



Published in final edited form as:

J Autism Dev Disord. 2008 April ; 38(4): 606–615. doi:10.1007/s10803-007-0427-8.

Diagnostic Stability in Very Young Children with Autism Spectrum Disorders

Jamie M. Kleinman,

Department of Psychology, University of Connecticut, 406 Babbidge Rd., Storrs, CT 06269-1020, USA

Pamela E. Ventola,

Department of Psychology, University of Connecticut, 406 Babbidge Rd., Storrs, CT 06269-1020, USA

Juhi Pandey,

Department of Psychology, University of Connecticut, 406 Babbidge Rd., Storrs, CT 06269-1020, USA

Alyssa D. Verbalis,

Department of Psychology, University of Connecticut, 406 Babbidge Rd., Storrs, CT 06269-1020, USA

Marianne Barton,

Department of Psychology, University of Connecticut, 406 Babbidge Rd., Storrs, CT 06269-1020, USA

Sarah Hodgson,

Department of Psychology, University of Connecticut, 406 Babbidge Rd., Storrs, CT 06269-1020, USA

James Green,

Department of Psychology, University of Connecticut, 406 Babbidge Rd., Storrs, CT 06269-1020, USA

Thyde Dumont-Mathieu,

Department of Psychology, University of Connecticut, 406 Babbidge Rd., Storrs, CT 06269-1020, USA

Diana L. Robins, and

Department of Psychology, University of Connecticut, 406 Babbidge Rd., Storrs, CT 06269-1020, USA

Deborah Fein

Department of Psychology, University of Connecticut, 406 Babbidge Rd., Storrs, CT 06269-1020, USA

Jamie M. Kleinman: jamie_kleinman@yahoo.com

Abstract

Autism Spectrum Disorders (ASD) diagnosis in very young children may be delayed due to doubts about validity. In this study, 77 children received a diagnostic and developmental

evaluation between 16 and 35 months and also between 42 and 82 months. Diagnoses based on clinical judgment, Childhood Autism Rating Scale, and the Autism Diagnostic Observation Schedule were stable over time. Diagnoses made using the Autism Diagnostic Interview were slightly less stable. According to clinical judgment, 15 children (19%) moved off the autism spectrum by the second evaluation; none moved onto the spectrum. Results indicate diagnostic stability at acceptable levels for diagnoses made at age 2. Movement off the spectrum may reflect true improvement based on maturation, intervention, or over-diagnosis at age 2.

Keywords

Autism; PDD-NOS; Diagnostic stability; Early detection

Introduction

For children with Autism Spectrum Disorders (ASD), early detection and subsequent early intervention can lead to substantially better prognosis, including improved language, social relationships, and adaptive functioning, and fewer maladaptive behaviors, all of which increase the chance of successful inclusion in public education (Harris and Handleman 2000; Jocelyn et al. 1998; Lord 1995; Mays and Gillon 1993; Prizant and Wetherby 1988; Siegel et al. 1988). However, such children are often referred for evaluation later than would be optimal. The average age at which caregivers first express concern to their pediatrician is 17 months, but the average age of diagnosis is significantly later, often as late as 4 years or even older (De Giacomo and Fombonne 1998; Flannagan and Nuallain 2001; Siegel et al. 1988). This may be due to several factors, including the fact that symptom presentation varies from case to case, social and language deficits may not be noted until children are observed interacting with peers, and screening and diagnostic instruments for young children are still relatively new and under development (Baron-Cohen et al. 1992, 1996; Bryson et al. 2003; Robins et al. 2001).

Another key element in the pattern of delayed diagnosis is doubt about the validity of very early diagnosis. Previously, the age for earliest reliable diagnosis was thought to be between 4 and 5 years (Howlin and Ashgarian 1999). However, a number of recent studies have shown that ASD can be accurately detected in children under the age of 3 years (Charman and Baird 2002; Cox et al. 1999; Eaves and Ho 2004; Lord 1995; Moore and Goodson 2003; Stone et al. 1999). Such studies have focused on: (1) stability of very early diagnosis, (2) applicability of specific instruments to very young children, and (3) whether DSM-IV symptoms (American Psychiatric Association 1994) apply to very young children. Table 1 presents numbers of children who retained, lost, or gained an ASD diagnosis in various studies. The percent stability reported in the last column represents the number of children whose diagnosis remained stable across the two evaluations (either ASD or non-ASD).

All of the studies described had high rates of diagnostic stability, with few children moving onto or off of the autism spectrum. Although failure to diagnose children at a very young age who then later develop autism is a serious concern, the notion of children no longer meeting criteria for a diagnosis of ASD is also of interest. Because of the small number of children in each of the previously described studies who lost an ASD diagnosis, it is impossible to do any more than speculate as to the reasons why this may have occurred. It could be that the original diagnosis was in error; several studies have noted that it is very difficult to distinguish children with ASD from children with severe global developmental delay at this age (Lord 1995). However, it is also possible that the initial diagnoses were accurate and that through intervention or because of specific characteristics of the children,

they moved off the autism spectrum to either another developmental disorder (e.g., language delay and global developmental disorder) or to no diagnosis.

With regard to specific diagnostic instruments or methods, the currently accepted “gold standard” for diagnosing children under the age of 5 years is clinical judgment (Volkmar et al. 2005), despite a number of empirically validated diagnostic measures, which include the Autism Diagnostic Interview-Revised (ADI-R; Lord et al. 1994), the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 1999), and the Childhood Autism Rating Scale (CARS; Schopler et al. 1988).

