



HHS Public Access

Author manuscript

Autism. Author manuscript; available in PMC 2015 May 10.

Published in final edited form as:

Autism. 2011 September ; 15(5): 545–562. doi:10.1177/1362361309358332.

Longitudinal changes in Scores on the Autism Diagnostic Interview–Revised (ADI-R) in pre-school children with autism: Implications for diagnostic classification and symptom stability

GNAKUB NORBERT SOKE,

School of Medicine, University of Colorado Denver, Aurora, Colorado, USA

AMY PHILOFSKY,

School of Medicine, University of Colorado Denver, Aurora, Colorado, USA

CAROLYN DIGUISEPPI,

Colorado School of Public Health, University of Colorado Denver, Aurora, Colorado, USA

DENNIS LEZOTTE,

Colorado School of Public Health, University of Colorado Denver, Aurora, Colorado, USA

SALLY ROGERS, and

M.I.N.D. Institute, University of California Davis, Sacramento, California, USA

SUSAN HEPBURN

School of Medicine, University of Colorado Denver, Aurora, Colorado, USA

Abstract

We prospectively examined mean changes in Autism Diagnostic Interview–Revised (ADI-R) Total and Domains scores and stability of the ADI-R diagnostic classification in 28 children with autism initially assessed at age 2–4 years and reassessed 2 years later. Mean Total, Social Interaction, and Communication scores decreased significantly from Time 1 to Time 2. Restricted/repetitive Domain mean scores did not change over time. The ADI-R diagnostic classification was stable in 67% of children using the current published criteria. The stability increased to 78% when a modified criterion was used in the Restricted/repetitive Domain and to 88% when the broader ASD criteria were used. Among pre-schoolers with autism, parent-reported symptoms decreased significantly at two-year follow-up in Social and Communication Domains but not in the Restricted/repetitive Domain. However, ADI-R diagnostic classification remained relatively stable over time. Revising ADI-R diagnostic criteria in the Restricted/repetitive Domain or including the broader ASD criteria may improve its sensitivity and diagnostic stability in younger children.

Keywords

autism; Autism Diagnostic Interview–Revised; assessment; stability; diagnostic classification

Copyright © The Author(s), 2011

Reprints and permissions: <http://www.sagepub.co.uk/journalspermissions.nav>

ADDRESS Correspondence should be addressed to: SUSAN HEPBURN, Department of Psychiatry, University of Colorado Denver School of Medicine, 13121 East 17th Avenue, Campus Box C-234, Aurora, CO 80045, USA. susan.hepburn@ucdenver.edu.

The diagnosis of an autism spectrum disorder (ASD) requires multidisciplinary assessment combining information from diverse sources (Dover and Le Couteur, 2007; Filipek et al., 2000; Risi et al., 2006). The clinical diagnosis is based upon criteria published in the Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition – Text Revision (*DSM-IV-TR*; American Psychiatric Association, 2000). Highly trained clinicians can accurately diagnose ASD in children by the age of 2 years, with early behavioral signs manifest in difficulties with joint attention, eye contact, smiling and sharing affect, play skills and imitation (Baranek, 1999; Cox et al., 1999; Stone et al., 1999; Sutera et al., 2007; Werner et al., 2000). However, the diagnosis is still often delayed. In a recent population-based study, the mean age for diagnosis was approximately 5 years (Wiggins et al., 2006). Diverse factors may account for the delay in identification, including lower diagnostic stability, limitations in the use of standardized instruments, and the effect of the level of functioning on symptom expression in younger children; lower stability in the broader ranges of the spectrum (due to its heterogeneity) compared to Autistic Disorder; and a lack of qualified professionals to make the diagnosis (Wiggins et al., 2006).

The stability over time of the diagnosis of ASD based on standardized instruments has been examined in a number of studies. Longitudinal studies have evaluated the Autism Diagnostic Interview–Revised (ADI-R; Lord et al., 1994) for diagnostic classification or domain-specific symptom stability, as well as for factors that predicted changes (Charman et al., 2005; Cox et al., 1999; Fecteau et al., 2003; Kleinman et al., 2007; Lord, 1995; Lord et al., 2006; Moore and Goodson, 2003; Piven et al., 1996).

In general, these studies have shown that the stability of the ADI-R diagnostic classification was lower than that of the clinical diagnosis. For example, Kleinman et al. (2007) reported ADI-R classification stability in 67% of the cohort compared to 80% for the clinical diagnosis.

Studies of older children have generally reported higher stability compared to those involving younger children (e.g., Kleinman et al., 2007; Lord, 1995). Using an earlier version of the ADI-R, Lord (1995) reported that of seven children who met criteria for autism on the ADI-R but failed to meet the criteria for a clinical diagnosis of autism at age 2, none met the ADI-R criteria at the second assessment. On the other hand, Cox et al. (1999) reported good ADI-R diagnostic classification stability in a sample of young children. Some researchers have suggested that the ADI-R agreement with the clinical diagnosis and the stability of the classification over time in young children could be improved by reducing the cut-off for the Restricted/repetitive behaviors Domain from 3 to 2 (Cox et al., 1999; Richler et al., 2007; Ventola et al., 2006). Others have suggested excluding Restricted/repetitive Domain scores from the diagnostic algorithm altogether (Chawarska et al., 2007; Ventola et al., 2006; Wiggins and Robbins, 2008) or including the broader ASD criteria for children who fail to meet the current published criteria for autism (Risi et al., 2006).

