

Autism Spectrum Disorders at 20 and 42 Months of Age: Stability of Clinical and ADI-R Diagnosis

Antony Cox

Bloomfield Clinic and Newcomen Centre,
Guy's, King's College, and St Thomas'
Medical School, London, U.K.

Kate Klein

West London Healthcare NHS Trust,
Ealing Hospital, London, U.K.

Tony Charman

Institute of Child Health,
University College, London, U.K.

Gillian Baird

Bloomfield Clinic and Newcomen Centre,
Guy's, King's College, and St Thomas'
Medical School, London, U.K.

Simon Baron-Cohen

University of Cambridge, U.K.

John Swettenham

University College London, U.K.

Auriol Drew

Bloomfield Clinic and Newcomen Centre,
Guy's, King's College, and St Thomas'
Medical School, London, U.K.

Sally Wheelwright

University of Cambridge, U.K.

The association between, and stability of, clinical diagnosis and diagnosis derived from the Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) was examined in a sample of prospectively identified children with childhood autism and other pervasive developmental disorders assessed at the age of 20 months and 42 months. Clinical diagnosis of autism was stable, with all children diagnosed with childhood autism at age 20 months receiving a diagnosis of childhood autism or a related pervasive developmental disorder (PDD) at age 42 months. Clinical diagnosis of childhood autism was also reasonably sensitive, with all children who went on to receive a clinical diagnosis of childhood autism at 42 months being identified as having autism or PDD at 20 months. However, clinical diagnosis for PDD and Asperger's syndrome lacked sensitivity at 20 months, with several children who subsequently received these diagnoses at 42 months receiving diagnoses of language disorder or general developmental delay, as well as in two cases being considered clinically normal, at the earlier timepoint. The ADI-R was found to have good specificity but poor sensitivity at detecting childhood autism at 20 months; however, the stability of diagnosis from 20 to 42 months was good. In addition, the ADI-R at age 20 months was not sensitive to the detection of related PDDs or Asperger's syndrome. The continuity and discontinuity between behavioural abnormalities identified at both timepoints in the three domains of impairment in autism was examined, both in children who met final clinical criteria for an autistic spectrum disorder, and for children with language disorder who did not, as well as for a small sample of typically developing children.

Keywords: Autism, pervasive developmental disorder, diagnosis, Autism Diagnostic Interview-Revised, stability.

Abbreviations: ADI-R: Autism Diagnostic Interview-Revised; CHAT: Checklist for Autism in Toddlers; LD: language disorder; NVMA: nonverbal mental age; PDD: pervasive developmental disorder.

Introduction

Disorders on the autistic spectrum are increasingly well established in clinical practice, but continue to elude precise neurobiological definition (Bailey, Phillips, & Rutter, 1996). Clinical assessment and classification has improved over the last decade (DSM-IV; American Psychiatric Association, 1994; ICD-10; World Health Organisation, 1993) but for most children with autism diagnosis is still delayed until the age of 3 years or later, even though it is now widely accepted that abnormalities in social behaviour are present in these children before the end of the second year of life (Gillberg et al., 1990; Rogers & DiLalla, 1990). The opportunities offered by early identification include earlier treatment, educational planning, implementation of professional support services, and genetic counselling, but the theoretical benefits of these strategic interventions have not until recently been demonstrated. However, there are now a number of studies reporting substantial benefits from early treatment programmes in terms of improved social behaviours and reduced abnormal behaviour (Hoysen, Jamieson, & Strain, 1984; Lovaas, 1987; Ozonoff & Cathcart, 1998; Rogers & Lewis, 1989; Sheinkopf & Siegel, 1998; see Dawson & Osterling, 1997; Rogers, 1996, for reviews).

In the light of these findings, the stability of early diagnosis of autistic spectrum disorders has become a key focus for research. This is important because the benefits of early intervention must be weighed against the potential hazards of a false positive diagnosis. A number of authors have now identified early markers of autism in retrospective studies using parental report (Dahlgren & Gillberg, 1989; Gillberg et al., 1990; Stone, Hoffman, Lewis, & Ousley, 1994) and video recordings of children made before diagnosis (Adrien et al., 1992; Osterling & Dawson, 1994). These findings converge on abnormalities in the use of eye contact and other early nonverbal communication behaviours, often characterised as "joint attention" behaviours, which are universally present in normally developing children by the age of 18 months (e.g. Butterworth, 1991), as key early indicators of autistic disturbance. For example, Dahlgren and Gillberg (1989) found that empty gaze and deficits in directing attention discriminated between children with and without autism in the first 2 years of life. Osterling and Dawson (1994), looking at videos of infants' first birthday parties, found that children later diagnosed as having autism showed differences in several social behaviours compared with other children, including looking at faces, pointing, orienting to name, and showing objects. These findings offer endorsement of the developmental continuity of joint attention impairments in children with autism from an early age (see Charman, 1998, for a review).

In contrast, several authors have found the unusual motor mannerisms and rigid and repetitive behaviours, characteristic of older children with autism, to be notably absent in more than half of the young autistic children studied, even by the age of 4 years (Dahlgren & Gillberg, 1989; Siegel, Pliner, Eschler, & Elliott, 1988; Stone et al., 1994; Stone & Hogan, 1993). However, in another study, abnormalities in these behaviours have been reported in children aged $2\frac{1}{2}$ years (Lord, 1995), and other retro-

spective studies have identified abnormal responses to sensory stimulation (Dahlgren & Gillberg, 1989) and an increased incidence of sensory and repetitive abnormalities (Ornitz, Guthrie, & Farley, 1977) to be present in the first few years of life (see Charman, in press; Stone, 1997, for reviews).

Clinical diagnosis of autism in very young children has been reported previously (Gillberg et al., 1990), and one recent study (Stone et al., 1999) has demonstrated both good (between-clinician) reliability of diagnosis at age 2 and high stability of diagnosis from age 2 to age 3 years, at least when an autism spectrum disorder approach is adopted. However, the reliability of standardised diagnostic schedules at this age remains to be demonstrated. Standardised instruments offer a rich and systematic source of data from which to gain information about the development of autism spectrum disorders in children. In a longitudinal study of thirty 2-year-olds referred for possible autism, Lord (1995) found that the Autistic Diagnostic Interview-Revised (ADI-R; Lord et al., 1994) both over-diagnosed and under-diagnosed autism, producing a false positive rate of 9/30 (30%), and a similar proportion of false negatives 8/30 (27%) compared to clinical diagnosis at follow-up 12 to 15 months later. Clinical assessment of the same group at this age proved more reliable, with no false positives, and only 2/30 (7%) false negatives. In a separate study, Lord and colleagues (Lord, Storoschuk, Rutter, & Pickles, 1993) found that the ADI-R had a high sensitivity and specificity in differentiating preschool children with autism from those with developmental delays when the children had mental ages above 18 months (all but 1 of the 51 children with autism were identified and only 2 of 30 with developmental delay were incorrectly identified). However, in children with mental ages below 18 months, the ADI-R incorrectly identified 60% of children with developmental delay as meeting established ADI-R criteria for autism. In contrast, Pilowsky, Yirmiya, Shulman, and Dover (1998) found that the ADI-R under-diagnosed autism in the youngest and lowest mental age subjects in their sample. These studies suggest that the ADI-R can only be used with caution with children in the first few years of life since it may both over- and under-diagnose autism in very young or low-functioning subjects (see also Lord et al., 1997).

Further developing the study of early indicators with a prospective design, Baron-Cohen, Allen, and Gillberg (1992) correctly identified 4 children with autism from a total sample of 91 using a screening questionnaire, the Checklist for Autism in Toddlers (CHAT) at 18 months of age. The sample included a genetic high-risk group of younger siblings of children already diagnosed as autistic and a group of controls. The questionnaire was designed to identify delay or impairment in joint attention and pretend play skills. The four children later diagnosed as autistic were the only ones who failed two or more of the five items measuring *interest in other children (A2)*, *playing social games such as peek-a-boo (A4)*, *pretend play (A5)*, *protodeclarative pointing (A7)*, and *bringing objects to show others (A9)* in the parent report section of the CHAT (Section A). These children also failed the *pretend play* and *protodeclarative pointing* items in the practitioner observation section of the CHAT (Section B). Of the

remaining 87 children, a substantial minority failed *one* of these behaviours at 18 months, most commonly *pretend play* (14%), followed by *protodeclarative pointing* (8%) and social interest (6%), but none failed more than one of these, and all were developing normally when reviewed at 30 months of age.

Using the same instrument in a population study, we demonstrated that this procedure could be applied to community screening (Baron-Cohen et al., 1996). The use of the instrument in this study differed slightly from that in the earlier high-risk study (Baron-Cohen et al., 1992). In the 1996 study the five key items of the CHAT—that is, those that were considered the best indicators of risk for developing autism—were those that assessed *pretend play* (A5 + Biii), *protodeclarative pointing* (A7 + Biv), and *gaze-monitoring* (Bii) across *both* the parent report and practitioner observation sections of the CHAT, since it was considered that confirmation of parental report was likely to reduce the incidence of screen false positives. In fact, all four of the children with autism in the Baron-Cohen et al. (1992) study failed all five of these key items. Twelve children out of 17,000 screened at 18 months were identified as being at high risk for autism by failing all 5 key items, and 11 of these were confirmed as having childhood autism (8) or other PDD (ICD-10) (3) at 42 months, with the remaining child having a language disorder (Baron-Cohen et al., 1996). In the Baron-Cohen et al. (1996) paper we reported that 10 out of the high risk for autism toddlers received an autism (spectrum) diagnosis at age 20 months, confirmed at age 42 months. Note that the criteria for diagnosis differ between the present study and Baron-Cohen et al. (1996) paper. In the latter, diagnosis at 20 months was arrived at from consensus between two out of three independent methods (two independent clinician ratings and a modified ADI-R algorithm). In addition, one diagnosis made at 42 months was subsequently revised in the light of a full review of clinical and historical information available at this time to arrive at the figures we now present (see below). These changes account for the slight discrepancies between the two papers.