A number of current studies have addressed the applicability of these instruments, as well as the diagnostic criteria from the DSM-IV and ICD-10, to the diagnosis of very young children (see Table 1). One major criticism of using the ADI-R to diagnose toddler age children is that it only diagnoses Autistic Disorder, and not PDD-NOS. Since to show Autistic Disorder, a child must show deficits in all three areas (socialization, communication, and stereotyped/repetitive behaviors), and many young children only show problems in the first two areas, using the ADI-R may fail to diagnose such children with an ASD (Charman and Baird 2002; Cox et al. 1999; Ventola et al. 2006; Lord 1995; Stone et al. 1999). Conversely, many children with severe global developmental delay meet diagnostic criteria for autism on this measure because they are engaging in a number of repetitive mannerisms, even at a very young age (Lord 1995). A toddler version of the ADI-R is in use that contains additional questions relating to early childhood behaviors, however, the scoring criteria remain the same as for the standard ADI-R (Lord et al. 2004). Finally, there are no current sets of criteria for diagnosing either PDD-NOS or Asperger syndrome with the ADI-R. Probably for all of these reasons, a recent study on diagnosis of 2-year old (Ventola et al. 2006) found good agreement among clinical judgment, ADOS, and CARS, but insufficient agreement between the ADI and the other three measures. Cox et al. (1999) found that early diagnosis based on the ADI-R showed good specificity, in that all children diagnosed with autism at 20 months met diagnostic criteria at 42 months. However, it showed poor sensitivity in detecting autism at 20 months, in that a high proportion of children later found to have ASD were missed at 20 months.

This issue is less of a problem with other diagnostic measures, such as the CARS and the ADOS, because these measures allow for a diagnosis of PDD-NOS that does not require repetitive and stereotyped behaviors. This may help increase the negative predictive power of these measures for children in this age range as the presence of repetitive behaviors does not differentiate children with autism and children with other developmental delays (Baranek 1999; Lord et al. 1993; Osterling and Dawson 1994; Stone and Hogan 1993). Eaves and Ho (2004) reported that despite a change in individual scores on the CARS ranging from +14 to -14 points between the two evaluations, there was no change in mean scores and a correlation between CARS scores in the two evaluations was 0.76.

Several groups have found that symptoms from the DSM-IV diagnostic criteria may be either developmentally irrelevant in very young children, such as impaired conversational ability or stereotyped and repetitive use of language, or not often seen in very young children, such as inflexible adherence to routines and rituals (Charman and Baird 2002; Cox et al. 1999; Eaves and Ho 2004; Lord 1995; Moore and Goodson 2003; Stone et al. 1999). It is likely that these results reflect the developmental level of toddlers, and many of these behaviors may emerge later in the developmental course of autism. Despite the inappropriateness of these items for very young children, the DSM-IV diagnostic criteria remain the standard against which all other measures are compared. This may not be problematic for diagnoses of Pervasive Developmental Disorder-Not Otherwise Specified or ASD since a symptom from the category of restricted and repetitive behaviors is not

required. It may, however, create problems for the validity of Autistic Disorder since a behavior from this domain is required.

In summary, a number of studies have shown that diagnoses of Autistic Spectrum Disorder made at the age of 2 years are generally reliable and stable throughout the preschool years. Both the ADI-R and the DSM-IV diagnostic criteria are not fully applicable to children in this age range and may need to be adjusted based on clinical judgment of age appropriate behaviors. Samples have been rather small, and there are marked inconsistencies among studies in whether the source of instability is children moving onto the autism spectrum or children moving off the spectrum. In addition, they have not directly compared the diagnostic stability of different instruments.

The aim of the current study is to investigate the diagnostic stability of DSM-IV based clinical judgment between the ages of 2 and 4, and compare it to the stability of three additional diagnostic tools (ADI-R, ADOS, and CARS) in a large sample of children recruited from a study of early identification of ASD.

Methods

Participants

A total of 77 children participated in this study (66 males, 11 females; mean age Time 1: 2 years, 3 months, SD 5 months, range 1 year, 4 months–2 years, 11 months; mean age Time 2: 4 years, 5 months, SD 8 months, range 3 years, 5 months–6 years, 10 months). All children were part of a larger study aimed at developing an effective screening tool for ASD for toddler age children in both the general population and in high-risk populations, the Modified Checklist for Autism in Toddlers (M-CHAT; Robins et al. 2001). Individuals were screened between the ages of 16 and 30 months at (1) well-child visits with their primary-care provider ($n = 9$), (2) intake visits with an early intervention agency ($n = 67$), or (3) if they were the younger sibling of a child diagnosed with an ASD ($n = 1$). Children in the study came from Connecticut, Massachusetts, Rhode Island, and New York. Seventy-four children were Caucasian, one Asian, one African-American and one Puerto-Rican. All children lived in homes where English was the native language spoken.

Exclusionary criteria were: (a) they had already received a diagnosis of an ASD or other disorder (e.g., global developmental delay) prior to screening, (b) they were older than 30 months or younger than 16 months when their caregiver filled out the screener, (c) they had severe physical impairments that prevented the use of standardized evaluation instruments (e.g., blind, deaf, unable to sit independently), and (d) the family were not fluent in English (a sample of Spanish-speaking families is currently being collected).

Materials

Clinical judgment by experienced clinicians is considered to be the “gold standard” for autism diagnosis (Volkmar et al. 2005). In the current study, the clinicians used the DSM-IV criteria for Autistic Disorder (APA 1994) on which to base their clinical judgments. A child could be diagnosed with Autistic Disorder, with PDD-NOS, or as not on the autism spectrum; if the last, they could be given no diagnosis, or a diagnosis of language disorder, global developmental delay, or other condition. Asperger’s Disorder was not considered for the Time 1 diagnosis, given the young age of the children, and no children at age 4 received a diagnosis of Asperger’s.

