

Inter-rater reliability and stability of diagnoses of autism spectrum disorder in children identified through screening at a very young age

Daalen, Emma; Kemner, Chantal; Dietz, Claudine; Swinkels, Sophie H. N.; Buitelaar, Jan K.; Engeland, Herman

Postprint / Postprint

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:

Daalen, E., Kemner, C., Dietz, C., Swinkels, S. H. N., Buitelaar, J. K., & Engeland, H. (2009). Inter-rater reliability and stability of diagnoses of autism spectrum disorder in children identified through screening at a very young age. *European Child & Adolescent Psychiatry*, 18(11), 663-674. <https://doi.org/10.1007/s00787-009-0025-8>

Nutzungsbedingungen:

Dieser Text wird unter dem "PEER Licence Agreement zur Verfügung" gestellt. Nähere Auskünfte zum PEER-Projekt finden Sie hier: <http://www.peerproject.eu> Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

gesis
Leibniz-Institut
für Sozialwissenschaften

Terms of use:

This document is made available under the "PEER Licence Agreement". For more information regarding the PEER-project see: <http://www.peerproject.eu> This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der

Leibniz-Gemeinschaft

Inter-rater reliability and stability of diagnoses of autism spectrum disorder in children identified through screening at a very young age

Emma van Daalen · Chantal Kemner · Claudine Dietz ·
Sophie H. N. Swinkels · Jan K. Buitelaar ·
Herman van Engeland

Received: 29 September 2007 / Accepted: 15 April 2009 / Published online: 7 May 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract To examine the inter-rater reliability and stability of autism spectrum disorder (ASD) diagnoses made at a very early age in children identified through a screening procedure around 14 months of age. In a prospective design, preschoolers were recruited from a screening study for ASD. The inter-rater reliability of the diagnosis of ASD was measured through an independent assessment of a randomly selected subsample of 38 patients by two other psychiatrists. The diagnoses at 23 months and 42 months of 131 patients, based on the clinical assessment and the diagnostic classifications of standardised instruments, were compared to evaluate stability of the diagnosis of ASD. Inter-rater reliability on a diagnosis of ASD versus non-ASD at 23 months was 87% with a weighted κ of 0.74 (SE 0.11). The stability of the different diagnoses in the autism spectrum was 63% for autistic disorder, 54% for pervasive developmental disorder, not otherwise specified (PDD-NOS), and 91% for the whole category of ASD. Most diagnostic changes at

42 months were within the autism spectrum from autistic disorder to PDD-NOS and were mainly due to diminished symptom severity. Children who moved outside the ASD category at 42 months made significantly larger gains in cognitive and language skills than children with a stable ASD diagnosis. In conclusion, the inter-rater reliability and stability of the diagnoses of ASD established at 23 months in this population-based sample of very young children are good.

Keywords Autistic disorder ·
Inter-rater reliability and stability of diagnosis ·
Preschool · Screening

Introduction

Autism spectrum disorders (ASD), which include autistic disorder or autism, Asperger syndrome, and pervasive developmental disorder, not otherwise specified (PDD-NOS), are characterised by deviant and delayed development of reciprocal social interaction, and of verbal and non-verbal communication, in combination with stereotyped and restricted behaviours, interests and activities, that lead to lifelong impairments. A further requirement for a classification of autistic disorder is that the delay or abnormal functioning starts before the child is 3 years [1]. However, in most of the children the diagnosis is made later [5, 9, 20], even though most parents report concerns about the development of their child as early as the second year of life or even earlier [10, 18–20, 25, 34]. Problems in children with Asperger syndrome and in children with autistic symptoms presenting after 30 months of age, therefore, diagnosed as PDD-NOS, are identified a later age than they are in autistic disorder [16, 20].

E. van Daalen (✉) · C. Kemner · C. Dietz · H. van Engeland
Rudolf Magnus Institute of Neuroscience, Utrecht,
The Netherlands
e-mail: e.vandaalen@umcutrecht.nl

E. van Daalen · C. Kemner · C. Dietz · H. van Engeland
Department of Child and Adolescent Psychiatry,
University Medical Centre, Utrecht, The Netherlands

C. Kemner
Department of Developmental Psychology,
University Medical Centre, Utrecht, The Netherlands

S. H. N. Swinkels · J. K. Buitelaar
Department of Psychiatry, Radboud University Nijmegen
Medical Centre, Nijmegen, The Netherlands

Diagnosing ASD at an early age has several advantages. First, it facilitates starting early intervention, educational planning and development of a professional support system. Several early treatment programmes report improved communication skills and social behaviour and diminished abnormal behaviour [17, 28, 29]. Second, early diagnosis enables professionals to learn about the developmental trajectories of ASD in the early years and to identify predictors of outcome [46]. Lastly, given the importance of genetic factors in the aetiology of ASD, early diagnosis enables early genetic counselling for parents and other relatives.

Recognition of the importance of the early identification of ASD has spurred researchers to improve diagnostic procedures in the preschool years [3, 7, 15, 16, 18, 21, 25, 27, 31, 33, 39, 40, 42, 43]. However, lowering the age of initial diagnosis presents new challenges [5]. For example, the phenotypic expression of autistic disorder at 2 years of age or younger may differ from that at 3 years or older. Thus, the severity and pattern of symptoms of ASD at a young age need to be established, as do the inter-rater reliability and stability of the early diagnosis.

The inter-rater reliability of a diagnosis made by clinicians refers to the consensus on the diagnosis between different psychiatrists. The stability of the diagnosis refers to the likelihood that the diagnosis at initial evaluation is the same as the diagnosis at the time of follow-up. The inter-rater reliability and stability of the diagnosis of autistic disorder have been examined in clinically referred samples of children older than 5 years and found to be excellent [30, 45]. Studies that investigated inter-rater reliability and stability in clinically referred children, younger than 5 years of age, with autistic disorder are summarised in Table 1 [7, 9, 15, 18, 25, 27, 31, 36, 39, 40, 42, 43]. Overall, these studies indicate that a diagnosis of autistic disorder made at 2 years is stable in clinically referred samples measured at 3 years, and even up to 9 and 12 years. Diagnostic stability, however, is less strong for PDD-NOS. Another result of these studies is that clinical judgement, when a child is 2 years of age, proved to be superior to the diagnostic algorithm of a standardised interview, the Autism diagnostic interview-revised (ADI-R) [24] or standardised observation, i.e. the Autism diagnostic observation schedule-generic (ADOS-G) [26] in predicting children's later diagnostic classification [9, 25, 27]. Diagnoses based on the ADI-R, appear to change significantly, particularly in younger and more intellectually disabled children [25]. Diagnostic thresholds from the ADI-R were crossed and recrossed between ages 2 and 7 years [9].

Although, standardised research instruments at age 2 years are inferior to the insight in the decision whether autism is present or not made by experienced, well-trained clinicians, this clinical insight proves not to be sufficient by

itself. In conclusion, scores on these standardised research instruments also make real contributions beyond their influence on informing and structuring clinical diagnosis [27].