As part of the assessment following screening, the ADI-R was administered at two timepoints, when the children were approximately 20 and 42 months old. The present paper presents data from the ADI-R at both timepoints and this data is used to address four questions. First, how sensitive are clinical diagnosis and the ADI-R in identifying autism spectrum disorders in our sample at 20 months? Second, how stable are ADI-R and clinical diagnoses in our sample from 20 to 42 months? Third, are there differences between the sensitivity, specificity, and stability, using both clinical judgement and the ADI-R, in the diagnosis of “core” childhood autism and of related PDDs at 20 months? Fourth, are the predictive features on the ADI-R at 20 months in our sample, obtained prospectively from screening a community population, the same as those reported previously for referred 2-year-olds (Lord, 1995)?

Method

Participants

A population of 17,173 children from 9 districts in the South

East Thames Health Region, U.K., were screened by Health Visitors using the CHAT at their routine 18-month developmental check (mean age at screening 18.7 months, *SD* 1.1 months). The social class distribution of this population was broadly representative of the U.K. (Economic Activity of Great Britain, 1981), and the sex ratio was 1.05:1 (male:female). Children with profound developmental delay, gross physical disability, or those already recognised as having a mental handicap were excluded from the screening sample, since Health Visitors were reluctant to impose additional assessment on parents whose children were already identified as having significant developmental difficulties. If a child failed two to five of the key items a second CHAT screening was conducted 1 month later, and children who consistently failed the key items on both administrations were subdivided into three groups: one predicted to be at high risk for autism (Group 1), one predicted to be at medium risk for autism (Group 2), and the third predicted to be developmentally normal (Group 3).

Group 1 included those children failing all five key items on the CHAT, i.e. protodeclarative pointing (A7 + Biv), gaze monitoring (Bii), and pretend play (A5 + Biii); *N* = 12.

Group 2 included those children failing the two key items relating to protodeclarative pointing (A7 + Biv), but who did not meet criteria for Group 1 (i.e. they passed at least one of the other items; A5, Bii, Biii); *N* = 44.

Group 3 included those children who did not meet criteria for Groups 1 + 2; *N* = 17,117.

Due to resource limitations only a small sample of children could be seen for detailed clinical assessment, and for pragmatic reasons we concentrated resources on those children who were considered at most risk of developing autism or a PDD. The sample was selected as follows: all 12 children in the high risk for autism group (Group 1) were seen, half of the children (22) selected at random from the medium risk for autism group (Group 2), and 16 children randomly selected from the no risk for autism group (Group 3) were seen for diagnostic assessment at age 20 months, and 49 of the 50 were followed up at age 42 months. The present paper is concerned solely with the sensitivity and stability of clinical diagnosis and the use of the ADI-R with the prospectively identified sample. See Baird et al. (1999) for details of the follow-up of the screened population and for data on the sensitivity and specificity of the CHAT screen itself.

Assessment Procedure

All 50 children underwent the same assessment procedure at age 20 months, which included a parental interview using the ADI-R (Lord et al., 1994), clinical assessment using a structured schedule of elicited child–investigator interaction, psychometric assessment using the Griffiths Scale of Infant Development (Griffiths, 1986) or Leiter International Performance Scale (Leiter, 1952), and language assessment using the Reynell Developmental Language Scales (Reynell, 1985). The same assessment procedure was repeated at 42 months.

The ADI-R is a semistructured, standardised diagnostic interview that includes questions relevant to past and current functioning of preschool children referred for possible autism (Lord, 1995; Lord et al., 1993, 1994). An ADI-R diagnosis of autism is conferred on the basis of an algorithm that is scored on three dimensional clusters of items: *qualitative impairments in reciprocal social interaction* (Dimension B), *impairments in verbal and nonverbal communication* (Dimension C), and *repetitive behaviours and stereotyped patterns* (Dimension D) (see Lord et al., 1994, for details). The algorithm specifies that a child must reach cutoff scores of 10 on Dimension B, a verbal score of 8 or a nonverbal score of 7 on Dimension C (verbal scores are only used for subjects with a sufficient overall level of

language), a score of 3 on Dimension D, and show evidence of abnormality before age 36 months, to receive an ADI-R diagnosis of autism.

The ADI-R was not taken as the sole arbiter for the diagnosis of autism, since the instrument had been little used in this age group previously, and has been shown to have only moderate specificity and sensitivity for children with mental ages below 2 years (Lord, 1995; Lord et al., 1993, 1997; Pilowsky et al., 1998). Indeed, the ADI-R algorithm for children aged 5 and over specifies that on some items the most prototypical autistic behaviour is seen at the ages 4 to 5 years, and thus below this age the instrument may be less sensitive or specific (Lord et al., 1994). At 20 months clinical diagnoses were reached according to ICD-10 criteria, based on ratings of communicative interaction by the clinicians conducting the assessment (GB, AD). Thus, clinical diagnosis at 20 months was independent of information on developmental history elicited from the parents during the ADI-R interview. At 20 months children were given clinical diagnoses of childhood autism or other PDD; expressive, receptive, or mixed receptive-expressive language disorder; general developmental delay; or clinically normal. At 42 months all children were assigned ICD-10 diagnoses, achieved as a result of the compilation of *all* available clinical, historical, and psychometric evidence. This diagnosis was a consensus diagnosis reached by the three experienced clinicians who conducted the assessments at both 20 and 42 months (AC, GB, AD). Thus clinical diagnosis at 42 months was not independent from clinical diagnosis at 20 months, nor from information collected via the ADI-R interview at this agepoint.

At 42 months children were given ICD-10 clinical diagnoses of childhood autism, atypical autism, Asperger's syndrome, or other PDD; expressive, receptive, or mixed receptive-expressive language disorder; general developmental delay; or clinically normal. A diagnosis of autism, or no autism, was also generated by the concurrent ADI-R algorithm (Lord et al., 1994) at both timepoints. We recognise that there is some difference in clinical and research practice between the application of DSM-IV and ICD-10 criteria for the related pervasive development disorders included in both schedules. We applied strict ICD-10 criteria for childhood autism (F84.0), atypical autism (in both cases this was due to atypicality in symptomatology, not onset, in that abnormalities were present in an insufficient number of areas to meet criteria for childhood autism, i.e. F84.11), and for Asperger's syndrome (F84.5). We also employed the category other PDD (F84.8) for cases in which the *degree or severity* of abnormality was less than that required to meet criteria for childhood autism or atypical autism, rather than merely that criteria were met in insufficient number of areas. Although there continues to be justifiable concern regarding the reliability and validity of the diagnosis of related PDDs in both the ICD-10 (other PDD and PDD-unspecified) and DSM-IV (PDD-not otherwise specified) diagnostic systems (e.g. Mahoney et al., 1998; Volkmar et al., 1994), children who show these profiles are often described in clinical practice as having an "autism spectrum" disorder (Wing, 1988) or "PDD". Unless children with these disorders are included in research designs, questions regarding the reliability, stability, and validity of these diagnoses, as well as their relationship to the well-agreed "core" condition of childhood autism, will remain unanswered, and for this reason these subjects are included in the present study. However, results for children with "core" childhood autism are presented separately from those of children with related PDDs.

Results

Fifty children were seen for assessment after CHAT screening at 20 months of age, and 49 were seen again at 42 months (1 child from CHAT risk Group 2 was not seen

at follow up). Of these 49 children, 2 (from risk Group 3) had incomplete ADI-Rs at 42 months, 1 child with cerebral palsy (from risk Group 2) was excluded from the study, and data is presented for the remaining 46 children for whom complete ADI-R and clinical diagnosis was available at both timepoints (12 from risk Group 1, 20 from risk Group 2, 14 from risk Group 3). The data will be presented in the following ways: first, a summary of the clinical diagnoses at 20 months and 42 months will be given. Second, ADI-R dimension and overall algorithm ADI-R scores will be compared to clinical diagnosis at both timepoints. Next, stability of clinical and ADI-R diagnosis between 20 and 42 months will be considered. Fourth, individual item scores on the ADI-R will be compared with the participants grouped according to final clinical (ICD-10) diagnosis at 42 months.

For the purpose of data comparison and statistical analyses, participants are grouped into four groups according to final ICD-10 diagnosis at 42 months:

- (1) Eight children with childhood autism (hereafter autism) (all boys).
- (2) Thirteen children with a related PDD (2 atypical autism, 2 Asperger's syndrome, 9 other PDD) (11 boys, 2 girls).
- (3) Nine children with language disorder (hereafter LD), comprising 1 child with receptive language disorder, 3 with expressive language disorder, and 5 with mixed receptive-expressive language disorder (4 boys, 5 girls).
- (4) Fifteen clinically normal children (12 boys, 3 girls).