The ADI-R (Lord et al. 1994) is a semi-structured clinician-based interview for caregivers that evaluates the child’s communication, social development, play, and restricted, repetitive, and stereotyped behaviors. The interview yields separate scores for each of the

three diagnostic domains. The child must meet the scoring criteria in each of the three domains in addition to demonstrating symptoms before age three, in order to meet diagnostic criteria for autism. The ADI-R has a scoring algorithm that is based on the DSM-IV criteria for autism that yields a classification of either autism or non-autistic; it does not diagnose PDD-NOS or other ASDs. For some children in the study, the Toddler version of the ADI replaced the ADI-R at Time 1 and the Short version of the ADI-R replaced the ADI-R at Time 2. See Tables 2 and 3 for the number of children receiving each version. The algorithm items remain the same for all three versions, and as only the algorithm items are used when making diagnostic judgments, the children receiving the different versions were collapsed for data analysis. The Toddler version contains additional questions relating to early childhood behaviors and eliminates questions designed for use with older children, for example those relating to certain peer interactions. The Short version contains a subset of questions from the ADI-R including all of the algorithm items.

The ADOS (Lord et al. 1999) is a semi-structured assessment of communication, social interactions and relatedness, play, imagination, and stereotyped or repetitive behaviors. The assessment consists of planned social interactions to encourage social initiations and responses. There are also opportunities to engage in imaginative play. This measure yields scores in the social domain, communication domain, and a combined score. Diagnostic classification is made by exceeding cut-off scores in these three areas (social, communication, and combined). A child can be classified as having Autistic Disorder or PDD-NOS, or as non-autistic.

The CARS (Schopler et al. 1988) consists of 15 items intended to measure the presence and severity of ASD. The child is rated on each item based on the clinician's observation of the child's behavior throughout the evaluation as well as on the parent's report. The CARS includes items on socialization, communication, emotional responses, and sensory sensitivities. The CARS classifies a child as having mild, moderate, or severe autism, or no autism, with a cutoff of 30 for the presence of autism. The cutoff of 30 was used in the present study to calculate diagnostic agreement between Times 1 and 2, but scores in the high 20s were considered consistent with a clinical judgment of PDD-NOS.

Vineland Adaptive Behavior Scales (VABS; Sparrow et al. 1984) is a widely used parent interview scale that assesses adaptive functioning in the areas of communication, daily living, socialization, and motor skills, as well as yielding an adaptive behavior composite score.

Mullen Scales of Early Learning (Mullen 1995) is a test of developmental functioning. The subscales of visual problem solving, fine motor skill, receptive and expressive language were administered. It also yields an early learning composite which is an estimate of overall developmental quotient.

Bayley Scales of Infant Development, Second Edition (Bayley 1993) is a test that measures mental and psychomotor development. It yields a developmental index score (mental developmental index) of the child's overall development.

Differential Abilities Scales (DAS; Elliot 1990) tests general cognitive functioning and generates scaled scores in the areas of non-verbal and Verbal abilities as well as a composite score (GCA) which is an estimate of full-scale intelligence.

Procedure

All children received the VABS, a semi-structured interview to review DSM-IV criteria, and a developmental, medical, and intervention history, at both time points. The CARS was also

administered at both time points. Seven children were missing the CARS at Time 1 and 12 children at Time 2; a total of 58 children had the CARS at both time points.

For the first 42 Time 1 evaluations, and 17 Time 2 reevaluations, children received the Bayley or the DAS, depending on age and functioning level. For the remaining 35 Time 1 evaluations and 60 Time 2 re-evaluations, the Bayley was replaced by the Mullen for all children within its age range (older children were given the DAS). For these 35 Time 1 and 60 Time 2 evaluations, the ADOS and the ADI-R were also added. Thirty children received the ADOS and ADI-R at both time points.

Families were recruited from one of the three sources listed previously. The child's caregiver filled out the M-CHAT between the ages of 16 and 30 months. If the child failed the screener and subsequent phone follow-up interview, they were invited to receive a free developmental and diagnostic evaluation. All children who received an initial evaluation were invited back to receive a follow-up evaluation between the ages of 42 and 54 months. However, some children did not receive the follow-up evaluation until after 54 months ($n = 28$) due to difficulties in scheduling, difficulty locating families, and parents initially declining the re-evaluation and then deciding to participate.

A team of investigators performed the evaluations at the University of Connecticut Psychological Services Clinic, including one licensed clinical psychologist (DF, SH, or MB), or developmental pediatrician (TDM) specializing in autism and one graduate student also experienced in autism assessment. One of the team members collected history information and completed the caregiver interviews while the other member of the team evaluated the child, usually in the same room. A blind assessment was not considered possible, as all children presenting for an evaluation had failed the M-CHAT, indicating some developmental concern, and at Time 2 many parents requested to see the same clinician as at Time 1. However, the graduate student testing and playing with the child at Time 2 was kept blind to Time 1 diagnosis whenever possible. Clinical DSM-IV based diagnosis was arrived at by discussion of the team.

Results

At Time 1, children were diagnosed as having (a) Autistic Disorder, $n = 46$, (b) Pervasive Developmental Disorder-Not Otherwise Specified, $n = 15$ or (c) having a non-autistic developmental delay or other condition (language delay $n = 10$; global developmental delay $n = 3$, or other $n = 3$). All children met criteria for some type of developmental disorder at the initial evaluation. Group characteristics are found in Table 2.

Analysis of variance (ANOVA) indicated significant group differences on all items except the Mullen scores and the VABS Motor scaled score. The Autistic Disorder group showed the most impairment and severity of symptoms.

At Time 2, children were diagnosed as having (a) Autistic Disorder $n = 34$, (b) PDD-NOS, $n = 12$, (c) a non-autistic developmental delay or other condition (language delay, $n = 7$; global developmental delay, $n = 6$, or other, $n = 8$), or (d) not meeting diagnostic criteria for any disorder, $n = 10$.

Analysis of variance indicated significant group differences on all measures; with the Autistic Disorder group showing the most impairment and severity of symptoms (see Table 3 for characteristics).

