schmeck, K., & Berner, W. (1994). Handbuch: Elternfragebogen über das Verhalten von Kindern und Jugendlichen. Forschungsergebnisse zur deutschen Fassung der Child Behavior Checklist (CBCL/4-18). Köln: Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik (KJFD).

Duku, E., Szatmari, P., Vaillancourt, T., Georgiades, S., Thompson, A., Liu, X.-Q., et al. (2013). Measurement equivalence of the autism symptom phenotype in children and youth. *Journal of Child Psychology and Psychiatry, 54*(12), 1346–1355.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test* (4th ed.). Pearson Assessments: San Antonio, TX.

Eaves, L. C., Ho, H. H., & Eaves, D. M. (1994). Subtypes of autism by cluster analysis. *Journal of Autism Developmental Disorders, 24*(1), 3–22.

Fein, D., Stevens, M., Dunn, M., Waterhouse, L., Allen, D., & Rapin, I. (1999). Subtypes of pervasive developmental disorder: Clinical characteristics. *Child Neuropsychology, 5*, 1–23.

Georgiades, S., Szatmari, P., Zwaigenbaum, L., Bryson, S., Brian, J., Roberts, W., et al. (2013). A prospective study of autistic-like traits in unaffected siblings of probands with autism spectrum disorders. *JAMA Psychiatry, 70*(1), 42–48.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardized ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism Developmental Disorders, 39*(5), 693–705.

Greaves-Lord, K., Eussen, M. L., Verhulst, F. C., Minderaa, R. B., Mandy, W., Hudziak, J. J., et al. (2012). Empirically based phenotypic profiles of children with pervasive developmental disorders: Interpretation in the light of the DSM-5. *Journal of Autism Developmental Disorders, 43*(8), 1784–1797.

Hair, J. F., & Black, W. C. (2000). Cluster analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 147–205). Washington, DC: American Psychological Association.

Horn, R. (Ed.). (2009). *Standard progressive matrices (SPM). Deutsche Bearbeitung und Normierung nach J. C. Raven* (2nd ed.). Frankfurt: Pearson Assessment.

Hu, V. W., & Lai, Y. (2013). Developing a predictive gene classifier for autism spectrum disorders based upon differential gene expression profiles of phenotypic subgroups. *North American Journal of Medicine and science,*. doi:10.7156/najms.2013.0603107.

Hu, V. W., & Steinberg, M. E. (2009). Novel clustering of items form the Autism diagnostic interview-revised to define phenotypes within autism spectrum disorders. *Autism Research, 2*(2), 67–77.

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*(10), 2400–2412.

Hus, V., & Lord, C. (2013). Effects of child characteristics on the autism diagnostic interview-revised: Implications as a measure of ASD severity. *Journal of Autism and Developmental Disorders, 43*(2), 371–381.

Hus, V., & Lord, C. (2014). The autism diagnostic observation schedule, module 4: Revised algorithm and standardized severity scores. *Journal of Autism and Developmental Disorders, 44*(8), 1996–2012.

Hus, V., Pickles, A., Cook, E. H., Risi, S., & Lord, C. (2007). Using the autism diagnostic interview-revised to increase phenotypic homogeneity in genetic studies of autism. *Biological Psychiatry, 61*, 438–448.

Jansen, R., Ceulemans, E., Grauwels, J., Maljaars, J., Zink, I., Steyaert, J., & Noens, I. (2013). Young children with language difficulties: A dimensional approach of subgrouping. *Research in Developmental Disabilities, 34*(11), 4115–4124.

Kim, S. H., & Lord, C. (2012). Combining information from multiple sources for the diagnosis of autism spectrum disorders for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Child Psychology and Psychiatry, 53*(2), 143–151.

Kim, S. H., Thurm, A., Shumway, S., & Lord, C. (2013). Multisite study of new autism diagnostic interview-revised (ADI-R) algorithms for toddlers and young preschoolers. *Journal of Autism Developmental Disorders, 43*(7), 1527–1538.

Lord, C., & Bishop, S. L. (2015). Recent advances in autism research as reflected in DSM-5 criteria for autism spectrum disorder. *Annual Review of Clinical Psychology, 11*, 53–70.

Lord, C., Petkova, E., Hus, V., & Risi, S. (2012). A multi-site study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry, 69*(3), 306–313.

Mandy, W., Charman, T., Gilmour, J., & Skuse, D. (2011). Toward specifying pervasive developmental disorders-not otherwise specified. *Autism Research, 4*(2), 121–131.

Melchers, P., & Preuß, U. (2009). *Kaufman Assessment Battery for Children, dt. Version (K-ABC)*. Frankfurt/M: Pearson Assessment.