The one child with a 42-month diagnosis of general developmental delay was excluded from the analysis. The results of psychometric assessment of nonverbal mental age (NVMA) using the mean of the Griffiths subscales D and E at age 20 months, and either mean D and E subscales scores or the estimated mental age from the Leiter at age 42 months, raw Reynell Language Scale scores, as well as chronological age, at both timepoints of participants grouped by final clinical diagnosis, is presented in Table 1. None of the groups differed on chronological age at either 20 months or 42 months [ANOVA, $F(3,41) = 1.19$, $p = \text{n.s.}$; and $F(3,41) = 0.92$, $p = \text{n.s.}$, respectively]¹. At age 20 months, the typically developing children had a higher NVMA than all three clinical groups [$F(3,41) = 11.65$, $p < .001$; post hoc Scheffé tests $p < .001$, $p < .001$, $p < .05$ for the autism, PDD, and LD groups, respectively]. A similar pattern emerged for the raw expressive [$F(3,41) = 10.1$, $p < .001$; post hoc tests $p < .001$, $p < .01$, $p < .05$ for the autism, PDD, and LD groups, respectively] and receptive raw language scores [$F(3,41) = 10.9$, $p < .001$; post hoc tests $p < .001$, $p < .001$, $p = .05$ for the autism, PDD, and LD groups, respectively]. However, at 20 months the three

¹Although there were no statistically significant differences between the groups at either timepoint, it should be noted that children from the high risk for autism group were invited back for re-assessment as early as possible in the cycle for clinical reasons, since on the basis of the high-risk study (Baron-Cohen et al., 1992) we had most reason to suspect that these children might have a pervasive developmental disorder.

Table 1
Mean (SD) Age, NVMA, and Language Scores of Participants at Both Timepoints

Clinical diagnosis at 42 months:	At 20 months				At 42 months			
	Autism (N = 8)	PDD (N = 13)	LD (N = 9)	Normal (N = 15)	Autism (N = 8)	PDD (N = 13)	LD (N = 9)	Normal (N = 15)
Age in months	21.4 (1.8)	20.5 (1.7)	20.7 (1.3)	20.3 (0.9)	41.9 (4.6)	45.2 (6.5)	46.1 (7.9)	46.6 (6.0)
NVMA in months ^a	16.9 (2.9)	16.9 (2.0)	18.0 (2.7)	21.4 (1.4) ^d	28.0 (8.0) ^e	38.6 (6.8) ^f	43.6 (3.9)	47.3 (5.6)
EL raw score ^b	6.5 (2.8)	8.4 (3.8)	9.4 (1.1)	15.2 (3.9) ^d	16.6 (8.3) ^e	29.9 (8.9) ^f	31.9 (3.7) ^g	44.1 (5.9)
RL raw score ^c	3.4 (1.2)	6.4 (3.8)	6.9 (3.7)	15.7 (6.8) ^d	16.6 (7.3) ^e	32.9 (6.7) ^f	34.0 (4.9) ^g	47.3 (8.3)

^a Griffiths D and E scales, Leiter.

^b Reynell Expressive Language Score.

^c Reynell Receptive Language Score.

^d N > A, PDD, LD; ^e A < PDD, LD, N; ^f PDD < N; ^g LD < N.

clinical groups did not differ from each other in terms of age, NVMA, or language scores. At age 42 months, the pattern was somewhat different: the autism group had a lower NVMA than the three other groups [$F(3,41) = 11.4, p < .001$; post hoc tests $p < .05, p < .01, p < .001$ for the PDD, LD, and typically developing groups, respectively], and the PDD group also had a lower NVMA than the typically developing group (post hoc $p < .05$). In terms of expressive and receptive language scores at 42 months, the typically developing group had higher scores than all three clinical groups, and the PDD and LD groups had higher scores than the autism group [$F(3,41) = 25.1, p < .001$ and $F(3,41) = 29.3, p < .001$; all post hoc tests $p < .01$ for the expressive and receptive Reynell raw score, respectively]. Although the clinical groups and the typically developing group, and indeed the clinical groups themselves, differed on some of these measures at one or both timepoints, and thus are not matched for NVMA and language abilities, between-group comparisons of ADI-R scores will be performed. Within the quasi-experimental and clinically oriented design of the screening study it was not possible to recruit matched groups, and though the exact lack of matching limits interpretation of the data, the opportunity to report sensitivity and stability of clinical and ADI-R diagnosis, as well as more detailed information from the ADI-R, on this young cohort of children with autism and related PDDs was considered worthwhile.

Overall ADI-R Scores at 20 Months and 42 Months in Relation to 42-month Clinical Diagnosis

Of the eight children in the 42-month clinical diagnosis of autism group, four reached the traditional ADI-R threshold for all three domains (see Lord et al., 1994) at 20 months, although a further two children reached threshold when a modified criteria, using a lowered threshold of 2 for repetitive behaviours and stereotyped patterns (Dimension D), was applied. This single adjustment to the ADI-R criteria therefore resulted in the correct identification of six of these eight children as having autism at 20 months, but this modification also led to the inclusion of four children who did not meet full ICD-10 criteria for autism at 42 months (one atypical autism, one other PDD, one Asperger's syndrome,

and one LD—see below). Altering the threshold for Dimension D, therefore, increased the sensitivity of the instrument at this age for detecting all disorders in the autistic spectrum, but reduced its specificity for discriminating between classical autism and related pervasive developmental disorders. At 42 months, seven of the eight children with autism were above cutoffs on all three dimensions, and thus met established ADI-R criteria, with the remaining child scoring above cutoff on Dimension C only. At 42 months only three children who did not meet full ICD-10 criteria for autism met the ADI-R algorithm for autism, and all three met criteria for related PDDs disorders (one atypical autism, one Asperger's syndrome, and one other PDD). Lowering the cutoff on the third axis—Dimension D—from 3 to 2 did not identify any other children.

Of the remaining 10 children with a final clinical diagnosis of PDD, 1 scored above cutoff on two dimensions and 2 above cutoff on one dimension only at 20 months, and 4 scored above cutoff on two dimensions and 2 above cutoff on one dimension only at 42 months. Of the 9 children with a final clinical diagnosis of language disorder, 2 scored above cutoff on two dimensions and 1 above cutoff on one dimension only at 20 months, and only 2 children remained above cutoff on one dimension only at 42 months. Of the 15 children in the clinically normal final diagnosis group, only 1 reached threshold for any one of the three ADI-R dimensions at 20 months, and similarly, only 1 (different) clinically normal child met cutoff on one dimension of the ADI-R at 42 months. The proportion of each group scoring above the cutoff on each dimension at each timepoint, and the means and standard deviations, are shown in Table 2. Note that no children had a sufficient level of language for the verbal communication items to be scored at 20 months, and only 2 of the children with autism and 11 with PDD, but all of the children with LD and all the typically developing children, did so at 42 months (see Lord et al., 1994, for details).

Differences between the groups on ADI-R dimension scores were analysed by ANOVA and post hoc Scheffé test to identify group-by-group differences. At 20 months the groups differed on all three dimensions [$F(3,41) = 10.3, p < .001$; $F(3,41) = 13.9, p < .001$; and $F(3,41) = 5.05, p < .01$ for the reciprocal social interaction dimension, communication (nonverbal), and repetitive and stereotyped behaviours dimensions, respect-

Table 2
 Mean (SD), and Percentage of Participants Above Cutoff on the ADI-R Domains at 20 Months and 42 Months

Clinical diagnosis at 42 months:	At 20 months				At 42 months			
	Autism (N = 8)	PDD (N = 13)	LD (N = 9)	Normal (N = 15)	Autism (N = 8)	PDD (N = 13)	LD (N = 9)	Normal (N = 15)
Reciprocal social interaction								
Mean (SD)	15.5 (6.7) ^{e,e}	9.2 (4.7)	6.9 (4.0)	4.1 (3.9) ^c	18.3 (7.2) ^{e,f}	8.2 (4.2) ^g	2.4 (2.0)	2.1 (2.5) ^{e,d}
% above cutoff	88%	31%	22%	13%	88%	54%	0%	0%
Communication (nonverbal)								
Mean (SD)	12.2 (1.5) ^{e,f}	7.3 (3.6)	5.8 (3.8)	3.7 (2.4) ^{e,d}	11.9 (1.7) ^{e,f}	6.4 (4.1) ^g	2.2 (1.3)	1.9 (2.4) ^{e,d}
% above cutoff	100%	62%	44%	13%	100%	50%	0%	7%
Communication (inc. verbal)								
Mean (SD)	—	—	—	—	(N = 2) 19.0 (2.8) ^{e,f}	(N = 11) 8.4 (4.2) ^g	(N = 9) 4.0 (2.3)	(N = 15) 2.9 (3.0) ^{e,d}
% above cutoff	—	—	—	—	100%	64%	11%	7%
Repetitive & stereotyped behaviour								
Mean (SD)	2.6 (1.5) ^e	1.2 (0.8)	0.6 (0.7)	0.7 (1.6) ^c	3.5 (1.7) ^{e,f}	2.1 (1.7) ^g	0.9 (1.1)	0.7 (0.8) ^{e,d}
% above cutoff ^a	50%	0%	0%	0%	88%	39%	11%	0%
% above cutoff ^b	75%	39%	11%	7%	88%	46%	22%	20%

^a Cutoff = 3 (Lord et al., 1994); ^b Adjusted cutoff = 2.