Diagnostic Stability

Clinical diagnoses based on *DSM-IV criteria* were stable over time. After combining the AD and PDD-NOS groups into an ASD group, 80% of the children remained in the same diagnostic category, with 46 of 61 children retaining an ASD diagnosis (see Table 4). When looking at specific ASD diagnoses, 46 were initially diagnosed with Autistic Disorder. Of these, 32 maintained their diagnosis (70%), whereas seven children moved from AD to PDD-NOS (15%), four from AD to a non-autistic developmental disorder or other condition (9%), and three from AD to no diagnosis (6%). Five of the 15 children who were initially diagnosed with PDD-NOS maintained their diagnosis (33%). Two children moved from PDD-NOS to AD (13%), five from PDD-NOS to a non-autistic developmental disorder or other condition (33%), and three from PDD-NOS to no diagnosis (20%). None of the children who were diagnosed with a non-autistic developmental disorder or other condition at the initial evaluation was diagnosed with an ASD at the re-evaluation.

Diagnostic stability for children who received the *ADI-R* was at 67% (see Table 5). Five children retained the diagnosis of autism between Times 1 and 2, six children no longer met criteria for the diagnosis, and four children gained the diagnosis. Fifteen children did not meet diagnostic criteria on the *ADI-R* at either time. Diagnostic stability using the *ADOS* was 83% when collapsing the categories of Autistic Disorder and PDD-NOS, with 21 children retaining an ASD diagnosis, five children no longer meeting criteria for the ASD diagnosis, zero children gaining an ASD diagnosis and four children receiving no autism diagnosis at either evaluation (see Table 6).

When breaking down the diagnostic groups, of the 22 children with an initial diagnosis of AD, 17 (77%) retained the diagnosis, 1 (5%) moved from AD to PDD-NOS and 4 (18%) moved from AD to no autism diagnosis. Of the four children with an initial diagnosis of PDD-NOS, three children (75%) moved from PDD-NOS to AD and 1 (25%) moved from PDD-NOS to no autism diagnosis.

On the *CARS*, diagnostic stability was 76% (see Table 7). Using the *CARS* cutoff of 30 to separate autism from non-autism, 42 were initially diagnosed with autism and 16 with non-autism. At Time 2, 30 children initially diagnosed (71%) retained a diagnosis of autism, and 12 children (20%) no longer met criteria for the diagnosis. Of those diagnosed with non-autism, 2 (12.5%) gained the diagnosis and 14 (87.5%) continued not to meet criteria for autism.

Discussion

In this study, 77 children were evaluated at two times: initially around the age of 2 years, and then again around 4 years. Diagnostic stability based on *DSM-IV* clinical judgment was high, with 80% of children remaining in the same diagnostic category at follow-up; this figure is quite consistent with what has been reported in previous studies with smaller samples. Fifteen children (20%) changed from having a diagnosis on the autism spectrum to either having a diagnosis of a non-autistic developmental disorder or having no diagnosis. Also consistent with other studies, a diagnosis of Autistic Disorder was more likely to persist than a diagnosis of PDD-NOS: 85% of children who received an initial diagnosis of Autistic Disorder retained a diagnosis on the autism spectrum at age 4, whereas 47% of children initially diagnosed with PDD-NOS retained an ASD diagnosis. Similarly, fewer children with an initial diagnosis of Autistic Disorder moved to meeting criteria for no developmental disorder than children initially diagnosed with PDD-NOS (6% vs. 20%). No children who were diagnosed as non-ASD at age 2 were diagnosed with ASD at follow-up, as we did not expect to identify any children who demonstrated regression after initial screening. Of the 46 children who retained a diagnosis of ASD, only two moved from PDD-

NOS to AD, whereas seven moved from AD to PDD-NOS. Thus, both in movement on and off the ASD spectrum, and movement within the spectrum, there was a general trend toward milder impairment between the ages of 2 and 4; symptomatology was more likely to improve than to worsen over time Fig. 1.

The exception to this trend was the emergence of repetitive behavior and resistance to change in some children. For the two children who moved from PDD-NOS to AD, both had no reported/observed problems with stereotyped/repetitive behaviors at the Time 1 evaluation but by the Time 2 evaluation began engaging in these types of behaviors, prompting the change to the AD diagnosis. Similarly, the children who moved from non-autistic to Autistic Disorder on the ADI-R generally did so because of the emergence of repetitive behaviors or resistance to change. In the studies summarized in Table 1, it is generally the case that studies that found a significant number of children moving onto the ASD spectrum used the ADI-R as a key diagnostic instrument, whereas the other studies found few or no children moving in this direction. Probably for this reason, the ADI-R had lower overall stability than the other three measures in the current study.

Although the children who gained an AD diagnosis generally did so because of the development of repetitive behaviors, the seven children who moved from an AD diagnosis to a diagnosis of PDD-NOS were heterogeneous in their behaviors and no clear pattern emerged to suggest a specific way in which symptoms were abating.

On the CARS, 2 children moved from having a score that was not on the spectrum at Time 1 to a score that was on the spectrum at Time 2. In each case, the Time 1 scores were within a few points of cutoff criteria for ASD and moved to just at cutoff criteria (from 26 to 30 and 27 to 30, respectively). This was not the case with the ADI-R; more children gained a diagnosis of autism for the specific reason that they developed repetitive and stereotyped behavior. Until a specific scoring algorithm is created for use with children in this age range, or scoring criteria for PDD-NOS are developed, it may be prudent to use the ADI-R in conjunction with other measures for children in this youngest age group.

Another factor that differentiated the ADI-R from the other instruments was that it was based on parent report only. The ADOS, in contrast, is based on direct child observation, whereas the CARS and DSM-IV-based clinical judgment incorporate information gleaned both from child observation and parent report. Although parents are generally excellent reporters, especially when asked about specific behaviors, and although parent report is crucial in providing information that may not be directly observable in an evaluation (e.g., reactions to peers), there were a few instances in this study where parents denied the existence of either a pathological behavior or a skill that was directly observed during the evaluation. Therefore, supplementing or modifying the ADI-R results with direct observation will result in a more accurate diagnosis. We conclude, therefore, that diagnostic stability of ASD diagnoses is high between ages 2 and 4, especially when using DSM-IV-based clinical judgment, ADOS, or CARS. Stability is also good for the ADI-R, except for children who do not meet criteria for Autistic Disorder because of few or no symptoms in the repetitive behavior domain. Were an algorithm for PDD-NOS to be derived from the ADI-R, stability would likely be high. When children move diagnostically, it is likely to be in a positive direction, both with and across the boundaries of the autism spectrum.