Obafemi-Ajayi, T., Miles, J. H., Takahashi, T. N., Qi, W., Aldridge, K., Zhang, M., et al. (2014). Facial structure analysis separates autism spectrum disorders into meaningful clinical subgroups. *Journal of Autism Developmental Disorders, 45*(5), 1302–1317.

Pearson, D. A., Loveland, K. A., Lachar, D., Lane, D. M., Reddoch, S. L., Mansour, R., & Cleveland, L. A. (2006). A comparison of behavioral and emotional functioning in children and adolescents with Autistic Disorder and PDD-NOS. *Child Neuropsychology, 12*(4–5), 321–333.

Petermann, F., & Petermann, U. (2010). *Hamburg-Wechsler-Intelligenztest für Kinder-IV*. Bern: Hans Huber.

Ring, H., Woodbury-Smith, M., Watson, P., Wheelwright, S., & Baron-Cohen, S. (2008). Clinical heterogeneity among people with high functioning autism spectrum conditions: evidence favouring a continuous severity gradient. *Behavioral and Brain Functions, 4*(11), 1–6.

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., et al. (2006). Combining information of multiple sources in the diagnosis of autism spectrum disorders. *Journal of American Academy of Child and Adolescent Psychiatry, 45*(9), 1094–1103.

Rühl, D., Bölte, S., Feineis-Matthews, S., & Poustka, F. (2004). *Diagnostische Beobachtungsskala für Autistische Störungen. Deutsche Fassung der Autism Diagnostic Observation Schedule*. Bern: Hans Huber, Hogrefe AG.

Rutter, M., Bailey, A., & Lord, C. (2001). *Social communication questionnaire (SCQ)*. Los Angeles, CA: Western Psychological Services.

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

Schopler, E., Reichler, R., & Rochen, Renner B. (1988). *The childhood autism rating scale*. Los Angeles, CA: Western Psychological Services.

Sevin, J., Matson, J., Coe, D., Love, S., Matese, M., & Benavidez, D. (1995). Empirically derived subtypes of pervasive developmental disorders: A cluster analytic study. *Journal of Autism Developmental Disorders, 25*, 561–578.

Siegel, B., Anders, T., Ciaranello, R., Bienenstock, B., & Kramer, H. (1986). Empirically derived sub-classification of the autistic syndrome. *Journal of Autism Developmental Disorders, 16*, 275–294.

Snow, A. V., & Lecavalier, L. (2011). Comparing autism, PDD-NOS, and other developmental disabilities on parent-reported behavior problems: Little evidence for ASD subtype validity. *Journal of Autism Developmental Disorders, 41*, 302–310.

Snow, A. V., Lecavalier, L., & Houts, C. (2008). The structure of the Autism Diagnostic Interview-Revised: Diagnostic and phenotypic implications. *Journal of Child Psychology and Psychiatry, 50*(6), 734–742. http://www.ncbi.nlm.nih.gov/pubmed/?term=Snow%2CLecavalier%2C+Houts%282008%29.

Spiker, D., Lotspeich, L. J., Dimiceli, S., Myers, R. M., & Risch, N. (2002). Behavioral phenotypic variation in autism multiplex families: Evidence for a continuous severity gradient. *American Journal of Medical Genetics, 114*, 129–136.

Steinhausen, D., & Langer, K. (1977). *Clusteranalyse: Einführung in Methoden und Verfahren der automatischen Klassifikation*. Berlin: Walter de Gruyter.

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 and Adolescent Psychiatry, 39*(3), 346–352.

Veatch, O. J., Veenstra-Vanderweele, J., Potter, M., Pericak-Vance, M. A., & Haines, J. L. (2014). Genetically meaningful phenotypic subgroups in autism spectrum disorders. *Genes Brain Behavior, 13*(3), 276–285.

Verté, S., Geurts, H. M., Roeyers, H., Rosseel, Y., Oosterlaan, J., & Sergeant, J. A. (2006). Can the children's communication checklist differentiate autism spectrum disorders? *Autism, 10*(3), 266–287.

Weiss, R. H. (2006). *CFT 20-R: Grundintelligenz Skala 2-Revision*. Göttingen: Hogrefe.

Wiggins, L. D., Robins, D. L., Adamson, L. B., Bakeman, R., & Henrich, C. C. (2012). Support for a dimensional view of autism spectrum disorders in toddlers. *Journal of Autism Developmental Disorders, 42*(2), 191–200.

World Health Organisation. (1992). *International statistical classification of diseases and related health problems Rev. 10*. Geneva: World Health Organisation.