Inter-rater reliability for ASD diagnoses below age 3 years has been examined in only two studies and found to be good to excellent for the distinction between ASD and non-ASD, and between presence and absence of autistic disorder, but poor for the distinction between autistic disorder and PDD-NOS (Table 1). A factor associated with more accuracy in an early ASD diagnosis is the experience of the clinician [39]. Less is known about the reliability and stability of ASD diagnoses in population-based samples. In a population-based screening study of 17,173 children, using the Checklist for Autism in Toddlers (CHAT), in the United Kingdom, the stability of a clinical diagnosis of autistic disorder made at 20 months was very good, with no false positives for ASD at 42 months. The diagnosis of autistic disorder appeared to be more stable than that of PDD-NOS, see Table 1 [9]. In a follow-up sample of children recruited using the CHAT, to a randomized control trial of a parent training early intervention [14], the stability of a clinical diagnosis of autistic disorder made at 20 months consistently proved to be good at 7 years of age. Almost all of these children met ADOS-G algorithm criteria for ASD and half of these children met the full ADI-R algorithm cut off for autistic disorder at age 7 [6].

The focus in recent studies has been on variability in outcome for children with an early diagnosis of ASD [6, 37, 38, 40]. Although differences between children with an early diagnosis of ASD who retain the diagnosis and who lose the diagnosis as a toddler do exist, the two groups are very difficult to differentiate when diagnosed initially [40]. Diagnostic stability has shown to be significantly higher for children who were initially diagnosed after 30 months (87%) than for those who were initially diagnosed at 30 months or younger (52%) [42].

The aims of the present study were as follows. First, we set out to evaluate the inter-rater reliability and stability of ASD diagnoses in children identified through a screening procedure applied at 14 months of age [11, 41]. Unlike the UK study [2], this population-based sample included children with intellectual disability. Second, we examined the cognitive and language correlates of children with a stable versus an unstable diagnosis of ASD.

Method

Design

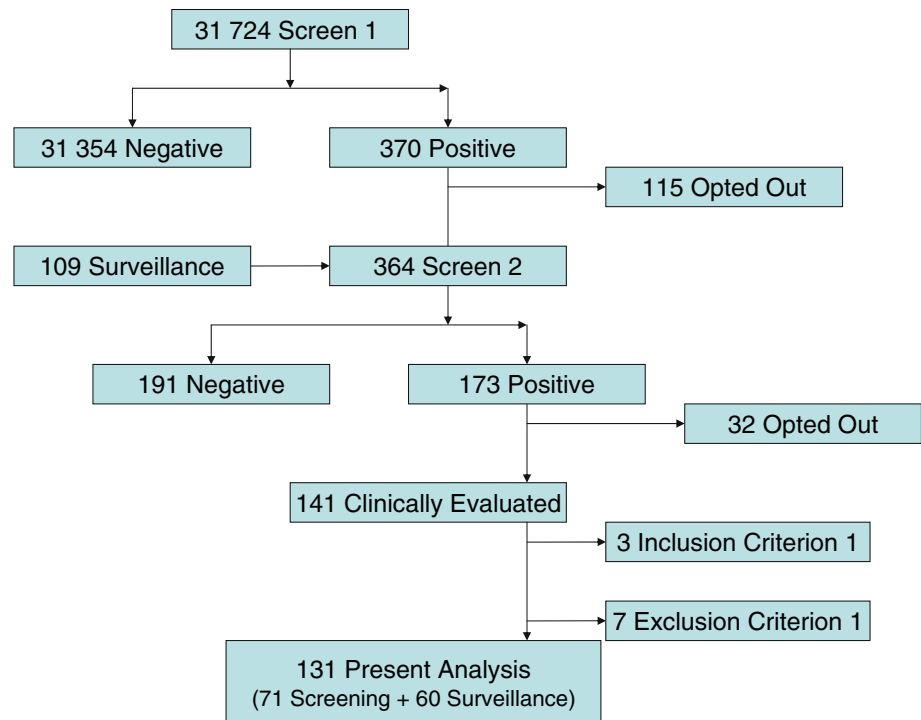
From October 1999 to April 2002, 31,724 children from the general population were screened by physicians at all

Table 1 Descriptive characteristics of best-estimate diagnoses, reliability and stability of clinical diagnoses

| Sample | N | Type of sample | Age at 't1', months (SD or range) | Age at 't2', months (SD or range) | Reliability at 't1', percentage agreement (κ , P), percentage agreement, certainty rating | Reliability at 't2', percentage agreement (κ , P) | Stability of diagnosis between 't1' and 't2', percentage agreement (κ , P) |
|------------------------|-----|----------------|--|---|---|---|--|
| Gillberg et al. [18] | 28 | CR | 23.96 (9.3) | 58 | – | 100% | AD 95% PDD-NOS 50% ASD 88% |
| Lord [25] | 30 | CR | 25–35 | 38–52 | – | 97% | AD 87.5% PDD-NOS |
| Sigman and Ruskin [36] | 50 | CR | 47.2 (12.1) | 154 (3.74) | – | – | AD 98% |
| Stone et al. [39] | 65 | CR | 31.4 (3.4) | 45 (4.3) | ASD/non-ASD 88%, 0.67, <0.5 | – | AD 80% PDD-NOS 42% |
| Moore and Goodson [31] | 20 | CR | 34 (29–42) | 53 (48–58) | AD/PDD-NOS 64%, 0.28, <0.5 | – | AD 87.5% PDD-NOS 33.3% ASD 100% |
| Eaves and Ho [15] | 49 | CR | 33 (4.6) | 59 (7.47) | – | – | AD 91% PDD-NOS 22% ASD 93% |
| Lord et al. [27] | 192 | CR | AD 29.1 (4.7) PDD-NOS 29.1 (5.6) Non-ASD 28.8 (5.5) | AD 110.1 (15.7) PDD-NOS 113.8 (17.1) Non-ASD 114.9 (11.8) | AD, yes/no 92%. Cert. Rat. 0.89 | AD > 90% PDD-NOS > 83% Non-ASD > 83% | Total agreement 67%, 0.47 ASD /non-ASD 90%, 0.72 AD/ non-AD 76%, 0.51 ASD 88% AD 89% PDD-NOS 29% AD 89% PDD-NOS 61% |
| Turner et al. [42] | 25 | CR | 31.0 (3.8) | 108.8 (7.9) | – | – | ASD 88% AD 89% PDD-NOS 29% AD 89% PDD-NOS 61% |
| Sutera et al. [40] | 90 | CR | ASD/ASD 27.6 (4.7) ASD/non-ASD 26.5 (4.9) Non-ASD/non-ASD 28 (3.9) | ASD/ASD 52.2 (6.6) ASD/Non-ASD 54.4 (10.1) Non-ASD/Non-ASD 28 (3.9) | – | – | ASD /non-ASD 63% AD/ non-AD 68%, AD 90% PDD-NOS 100% ASD 100% AD 67% PDD-NOS 33% ASD 100% |
| Turner and Stone [43] | 26 | CR | 28.8 (3.4) | 53.3 (3.5) | – | – | ASD /non-ASD 63% AD/ non-AD 68%, AD 90% PDD-NOS 100% ASD 100% AD 67% PDD-NOS 33% ASD 100% |
| Chawarska et al. [7] | 31 | CR | AD 21.6 (3.2) PDD-NOS 21.6 (2.5) | AD 34.8 (2.5) PDD-NOS 38.1 (8.3) | – | – | ASD /non-ASD 63% AD/ non-AD 68%, AD 90% PDD-NOS 100% ASD 100% AD 67% PDD-NOS 33% ASD 100% |
| Cox et al. [9] | 46 | PB | 20.7 | 42 | – | – | ASD /non-ASD 63% AD/ non-AD 68%, AD 90% PDD-NOS 100% ASD 100% AD 67% PDD-NOS 33% ASD 100% |