Sutera et al. (2007) examined in detail the children who moved off the autism spectrum between ages 2 and 4, to determine whether child characteristics at age 2 could predict this diagnostic change. Contrary to expectations, the children who retained the ASD diagnosis were very similar to those who moved off the spectrum on most variables, including receptive and expressive language, non-verbal problem solving, and number and severity of

DSM-IV autism symptoms. Motor development, however, both by parent report and child testing, was different, with the improved outcome children showing significantly better motor development at age 2.

Several limitations to the current study should be noted. First, a truly blind assessment at Time 2 would be preferable. This was not feasible in the current study, partly because parents sometimes requested to see the same clinician, and because feedback to parents involved comparing developmental and diagnostic results between the two assessments. However, developmental and ADOS testing was done by student clinicians who were generally blind to initial assessment results, and who followed specific behavioral rules for diagnosis. As reported by Ventola et al. (2006) for the current sample of children, reliability between the ADOS, scored by the student clinicians, and the DSM-IV-based clinical judgment of the experienced clinician, was very high.

Another limitation was that the sample was drawn from Connecticut and surrounding areas in Massachusetts, Rhode Island, and New York. Early intervention services are more intensive in these areas than in many parts of the US and other countries. Future analyses will examine the relationship between type and intensity of intervention and outcome in our longitudinal sample, but the extent to which intervention played a role in the generally positive movement of the children in our sample is not yet known. In addition, the extent to which our findings can be generalized to other geographic areas is not clear.

A final question which is raised by the present findings is the youngest age at which a reliable diagnosis can be made. The mean age of our sample was 27 months, with children as young as 16 months; can reliable diagnosis be made earlier than that, and what percent of children as young as 16–17 months can be reliably diagnosed? To some extent, this will depend on the degree of developmental delay; we feel that children with all skills (motor, language, and cognitive) under a 12-month developmental level cannot be diagnosed with ASD using DSM-IV criteria. Should children under the age of 16–17 months chronological age, or 12 months developmental age, be diagnosed with ASD, using criteria developed specifically for this age group, or should the question of possible autistic development be raised without a firm diagnosis being made? These are questions for ongoing longitudinal, prospective studies of children at high risk for developing ASD to consider.

Acknowledgments

We gratefully acknowledge the assistance of participating pediatricians and staff in their offices, as well as providers in the Connecticut and Massachusetts Birth-to-Three Early Intervention systems. We would like to thank Jillian Wood, Executive Director of the Hezekiah Beardsley Chapter of the American Academy of Pediatrics, who assisted in recruiting participating physician offices and promoted the study state-wide, as well our dedicated undergraduate research assistants, and the children and families who participated. We would also like to extend sincere thanks to members of the Early Detection Advisory Board, especially Ho-Wen Hsu, MD, and Mark Greenstein, MD, for their wise advice and support. Graduate students Leandra Berry, Hilary Boorstein, Emma Esser, Saasha Suter, and Mike Rosenthal, and Gail Marshia, Project Coordinator, were invaluable in conducting all aspects of the study. This paper was prepared from the masters thesis of Jamie M. Kleinman at the University of Connecticut. This study is supported by NIH grant R01 HD039961 and Maternal and Child Health Bureau grant R40 MC00270 and prior grants from the National Association for Autism Research, NIMH, and the Department of Education.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders (4th ed., DSM-IV). Washington, DC: Author; 1994.
- Baranek GT. Autism during infancy: A retrospective video analysis of sensory-motor and social behaviors at 9–12 months of age. *Journal of Autism and Developmental Disorders*. 1999; 29:213–224. [PubMed: 10425584]