Longitudinal changes in ADI-R domain scores have varied depending upon the study and the domain. In the Social Interaction and Communication domains, four of seven studies reported decreases in scores over time. In contrast, in the Restricted/repetitive Domain, five of seven studies reported increases in scores. Studies that used retrospective cohort designs

(Fecteau et al., 2003; Piven et al., 1996) tended to report much larger declines in domain scores over time compared to those that used a prospective cohort design. Retrospective cohort design studies compare current scores to retrospectively recalled scores at an earlier age, using data collected during the same assessment. This type of study design may suffer from biased recall of the informant. Prospective cohort design studies compare current scores assessed during two different assessments and therefore, may be more accurate in assessing changes.

Some of the studies (Cox et al., 1999; Lord et al., 2006; Moore and Goodson, 2003) used the ADI-R in children with an ASD rather than those with autism only. It is important to note that the ADI-R diagnostic criteria have been validated only for children with Autistic Disorder (Risi et al., 2006; Wiggins and Robbins, 2008). This may have contributed to some of the variability in results between studies.

Overall, there remains a lack of consensus concerning longitudinal changes in ADI-R scores and their implication for diagnostic classification and symptom stability in young children with autism. The goal of this study was to assess the stability of ADI-R Domain and Total scores, and the stability of ADI-R diagnostic classification in toddlers with autism across a two year period. Specifically, in a sample of children aged 2–4 years at initial assessment, we examined changes in current Total and Domain scores and changes in the proportion of children meeting the cut-off for ADI-R diagnostic classification using both the current published criteria and other alternative criteria.

Method

Study design

In this prospective cohort study, children were assessed first when they were between 2 and 4 years old and reassessed an average of 2 years later when they were between 4 and 6 years old.

Study population

The study population was a community-based sample of children with autism aged 2 to 4 years, recruited between 1997 and 2003 from various health and early education agencies and parent support/advocacy groups (e.g., Autism Society of Colorado). These children were part of a larger study funded through the Collaborative Programs of Excellence in Autism (CPEA) of the National Institute of Child Health and Human Development (PI = Rogers). Research assessments were conducted by two or more experienced clinical psychologists, a speech pathologist, and/or advanced trainees in clinical psychology. Evaluations included standardized administration of the ADI-R, the Autism Diagnosis Observation Schedule–Generic (ADOS-G; DiLavore et al., 1995; Lord et al., 2000), the Mullen Scales of Early Learning (MSEL; Mullen, 1995), and other measures not reported here. The clinical diagnosis of autism was verified by one or two psychologists based upon a comprehensive review of case information, as well as direct observations of the child and parent interviews. A *DSM-IV* check list was completed to operationalize clinical diagnoses. Thirty-six children who satisfied criteria for autism on the clinical assessment and were evaluated at JFK

Partners in Denver, Colorado, were enrolled in this study. These children had no uncorrected visual or hearing impairments, had walked by 15 months of age, and were free from any other identified medical conditions at study enrollment. All had been diagnosed previously with Autistic Disorder by a community-based provider.

Of the 36 enrolled children, 32 were reassessed approximately two years after the initial assessment. However, data on three of these children were not usable because of problems with reliability of test administration. A fourth child was excluded after the clinical psychologist concluded that the child's mental age was too low for a reliable ADI-R classification and she had been recently diagnosed with a degenerative neurological disorder. Therefore, the final study sample consisted of 28 children (78% of those originally enrolled in the cohort).

Measures

Autism Diagnostic Interview–Revised (ADI-R; Lord et al., 1994)

The ADI-R is a semi-structured, standardized parent interview developed to assess the presence of symptoms of autism in early childhood across all three main symptom domains: reciprocal social interaction, communication, and restricted/repetitive behaviors. The interview consists of over 100 questions. This instrument yields an algorithm score and cut-offs for a classification of autism. The algorithm differentiates autism from other developmental disorders with greater than 90% sensitivity and specificity for subjects with mental ages of 18 months and older (Lord et al., 1994). Items are scored as 0 (no definitive behavior of the type specified), 1 (behavior of the type specified is present in an abnormal form, but not sufficiently severe or frequent to meet the criteria for a 2), 2 (definite abnormal behavior), 3 (extreme severity of the specified behavior). In the algorithm, scores of 3 are recoded as 2. The ADI-R has been suggested to be less sensitive in children with mental ages below 18 months or with IQ below 20 (Cox et al., 1999; Lord, 1995), and those with a chronological age of less than 24 months (Cox et al., 1999; Ventola et al., 2006). It is also limited by reliance on an informant who is familiar with the child's early developmental history (Ventola et al., 2006).

The ADI-R was administered to parents or legal guardians of all enrolled children by interviewers who were fully trained in ADI-R to the standards of research reliability set by the developer. Inter-observer reliability of 85% or better was maintained across the duration of the study and systematically evaluated for more than 20% of the subjects.