^c N < A; ^d N < PDD; ^e A > PDD, LD; ^f A > LD; ^g PDD > LD.

ively]. Post hoc group-by-group comparisons showed that the typically developing group scored significantly lower than the autism group on all three dimensions ($p < .001$, $p < .001$, and $p < .05$, respectively), and lower than the PDD group on the (nonverbal) communication dimension ($p < .05$). Amongst the clinical groups, the autism group scored significantly higher than the LD group on all three dimensions ($p < .01$, $p < .001$, $p < .05$, respectively) but higher than the PDD group on the first two dimensions only (both $p < .05$). Similarly, at 42 months, the groups differed on all three dimensions [$F(3,41) = 31.1$; $F(3,41) = 26.3$; $F(3,41) = 9.24$, all $p < .001$, respectively), and the subgroup of children who had sufficient language ability to score the verbal items also differed on the verbal communication dimension [$F(3,36) = 18.3$, $p < .001$]. The typically developing group scored lower than the autism (all $p < .001$) on all three dimensions, lower than the PDD group on the first two dimensions only (both $p < .01$), but no differently to the LD group on all three dimensions. The autism group scored more highly than the other two clinical groups on the reciprocal social interaction and communication (verbal and nonverbal) dimensions (all $p < .001$), but higher than the LD group but not the PDD group on the repetitive behaviour dimension ($p < .01$). At 42 months the PDD group scored more highly than the LD group on the reciprocal social interaction and communication (verbal and nonverbal) dimensions (all $p < .05$), but not on the repetitive behaviour dimension.

Stability of Clinical and ADI-R Diagnoses between 20 Months and 42 Months

Of the nine children clinically diagnosed to have autism at 20 months, none had a diagnosis outside the broader

autism spectrum at 42 months, with six meeting DSM-IV criteria for autism, two for atypical autism, and one for other PDD. Thus, if an autism spectrum approach (Wing, 1988) is taken, there were no false positives for clinical diagnosis at 20 months. There were only two false negative diagnoses of autism at 20 months in that two children who received a clinical diagnosis of other PDD at 20 months went on to receive a DSM-IV diagnosis of autism at 42 months, and once again if a spectrum approach is taken these would not be counted as false negative diagnoses. In addition, one other child received a clinical diagnosis of other PDD at both 20 months and 42 months. However, clinical diagnosis of PDDs showed poor sensitivity at 20 months—seven children thought at 20 months to have a developmental problem outside the autistic spectrum (three language disorder, four general developmental delay) received a diagnosis of Asperger's syndrome (one) or other PDD (six) at 42 months, and two children who received no clinical diagnosis at 20 months were identified as having other PDD (one) and Asperger's syndrome (one) at 42 months.

Diagnosis of autism according to traditional ADI-R criteria was less sensitive than clinical diagnosis at 20 months. Of the eight children receiving a diagnosis of classical autism at 42 months, only four met ADI-R criteria at 20 months, although two others met our modified criteria with a lowered threshold for repetitive behaviours. However, seven of the eight met the standard ADI-R criteria at 42 months. The remaining child fell below threshold on both the reciprocal social interaction and repetitive behaviours dimensions at 42 months. Sensitivity of the ADI-R at 20 months to related PDDs was even lower. Of the 13 children with a final diagnosis of atypical autism, Asperger's syndrome, or other PDD, none met ADI-R criteria at 20 months, and only three

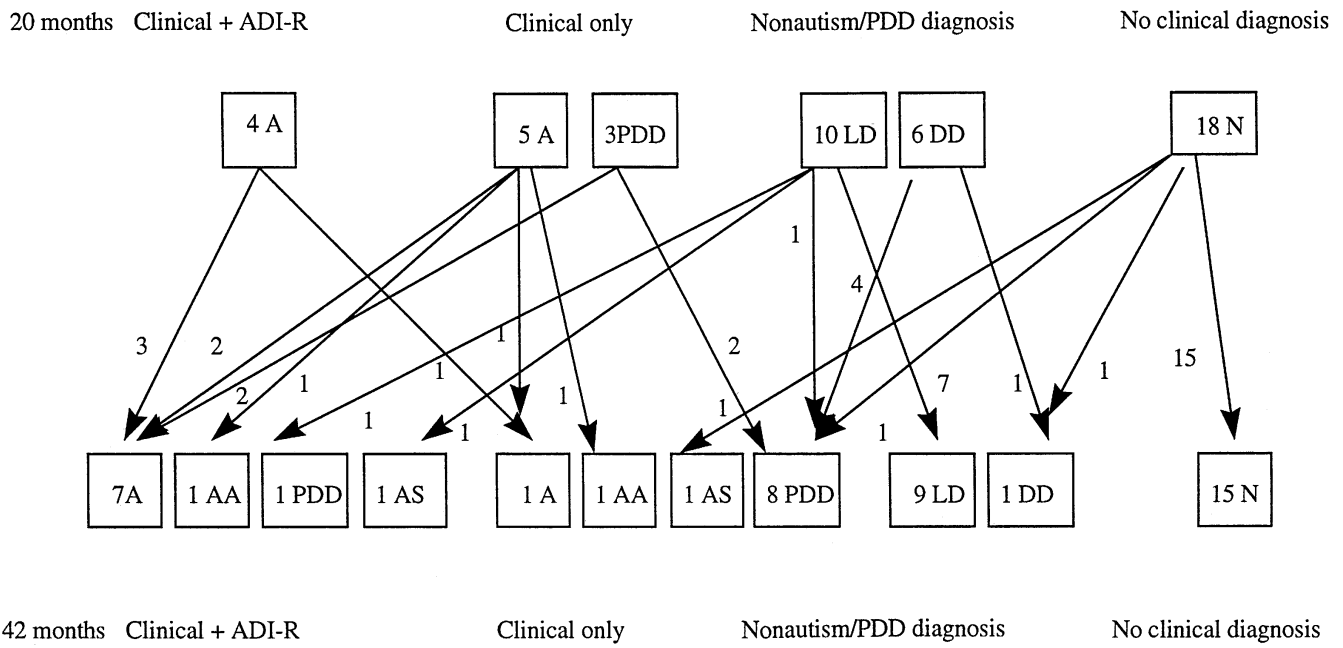


Figure 1. Diagnosis of autism and PDD by clinical judgement and ADI-R at 20 months (A = autism, AS = Asperger’s syndrome, AA = atypical autism, PDD = other PDD, LD = language disorder, DD = developmental delay, N = normal).

Table 3
Percentage of Children Scoring for Definite Abnormality on ADI-R Items at 20 Months and 42 Months: Reciprocal Social Interaction

	At 20 months				At 42 months			
	Autism (N = 8)	PDD (N = 13)	LD (N = 9)	Normal (N = 15)	Autism (N = 8)	PDD (N = 13)	LD (N = 9)	Normal (N = 15)
Clinical diagnosis at 42 months:								
Direct gaze	38	8	0	0	25	15	0	0
Social smile	13	0	0	0	25	0	0	0
Range of facial expressions	63 ^a	8	0	0	50	8	0	0
Interest in children	88 ^a	39	11	20	88	31	22	7
Response to approaches	38	8	0	0	75	15	22	0
Showing/directing attention	38	39	44	13	63	23	0	0
Offering to share	75	62	55	27	75	23	11	13
Share enjoyment with others	38	15	33	0	75 ^{a,b}	8	0	0
Offers comfort	63	39	55	33	75 ^a	23	0	7
Use other’s body to communicate	13	15	22	13	50	8	0	0
Quality of social overtures	75	31	11	7	50	0	0	0
Inappropriate facial expression	13	0	0	0	25	0	0	0
Appropriate social response	38	31	0	7	38	23	0	0

^a A > LD; ^b A > PDD.

met our modified criteria (one atypical autism, one Asperger’s syndrome, one other PDD). Sensitivity of the ADI-R to related PDDs was also low in that only three children met standard criteria for autism on the ADI-R at 42 months (one atypical autism, two other PDD), and this number did not increase with our modified criteria. No child with a final diagnosis outside the autism spectrum met ADI-R criteria at either 20 months or 42 months, and only one child with a language disorder met our modified criteria at 20 months only. The stability and association between clinical ICD-10 and ADI-R diagnosis at both timepoints is summarised in Figure 1.