- Baron-Cohen S, Allen J, Gillberg C. Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *British Journal of Psychiatry*. 1992; 161:839–843. [PubMed: 1483172]
- Baron-Cohen S, Cox A, Baird G, Swettenham J, Nightengale N, Morgan K, Drew A, Charman T. Psychological markers in the detection of autism in infancy in a large population. *British Journal of Psychiatry*. 1996; 168:158–163. [PubMed: 8837904]
- Bayley, N. *The Bayley scales of infant development*. 2nd ed. San Antonio, TX: Harcourt Brace; 1993.
- Bryson S, Rogers S, Fombonne E. Autism spectrum disorders: Early detection, intervention, education, and psychopharmacological management. *Canadian Journal of Psychiatry*. 2003; 48:506–516.
- Charman T, Baird G. Practitioner review: Diagnosis of autism spectrum disorder in 2- and 3-year-old children. *Journal of Child Psychology and Psychiatry*. 2002; 43:289–305. [PubMed: 11944873]
- Cox A, Klein K, Charman T, Baird G, Baron-Cohen S, Swettenham J, Drew A, Wheelwright S. Autism spectrum disorders at 20 and 42 months of age: Stability of clinical and ADI-R diagnosis. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. 1999; 40:719–732.
- De Giacomo A, Fombonne E. Parental recognition of developmental abnormalities in autism. *European Journal of Child and Adolescent Psychiatry*. 1998; 7:131–136.
- Eaves L, Ho H. Brief report: Stability and change in cognitive and behavioral characteristics of autism through childhood. *Journal of Autism and Developmental Disorders*. 2004; 26:557–569. [PubMed: 8906456]
- Elliott, CD. *Differential ability scales (DAS)*. San Antonio, TX: The Psychological Corporation; 1990.
- Flannagan O, Nuallain S. A study looking at the effectiveness of developmental screening in identifying learning disabilities in early childhood. *Irish Medical Journal*. 2001; 94:148–150. [PubMed: 11474856]
- Gillberg C, Ehlers S, Schaumann H, Jakobsson G, Dahlgren S, Lindblom R, Bagenholm A, Tjuus T, Glidner E. Autism under age 3 years: A clinical study of 28 cases referred for autistic symptoms in infancy. *Journal of Child Psychology and Psychiatry*. 1990; 31:921–934. [PubMed: 2246342]
- Harris S, Handleman J. Age and IQ at intake as predictors of placement for young children with autism: A four to six-year follow-up. *Journal of Autism and Developmental Disorders*. 2000; 30:137–142. [PubMed: 10832778]
- Howlin P, Ashgarian A. The diagnosis of autism and Asperger syndrome: Findings from a systematic survey. *Developmental Medicine and Child Neurology*. 1999; 41:834–839. [PubMed: 10619282]
- Jocelyn L, Casiro O, Beattie D, Bow J, Kneisz J. Treatment of children with autism: A randomized controlled trial to evaluate a caregiver-based intervention program in community day-care center. *Journal of Developmental and Behavioral Pediatrics*. 1998; 19:326–334. [PubMed: 9809262]
- Lord C. Follow-up of two-year-olds referred for possible autism. *Journal of Child and Adolescent Psychiatry*. 1995; 36:1365–1382.
- Lord C, Shulman C, DiLavore P. Regression and word loss in autistic spectrum disorders. *Journal of Child Psychology and Psychiatry*. 2004; 45:936–955. [PubMed: 15225337]
- Lord C, Storoschuk S, Rutter M, Pickles A. Using the ADI-R to diagnose autism in preschool children. *Infant Mental Health Journal*. 1993; 14:234–252.
- Lord C, Rutter M, Le Couteur A. Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*. 1994; 24:659–685. [PubMed: 7814313]
- Lord, C.; Rutter, M.; DiLavore, P.; Risi, S. *Autism diagnostic observation schedule-WPS edition*. Los Angeles, CA: Western Psychological Services; 1999.
- Mays R, Gillon J. Autism in young children: An update. *Journal of Pediatric Health Care*. 1993; 7:17–23. [PubMed: 8421239]
- Moore V, Goodson S. 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*. 2003; 7:47–63. [PubMed: 12638764]
- Mullen, E. *Mullen scales of early learning*. Circle Pines, MN: American Guidance Service; 1995.
- Osterling J, Dawson G. Early recognition of children with autism: A study of first birthday home videotapes. *Journal of Autism and Developmental Disorders*. 1994; 24:247–257. [PubMed: 8050980]

- Prizant B, Wetherby A. Providing services to children with autism (ages 0 to 2 years) and their families. *Focus on Autistic Behavior*. 1988; 4:1–16. (Reprinted from (1988). *Topics in Language Disorders*, 9, 1–23).
- Robins D, Fein D, Barton M, Green J. The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*. 2001; 31:131–144. [PubMed: 11450812]
- Schopler, E.; Reichler, R.; Renner, B. *The childhood autism rating scale*. Los Angeles: Western Psychological Services; 1988.
- Siegel B, Pliner C, Eschler J, Elliott G. How children with autism are diagnosed: Difficulties in identification of children with multiple developmental delays. *Journal of Developmental and Behavioral Pediatrics*. 1988; 9:199–204. [PubMed: 2464003]
- Sparrow, S.; Balla, D.; Cicchetti, D. *The vineland adaptive behavior scales*. Circles Pines, MN: American Guidance Services; 1984.
- Stone W, Hogan K. A structured parent interview for identifying young children with autism. *Journal of Autism and Developmental Disorders*. 1993; 23:639–652. [PubMed: 8106304]
- Stone W, Lee E, Ashford L, Brissie J, Hepburn S, Coonrod E, Weiss B. Can autism be diagnosed accurately in children under 3 years? *Journal of Child Psychology and Psychiatry and Allied Disciplines*. 1999; 40:219–226.
- Sutera S, Pandey J, Esser E, Rosenthal M, Wilson L, Barton M, Green J, Hodson S, Robins R, Dumont-Mathieu T, Fein D. Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2007; 37:98–107. [PubMed: 17206522]
- Ventola P, Kleinman B, Pandey J, Barton M, Allen S, Green J, Robins D, Fein D. Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders*. 2006; 36:839–847. [PubMed: 16897398]
- Volkmar F, Chawarska K, Klin A. Autism in infancy and early childhood. *Annual Review of Psychology*. 2005; 56:315–336.