Autism Diagnostic Observation Schedule–Generic (ADOS-G; DiLavore et al., 1995; Lord et al., 2000)

The ADOS-G is a semi-structured standardized direct child assessment-based interview that employs developmentally appropriate social and play-based interactions in a 45–60 minute session to elicit symptoms of autism in social interaction, communication, play, and repetitive behaviors. The ADOS-G consists of four different modules, each directed at a particular level of language ability. Items are typically scored on a 3-point scale from 0 (no evidence of abnormality related to autism) to 2 (definite evidence). Some items include a code of 3 to indicate abnormalities so severe as to interfere with the observation.

The ADOS-G was administered to all subjects in the study as part of the diagnostic qualification process. Raters were trained with reliability of 80% or better item agreement on three consecutive administrations using the full range of scores (0–3). Inter-observer reliability was maintained at 85% and checked for 20% of participants across the period of data gathering.

Mullen Scales of Early Learning (MSEL; Mullen, 1995)

The MSEL is a standardized developmental test for children ages birth to 68 months that yields five subscale scores: Gross Motor, Fine Motor, Visual Reception, Expressive Language, and Receptive Language. The MSEL provides standardized t-scores and age equivalents for all five domains. To reduce floor effects, overall, verbal and nonverbal developmental quotients (DQ) can be constructed by dividing overall, verbal and nonverbal mental ages by chronological age (Munson et al., 2008). This conversion reduces floor effects.

The MSEL was administered according to standardized instructions to all subjects by raters with advanced degrees who were trained in assessing young children with autism and other developmental disorders. Meetings were held regularly among clinicians administering the Mullen to ensure reliability in the administration of the assessment.

Data collection

During the initial visit after enrollment (Time 1, or T1), a psychologist or Masters' level clinician administered the ADI-R to the parent or legal guardian of the child. The ADI-R was administered again to the same parent or legal guardian an average of two years later (Time 2, or T2), when the child was 4 to 6 years old. The interviewers were not blind to the clinical diagnosis and were not randomly assigned. The session typically lasted between 120 and 180 minutes. At the end of the visit, the interviewer completed the ADI-R algorithm for each domain and computed ADI-R Total Scores for current behaviors (i.e., Social Interaction + Communication + Restricted/repetitive scores).

At T1, all children had a chronological age less than 48 months; therefore the algorithm for younger children was used. At T2, all but one child had a chronological age over 48 months, so the algorithm for children who were older than 48 months was used in all but the one younger child. When assessing dimensional changes in Domain and Total scores, only 'current' scores at both times (T1 and T2) were used. To account for the differences in the number of items in each algorithm, McGovern and Sigman (2005) and others have proposed using only the items which are common in both algorithms. However, this approach could introduce a bias by eliminating items that could be informative about the current level of autistic symptoms. Therefore, 'weighted' domain scores were instead computed by dividing the crude scores by the number of items used at each assessment. Children were classified as 'verbal' or 'nonverbal' based on the response provided by the parent or legal guardian on item # 30 on the ADI-R (overall level of language).

In order to meet the ADI-R classification criteria for autism, a child must obtain scores above the cut-off in all three domains, with parental report of and parents developmental concerns prior to the age of 3 years. The recommended cut-offs are 10 for the Social

Interaction Domain, 8 for the Communication Domain for verbal children and 7 for nonverbal children, and 3 for the Restricted/repetitive Domain (Lord et al., 1994). For ADI-R diagnostic classification, appropriate scores (i.e., 'current,' 'ever,' or 'most abnormal') were included based on the type of algorithm used during the assessment (T1 or T2).

The ADOS-G was administered by any one of the several trained clinicians, who were not randomly assigned. All subjects received Module 1 at T1. The ADOS-G was readministered on average 2 years later. At T2, 17 children received Module 1, seven children Module 2, and four children Module 3.

The MSEL was administered at T1 by any one of a group of trained clinicians who were not randomly assigned. The same instrument was repeated on average 2 years later (T2).

Two children at T2 had a chronological age over 68 months (specifically 70 and 74 months) and both children were significantly delayed at T1, so the clinicians chose to readminister the MSEL at T2 to provide a measure of change over time. In addition, a larger study funded through the CPEA also planned to assess the stability of Mullen scores over time. The MSEL provided multiple measures of global development: nonverbal mental age, verbal mental age, overall mental age, NVDQ, VDQ, and overall DQ.

Child and family characteristics

Mother's level of education was used as a proxy for the family's socioeconomic status (SES), and was categorized into 3 levels: those with no high school degree, those with at least a high school degree but not a college degree, and those with a college degree or higher. Self-reported race/ethnicity was categorized into the following four groups: Non-Hispanic White, Non-Hispanic African American, Hispanic, and Multiracial. For analytic purposes, Non Hispanic White ('White') children were compared to children of all other races and ethnicities ('Minority'). Level of functioning was based on the DQ, categorized as developmentally delayed ($DQ \leq 70$) and not delayed ($DQ > 70$).