Discrimination between Groups Using Individual ADI-R Items at 20 Months and 42 Months

On scoring the ADI-R for individual items, scores of 0 represent definite normality, 1 possible abnormality, and scores of 2 and 3 represent definite or unequivocal abnormality (see Lord et al., 1994, for details). In order to identify which behaviours at age 20 and 42 months best predict an ICD-10 clinical diagnosis of autism or related PDDs (atypical autism, Asperger’s syndrome, other PDD) at 42 months we compared the proportion of children grouped by final clinical diagnosis who scored

Table 4

Percentage of Children Scoring for Definite Abnormality on ADI-R Items at 20 Months and 42 Months: Communication

Clinical diagnosis at 42 months:	At 20 months				At 42 months			
	Autism (N = 8)	PDD (N = 13)	LD (N = 9)	Normal (N = 15)	Autism (N = 8)	PDD (N = 13)	LD (N = 9)	Normal (N = 15)
Nonverbal items								
Point for interest	100 ^{a,b}	23	33	7	75 ^a	23	0	13
Conven, instrumental gestures	88 ^a	54	11	20	88 ^a	30	0	7
Nodding	88	62	67	40	100 ^{a,b}	39	11	7
Headshaking	75	23	22	7	50	15	22	7
Spontaneous imitation	88	46	22	7	88	54	22	0
Imaginative play	100	62	55	27	100 ^a	62 ^c	0	7
Imitative play	13	8	0	7	50	39	0	20
Verbal items								
Reciprocal conversation	—	—	—	—	(N = 2) 100	(N = 11) 55	(N = 9) 11	(N = 15) 0
Social chat	—	—	—	—	50	18	11	7
Stereotyped utterances	—	—	—	—	100	9	0	0
Inappropriate questions	—	—	—	—	0	0	11	7
Pronomial reversal	—	—	—	—	50	18	11	7
Neologisms	—	—	—	—	0	0	0	0

^a A > LD; ^b A > PDD; ^c PDD > LD.

Table 5

Percentages of Children Scoring for Definite Abnormality on ADI-R Items at 20 Months and 42 Months: Repetitive Behaviours and Stereotyped Patterns

Clinical diagnosis at 42 months:	At 20 months				At 42 months			
	Autism (N = 8)	PDD (N = 13)	LD (N = 9)	Normal (N = 15)	Autism (N = 8)	PDD (N = 13)	LD (N = 9)	Normal (N = 15)
Unusual preoccupations	0	8	0	0	13	0	0	7
Verbal rituals	0	0	0	0	0	0	0	0
Compulsions	13	0	0	7	13	0	11	0
Hand/finger mannerisms	25	0	0	7	50	31	0	0
Complex mannerisms	25	0	11	13	38	0	11	0
Repetitive use of objects	13	0	0	0	50	23	0	0
Unusual sensory interests	0	0	0	0	0	8	0	0

for definite abnormality of the type indicated on the ADI-R schedule (a score of 2 or 3) versus no or questionable abnormality (a score of 0 or 1) on the ADI-R algorithm items at age 20 months and age 42 months. This data is presented in Tables 3 to 5. For reasons of parsimony and because the main clinical concern is how well particular items discriminate the autism, PDD, and LD clinical groups from each other, statistical analyses at this item-by-item level are conducted between the three clinical groups only. However, data for the normally developing children are presented in the accompanying tables for comparison, and comment will be made in the discussion regarding the presence of these behaviours in this typically developing group of children.

Items that Discriminated Autism from Language Disorder

Item-by-item ADI-R scores were compared between the three clinical groups using a $3 \times 2 \chi^2$ analysis and Fisher Exact post hoc tests (to take account of low expected cell frequencies), with an alpha level for stat-

istical significance set at .01 to guard against type I errors given the number of comparisons made. The ADI-R items that best discriminated between the children with a 42-month diagnosis of childhood autism and those with LD in our sample at age 20 months were: *range of facial expression* [$\chi^2(2,30) = 12.5, p < .01, \text{ post hoc } p < .01$] and *interest in other children* [$\chi^2(2,30) = 10.3, p < .01, \text{ post hoc } p < .01$] from the reciprocal social interaction dimension; *point for interest* [$\chi^2(2,30) = 12.7, p < .01, \text{ post hoc } p < .01$], *use of conventional gestures* [$\chi^2(2,30) = 10.0, p < .01, \text{ post hoc } p < .01$] from the communication dimension. No items from the repetitive behaviours and stereotyped patterns dimension differentiated between the children with autism and those with LD at this age. At 42 months of age the items that significantly differed between the autism and language disorder groups were: *seeking to share enjoyment with others* [$\chi^2(2,30) = 16.5, p < .001, \text{ post hoc } p < .01$] and *offering comfort* [$\chi^2(2,30) = 11.9, p < .01, \text{ post hoc } p < .01$] from the reciprocal social interaction dimension; *point for interest* [$\chi^2(2,30) = 11.9, p < .01, \text{ post hoc } p < .01$], *use of conventional gestures* [$\chi^2(2,30) = 14.3,$

$p < .001$, post hoc $p < .001$], *nodding* [$\chi^2(2,30) = 14.1$, $p < .001$, post hoc $p < .001$] and *imaginative play* [$\chi^2(2,30) = 17.7$, $p < .001$, post hoc $p < .001$] from the communication dimension. Once again no items from the repetitive behaviours and stereotyped patterns dimension discriminated between the children with autism and those with language disorder at this age.

Items that Discriminated Autism from PDD

The ADI-R items that best discriminated between the children with autism and those with PDD in our sample at age 20 months were: no items from the reciprocal social interaction dimension; *point for interest* [$\chi^2(2,30) = 13.0$, $p < .01$, post hoc $p < .001$] from the communication dimension; and no items from the repetitive behaviours and stereotyped patterns dimension. At 42 months of age the items that significantly discriminated between the autism and PDD groups were: *seeking to share enjoyment with others* [$\chi^2(2,30) = 16.5$, $p < .001$, post hoc $p < .01$] items from the reciprocal social interaction dimension; *nodding* [$\chi^2(2,30) = 13.5$, $p < .01$, post hoc $p < .001$] from the communication dimension; and no items from the repetitive behaviours and stereotyped patterns dimension.

Items that Discriminated PDD from LD

No individual ADI-R items significantly discriminated between the children with PDD and those with LD in our sample at age 20 months, although at age 42 months the children with PDD were rated as showing more definite abnormality in *imaginative play* [$\chi^2(2,30) = 17.7$, $p < .001$, post hoc $p < .01$] from the communication dimension.

A Developmental Picture of Items that Discriminated between the Groups at the Two Timepoints

Looking at the items that discriminated between the autism, PDD, and LD groups at 20 months or 42 months, it is apparent that at least four distinct patterns account for the differences found:

- (1) A few items discriminated clearly between the autism group and the language disorder group at both 20 months and 42 months, and as such these items are highly specific to autism (but not PDD): *point for interest* and *use of conventional gestures*. However, it is important to note that at age 20 months (but not at 42 months) a significant proportion of children in the language disorder group scored for possible abnormality for these items. In addition, *range of facial expression* discriminated between the children with autism and those with language disorder at the level of definite abnormality at 20 months and at the level of possible (but not definite) abnormality at 42 months.
- (2) Other items did not discriminate between the children with autism or PDD and the children with LD at 20 months, but do so at 42 months because

of the relatively high proportion of children with LD without autism who show abnormalities in these behaviours at 20 months but do not still do so at 42 months. These items included *offering comfort*, *nodding*, *imaginative play*, and *seeking to share enjoyment with others*. “Abnormalities” in these items appear to be related to developmental level—and it is striking that on some items in the reciprocal social interaction and communication domains of the ADI-R, such as *offering to share*, *offering comfort*, *imaginative play*, and *nodding*, a high proportion of typically developing children, as well those with language disorder, score for definite (as well as possible) abnormality on these items at 20 months but not at 42 months.

- (3) A third pattern concerns items on the third dimension of the ADI-R, that of repetitive behaviours and stereotyped patterns. On this dimension very few children from any of the groups show *definite* abnormality at 20 months, although some children with autism (but less with PDD) show *possible* abnormality at the younger age, with more showing definite abnormality particularly on *hand and finger* and *complex mannerisms* and *repetitive use of objects* at the older agepoint. However, at neither timepoint were these differences large enough for a statistically higher proportion of children with autism or PDD to show definite abnormality on any items on this axis, compared to children with LD.
- (4) A fourth pattern that is notable is the low number of items that discriminated between the children with PDD and those with LD. At 20 months no items at either level of abnormality discriminated the groups. At 42 months of age only *imaginative play* at the level of definite abnormality and *offering comfort* at the level of possible abnormality discriminated between children with PDD and children with LD. Indeed, on many items the children with PDD scored similarly to those with LD. There were, however, some clear differences in developmental trajectory between the two groups, with the children with PDD continuing to show abnormalities over time—in fact the proportion of children in the PDD group who showed abnormalities at both the definite and the possible level of certainty increased on many items across all three dimensions over time. This contrasts to a general reduction in abnormal behaviour over time in the children with LD, who had “grown out of” many of the abnormalities reported at age 20 months by the time they were 3½ years old.

Abnormalities on Other Dimension D Behaviours and Level of Reported Behavioural Problems²

In the light of the finding that few of the algorithm repetitive and stereotyped behaviours discriminated between the groups at either age, we also analysed other

²We are grateful to Marian Sigman for suggesting these analyses to us.