ASD autism spectrum disorders, Non-ASD no autism spectrum disorder, AD autistic disorder, PDD-NOS pervasive developmental disorder, not otherwise specified, CR clinically referred, PB population based, Cert. Rat certainty rating

Fig. 1 Design: two level screening for ASD. *Screen 1* 4-item early screening of autistic traits (ESAT) scale at routine 14-month developmental check, *Screen 2* 14-item ESAT scale, Inclusion criterion 1 a first psychiatric evaluation before the age of 37 months and a second evaluation at approximately the age of 42 months, and no sooner than 12 months after the first evaluation, exclusion criterion 1 presence of a genetic or medical disorder that could be associated with specific phenotypes of psychiatric disorders



well-baby clinics in the province of Utrecht using the four-item early screening of autistic traits (ESAT) scale at their routine 14-month developmental check (Screen 1)¹ [41], see Fig. 1. Parents were advised by the physician to continue with the screening procedure if their child failed at least one of four items of the ESAT and was considered screen positive. Children who scored positive at Screen 1 (population screening) and whose parents did consent ($n = 255$) and children aged up to 36 months identified by surveillance ($n = 109$) underwent Screen 2 [11]. Screen 2 consisted of the 14-item ESAT scale [41] and was done at a home visit by an experienced psychologist (C.D.), a member of our research team. Also, the cognitive development of the child was examined by the Mullen scales of early learning (MSEL) [32]. Children who failed at least three items of the 14-item ESAT scale were considered screen positive. The average (SD) age at Screen 2 was 16 (2) months for children recruited by the population screening and 27 (6) months for the group detected by surveillance. Children who scored positive at Screen 2 were invited for a first comprehensive psychiatric evaluation at the Department of Child and Adolescent Psychiatry of University Medical Centre Utrecht. A second, follow-up evaluation was performed when the children were on average 43-month-old (range 34–64 months). Because of limited resources, only children with a preliminary clinical

diagnosis of ASD, intellectual disability, language or phonological disorder as a result of the first psychiatric evaluation, or at parental request were included in a follow-up evaluation. As a result, 141 young children received two comprehensive psychiatric evaluations, see Fig. 1.

Further details of the screening procedure can be found elsewhere [11, 13, 41].

Clinical measurements

The first psychiatric evaluation at t1 (at about 23 months) was scheduled in the preschool programme at the department of child and adolescent psychiatry. The preschool programme consisted of a parent interview and psychiatric evaluation of the child. The parent interview included a developmental history, the Vineland social emotional early childhood scales [38, 44], and the Wing autistic disorder interview checklist (WADIC), administered by the primary clinician [47]. The evaluation of the child consisted of an unstructured psychiatric evaluation by the primary clinician and an ADOS-G, a semi-standardised observation procedure, administered by a research associate, which were both videotaped.

The cognitive evaluation of the child was performed with the Mullen scales of early learning (MSEL) by trained psychologists. Some children with intellectual disability were evaluated with the psycho-educational profile revised (PEP-R) [35]. The first children in the project were assessed with the Bayley scales of infant development (BSID-II) [4], see Table 2. The MSEL and the BSID-II were used to

¹ The routine developmental check is part of the system of surveillance for infants and toddlers as it is performed in the well-baby clinics in the Netherlands [11–13, 41].

Table 2 Distribution of number of participants by instruments used for cognitive evaluation and by instruments used for standardised psychiatric evaluation at t1 and t2, number of participants at t1 and at t2 is 131

| | t1 | t2 |
|-------------------------------------|-----|-----|
| Cognitive instruments | | |
| BSID-II | 8 | 0 |
| MSEL | 117 | 88 |
| MSEL-NV | 5 | 2 |
| PEP-R | 0 | 23 |
| No IQ | 1 | 18 |
| Total IQ | 131 | 131 |
| Standardised diagnostic instruments | | |
| ADOS-G, module I | 126 | 65 |
| ADOS-G, module II | 3 | 59 |
| ADOS-G, missing | 2 | 7 |
| ADOS-G total | 131 | 131 |
| ADI-R | – | 98 |
| ADI-R, missing | – | 33 |
| ADI-R total | – | 131 |

BSID-II Bayley scales of infant development, *MSEL* Mullen scales of early learning, *MSEL-NV* Mullen scales of early learning-non-verbal subscales, *PEP-R* psycho-educational profile revised, *No IQ* no cognitive evaluation performed, *ADOS-G* autism diagnostic observation schedule-generic, *ADI-R* autism diagnostic interview-revised, *t1* first psychiatric evaluation, *t2* second psychiatric evaluation

calculate an overall cognitive score (CS), the PEP-R was used to calculate an age equivalent score. This last score was converted to an overall cognitive score (CS) to make the scores of the three instruments comparable.²

At t2, the parents of 18 children agreed to a psychiatric and an ADOS-G evaluation, but did not give consent for a cognitive evaluation. These were all children with a high level of intellectual disability. Eight of these children received a diagnosis of autistic disorder and three of these children a diagnosis of intellectual disability without an ASD. One of the children was diagnosed with ADHD and two of the children with a language disorder. Four children were diagnosed with a regulatory disorder and had been evaluated at the age of 24 months, and found to perform at an average cognitive level.

Children were given a preliminary clinical diagnosis at t1 on the basis of the judgement of the primary psychiatrist of whether the child was likely to meet the DSM-IV-TR criteria for autistic disorder, PDD-NOS, or another psychiatric diagnosis when he or she was 4 or 5 years of age. The child psychiatrist used all available written and videotaped information with the exception of the results of the

ADOS-G algorithm or individual item scores and classified the children according to the DSM-IV-TR diagnostic criteria they were likely to meet at 4 or 5 years of age. The same evaluation procedure was repeated at the second psychiatric evaluation at 42 months (t2). In addition, the parents were interviewed with the ADI-R by a research associate, see Table 2. The children were assigned DSM-IV-TR diagnoses, based on all the available clinical information, again with the exception of the results of the ADOS-G and ADI-R algorithms. The diagnosis autistic disorder was reserved for these children meeting the algorithm for autistic disorder of the DSM-IV, the other diagnoses of ASD were given to children with serious and pervasive symptoms of ASD, but who are not meeting the threshold for autistic disorder. The ADI-R Diagnostic Algorithm specifies that most of the prototypical autistic behaviour is seen at the ages 4–5 years, and that the ADI-R may be less specific or sensitive at younger ages [24]. Thus, because the mean age of the children at t2 was 43.07 months (SD = 5.15), the instrument was not used as sole arbiter in the diagnostic process [9].