Outcome variables

The primary outcomes of interest are 1) mean change in ADI-R Total and domain-specific current scores, and 2) change in the proportion of children meeting the cut-off for diagnosis between T1 and T2. Changes in the current scores were defined as the difference between weighted scores at T2 and T1. In this study, only the items found in the algorithm for diagnosis were used in assessing changes in scores. For ADI-R diagnostic classification, the recommended cut-offs (see above) for each domain were included. In a secondary analysis, two modified criteria for diagnostic classification were used: a) reducing the cut-off from 3 to 2 in the Restricted/repetitive behavior domain, and b) applying the broader ASD criteria to children who did not meet the current published ADI-R criteria for autism, that is, the total of the social and communication domain scores within two points of the total cut-off if younger than 3 years, or meeting the cutoff on either the Social Interaction or Communication Domains and coming within two points on the other, if the child was 3 years or older (Risi et al., 2006).

Based on the ADI-R scores at both assessments in relation to the above cut-offs in each domain, children were assigned to one of four ADI-R stability categories: children who had scores equal to or above the cut-off on both assessments, those who had scores below the cut-off on both assessments, those who had scores below the cut-off at T1 but equal to or above the cut-off at T2, and those who had scores equal to or above the cut-off at T1 but below the cut-off at T2. Children whose scores were above the cut-off at both assessments or below the cut-off at both assessments were considered to have a 'stable' diagnostic classification. Children whose scores changed from above to below the cut-off, or vice versa, between T1 and T2 were considered to have a 'changed' diagnostic classification.

When the broader ASD criteria were included, children were classified in one of the following three groups at each time: those meeting the current published ADI-R criteria (Autism), those meeting the broader ASD criteria (ASD), and those who did not meet either criteria (non-ASD).

Human subjects protection

The study was approved by the Colorado Multiple Institutional Review Board (COMIRB). Written consent was obtained from the parents or legal guardians of all enrolled children.

Statistical analysis

Paired t-tests were used to compare the changes in mean ADI-R current Total and domain-specific scores between T1 and T2. The change in the proportion of children meeting the cut-off for ADI-R diagnostic classification between T1 and T2 was examined using McNemar's Test.

For all statistical tests, a *p*-value of .05 was selected to assess statistical significance.

Results

Characteristics of the participants

Most of the 28 children were boys (78.6%), Non-Hispanic White (78.6%), and had a mother with a college degree or higher (76.9%) (Table 1). At T1, the mean chronological age was 33.0 months (*SD* = 4.2) and the mean overall mental age was 19.4 months (*SD* = 6.5). The mean overall DQ was 57.0 (*SD* = 18.7). The mean NVDQ was 68.0 (*SD* = 17.6) and the mean VDQ was 46.1 (*SD* = 21.6). Twenty-four children (85.7%) were nonverbal and 22 (78.6%) had a DQ in the developmentally delayed range. At T2, the mean chronological age was 58.0 months (*SD* = 5.2) and the mean overall mental age was 34 months (*SD* = 12.2). The mean overall DQ was 60.0 (*SD* = 22.8). The mean NVDQ was 67.5 (*SD* = 21.1) and the mean VDQ was 52.3 (*SD* = 27.2). Twelve children (42.9%) were nonverbal and 19 (67.9%) had a DQ in the developmentally delayed range.

Longitudinal changes in ADI-R scores

Mean Total scores, and scores in the Social Interaction and Communication Domains all decreased significantly (i.e., parents reported that the behaviors in these two domains

became less severe). However, mean scores in the Restricted/repetitive behaviors Domain did not change between the two assessments (Table 2).

Diagnostic classification stability

Using the current published criteria and relying upon clinical diagnosis as the gold standard, the ADI-R correctly identified 21 children (75%) at T1 and 19 children (67.9%) at T2 (Table 3). There was no significant difference in the proportion of children meeting the cut-off for the ADI-R diagnostic classification between the two assessments ($p = .77$). The ADI-R correctly identified 23 children (82.2%) at T1 and 21 children (75%) at T2 when the modified criterion for the Repetitive Behavior Domain was used. Using the current published criteria, among the 21 children identified at T1, 14 (67%) were also identified at T2; that is, 50% of all children tested were above the cut-off for the ADI-R at both T1 and T2 (Table 3). Using the modified criterion in the Restricted/Repetitive Domain, 18 of the 23 children (78%) who originally met the criteria at T1 also met the criteria at T2 (i.e., 64.3% of all children tested).

Using the current published criteria, the ADI-R missed seven children at T1: three children (42.8%) failed to meet the cut-off in the Restricted/Repetitive Domain, one child (14.3%) in the Communication domain, two children (28.6%) in both Social Interaction and Communication Domains, and one child (14.3%) in both Social Interaction and Restricted Behavior Domains. Of the seven children who failed to meet the current published ADI-R criteria at T1, five (71.4%) met the broader ASD criteria proposed by Risi et al. (2006). Therefore, a total of 26 children (92.9%) met either the current ADI-R criteria for autism or the broader ASD criteria (Table 3). Among these 26 children, 23 (88%) met either the current ADI-R criteria or the broader ASD criteria at T2.

Using the current published criteria, the ADI-R missed nine children at T2: five children (55.6%) in the Social Interaction Domain, two children (22.2%) in the Communication Domain and two children (22.2%) in the Restricted/Repetitive Domain. Of the nine children missed at T2, six (66.7%) met the broader ASD criteria. Therefore, a total of 25 children (89.3%) at T2 met either the current ADI-R criteria for autism or the broader ASD criteria (Table 3).