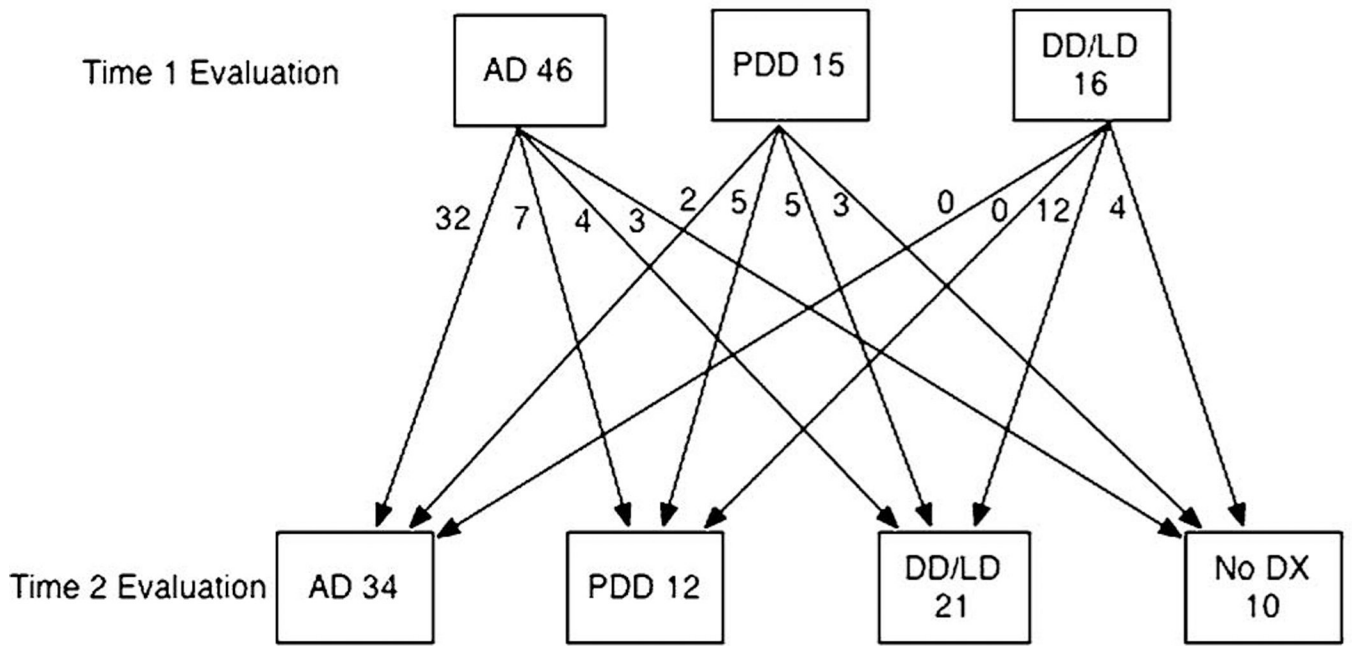


Fig. 1. Movement between diagnostic categories between Times 1 and 2 evaluations

Table 1

Summary of diagnostic stability studies in very young children

Study	Age at Time 1	Age at Time 2	Total N	Measures used	DX Time 1	DX Time 2	N	% stability
Gillberg et al. (1990) ^a	23.9	57.7	28	Autism behavior checklist	AD	AD	21	75%
Lord (1995) ^b	30.5	50.3	30	Griffiths scales of infant development DSM-III-R	PDD-NOS	PDD-NOS	2	
					PDD-NOS	Non-ASD	2	
					Non-ASD	Non-ASD	3	
					Autism	Autism	14	88%
Cox et al. (1999) ^b	20.6	45.3	44	ADJ-R DSM-IV	Autism	Non-autism	2	
					Non-autism	Autism	12	
					Non-autism	Non-autism	2	
					Autism	Autism	7	81%
					Autism	ASD	1	
					Autism	Non-ASD	1	
					ASD	Autism	1	
ASD	ASD	2						
Stone et al. (1999) ^b	31.4	45.0	45	Reynell developmental language scales	Non-ASD	ASD	7	
					Non-ASD	Non-ASD	24	
					Autism	Autism	22	81%
					Autism	PDD-NOS	1	
					Autism	Non-PDD	2	
					PDD-NOS	Autism	3	
					PDD-NOS	PDD-NOS	5	
					PDD-NOS	Non-PDD	4	
					Non-PDD	PDD-NOS	1	
					Non-PDD	Non-PDD	7	
Eaves and Ho (2004) ^c	33	59	49	PDDST CHAT Bayley (Time 1) WPPSI-R, Leiter, Bayley, SB, Mullen (Time 2)	Autistic	Autistic	31	79%
					Autistic	PDD-NOS	2	
					Autistic	Not autistic	1	
					PDD-NOS	Autistic	5	

Study	Age at Time 1	Age at Time 2	Total N	Measures used	DX Time 1	DX Time 2	N	% stability
Moore and Goodson (2004) ^d	34	53	20	Vineland	PDD-NOS	PDD-NOS	2	
				CARS	PDD-NOS	Not autistic	2	
				DSM-IV	Not autistic	Not autistic	6	
				Intervention HX				
				ADI-R	Autism	Autism	14	95%
				ICD-10	Autism	Atypical autism	2	
				Griffiths mental disabilities scale	Atypical autism	Autism	2	
				Observation	Atypical autism	Atypical autism	1	
					Non-autism	Non-autism	1	

^aFrom Table 1, Gillberg et al. (1990)

^bFrom Table 2, Charman and Baird (2002)

^cFrom Fig. 1, Eaves and Ho (2004)

^dFrom Table 1, Moore and Goodson (2004)

Table 2

Mean scores on measures at Time 1

	Autistic (<i>n</i> = 46)	PDD-NOS (<i>n</i> = 15)	Non-autistic DD (<i>n</i> = 16)	<i>F</i>
Age at evaluation (in months)	27.06 (6.43)	24.74 (2.89)	27.44 (3.95)	1.39
ADI-R total ^d (<i>n</i> = 32)	30.83 (6.77) ^b	28.57 (5.82)	18.33 (1.15) ^b	5.10 [*]
ADOS (AB score) (<i>n</i> = 34)	18.1 (2.98) ^{a,b}	10.87 (5.91) ^a	6 (4.24) ^b	20.13 ^{**}
CARS (<i>n</i> = 70)	36.05 (4.69) ^{a,b}	30.30 (5.44) ^{a,c}	22.64 (4.29) ^{b,c}	42.11 ^{**}
DSM-IV number of symptoms	7.33 (1.66) ^b	4.50 (1.41)	1.44 (1.74) ^b	9.50 ^{**}
FSIQ (Bayley or Mullen) ^e	53.24 (7.34) ^b	62.33 (16.19)	71.06 (25.20) ^b	8.56 ^{**}
Mullen VR	24 (6.24)	28 (10.84)	35.75 (18.40)	2.65
Mullen FM	25.4 (8.93)	27.57 (14.52)	31.25 (18.13)	0.45
Mullen RL	20 (0)	24.85 (8.47)	28.0 (13.85)	4.05
Mullen EL	21.55 (3.44)	28.57 (9.65)	24.66 (5.03)	4.20
Mullen ELC	54.31 (7.52)	62.84 (19.55)	68.50 (13.07)	1.76
Bayley MDI	52.31 (7.22) ^b	61.33 (13.60)	74.33 (26.81)	7.01 ^{**}
VABS Com	62.81 (5.76) ^b	67.50 (6.41) ^c	76.68 (13.25) ^b	17.50 ^{**}
VABS DL	66.34 (6.07) ^b	68.85 (6.54)	75.75 (12.35) ^b	8.316 ^{**}
VABS Soc	64.70 (7.60) ^b	68.71 (8.24) ^c	81.00 (8.80) ^b	24.41 ^{**}
VABS Mot	78.15 (12.25)	81.64 (9.70)	86.62 (12.74)	3.01
VABS ABC	62.06 (5.33) ^b	66.07 (5.18) ^c	74.93 (10.44) ^b	21.55 ^{**}