Children could have more than one diagnosis, but only the principal diagnosis, being the main focus of attention or treatment [1], was used for the scope of this article. For example, the diagnosis of autistic disorder took precedence in the case of a child with an autistic disorder and a phonological disorder. If only a phonological disorder was present, this was considered as being the principal diagnosis.

With regard to the treatment, all children with an ASD diagnosis or another developmental disorder in our cohort went to a facility for challenged toddlers or a facility for children with a mental handicap for 4 days a week. These facilities offer a day-care programme based on behavioural principles. The facilities for challenged toddlers offer this approach in a group especially for autistic children. Children receive speech and language therapy in the facility or externally. For most children, the frequency was limited to 1 h in every 2 weeks. One of the children received an intensive treatment, especially designed for autistic children in the facility for children with an intellectual disability. She was severely handicapped and later diagnosed with Rett's syndrome.

The effect of treatment was not assessed for the purpose of this article.

Statistics

To evaluate inter-rater reliability of diagnosis, Cohen's kappa was used. Kappa values were interpreted according to the criteria by Cicchetti and Sparrow [8]: excellent agreement (κ between 0.75 and 1.00); good agreement (κ between 0.60 and 0.74); fair agreement (κ between 0.40 and 0.59); and poor agreement ($\kappa < 0.40$).

² A cognitive score was calculated from the PEP-R by dividing the mental age in months by the chronological age in months and then multiply this by 100.

Contingency tables were applied to assess stability of diagnosis between t1 and t2. Differences in age and cognitive scores between the different diagnostic groups were tested with analysis of variance, and if significant, followed by Bonferroni corrected post hoc tests. Comparisons of changes in cognitive scores between the stable and unstable groups were done using Student's *t* test for independent samples. In all cases *P* values <0.05 were considered significant. All statistical analyses were performed using SPSS 12 for Windows.

Results

Participants

Children were only included for the present analyses if a first psychiatric evaluation, at t1, was performed before the age of 37 months and if a second evaluation, at t2, was carried out at approximately the age of 42 months, and no sooner than 12 months after the first evaluation. Accordingly, 138 children were selected from the 141 that were clinically evaluated after the screening procedures (Fig. 1). In addition, children in whom the presence of a genetic or medical disorder that could be associated with specific phenotypes of psychiatric disorders was confirmed were excluded [Rett's disorder (*n* = 1), tuberous sclerosis (*n* = 2), neurofibromatosis (*n* = 2), 22q11.2 deletion syndrome (*n* = 1), and fragile X syndrome (*n* = 1)].

As a result, 131 children were left to be included in the analysis. Of these, 131 children, 71 children originated from the population screening and 60 children originated from surveillance by the well-baby clinics. These 131 children were on an average 26 months (*SD* = 6.2) at t1, and on average 45 months (*SD* = 6.4) at t2. Accordingly, 53 out of the 80 children with a preliminary diagnosis of ASD were included for the present analyses.

Descriptive data for children at t1

The descriptive data for the remaining 131 children at t1 are reported in Table 3 by diagnostic category. Forty children were classified as having an autistic disorder by clinical judgement; 13 as having PDD-NOS, 20 as having an intellectual disability, without an ASD, 28 as having an expressive language disorder, 6 as having a mixed receptive–expressive language disorder, 7 as having ADHD, and 4 as having other axis I diagnoses of the DSM-IV-TR (i.e. sleeping disorder, separation anxiety disorder, stereotypic movement disorder, parent–child relational problem); 6 as having borderline intellectual functioning; and 7 children were not classified according to the DSM-IV-TR. These children had severe regulatory disorders.

The diagnostic groups differed in chronological age at t1 [ANOVA, *F* (8, 122) = 4.69, *P* < 0.01]; post hoc Bonferroni tests revealed significant higher ages for children with an autistic disorder than children with an expressive language disorder, other axis I diagnoses, borderline intellectual functioning or regulatory disorders; *P* < 0.03.

Children with an autistic disorder had a significantly lower cognitive score than the children in the other diagnostic groups (all *P* < 0.03), with the exception of the children with an intellectual disability without an ASD [ANOVA, *F* (8, 121) = 18.53, *P* < 0.01. In addition, children with PDD-NOS had a significantly lower cognitive score than children with ADHD and other axis I diagnoses (all *P* < 0.02). Ten children cognitively evaluated with the MSEL received the lowest possible score on the instrument and received a cognitive score of 49 (see Table 3). To correct for a possible floor effect, the one-way ANOVA for cognitive score was repeated without these ten children. Accordingly, children with an autistic disorder had a significant lower cognitive score than children in all the other diagnostic groups (all *P* < 0.03), with the exception of children with an intellectual disability without an ASD and children with PDD-NOS [ANOVA, *F* (8, 111) = 16.13, *P* < 0.01.

Descriptive data for children at t2

The descriptive data for the 131 children at t2 are reported in Table 3 by diagnostic category. Twenty-six children were classified as having an autistic disorder by clinical judgement, 22 as having PDD-NOS, 13 as having an intellectual disability without an ASD, 6 as having an expressive language disorder, 8 as having a mixed receptive–expressive language disorder, 16 as having a phonological disorder, 2 as having another developmental disorder (developmental coordination disorder), 7 as having ADHD, 3 as having other axis I problems of the DSM-IV-TR (i.e. 2 as having a parent–child relational problem; 1 as having selective mutism); 28 were not classified according to the DSM-IV-TR. These children had severe regulatory disorders.

The diagnostic groups did not differ in chronological age [ANOVA, *F* (9, 121) = 1.2, n.s.]. Children with an autistic disorder had a significantly lower cognitive score than the children in the other diagnostic groups (all *P* < 0.03), with the exception of the children with an intellectual disability without an ASD [ANOVA, *F* (9, 101) = 20.7, *P* < 0.01. In addition, children with PDD-NOS had a significantly lower cognitive score than the children with no axis I problems *P* < 0.02, and a significantly higher cognitive score than children with an intellectual disability, without an ASD, *P* < 0.01].