Table 6
Percentage of Children Scoring for Definite Abnormality on ADI-R Items at 20 Months and 42 Months: Behaviour Problems and Extra Items from Dimension D

Clinical diagnosis at 42 months:	At 20 months				At 42 months			
	Autism (<i>N</i> = 8)	PDD (<i>N</i> = 13)	LD (<i>N</i> = 9)	Normal (<i>N</i> = 15)	Autism (<i>N</i> = 8)	PDD (<i>N</i> = 13)	LD (<i>N</i> = 9)	Normal (<i>N</i> = 15)
Difficulty with changes in routine	0	0	11	0	13	0	0	0
Resistance to trivial changes	0	0	0	7	0	0	0	0
Unusual attachment to object	13	0	0	0	0	0	0	0
Negative response to stimuli	0	0	11	0	0	0	0	0
Unusual fears	13	0	11	0	0	0	0	7
Problems getting to bed	0	8	11	0	13	8	22	0
Sleep problems	13	8	0	0	13	15	11	7
Tantrums	0	8	0	0	0	23	11	0
Eating	0	23	11	7	13	39	0	0

items from the third dimension of the ADI-R that are not included on the algorithm: *difficulties with changes in routine, resistance to trivial changes, unusual attachment to object, negative response to sensory stimuli, and unusual fears*. This was to investigate whether abnormalities in this domain not included on the original ADI-R algorithm for older children might discriminate between the groups at this younger age. However, very few children with autism or children in any of the groups showed examples of definite abnormality in these areas at either timepoint. In addition, we looked at the behaviour problems listed in the ADI-R but also omitted from the final algorithm: *problems getting to bed, problems sleeping, tantrums, and feeding problems*. Although some of these behaviour problems were present in some individual children from all four groups at both timepoints, there were no statistically significant differences between the proportion of children in the clinical groups reported to show such behaviours (see Table 6).

Discussion

Clinical diagnosis of childhood autism at 20 months proved to be highly sensitive and stable if an autism spectrum approach was adopted. In terms of strict DSM-IV criteria for “core” childhood autism only, six of the nine children who received a clinical diagnosis of autism at 20 months received the same diagnosis at 42 months. Adopting a spectrum approach increased the reliability as the other children all went on to receive the related diagnoses of atypical autism (two) and other PDD (one) at the latter timepoint. There were only two false negative diagnosis of autism, with two of the eight children receiving a final diagnosis of autism being missed clinically at 20 months, although once again even these children’s diagnoses were on the autistic spectrum at 20 months (other PDD). However, clinical diagnosis of related PDDs at 20 months was less sensitive. Nine children who had a final clinical ICD-10 diagnosis of other PDD (seven) and Asperger’s syndrome (two) were not correctly identified at age 20 months, and while three were thought to have a developmental language disorder, and four general developmental delay, two children who went on to receive a diagnosis on the autistic spectrum (one Asperger’s syndrome, one other PDD) were thought

to be developmentally normal at age 20 months (or at least their development was not sufficiently abnormal to meet ICD-10 criteria for a developmental disorder).

One possibility following on from this under-diagnosis of PDD is that these cases are examples of “late-onset” autism in which autistic abnormalities were not notable at 20 months but became apparent between this age and 42 months. Within the present design we are not able to comment much on whether this was the case. Certainly it is clear that as assessed by clinical judgement following an interactive assessment session, by a series of experimental tasks assessing play, empathy, joint attention, and imitation skills (Charman et al., 1998), and by parental report from the ADI-R interview, these children who went on to receive a diagnosis of other PDD and Asperger’s syndrome at age 42 months were more similar at 20 months to the children who went on to receive a diagnosis of LD than they were to the children who went on to receive a diagnosis of childhood autism or atypical autism. Whether this is a question of what is an appropriate threshold for judging autistic symptomatology in children so young—in interaction and clinical decision-making, in the design of experimental tasks and in the definitions of abnormality and severity employed by the ADI-R—or whether it is a question of autistic symptomatology simply being absent at this age and emerging subsequently, we do not know. Both parents and the interviewer are asked to make a retrospective judgement with hindsight about when the child first showed problems in the ADI-R interview. On this measure those children with final PDD diagnoses who were “missed” at 20 months did not differ from the ones who were identified earlier, with almost all children being judged by both their parent and the interviewer at the second timepoint to have problems before the age of 18 months, although a retrospective bias may be operating here.

In terms of clinical diagnosis of childhood autism the present findings are fairly consistent with those obtained by Lord (1995) in a clinically referred sample. Lord also found that clinical diagnosis of autism in 2-year-olds was sensitive and stable over time on follow-up, as had Gillberg et al. (1990) in a previous study. Similarly, Stone et al. (1999) report high stability across time in the clinical diagnosis of autism in a sample of 2½-year-olds followed up approximately 1 year later, particularly when

an autism spectrum approach is adopted, and also demonstrate high agreement between independent clinicians both within and across time, a factor not included in our present design. Stone et al. also found somewhat lower stability across time for an initial diagnosis of PDD-NOS, with 4/12 children receiving a nonspectrum diagnosis at Timepoint 2 from the same primary clinician (though only 1/12 from the secondary clinician). It is not clear from the Stone et al. paper whether there were any false negatives—in terms of children who received autism spectrum diagnosis at Timepoint 2 who were given nonspectrum diagnoses at timepoint 1—in their study since they do not report on follow-up of the nonspectrum sample. Reduced reliability of diagnosis of related PDDs on the autism spectrum, compared to diagnosis of the “core” autistic disorder, has also been reported for older clinic samples (Mahoney et al., 1998; Volkmar et al., 1994).

The ADI-R demonstrated good specificity in diagnosing childhood autism at 20 months of age. All four of the children diagnosed as having autism by ADI-R criteria at 20 months received a clinical ICD-10 diagnosis of childhood autism on follow-up age 42 months. However, using these criteria at this age, the ADI-R showed only moderate sensitivity as a diagnostic instrument. Four of the eight children with a final clinical diagnosis of childhood autism did not meet ADI-R criteria at age 20 months. In terms of identifying children with related PDD disorders, the ADI-R at 20 months was shown to be less sensitive still. None of the 13 children with final clinical diagnoses of atypical autism, Asperger’s syndrome, or other PDD met standard ADI-R criteria (Lord et al., 1994) at 20 months, and only three children from this group met ADI-R criteria at 42 months (one atypical autism, one Asperger’s syndrome, and one other PDD). The sensitivity of the ADI-R at 20 months was increased somewhat by reducing the threshold for the repetitive behaviours and stereotyped patterns axis (Dimension D) from 3 to 2, reducing the false negative rate for core autism to 2 out of 8. In addition, with this adjusted threshold 3 of 13 children with final clinical diagnoses of a related PDD were identified at 20 months. However, the modification reduced the instrument’s specificity in that one child with a clinical diagnosis outside the autism spectrum (a children with a mixed expressive-receptive language disorder) was falsely identified at 20 months. The modification made no difference at 42 months.

The present findings in terms of the sensitivity and stability of ADI-R diagnosis are partly consistent with and partly contrasting with those obtained by Lord (1995). Lord found that the ADI-R both over- and under-diagnosed autism in her sample, in contrast to the clear under-diagnosis in the present study. However, the findings appear less disparate when differences in the IQ of the samples in the studies are considered, in that in Lord’s study under-diagnosis occurred in children with less severe cognitive deficits (IQ approximately 60—similar to the present study) and over-diagnosis occurred in a group of more severely impaired 2-year-old children (IQ approximately 40). In line with this, Lord and colleagues (1993) have previously found that although the sensitivity and specificity of ADI-R were both high in diagnosing autism amongst a group of 3½ to 4-year-old

children with a mean mental age of 26 months, it incorrectly diagnosed 2 nonautistic mentally handicapped 2-year-olds with single words and 8 out of 14 children with mental handicap with no speech and mental ages below 18 months. However, Pilowsky et al. (1998) found that the ADI-R under- (but not over-) diagnosed autism in some (but not all) subjects with low age and mental age (below 18 months). The authors of the ADI-R (Lord et al., 1997) caution that the ADI-R becomes less reliable as an instrument for detecting autism in children with developmental levels below 2 years of age, and the present findings confirm that such caution is warranted.

In terms of the use of the ADI-R at both timepoints with children who do not meet full diagnostic criteria for autism, but do so for the related PDDs, atypical autism, Asperger’s syndrome, and other PDD, the present study found the instrument to have low sensitivity at both 20 and 42 months in identifying children with related PDD. The ADI-R has previously primarily been used to differentiate between clinic-referred children who meet full diagnostic criteria for autism and those who do not (Le Couteur et al., 1989; Lord, 1995; Lord et al., 1993, 1997). The authors of these papers do not discuss the ADI-R scores of children who do not meet full criteria for autism but who would meet criteria for the related PDDs³. However, it is our experience that clinicians are increasingly using the related PDD diagnoses and if the ADI-R is to be widely adopted as a clinical and research tool it would be helpful to know how it performs with a sample of children with PDD who are more representative of those seen in clinical settings than are the present sample. One example is that evidence from family genetic studies has shown that subjects with quite low ADI-R dimension scores can be considered to fit the “broader phenotype” of autism presentation which, while not consistent with ICD-10 criteria for a pervasive developmental disorder, does indicate significant social and communicative difficulties (Le Couteur et al., 1996). However, it is worth noting that the ADI-R is intended as a standardised clinical tool and not a diagnostic “gold standard” for diagnosing autism spectrum disorders, and that for a clinical diagnosis to be made—whether of autism or a related spectrum disorder—additional clinical information is taken into account besides the algorithm scores (Lord et al., 1994).