Compared to the current published criteria, the ADI-R missed fewer children: five at T1 and seven at T2 when the criterion in the Restricted/Repetitive Domain was modified (Table 3).

Figure 1 shows the proportion of children classified in each the three groups (Autism, ASD and non-ASD) at T1 and T2.

When the current published criteria were used, five children (17.9%) moved from below the cut-off at T1 to above the cut-off at T2, while seven children (25%) moved from above the cut-off at T1 to below the cut-off at T2. Hence, 42.9% of children changed ADI-R diagnostic classification between T1 and T2. Two children (7.1%) were below the cut-off at both T1 and T2. After modifying the criterion in the Restricted/Repetitive Behavior Domain, the proportion of children moving from below to above and vice-versa decreased to 28.6% but the proportion of children who were below the cut-off at both times did not change

(Table 3). When the broader ASD criteria were included, the proportion of children moving from below to above and vice-versa decreased to 17.8% and no child was classified as ‘non-ASD’ at both times (Table 3).

Reanalyzing the data to eliminate four children who had a nonverbal mental age less than 18 months at T1, or all 14 children who had an overall mental age less than 18 months at T1, did not change the results (data not shown).

Discussion

The purpose of this study was to examine longitudinal changes in ADI-R current scores and the stability of the ADI-R diagnostic classification over time in a sample of young children with autism.

The proportions of children meeting the cut-offs for autism and the mean ADI-R scores in the Restricted/Repetitive Domain remained relatively stable over a two-year period. Mean scores in Total and in the Communication and Social Interaction Domains declined significantly (i.e., behaviors became less severe) from T1 to T2, indicating an improvement in autistic symptoms as reported by the parents. The changes in mean communication scores correlated with the considerable movement of the children from ‘nonverbal’ to ‘verbal’ status between T1 and T2. However, the improvements in communication and social interaction symptoms did not significantly reduce the proportion of children meeting the cut-off for a ADI-R diagnostic classification of autism. These apparently contradictory findings demonstrate the importance of evaluating changes in ADI-R scores both dimensionally and categorically. Dimensional changes in ‘current’ ADI-R scores provide information concerning the level of autistic symptoms at the time of the interview. However, categorical changes provide insight concerning the ADI-R diagnostic classification, which use ‘current’ scores only for children who are younger than 48 months in the Social Interaction and communication domains. Even if the proportion of children meeting the cut-off for an autism classification did not change significantly overall, there were still some individual changes, as shown by the proportions of children who moved from ‘above to below’ and ‘below to above’ the diagnostic cut-offs between T1 and T2.

The improvement of autistic symptoms over time is a phenomenon described in previous studies (Kelley et al., 2006; Starr et al., 2003, for review). However, these authors also reported that despite improvements in symptoms, the children continued to present with a number of difficulties in different domains compared to typically developing children and retained their clinical diagnoses of autism. Our findings are comparable to theirs, in that all the children who ‘improved’ in our sample based on the ADI-R diagnostic classification still maintained a clinical diagnosis of autism. This supports the need to base the diagnosis of autism on information from multiple sources rather than relying on the results from a single assessment, as suggested by Risi et al (2006) and others.

Two children determined by two independent, experienced clinicians to have autism at both visits were missed by the ADI-R at both assessments. This may be related to factors that can impact parent report of symptoms. One informant was a very young mother who had some

limitations in parenting experience. The other was a mother who already had two children affected by ASD. She may have under-reported symptoms given that this child was less affected than her other children. However, neither child was missed when the broader ASD criteria were used.

The sensitivity and stability of the ADI-R diagnostic classification increased when a modified criterion in the Repetitive Behavior Domain was used. The modified criterion for the repetitive behavior domain should be considered when using the ADI-R in younger children, who may fail to show enough repetitive behaviors to meet cut-offs. The broader ASD criteria proposed by Risi et al. (2006) also correctly identified a higher proportion of children than did the current ADI-R criteria. Unlike the current criteria, the broader ASD criteria use lower cut-offs and place more emphasis on Social Interaction and Communication Domains. Therefore, researchers should consider including the broader ASD criteria when using the ADI-R in younger samples.

Unlike some previous studies, this study evaluated longitudinal changes in ADI-R scores both dimensionally and categorically in a sample of young children with autism. In addition, we used a prospective cohort design, the same informant at both assessments, and 'weighted' scores to account for the difference in the number of items included in the ADI-R algorithms used at different ages. The findings of this study were similar to those published by Kleinman et al. (2007) concerning the magnitude of the ADI-R diagnosis stability. As reported by Kleinman et al. (2007), the stability of the ADI-R was lower compared to the clinical diagnosis. This is a common finding for instruments that do not include observations of the affected child.