The ADI-R and ADOS-G domain scores per diagnostic group at t1 and t2 are presented in Table 4. Children with a

Table 3 Demographic data for children at t1 and t2

| Diagnosis | N | Gender | | Chronological age in months (SD) | N (CSS) ^a | CSS ^a (SD) |
|---------------|-----|--------|----|-------------------------------------|----------------------|-----------------------|
| | | M | F | | | |
| t1 | | | | | | |
| AD | 40 | 34 | 6 | 29.4 (5.6) | 39 | 57.4 (15.0) |
| PDD-NOS | 13 | 9 | 4 | 28.2 (5.2) | 13 | 72.3 (18.5) |
| ID | 20 | 15 | 5 | 26.4 (6.1) | 20 | 60.4 (11.6) |
| ELD | 28 | 24 | 4 | 24.1 (1.1) | 28 | 83.8 (10.7) |
| MR-ELD | 6 | 6 | 0 | 24.6 | 6 | 83.0 (10.4) |
| PhD | 0 | 0 | 0 | – | – | – |
| Other DD | 0 | 0 | 0 | – | – | – |
| ADHD | 7 | 7 | 0 | 25.4 (5.7) | 7 | 94.9 (8.5) |
| Other axis I | 4 | 2 | 2 | 19.1 (1.0) | 4 | 102.0 (6.2) |
| BIF/no axis I | 13 | 7 | 6 | 20.8 (5.6) | 13 | 88.5 (13.3) |
| Total | 131 | 104 | 27 | | 130 | |
| t2 | | | | | | |
| AD | 26 | 22 | 4 | 46.2 (5.0) | 21 | 50.7 (18.5) |
| PDD-NOS | 22 | 18 | 4 | 46.3 (9.8) | 19 | 88.5 (22.6) |
| ID | 13 | 8 | 5 | 44.6 (2.9) | 10 | 54.0 (14.4) |
| ELD | 6 | 6 | 0 | 43.0 (5.4) | 5 | 93.6 (11.5) |
| MR-ELD | 8 | 7 | 1 | 43.8 (6.1) | 6 | 86.0 (5.2) |
| PhD | 16 | 15 | 1 | 43.7 (2.8) | 16 | 100.3 (16.9) |
| Other DD | 2 | 2 | 0 | 40.1 (5.3) | 2 | 103.0 (11.3) |
| ADHD | 7 | 5 | 2 | 49.5 (9.2) | 6 | 99.8 (20.0) |
| Other axis I | 3 | 3 | 0 | 40.2 (2.8) | 2 | 95.5 (9.2) |
| BIF/no axis I | 28 | 18 | 10 | 43.9 (6.3) | 25 | 106.8 (12.4) |
| Total | 131 | 104 | 27 | | 112 | |

t1 first psychiatric evaluation, t2 second psychiatric evaluation, M male, F female, SD standard deviation, N (CSS) number of children with an available cognitive standard score, N (CSS)^a number of children with an available cognitive standard score, CSS cognitive standard score. CSS^a cognitive standard score without correction for floor effect. For correction for floor effect, see text

AD autistic disorder, PDD-NOS pervasive developmental disorder, not otherwise specified, ID intellectual disability without ASD, ELD expressive language disorder, MR-ELD mixed receptive-expressive language disorder, PhD phonological disorder, Other DD other developmental disorder, ADHD attention-deficit/hyperactivity disorder, Other axis I, DSM-IV-TR other axis I diagnosis, BIF/no axis I borderline intellectual functioning and no diagnosis on axis I, DSM-IV-TR

clinical diagnosis of autistic disorder received higher scores on all domains than children diagnosed with PDD-NOS or no ASD, indicating more or severe symptoms. The mean score on the repetitive domain of the ADOS-G, module I, at 2 years of age for children diagnosed with an autistic disorder is 2.9 (SD 1.5), indicating a high prevalence of restrictive and repetitive behaviours (RRBs) in our sample in this diagnostic group. In our sample, children with PDD-NOS and no ASD show a much lower prevalence at 2 years of age, 0.8 (SD 0.9) and 0.7 (SD 1.2), respectively.

Inter-rater reliability

The inter-rater reliability of the diagnosis established at t1 was measured in 38 children. Two psychiatrists, who had not conducted the psychiatric evaluations and parent

interviews, assessed the children independently by reviewing the videotape of the psychiatric evaluation and the written reports of the parent interview and the evaluation of the cognitive development. They were not aware of the diagnosis made by the psychiatrist who conducted the initial evaluation.

The agreement amongst psychiatrists regarding ASD diagnoses at t1 was 87%, 33 out of 38 cases, Cohen's kappa (κ), was 0.74 (SE 0.11). The differentiation between ASD, intellectual disability without ASD, and other diagnostic categories was in 79%, 30 out of 38 cases, in conformity ($\kappa = 0.66$, SE 0.10). Disagreement was for about 37.5%, three out of eight cases, due to the distinction between ASD and an intellectual disability without ASD. Agreement regarding the distinction between autistic disorder and PDD-NOS was 75% ($\kappa = 0.51$, SE 0.21).

Table 4 ADI-R and ADOS scores by clinical diagnoses at t1 and t2

| Variable | Clinical diagnosis, t1 | | | | | | Clinical diagnosis, t2 | | | | | |
|--------------------------------------|------------------------|----|--------------|----|--------------|----|------------------------|----|--------------|----|--------------|----|
| | AD (SD) | N | PDD-NOS (SD) | N | Non-ASD (SD) | N | AD (SD) | N | PDD-NOS (SD) | N | Non-ASD (SD) | N |
| ADOS social domain, module 1 | 11.8 (2.6) | 38 | 3.8 (2.7) | 12 | 2.6 (2.6) | 76 | 10.8 (3.5) | 22 | 2.9 (2.0) | 13 | 1.7 (2.2) | 30 |
| ADOS social domain, module 2 | 8.0 (7.1) | 2 | 5.0 (0.0) | 1 | – | 0 | 11.5 (0.7) | 2 | 3.4 (3.6) | 8 | 1.0 (1.4) | 49 |
| ADOS communication domain, module 1 | 5.5 (1.6) | 38 | 2.3 (1.6) | 12 | 2.3 (1.7) | 76 | 5.9 (1.6) | 22 | 1.8 (1.8) | 13 | 1.3 (1.2) | 30 |
| ADOS communication domain, module 2 | 5.5 (5.0) | 2 | 4.0 (0.0) | 1 | – | 0 | 5.0 (0.0) | 2 | 2.4 (1.9) | 8 | 1.3 (1.2) | 49 |
| ADOS repetitive domain, module 1 | 2.9 (1.5) | 38 | 0.8 (0.9) | 12 | 0.7 (1.2) | 76 | 2.6 (1.4) | 22 | 0.8 (1.1) | 13 | 0.6 (1.2) | 30 |
| ADOS repetitive domain, module 2 | 1.0 (1.4) | 2 | 3.0 (0.0) | 1 | – | 0 | 2.0 (1.4) | 2 | 0.9 (1.1) | 8 | 0.2 (0.5) | 49 |
| ADI-R social domain | – | – | – | – | – | – | 8.9 (3.4) | 22 | 6.1 (3.4) | 19 | 3.4 (3.4) | 57 |
| ADI-R nonverbal communication domain | – | – | – | – | – | – | 5.3 (1.8) | 22 | 3.9 (2.4) | 19 | 2.7 (2.3) | 57 |
| ADI-R repetitive domain | – | – | – | – | – | – | 3.9 (1.3) | 22 | 2.7 (2.1) | 19 | 2.1 (1.4) | 57 |

ADOS-G autism diagnostic observation schedule-generic, *ADI-R* autism diagnostic interview-revised, *ASD* autism spectrum disorders, *Non-ASD* no autism spectrum disorder, *AD* autistic disorder, *PDD-NOS* pervasive developmental disorder, not otherwise specified, *N* number of children of whom data are available on every separate domain of the ADOS-G or the ADI-R, *SD* standard deviation, *t1* first psychiatric evaluation, *t2* second psychiatric evaluation

Stability

Of the 40 children diagnosed with an autistic disorder at t1, 25 received the same diagnosis at t2 (see Fig. 2), giving a stability of 63%. Of the 13 children diagnosed with PDD-NOS at t1, 7 had the same diagnosis at t2 (stability of 54%). The stability of a diagnosis of ASD between t1 ($n = 54$) and t2 ($n = 47$) was 87%.