Examining the pattern of scores on individual items of the ADI-R at both timepoints, some behaviours discriminated between children with autism and those with a LD in our sample at both 20 months and 42 months, including *point for interest* and *use of conventional gestures*. However, other items including *seeking to share enjoyment with others*, *nodding*, as well as *imaginative play* discriminated between the children with autism and LD at age 42 months but not at age 20 months. This reflects the fact that a significant proportion of individuals with LD showed abnormalities in these behaviours at 20 months, but went on to demonstrate more typical

³Lord (1995) does mention that three children who met ADI-R but not clinical criteria for autism at age 2½ went on to receive other PDD diagnoses (and fall outside the ADI-R criteria) at age 3½ years.

behaviour in these areas between the initial assessment at 20 months and follow-up 22 months later. This developmental change is seen clearly by comparing the mean ADI-R dimension scores at 20 and 42 months for the three clinical groups. At both timepoints the subjects with autism scored more highly on the ADI-R dimensions than those with related PDDs. However, whereas autism vs. language disorder comparisons were significant at both the earlier and the later timepoint, this was only true for the PDD vs. LD comparisons at 42 months. This difference is because for the autism and PDD groups the mean scores change little (and in no consistent direction) between 20 and 42 months, but for the children with LD (and the clinically normal children), mean scores on the reciprocal social interaction and communication dimensions reduce significantly between the two timepoints. Further to this point, it is noteworthy that a significant proportion of the clinically normal (as well as language disordered) children showed some “abnormalities”—both at the definite and the possible level of abnormality—in behaviours such as *offering comfort*, *offering to share*, and *nodding* at age 20 months but not at 42 months. It appears that abnormalities in development that the ADI-R is trying to assess may be present as a passing phase in the repertoire of many typically developing children in the second and perhaps third year of life (see Evans et al., 1997, for a similar account).

The changing ADI-R profiles of our community-derived group of children show some similarities and some differences from the clinically referred group studied by Lord (1995). Both studies show that abnormalities in reciprocal social interaction and communication are present at a young age and continue to be present at follow-up, and that these behaviours become less notable over time in the nonautism comparison groups—thus increasing the discrimination between the groups. However, in the present study abnormalities on the third dimension of repetitive and stereotyped behaviours were not reported at age 20 months in many children with autism (and very few with related PDDs), at least at the definite level of abnormality, although they were present in most individuals with autism (though again sometimes only at the possible level of abnormality and less so in those children with related PDDs) by age 42 months. In the Lord (1995) study abnormalities such as *hand and finger mannerisms*, *unusual sensory behaviours*, *unusual preoccupations*, and *whole body mannerisms* were present at *both* the younger and older timepoints, although more autism vs. nonautism group differences emerged at follow-up due to a reduction in nonautistic children showing such behaviours. Differences between the two studies most probably reflect the differences in age, IQ, and recruitment between the two samples (see below).

Although the present findings of lower rates of abnormality on this third axis of autistic impairments contrasts to Lord's (1995) study, there is some independent evidence that in younger children (Pilowsky et al., 1998), or in children with an autism spectrum but not “core” autistic disorder diagnosis, such behaviours may be less common (Tanguay, Robertson, & Derrick, 1998). Remembering that this prospectively identified sample of children with autism and related PDDs are the youngest studied to date, we cannot rule out that in at least some

children with autism spectrum disorders abnormalities on this third dimension only begin to emerge in children with autism *after* infancy, later than the social and communication deficits are apparent. One question that arises is whether this reflects a real absence of difficulties in this domain during the first 3 years of life in children with autism, or whether the nature and extent of difficulties in this domain differs in young preschool children with autism from that seen in older 4- and 5-year-old children with autism (the age at which the authors of the ADI-R indicate that autistic symptomatology is most prototypical). Our investigation of non-algorithm items from Dimension D and reported behavioural problems did not identify any candidate behaviours, but it may be that at least in a subsample of children with autism, such as those identified by the prospective community screening method employed here, repetitive and stereotyped behaviours do not appear until the third or even fourth year of life. Clarification of the association between the emergence of the social, communicative, and repetitive behaviours in autism is important not only for refining clinical diagnosis, perhaps particularly at the subgroup level of autism spectrum disorders, but is also potentially important as an empirical test of different psychological and neurobiological accounts of the underlying pathology in autism (see Bailey et al., 1996; Bishop, 1993, for reviews), and further work in this regard is encouraged.

One critical and cautionary point for the interpretation and generalisation of the present findings, particularly in relation to how much they may be able to inform our knowledge of the early development of individuals with autism and related PDDs through the second and third years of life, is the way in which this prospectively identified sample were identified. Joint attention and play behaviours as recorded on the CHAT at 18 months formed the basis for selection of our high and medium autism risk groups, so that not only children who went on to receive a final clinical diagnosis of autism or a related PDD *but also* those who were diagnosed as having a LD at age 42 months had already been identified as demonstrating poor basic joint attention skills (and in some individuals poor play skills), at least as measured by the CHAT at age 18 and 19 months. Thus, the findings in this sample might apply to perhaps only a subgroup of individuals with autism—those with particularly serious or early joint attention and play impairments. This, for example, might explain the relatively low rate of repetitive and stereotyped behavioural problem reported by the parents of this (possibly unrepresentative) subgroup. Another limitation is that children with identified profound mental handicap, sensory impairment, and physical disability, and children with pre-identified disabilities such as Downs syndrome, were excluded from our study at the community screening stage, as Health Visitors were reluctant to submit parents to further screening procedures if their children were already identified as developmentally delayed and receiving services. Thus, the IQ (nonverbal IQ approximately 65 at age 42 months) of our sample of children with autism is above the average for many samples previously studied, and probably in comparison to the population of individuals with autism as a whole (Sigman, Dissanayake, Arbelle, & Ruskin, 1997), although the range of IQs in our very small sample

is considerable (33 to 98 in eight subjects as measured at 42 months), limiting comparison to other studies and generalisation from the present findings. The very small sample sizes themselves are also limiting factors on the generalisability of the present findings. Once again, the low rate of repetitive and stereotyped behaviours found in the children with autism (particularly at the younger timepoint) in the present study may be characteristic of our mild to moderately handicapped sample, in contrast to more severely handicapped samples (e.g. Lord, 1995) who show such behaviours from a younger age.

Another important caution in interpreting the ADI-R data is the accuracy of parental observation and reporting. Many of the parents of our autistic and PDD group reported *no* significant concerns about their children at 20 months or only mild concern about delayed spoken language, prior to the clinical assessment. Furthermore, it was apparent in at least two cases that parents were making inappropriately positive attributions for their children's abnormal behaviours as witnessed during the clinical assessment, and these parents also reported spontaneous, communicative overtures by their child, when it appeared clinically that the parent was scaffolding the interaction in a very clear and directed way. It is possible that the observations of parents in the present study about their child's behaviour may have been less focused or less accurate than those of parents who had already noted problems and initiated or acquiesced to a clinical assessment, as in Lord's earlier studies (Lord, 1995; Lord et al., 1993).

One further limitation of the present study is the lack of independent clinical diagnosis across time (although clinical diagnosis at 20 months was independent of the ADI-R) and lack of information on inter-rater reliability of diagnosis at either the younger or older timepoint. This is particularly limiting due to the fact that diagnosis of non "core" childhood autism PDDs is less reliable than of childhood autism (Mahoney et al., 1998; Volkmar et al., 1994), and our use of the ICD-10 category of "other PDD" would be considered a more liberal diagnosis than those of atypical autism on the ICD-10 system and (arguably) PDD-NOS on the DSM-IV system. As we stated above, the use of a diagnosis of "autism spectrum disorder" is increasing in clinics in the U.K. and we included such children in the current study—by the application of the ICD-10 Research Criteria category that we felt most closely fitted with this use. Further empirical work is needed to validate these related PDDs and to measure the reliability and stability of their use.

Despite these considerable limitations to the generalisability of the present findings, the opportunity afforded by the prospective screening study (Baron-Cohen et al., 1996) to examine in some detail the sensitivity and stability of clinical diagnosis and the use of the ADI-R with such a young sample was considered worthwhile. In summary, the present data show that the clinical diagnosis of autism was sensitive and stable in children as young as 20 months of age. Clinical diagnosis of a broader spectrum of related PDDs was less sensitive at this age. The ADI-R showed high specificity but poor sensitivity for identifying autism at age 20 months, and although the instrument's sensitivity increased at age 42 months, few children with related PDDs met ADI-R thresholds on all three dimensions at this timepoint.

Although the present community-derived sample differs in important ways from the referred samples commonly seen in many clinic settings, the identification of symptoms that are highly specific to autism and PDD at both timepoints, and that are shared in common with children with LD and typically developing children in infancy and the preschool years, is an important reference point for refining future clinical diagnosis and practice.

Acknowledgements—This research was supported by two MRC Project Grants to SBC, AC, and GB (1992–96). We are very grateful to the 400 Health Visitors and 50 GPs in the following 9 Districts who took part in the initial screening: South Downs Health NHS Trust, Eastbourne and County Trust, Hastings and Rother, Thameslink, Mid Downs (Crawley and Chichester), and Mid Kent Healthcare Trust. We especially want to thank all the families who took part in the screening and clinical assessment aspects of the study.