Our findings on changes in domain scores differed somewhat from those reported by some other investigators (Charman et al., 2005; Fecteau et al., 2003; Lord, 1995). In particular, Fecteau et al. (2003) reported a decrease in scores (i.e., improved symptoms) in all three domains, although the decrease in the Restricted/Repetitive Behavior Domain was smaller than the decreases in the other two domains. Charman et al. (2005) and Lord (1995) both reported an increase in Restricted/Repetitive Behavior Domain scores, whereas we found no major change in this domain. Lord (1995) reported an increase in Communication Domain scores, whereas we reported a decline. Factors in our study design, methodology, and implementation that might explain at least some of the differences between our study and these others include use of a prospective cohort design, younger age of the children, short duration of follow-up, use of 'weighted' scores, and over-representation of children with lower DQ. Charman et al. (2005) also did not adjust scores to account for the difference in items between the ADI-R versions used for younger and older children. On the other hand, our findings on all three domain score changes were similar to those published by Piven et al. (1996). We found that while scores from T1 to T2 in the Social and Communication domains decreased significantly, there were no major changes in the Restricted/Repetitive Domain scores. Charman et al. (2005) and Piven et al. (1996) have suggested that the development of restricted/repetitive behaviors follows a third axis that is different from those of social and communication domains.

Limitations of the study

We included a sub-sample of children recruited for a larger study who were evaluated locally, and this could limit the generalizability of our findings. Parents were not blind to the clinical diagnosis during the interview, since all children came into the study with a pre-existing diagnosis of autism, and this could be a potential source of recall bias. The interviewers were not blind to the clinical diagnosis of children and were not randomly assigned, meaning that sometimes the same interviewer administered the ADI-R at both times to the same child, which could have introduced an interviewer bias. The relatively short duration of follow-up may have limited the possibility of finding substantial changes in the Restricted/Repetitive Domain, as some of these behaviors require time to manifest. The small size of the sample could have limited our ability to identify significant changes in the Restricted/Repetitive Domain scores.

Conclusion

This study has confirmed findings from other studies that ADI-R domain scores can change over time and that although the ADI-R diagnostic classification is relatively stable after the age of 2, it is not as stable as a clinical diagnosis. Therefore, the ADI-R cannot be used alone for diagnostic purposes, as has been suggested by others, but instead should be used with the clinical diagnosis and other instruments (Risi et al., 2006). Our study has demonstrated that the sensitivity of the ADI-R diagnostic classification in younger children with autism can be improved by lowering the cut-off in the Restricted/Repetitive T Domain or by using broader ASD criteria as proposed by Risi et al. (2006). The ADI-R total scores, and all ADI-R domain-specific scores except in the Restricted/Repetitive Domain, declined significantly over time, indicating improvements in symptoms. However, the proportion of children who met the cut-off for ADI-R classification of autism did not significantly vary between the two assessments. This shows that improvements in symptoms reported by parents do not necessarily translate into changes in diagnosis, so clinicians should be cautious when interpreting the results from one assessment.

Despite important progress in the understanding of autism and in the instruments developed for diagnosis, there is still a need for rigorous longitudinal studies with enough power to evaluate the diagnostic utility of these instruments in younger children and the profile of changes in domain scores. Because of the condition's rarity, large multi-center studies, or similarly designed studies from multiple centers that can be combined using meta-analysis, may be required for thorough evaluation of these issues. The effects of different therapies on test scores, and of the information available to parents concerning autism on parental reporting, should be examined in future studies.

Acknowledgements

This study was supported by Grant # U19HD35468 from the National Institute for Child Health and Human Development (NICHD). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NICHD. We would like to thank all the families who participated in this study. This paper was prepared in partial fulfillment of the requirements for the degree of Master of Sciences in Public Health at the University of Colorado Denver.

References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR®). APA Press; Washington, DC: 2000. Pervasive Developmental Disorders.
- Baranek G. 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; 29:213–220. [PubMed: 10425584]
- Charman T, Taylor E, Drew A, Cockerill H, Brown J, Baird G. Outcomes at 7 Years of Children Diagnosed with Autism at Age 2: Predictive Validity of Assessments Conducted at 2 and 3 Years of Age and Pattern of Symptom Change over Time. *Journal of Child Psychology and Psychiatry*. 2005; 46:500–513. [PubMed: 15845130]
- Chawarska K, Klin A, Volkmar F. Autism Spectrum Disorder in the Second Year: Stability and Change in Syndrome Expression. *Journal of Child Psychology and Psychiatry*. 2007; 48:128–138. [PubMed: 17300551]
- Cox A, Klein K, Charman T, Baird G, Baron-Cohen S, Swettenham J, et al. Autism Spectrum Disorders at 20 and 42 Months of Age: Stability of Clinical and ADI-R Diagnosis. *Journal of Child Psychology and Psychiatry*. 1999; 40:719–732. [PubMed: 10433406]
- DiLavore PC, Lord C, Rutter M. The Pre-Linguistic Autism Diagnostic Observation Schedule. *Journal of Autism and Developmental Disorders*. 1995; 25:355–379. [PubMed: 7592249]
- Dover C, Le Couteur A. How to Diagnose Autism. *Archives of Disease in Childhood*. 2007; 92:540–545. [PubMed: 17515625]
- Fecteau S, Mottron L, Berthiaume C, Burack J. Developmental changes of autistic symptoms. *Autism*. 2003; 7:255–268. [PubMed: 14516059]
- Filipek P, Accardo P, Ashwal S, Baranek G, Cook EJR, Dawson G, et al. Practice Parameter: Screening and Diagnosis of Autism”. A Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. *Neurology*. 2000; 55:468–479. [PubMed: 10953176]
- Kelley E, Paul J, Fein D, Naigles L. Residual Language Deficits in Optimal Outcome Children with a History of Autism. *Journal of Autism and Developmental Disorders*. 2006; 36:807–828. [PubMed: 16897404]
- Kleinman J, Ventola P, Pandey J, Verbalis A, Barton M, Hodgson S, et al. Diagnostic Stability in Very Young Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*. 2007; 38:606–615. [PubMed: 17924183]
- Lord C. Follow-up of Two-Year-Olds Referred for Possible Autism. *Journal of Child Psychology and Psychiatry*. 1995; 36:1365–1382. [PubMed: 8988272]
- Lord C, Risi S, DiLavore P, Shulman C, Thurm A, Pickles A. Autism from 2 to 9 Years of Age. *Archives of General Psychiatry*. 2006; 63:694–701. [PubMed: 16754843]
- Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore P, et al. The Autism Diagnostic Observation Schedule–Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. *Journal of Autism and Developmental Disorders*. 2000; 30:205–223. [PubMed: 11055457]
- Lord C, Rutter M, Le Couteur A. Autism Diagnosis 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]
- McGovern C, Sigman M. Continuity and Change from Early Childhood to Adolescence in Autism. *Journal of Child Psychology and Psychiatry*. 2005; 46:401–408.
- 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, EM. Mullen Scales of Early Learning. AGS edn. American Guidance Service Inc; Circle Pines, MN: 1995.