In turn, sensitivity, that is, the probability of a diagnosis of a specific disorder at t1 if the disorder is present at t2, was 96% for autistic disorder, 32% for PDD-NOS, and 96% for ASD. There were 7 false positives for ASD at t1. Only two children not diagnosed with an ASD at t1 were diagnosed with PDD-NOS at t2 (see Fig. 2). Thirteen children (59%) diagnosed with PDD-NOS at t2 were classified as having an autistic disorder at t1, and one child (4%) diagnosed with an autistic disorder at t2 was diagnosed with PDD-NOS at t1.

Characteristics of children with an unstable ASD diagnosis

Forty-six children diagnosed with ASD at t1 had a stable ASD diagnosis at t2 (38 boys and 8 girls), and seven other children (5 boys and 2 girls), diagnosed with ASD at t1 had a diagnosis other than ASD at t2, i.e. children with an unstable ASD diagnosis. The changes in cognitive scores between t1 and t2 of the children with a stable ASD diagnosis and of the children with an unstable ASD diagnosis were compared. Information about cognitive scores at both t1 and t2 were available for 35 children with a stable ASD diagnosis and six children with an unstable ASD diagnosis. The children with an unstable ASD diagnosis

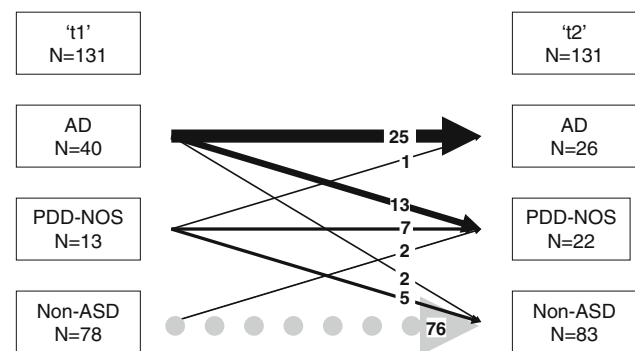


Fig. 2 Stability of diagnoses between 't1' and 't2'. *AD* autistic disorder, *PDD-NOS* pervasive developmental disorder not otherwise specified, *Non-ASD* no autism spectrum disorder

showed a significantly higher increase in cognitive scores [mean (M) = 37.2, $SD = 13.1$] than those with a stable ASD diagnosis ($M = 7.4$, $SD = 22.4$) [$t(39) = 3.1$, $P = 0.003$]. The effect size of this difference is large (Cohen's $d = 1.39$). The change in cognitive scores between t1 and t2 on the different subscales of the Mullen scales of early learning for the two groups was also compared. The number of children with an evaluation with the Mullen scales at both t1 and t2 was 14 for the stable ASD group, and 6 for the unstable ASD group. The children with an unstable ASD diagnosis ($M = 25.8$, $SD = 7.9$) showed a higher increase in scores on the expressive language subscale than those with a stable ASD diagnosis ($M = 8.6$, $SD = 15.2$). The difference is significant: $t(18) = 2.6$, $P = 0.018$. The effect size of this difference is large (Cohen's $d = 1.27$). The gender of the children in the stable and unstable group was compared and showed no significant difference.

Discussion

We found a good agreement ($\kappa = 0.74$) between psychiatrists in deciding whether 2-year-old children had an ASD or non-ASD diagnosis. This is in concordance with inter-rater reliability measurements of the distinction between an ASD or non-ASD diagnosis in very young clinically referred children [39], see Table 1. In our study, overall agreement for the finer distinction between autistic disorder and other ASD was fair, and also comparable with the agreement obtained by experienced clinicians in a sample of clinically referred children [39], see Table 1. The inter-rater reliability in the DSM-IV field trial for autistic disorder was excellent ($\kappa = 0.95$) for clinically referred, older children, in deciding whether a child had an autistic disorder or a non-ASD diagnosis [22, 45]. In contrast to our findings, we expected that clinician's ability to distinguish between ASD and non-ASD would be lower in very young children, given the possible diagnostic instability and the lack of age-appropriate diagnostic criteria for 2-year-old children. Also, we expected a lower inter-rater reliability in a population-based sample in comparison with a clinical referred sample of ASD children, a lower inter-rater reliability for the finer distinction between autistic disorder and PDD-NOS. The DSM-IV autistic disorder field trial reported a kappa of 0.85 regarding the differentiation between autistic disorder and other ASD in older children for experienced clinicians, and reported a kappa of 0.59 for inexperienced clinicians [22, 45]. Our findings show that the agreement between psychiatrists in deciding whether 2-year-old children have an ASD or non-ASD diagnosis is good, also in children, identified through screening and detected by surveillance [11]. Inter-rater reliability is lower, but still fair for the finer discrimination between an autistic disorder and PDD-NOS, as found earlier in clinically referred children. In our study, even experienced clinicians had most disagreement on the distinction between ASD and an intellectual disability without ASD. This illustrates that in the first 2 years of life the differentiation between delayed and deviant development remains clinically challenging.

The stability of the clinical diagnosis of autistic disorder between 26 and 45 months in our study was 63%, a figure comparable to that of 67% found in the CHAT study, the only other population-based study. These stability indices are lower than those obtained in clinically referred samples. This may be due to several factors, such as the older mean age of the clinically referred children at the first diagnostic evaluation in comparison with that of children in population-based studies, a factor of importance as found in recent studies [33, 42]. Another factor might be that symptom severity usually is higher in clinically referred children compared with very young children selected from

the population. The stability of the PDD-NOS diagnosis between 26 months and 45 months in our study was 54%, which is somewhat higher than the stability of PDD-NOS in the CHAT study, i.e. 33%. The lower stability of the diagnosis of PDD-NOS relative to that of autistic disorder may indicate that the diagnosis of autistic disorder is based on a more well-defined symptom cluster than that of PDD-NOS. It might also reflect that the diagnosis of autistic disorder is reserved for children with more severe symptoms and social handicaps, who are, therefore, less amenable to change [39]. This is indeed the case in our study, see Table 4. The stability of the diagnoses of ASD overall is lower in our study, i.e. 91%, than that reported in the CHAT study, i.e. 100%. This difference in overall stability of diagnosis of ASD can express that, unlike the CHAT study [2], our study included children with an intellectual disability. Differentiating autistic disorder with severe intellectual disabilities from equivalent degrees of severe intellectual disabilities without autistic disorder is much more difficult than differentiating autistic disorder from a generally less handicapped population [23, 25], as was also found in our inter-rater reliability data. Neither the ADOS-G nor the ADI-R shows a good specificity in diagnosing very young children with severe intellectual disability [25]. As it is likely that children with ASD who are referred at a young age to a diagnostic facility have intellectual disabilities as well, it is of great importance to improve specificity in diagnostic instruments for young children with autistic disorder with severe intellectual disabilities.