* $p < 0.05$ ** $p < 0.01$ ^aGroup difference between autistic and PDD-NOS^bGroup difference between autistic and non-autistic DD^cGroup difference between PDD-NOS and non-autistic DD^d16 children received the ADI-R and 16 children received the Toddler version of the ADI-R^e42 children received the Bailey and 35 children received the Mullen

Table 3

Mean scores on measures at Time 2

	Autistic (n = 34)	PDD-NOS (n = 12)	Non-autistic DD (n = 21)	No diagnosis (n = 10)	F
Age at evaluation	51.07 (6.80)	53.29 (5.53)	52.88 (7.77)	58.60 (12.24)	2.41*
ADI-R (n = 50)	25.33 (13.86) ^b	10.40 (9.07)	7.81 (7.50) ^b	14.28 (14.20)	5.92**
ADOS (n = 47)	17.40 (2.97) ^{a,b,c}	6.25 (4.11) ^a	4.00 (3.16) ^b	1.71 (2.05) ^c	72.37**
CARS (n = 66)	35.50 (4.61) ^{a,b,c}	26.61 (3.71) ^{a,d,e}	20.96 (4.01) ^{b,d}	16.72 (1.37) ^{c,e}	70.77**
DSM-IV number of symptoms	7.38 (1.38) ^{a,b,c}	5.50 (2.72) ^{a,d,e}	1.76 (1.48) ^{b,d}	1.28 (1.25) ^{c,e}	46.85**
FSIQ (Mullen or DAS) ^f	56.34 (16.58) ^{a,b,c}	86.37 (19.24) ^a	78.93 (17.64) ^b	107.6 (14.89) ^c	24.25**
Mullen VR	27.19 (13.54) ^c	46.66 (5.50)	42.00 (19.58) ^f	64.50 (6.59) ^{c,f}	11.70**
Mullen FM	27.30 (12.19) ^{a,c}	48.25 (19.92) ^a	32.22 (12.30) ^f	56.50 (11.25) ^{c,f}	9.41**
Mullen RL	25.76 (10.83) ^c	37.75 (20.56)	34.77 (12.36)	49.50 (9.02) ^c	6.50**
Mullen EL	24.66 (9.73) ^{b,c}	37.00 (16.39)	36.11 (9.54) ^b	45.83 (3.31) ^c	8.69**
Mullen ELC	58.04 (17.94) ^{a,c}	94.33 (19.55) ^a	75.33 (18.99) ^f	111.66 (12.51) ^{c,f}	16.17**
DAS V	51.20 (2.68) ^{a,b,c}	76.00 (20.46) ^{a,e}	82.00 (9.61) ^b	101.50 (8.58) ^{c,e}	13.13**
DAS NV	50.40 (8.29) ^{a,b,c}	90.40 (17.27) ^a	89.66 (20.76) ^b	99.75 (22.09) ^c	7.29**
DAS SS	49.60 (5.59) ^{a,b,c}	81.60 (19.52) ^a	84.33 (15.37) ^b	101.50 (17.93) ^c	9.19**
VABS Com	57.31 (10.69) ^{a,b,c}	81.72 (17.92) ^{a,e}	84.20 (18.95) ^b	99.40 (15.62) ^{c,e}	26.78**
VABS DL	56.12 (7.83) ^{b,c}	64.18 (7.48) ^e	70.80 (14.48) ^{b,f}	91.60 (17.37) ^{c,e,f}	26.43**
VABS Soc	58.87 (7.81) ^{a,b,c}	73.18 (10.28) ^{a,e}	81.00 (17.63) ^b	92.60 (9.98) ^{c,e}	28.40**
VABS Mot	66.93 (16.64) ^c	76.36 (14.26) ^e	78.80 (16.95) ^f	97.37 (10.26) ^{c,e,f}	8.49**
VABS ABC	54.85 (9.28) ^{a,b,c}	68.54 (10.21) ^{a,e}	73.84 (14.93) ^{b,f}	92.70 (12.66) ^{c,e,f}	29.23**

* ns

** p < 0.01

^a Group difference between autistic and PDD-NOS

^b Group difference between autistic and non-autistic DD

^c Group difference between autistic and no diagnosis

d Group difference between PDD-NOS and non-autistic DD

e Group difference between PDD-NOS and no diagnosis

f Group difference between non-autistic DD and no diagnosis

g 22 children received the DAS and 55 children received the Mullen

Table 4

Diagnostic stability of DSM-IV (80%)

DSM-IV Time 1	DSM-IV Time 2	
	Non	ASD
Non	16	0
ASD	15	46

Table 5

Diagnostic stability of ADI-R (67%)

ADI-R Time 1	ADI-R Time 2	
	Non	Autism
Non	15	4
Autism	6	5

Table 6

Diagnostic stability of ADOS (83%)

ADOS Time 1	ADOS Time 2	
	Non	ASD
Non	4	0
ASD	5	21

Table 7

Diagnostic stability of CARS (76%)

CARS Time 1	CARS Time 2	
	Non	ASD
Non	14	2
ASD	12	30