- Munson J, Dawson G, Sterling L, Beauchaine T, Zhou A, Koehler E. Evidence of Latent Classes of IQ in Young Children with Autism Spectrum Disorder. *American Journal of Mental Retardation*. 2008; 113:439–452. [PubMed: 19127655]
- Piven J, Harper J, Palmer P, Arndt S. Course of Behavioral Change in Autism: A Retrospective Study of High-IQ Adolescents and Adults. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1996; 35:523–529. [PubMed: 8919715]
- Richler J, Bishop S, Kleinke J, Lord C. Restricted and Repetitive Behaviors in Young Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*. 2007; 37:73–85. [PubMed: 17195920]
- Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, et al. Combining Information from Multiple Sources in the Diagnosis of Autism Spectrum Disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2006; 45:1094–1103. [PubMed: 16926617]
- Starr E, Szatmari P, Bryson S, Zwaigenbaum L. Stability and Change among High-Functioning Children with Pervasive Developmental Disorders: A 2-Year Outcome Study. *Journal of Autism and Developmental Disorders*. 2003; 33:15–22. [PubMed: 12708576]
- Stone W, Lee E, Ashford L, Brissie J, Hepburn S, Coonrod E, et al. Can Autism Be Diagnosed Accurately in Children under 3 Years? *Journal of Child Psychology and Psychiatry*. 1999; 40:219–226. [PubMed: 10188704]
- Sutera S, Pandley J, Esser E, Rosenthal M, Wilson L, Barton M, et al. 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 J, Pandey J, Barton M, Allen S, Green J, et al. Agreement among Four Diagnostic Instruments for Autism Spectrum Disorders in Toddlers. *Journal of Autism and Developmental Disorders*. 2006; 36:839–847. [PubMed: 16897398]
- Werner E, Dawson G, Osterling J, Dinno N. Recognition of Autism Spectrum Disorders before One Year of Age: A Retrospective Study Based on Home Videotapes. *Journal of Autism and Developmental Disorders*. 2000; 30:157–161. [PubMed: 10832780]
- Wiggins L, Baio J, Rice C. Examination of the Time between First Evaluation and First Autism Spectrum Diagnosis in a Population-Based Sample. *Journal of Developmental & Behavioral Pediatrics*. 2006; 27:79–87.
- Wiggins L, Robbins D. Excluding the ADI-R Behavioral Domain Improves Diagnostic Agreement in Toddlers. *Journal of Autism Developmental Disorders*. 2008; 38:972–976. [PubMed: 17879150]

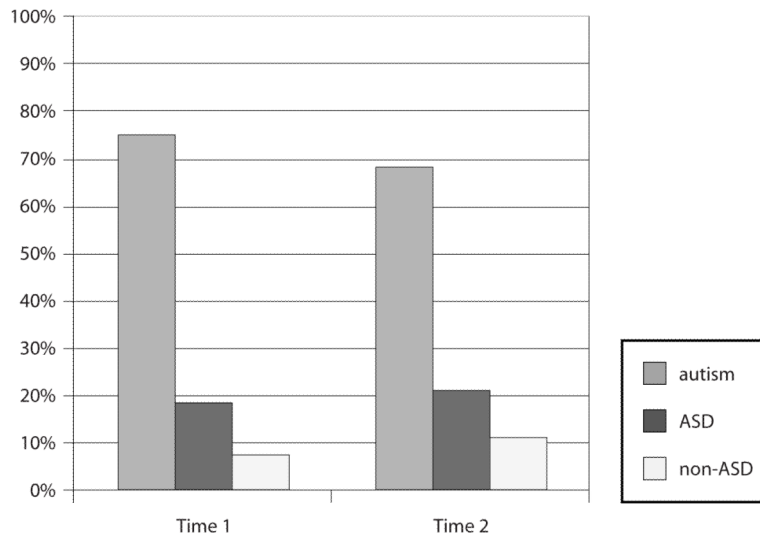


Figure 1.
Proportion of children in the three categories during each assessment

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Demographic characteristics at initial assessment and two-year follow-up in children with autism