Earlier studies observed transitions between the sub-categories autistic disorder and PDD-NOS, and found particularly that about 50% of children with an initial diagnosis of PDD-NOS around the age of 2 years received a diagnosis of autistic disorder at follow-up [27, 39]. In contrast, our study found a reverse pattern that about one-third of children with a first diagnosis of autistic disorder were diagnosed as having PDD-NOS at follow-up. This pattern was more consistent with another study with clinically referred children [42].

Our second aim was to explore the differences in cognitive and verbal scores between children with a stable and unstable ASD diagnosis. The children in our study diagnosed as ASD at t1, and diagnosed as non-ASD at t2, the unstable ASD group, showed a substantial improvement in cognitive scores, especially verbal scores, between t1 and t2, that was significantly larger than the gain in cognitive scores found in the stable group. An increase in cognitive functioning has been reported in young children with a stable ASD diagnosis in earlier studies [6, 15, 42, 48] and in our sample [12]. So far, there appear to be two groups of children with an early diagnosis of ASD identified with our screening instrument: a group of children who showed catch-up growth in

language and other cognitive abilities, but still received a diagnosis of ASD at t2, and another group of children who had an even larger improvement in cognitive abilities, especially in the expression of language, but no longer fulfilled criteria for ASD at follow-up. It is essential for our understanding of ASD to follow these children in their further development to be able to determine whether these changes in cognitive and language scores and social functioning are temporary or lasting. Further, it is an important issue to examine whether the improvements of social interaction and communication drive the improvements of cognitive and language skills, or vice versa, whether the speed-up of cognitive and language development drives the changes in social repertoire.

Some limitations of our study should be noted. By our design of a prospective cohort study of children selected by screening from the population, we may have identified children that differ in clinical characteristics from those who are clinically referred. For example, we have screened for children with an early onset of autistic symptoms and early intellectual disabilities. This may have increased the subgroup of children with intellectual disabilities in our selection. The diagnosis of ASD in children who are high functioning, in whom language milestones are not delayed, and whose cognitive skills are average or above average is likely to be delayed until school age [16, 20]. Also, we do not know the sensitivity of our screening instrument, the ESAT. It may well be that we have detected a subgroup of children with ASD, and this needs to be established. Further, our follow-up period of 2 years is rather short. It is important for our understanding of developmental trajectories of young children with ASD to follow their development over the school age period. Also, in our sample, especially the parents of children with severe intellectual disabilities did not always give consent for a cognitive evaluation at t2, although they did give consent for a psychiatric re-evaluation. This is a general problem encountered in studies on early detection of ASD. Probably, parents may be less likely to come in for an evaluation at t2 than at t1, since the child already has been diagnosed at t1 and might be receiving services, which are satisfactory to the parents [21]. In addition, we were not able to use the same measure of cognitive evaluation for all children at both moments of evaluation in time. Comparison of results from different instruments reduces the inter-rater reliability of these results. Also, means and standard deviations of the cognitive level of children in the different diagnostic subgroups show large differences. Differences in cognitive and language findings between the stable and the unstable ASD group in our cohort should be interpreted with care and regarded as an exploratory finding. This exploratory finding needs and awaits replication in other studies.

Conclusions

These results show that both autistic disorder and the broader category of autistic spectrum disorder can be reliably diagnosed in very young children selected by means of a population screening procedure, as was earlier shown for samples of very young, clinically referred children. The stability of autistic disorder is higher than that of PDD-NOS. Given (1) the lower inter-rater reliability for the distinction between autistic disorder and PDD-NOS in our study, and in earlier studies [39] in very young children, and (2) the transition rate between autistic disorder and PDD-NOS and vice versa between the first and later assessments observed in our study and earlier work [27, 39, 40, 42], one may question whether it is valid or useful to differentiate PDD-NOS from autistic disorder at the age of 2 years or below. For clinical practice, it might be more relevant to restrict prediction of a clinical diagnosis to ASD or non-ASD in children younger than 2 years and to be more careful in diagnosing ASD as a final diagnosis for all children at such a young age.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. American Psychiatric Association (2000) Diagnostic and statistical manual of mental disorders (fourth edition, text revision). American Psychiatric Association, Washington, DC
2. Baird G, Charman T, Baron-Cohen S, Cox A, Swettenham J, Wheelwright S, Drew A (2000) A screening instrument for autism at 18 months of age: a 6-year follow-up study. *J Am Acad Child Adolesc Psychiatry* 39:694–702
3. Baird G, Charman T, Cox A, Baron-Cohen S, Swettenham J, Wheelwright S, Drew A (2001) Current topic: screening and surveillance for autism and pervasive developmental disorders. *Arch Dis Child* 84:468–475
4. Bayley Scales of Infant Development, 2nd edn. Psychological Corporation, San Antonio, TX
5. Charman T, Baird G (2002) Practitioner review: diagnosis of autism spectrum disorder in 2- and 3-year-old children. *J Child Psychol Psychiatry* 43:289–305
6. Charman T, Taylor E, Drew A, Cockerill H, Brown J, Baird G (2005) Outcome at 7 years of children diagnosed with autism at age 2: predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *J Child Psychol Psychiatry* 46:500–513
7. Chawarska K, Klin A, Paul R, Volkmar F (2007) Autism spectrum disorder in the second year: stability and change in syndrome expression. *J Child Psychol Psychiatry* 48:128–138
8. Cicchetti DV, Sparrow SS (1981) Developing criteria for establishing inter-rater reliability of specific items in a given inventory. *Am J Ment Def* 86:127–137
9. Cox A, Klein K, Charman T, Baird G, Baron-Cohen S, Swettenham J, Drew A, Wheelwright S (1999) Autism spectrum disorders at 20 and 42 months of age: stability of clinical and ADI-R diagnosis. *J Child Psychol Psychiatry* 40:719–732