References

- Adrien, J. L., Perrot, A., Sauvage, D., Leddet, I., Larmande, C., Hameury, L., & Barthelemy, C. (1992). Early symptoms in autism from family home movies: Evaluation and comparison between 1st and 2nd year of life using IBSE Scale. *Acta Paedopsychiatrica*, *55*, 71–75.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed., DSM-IV). Washington, DC: Author.
- Bailey, A., Phillips, W., & Rutter, M. (1996). Autism: Towards an integration of clinical, genetic, neuropsychological and neurobiological perspectives. *Journal of Child Psychiatry and Psychology*, *37*, 89–126.
- Baird, G., Baron-Cohen, S., Charman, T., Cox, A., Swettenham, J., Wheelwright, S., & Drew, A. (1999). *Screening a large population of 18-month-olds with the CHAT (Checklist for Autism in Toddlers): A six-year follow-up*. Unpublished manuscript, University of Cambridge.
- Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack and the CHAT. *British Journal of Psychiatry*, *138*, 839–843.
- Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., Nightingale, N., Morgan, K., Drew, A., & Charman, T. (1996). Psychological markers in the detection of autism in infancy in a large population. *British Journal of Psychiatry*, *168*, 158–163.
- Bishop, D. (1993). Annotation: Autism, executive functions and theory of mind: A neuropsychological perspective. *Journal of Child Psychiatry and Psychology*, *34*, 279–293.
- Butterworth, G. (1991). The ontogeny and phylogeny of joint visual attention. In A. Whiten (Ed.), *Natural theories of mind*. Oxford: Basil Blackwell.
- Charman, T. (1998). Specifying the nature and course of the joint attention impairment in autism in the preschool years: Implications for diagnosis and intervention. *Autism: The International Journal of Research and Practice*, *2*, 61–79.
- Charman, T. (in press). Theory of mind and the early diagnosis of autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from autism and developmental cognitive neuroscience* (2nd ed.). Oxford: Oxford University Press.
- Charman, T., Baron-Cohen, S., Swettenham, J., Cox, A., Baird, G., & Drew, A. (1998). An experimental investigation of social-cognitive abilities in infants with autism: Clinical implications. *Infant Mental Health Journal*, *19*, 260–275.
- Dahlgren, S. O., & Gillberg, C. (1989). Symptoms in the first two years of life: A preliminary population study of infantile autism. *European Archives of Psychiatric and Neurological Science*, *283*, 169–174.

- Dawson, G., & Osterling, J. (1997). Early intervention in autism. In M. Guralnick (Ed.), *The effectiveness of early intervention*. Baltimore, MD: Brookes.
- Economic Activity of Great Britain*. (1981). London: HMSO.
- Evans, D. W., Leckman, J. F., Carter, A., Reznick, S., Henshaw, D., King, R. A., & Pauls, D. (1997). Ritual, habit, and perfectionism: The prevalence and development of compulsive-like behaviour in normal young children. *Child Development*, 68, 58–68.
- Gillberg, C., Ehlers, S., Schaumann, H., Jakobsson, G., Dahlgren, S. O., Lindblom, R., Bagenholm, A., Tjuus, T., & Blinder, E. (1990). Autism under age 3 years: A clinical study of 28 cases referred for autistic symptoms in infancy. *Journal of Child Psychology and Psychiatry*, 31, 921–934.
- Griffiths, R. (1986). *The abilities of babies*. London: University of London Press.
- Hoyson, M., Jamieson, B., & Strain, P. S. (1984). Individualising group instruction of normally developing and autistic-like children: The LEAP curriculum model. *Journal of the Division for Early Childhood*, Summer, 157–172.
- Le Couteur, A., Bailey, A., Goode, S., Pickles, A., Robertson, S., Gottesman, I., & Rutter, M. (1996). A broader phenotype of autism: The clinical spectrum in twins. *Journal of Child Psychology and Psychiatry*, 37, 785–801.
- Le Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holdgrafer, M., & McLennan, J. (1989). Autism Diagnostic Interview: A standardized investigator-based instrument. *Journal of Autism and Developmental Disorders*, 19, 363–387.
- Leiter, R. G. (1952). *Leiter International Performance Scale*. Wood Dale, IL: Stoetling.
- Lord, C. (1995). Follow-up of two-year-olds referred for possible autism. *Journal of Child Psychology and Psychiatry*, 36, 1365–1382.
- Lord, C., Pickles, A., McLennan, J., Rutter, M., Bregman, J., Folsten, S., Fombonne, E., Libya, M., & Minshew, N. (1997). Diagnosing autism: Analysis of data from the Autism Diagnostic Interview. *Journal of Autism and Developmental Disorders*, 27, 501–517.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Austin Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders*, 24, 659–686.
- Lord, C., Storoschuk, S., Rutter, M., & Pickles, A. (1993). Using the ADI-R to diagnose autism in preschool children. *Infant Mental Health Journal*, 14, 234–252.
- Lovaas, O. I. (1987). Behavioural treatment and normal education and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, 155, 3–9.
- Mahoney, W. J., Szatmari, P., MacLean, J. E., Bryson, S. E., Bartolucci, G., Walter, S. D., Jones, M. B., & Zwaigenbaum, L. (1998). Reliability and accuracy of differentiating pervasive developmental disorder subtypes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 278–285.
- McEvoy, R. E., Rogers, S. J., & Pennington, B. F. (1993). Executive function and social communication deficits in young autistic children. *Journal of Child Psychiatry and Psychology*, 34, 563–578.
- Ornitz, E. M., Guthrie, D., & Farley, A. J. (1977). The early development of autistic symptoms. *Journal of Autism and Childhood Schizophrenia*, 7, 207–229.
- Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday home videotapes. *Journal of Autism and Developmental Disorders*, 24, 247–259.
- Ozonoff, S., & Cathcart, K. (1998). Effectiveness of a home program intervention for young children with autism. *Journal of Autism and Developmental Disorders*, 28, 25–32.
- Pilowsky, T., Yirmiya, N., Shulman, C., & Dover, R. (1998). The Autism Diagnostic Interview-Revised and the Childhood Autism Rating Scale: Differences between diagnostic systems and comparison between genders. *Journal of Autism and Developmental Disorders*, 28, 143–151.
- Reynell, J. K. (1985). *Reynell Developmental Language Scales* (2nd ed.). Windsor, U.K.: NFER Nelson.
- Rogers, S. J. (1996). Brief report: Early intervention in autism. *Journal of Autism and Developmental Disorders*, 26, 243–246.
- Rogers, S. J., & DiLalla, D. L. (1990). Age of symptom onset in young children with pervasive developmental disorders. *Journal of the Academy in Child and Adolescent Psychiatry*, 29, 863–872.
- Rogers, S. J., & Lewis, H. (1989). An effective day treatment model for young children with pervasive developmental disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 28, 207–214.
- Sheinkopf, S. J., & Siegel, B. (1998). Home-based treatment of young children with autism. *Journal of Autism and Developmental Disorders*, 28, 15–23.
- Siegel, B., Pliner, C., Eschler, J., & Elliott, G. (1988). How children with autism are diagnosed: Difficulties in identification of children with multiple developmental delays. *Developmental and Behavioural Paediatrics*, 9, 199–204.
- Sigman, M., Dissanayake, C., Arbelle, S., & Ruskin, E. (1997). Cognition and emotion in children and adolescents with autism. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of autism and pervasive developmental disorders* (2nd ed.) (pp. 248–265). New York: Wiley.
- Stone, W. L. (1997). Autism in infancy and early childhood. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of autism and pervasive developmental disorders* (2nd ed.) (pp. 266–282). New York: Wiley.
- Stone, W., Hoffman, E., Lewis, S., & Ousley, O. (1994). Early recognition of autism: Parental reports versus clinical observation. *Archives of Pediatrics and Adolescent Medicine*, 148, 174–179.
- Stone, W. L., & Hogan, K. L. (1993). A structured parent interview for identifying young children with autism. *Journal of Autism and Developmental Disorders*, 23, 639–652.
- Stone, W. L., Lee, E. B., Ashford, L., Brissie, J., Hepburn, S. L., Coonrod, E. E., & Weiss, B. H. (1999). Can autism be diagnosed accurately in children under three years? *Journal of Child Psychology and Psychiatry*, 40, 219–226.
- Strain, P. S., Hoyson, M., & Jamieson, B. (1985). Normally developing preschoolers as intervention agents for autistic-like children. *Journal of the Division for Early Childhood*, Spring, 105–115.
- Tanguay, P. E., Robertson, J., & Derrick, A. (1998). A dimensional classification of autism spectrum disorder by social communication domains. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 271–277.
- Volkmar, F. R., Klin, A., Siegel, B., Szatmari, P., Lord, C., Campbell, M., Freeman, B. J., Cicchetti, D. V., Rutter, M., Kline, W., Buitelaar, J., Hattab, Y., Fombonne, E., Fuentes, J., Werry, J., Stone, W. L., Kerbeshian, J., Hoshino, Y., Bergman, J., Loveland, K., Szymanski, L., & Towbin, K. (1994). Field trial for autistic disorder in DSM-IV. *American Journal of Psychiatry*, 151, 1361–1367.
- Wing, L. (1988). The continuum of autistic characteristics. In E. Schopler & G. B. Mesibov (Eds.), *Diagnosis and assessment in autism*. New York: Plenum Press.
- World Health Organisation. (1993). *Mental disorders: A glossary and guide to their classification in accordance with the 10th Revision of the International Classification of Diseases (ICD-10)*. Geneva: Author.