10. De Giacomo A, Fombonne E (1998) Parental recognition of developmental abnormalities in autism. *Eur Child Adolesc Psychiatry* 7:131–136
11. Dietz C, Swinkels SHN, van Daalen E, van Engeland H, Buitelaar JK (2006) Screening for autistic spectrum disorder in children aged 14–15 months. II: population screening with the early screening of autistic traits questionnaire (ESAT). Design and general findings. *J Autism Dev Disord* 36:713–722
12. Dietz C, Swinkels SHN, Buitelaar JK, van Daalen E, van Engeland H (2007) Stability and change of IQ scores in preschool children diagnosed with autistic spectrum disorder. *Eur Child Adolesc Psychiatry* 16:405–410
13. Dietz C, Swinkels SH, van Daalen E, van Engeland H, Buitelaar JK (2007) Parental compliance after screening social development in toddlers. *Arch Pediatr Adolesc Med* 161:363–368
14. Drew A, Baird G, Baron-Cohen S, Cox A, Slonims V, Wheelwright S, Swettenham J, Berry B, Charman T (2002) A pilot randomised control trial of a parent training intervention for preschool children with autism. Preliminary findings and methodological challenges. *Eur Child Adolesc Psychiatry* 11:266–272
15. Eaves LC, Ho HH (2004) The very early identification of autism: outcome to age 4 1/2–5. *J Autism Dev Disord* 34:367–378
16. Filipek PA, Accardo PJ, Baranek GT, Cook EH Jr, Dawson G, Gordon B, Gravel JS, Johnson CP, Kallen RJ, Levy SE, Minshew NJ, Ozonoff S, Prizant BM, Rapin I, Rogers SJ, Stone WL, Teplin S, Tuchman RF, Volkmar FR (1999) The screening and diagnosis of autistic spectrum disorders. *J Autism Dev Disord* 29:439–484
17. Francis K (2005) Autism interventions: a critical update. *Dev Med Child Neurol* 47:493–499
18. Gillberg C, Ehlers S, Schaumann H, Jakobsson G, Dahlgren SO, Lindblom R, Bagenholm A, Tjuus T, Blidner E (1990) Autism under age 3 years: a clinical study of 28 cases referred for autistic symptoms in infancy. *J Child Psychol Psychiatry* 31:921–934
19. Gray KM, Tonge BJ (2001) Are there early features of autism in infants and preschool children? *J Paediatr Child Health* 37:221–226
20. Howlin P, Asgharian A (1999) The diagnosis of autism and Asperger syndrome: findings from a survey of 770 families. *Dev Med Child Neurol* 41:834–839
21. Kleinman JM, Robins DL, Ventola PE, Pandey J, Boorstein HC, Esser EL, Wilson LB, Rosenthal MA, Sutura S, Verbalis AD, Barton M, Hodgson S, Green J, Dumont-Mathieu T, Volkmar F, Chawarska K, Klin A, Fein D (2008) The modified checklist for autism in toddlers: a follow-up study investigating the early detection of autism spectrum disorders. *J Autism Dev Disord* 38:827–839
22. Klin A, Lang J, Cicchetti DV, Volkmar FR (2000) Brief report: interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: results of the DSM-IV autism field trial. *J Autism Dev Disord* 30:163–167
23. Krug DA, Arick J, Almond P (1980) Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior. *J Child Psychol Psychiatry* 21:221–229
24. Lord C, Rutter M, Le Couteur A (1994) Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 24:659–685
25. Lord C (1995) Follow-up of two-year-olds referred for possible autism. *J Child Psychol Psychiatry* 36:1365–1382
26. Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, Pickles A, Rutter M (2000) The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 30:205–223
27. Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Pickles A (2006) Autism from 2 to 9 years of age. *Arch Gen Psychiatry* 63:694–701
28. Matson JL (2007) Determining treatment outcome in early intervention programs for autism spectrum disorders: a critical analysis of measurement issues in learning based interventions. *Res Dev Disabil* 28:207–218
29. McConachie H, Diggle T (2006) Parent implemented early intervention for young children with autism spectrum disorder: a systematic review. *J Eval Clin Pract* 13:120–129
30. McGovern CW, Sigman M (2005) Continuity and change from early childhood to adolescence in autism. *J Child Psychol Psychiatry* 46:401–408
31. Moore V, Goodson S (2003) How well does early diagnosis of autism stand the test of time? Follow-up study of children assessed for autism at age 2 and development of an early diagnostic service. *Autism* 7:47–63
32. Mullen EM (1995) The Mullen scales of early learning: AGS edition. American Guidance Service, Circle Pines
33. Pandey J, Verbalis A, Robins DL, Boorstein H, Klin AM, Babitz T, Chawarska K, Volkmar F, Green J, Barton M, Fein D (2008) Screening for autism in older and younger toddlers with the modified checklist for autism in toddlers. *Autism* 12:513–535
34. Rogers SJ, DiLalla DL (1990) Age of symptom onset in young children with pervasive developmental disorders. *J Am Acad Child Adolesc Psychiatry* 29:863–872
35. Schopler E, Reichler RJ, Bashford A, Lansing MD, Marcus LM (1994) Psychoeducational profile revised: Pro-ED
36. Sigman M, Ruskin E, Arbeile S, Corona R, Dissanayake C, Espinosa M, Kim N, Lopez A, Zierhut C (1999) Continuity and change in the social competence of children with autism, down syndrome, and developmental delays. *Monogr Soc Res Child Dev* 64:1–114
37. Sigman M, McGovern CW (2005) Improvement in cognitive and language skills from preschool to adolescence in autism. *J Autism Dev Disord* 35:15–23
38. Sparrow SS, Balla DA, Cicchetti DV (1997) Vineland social-emotional early childhood scales: manual. American Guidance Service, Circle Pines
39. Stone WL, Lee EB, Ashford L, Brissie J, Hepburn SL, Coonrod EE, Weiss B (1999) Can autism be diagnosed accurately in children under 3 years? *J Child Psychol Psychiatry* 40:219–226
40. Sutura S, Pandey J, Esser EL, Rosenthal MA, Wilson LB, Barton M, Green J, Hodgson, Robins DL, Dumont-Mathieu T, Fein D (2007) Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *J Autism Dev Disord* 37:98–107
41. Swinkels SH, Dietz C, Van Daalen E, Kerkhof IH, van Engeland H, Buitelaar JK (2006) Screening for autistic spectrum in children aged 14 to 15 months. I: the development of the Early Screening of Autistic Traits Questionnaire (ESAT). *J Autism Dev Disord* 36:723–732
42. Turner LM, Stone WL, Pozdol SL, Coonrod EE (2006) Follow-up of children with autism spectrum disorders from age 2 to age 9. *Autism* 10:257–279
43. Turner WL, Stone LM (2007) Variability in outcome for children with an ASD diagnosis at age 2. *J Child Psychol Psychiatry* 48:793–802
44. Volkmar FR, Sparrow SS, Goudreau D, Cicchetti DV, Paul R, Cohen DJ (1987) Social deficits in autism: an operational approach using the Vineland adaptive behavior scales. *J Am Acad Child Adolesc Psychiatry* 26:156–161
45. Volkmar FR, Klin A, Siegel B, Szatmari P, Lord C, Campbell M, Freeman BJ, Cicchetti DV, Rutter M, Kline W et al (1994) Field trial for autistic disorder in DSM-IV. *Am J Psychiatry* 151:1361–1367

46. Werner E, Dawson G, Munson J, Osterling J (2005) Variation in early developmental course in autism and its relation with behavioral outcome at 3–4 years of age. *J Autism Dev Disord* 35:337–350
47. Wing L (1996) Wing autistic disorder interview checklist (WADIC). In: Rapin I (ed) *Preschool children with inadequate communication*. Mac Keith Press, London, pp 247–252
48. Yang P, Jong Y, Hsu H, Chen C (2003) Preschool children with autism spectrum disorders in Taiwan: follow-up of cognitive assessment to early school age. *Brain Dev* 25:549–554