Variable	Time 1	Time 2
Gender (<i>n, %</i>)		
Male	22 (78.6)	
Female	6 (21.4)	
Race/Ethnicity (<i>n, %</i>)		
Non Hispanic White	22 (78.6)	
Non Hispanic Black	2 (7.1)	
Hispanic	2 (7.1)	
Multi-racial	2 (7.1)	
Mother's education ^a (<i>n, %</i>)		
Less than high school degree	0 (0)	
High school degree with/without some college	6 (23.1)	
College degree or higher	20 (76.9)	
Chronological age (Months)		
Mean (<i>SD</i>)	33.0 (4.2)	58.0 (5.2)
Median	33.0	58.0
Range	22–42	47–74
Overall mental age (Months)		
Mean (<i>SD</i>)	19.4 (6.5)	34.0 (12.2)
Median	18.0	31.0
Range	9.3–41.8	12.3–56.3
Overall DQ ^b		
Mean (<i>SD</i>)	57.0 (18.7)	60.0 (22.8)
Median	54.2	56
Range	34–119	20–101
Nonverbal DQ (NVDQ)		
Mean (<i>SD</i>)	68.0 (17.6)	67.5 (21.1)
Median	65.5	64.1
Range	40.4–125.7	27.5–101.7
Verbal DQ (VDQ)		
Mean (<i>SD</i>)	46.1 (21.6)	52.3 (27.2)
Median	39.2	52.3
Range	17.2–104.3	13.3–105.7
Verbal Status (<i>n, %</i>)		
Nonverbal	24 (85.7)	12 (42.9)
Verbal	4 (14.3)	16 (57.1)
Level of functioning (<i>n, %</i>)		
DQ > 70	6 (21.4)	9 (32.1)
DQ ≤ 70	22 (78.6)	19 (67.9)

Note. DQ = developmental quotient; NVDQ = nonverbal developmental quotient; VDQ = verbal developmental quotient.

^a $n = 2$ missing.

^b DQ, NVDQ and VDIQ were obtained using the Mullen Scales of Early Learning (MSEL; Mullen, 1995).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

ADI-R total and domain-specific scores^a at initial assessment and two-year follow-up, and mean changes in scores between assessments, in children with autism

Variable	Time 1 Mean (SD)	Time 1 Median	Time 2 Mean (SD)	Time 2 Median	Mean Change T1 to T2 (95% CI)	<i>p</i> -value for Mean Change
ADI-R Total	3.73 (1.01)	3.96	3.15 (1.14)	3.53	-0.58 (-1.02, -0.15)	0.01
ADI-R Social Interaction Domain	1.24 (0.40)	1.34	1.01 (0.51)	1.06	-0.23 (-0.38, -0.07)	0.005
ADI-R Communication Domain	1.55 (0.46)	1.71	1.19 (0.47)	1.19	-0.36 (-0.55, -0.17)	0.001
ADI-R Restricted/Repetitive Behavior Domain	0.94 (0.48)	0.90	0.95 (0.47)	0.92	0.01 (-0.23, 0.23)	0.99

Note. ADI-R = Autism Diagnostic Interview–Revised.

^aScores weighted for number of items at each assessment.

Table 3

Proportions of children with autism who had scores above the ADI-R diagnostic cut-off at initial assessment and two-year follow-up, and proportions in the four ADI-R stability categories

Category	Above cut-off at T1 n (%)	Above cut-off at T2 n (%)	Above cut-off at both T1 and T2 n (%)	Below cut-off at both T1 and T2 n(%)	Below cut-off at T1; above cut-off at T2 n (%)	Above cut-off at T1; below cut-off at T2 n (%)
ADI-R all domains	21 (75.0)	19 (67.9)	14 (50.0)	2 (7.1)	5 (17.9)	1 (25.0)
ADI-R all domains – Modified ^a	23 (82.2)	21 (75.0)	18 (64.3)	2 (7.1)	3 (10.7)	5 (17.9)
ADI-R current criteria or broader ^b ASD criteria	26 (92.9%)	25 (89.3%)	23 (82.2%)	0 (0%)	2 (7.1%)	3 (10.7%)
ADI-R Social Interaction Domain	25 (89.3)	23 (82.2)	21 (75.0)	1 (3.6)	2 (7.1)	4 (14.3)
ADI-R Communication Domain	25 (89.3)	26 (92.9)	23 (82.2)	0 (0.0)	3 (10.7)	2 (7.1)
ADI-R Restricted/Repetitive Behavior Domain	24 (85.7)	26 (92.9)	22 (78.7)	0 (0)	4 (14.2)	2 (7.1)
ADI-R Restricted/Repetitive Behavior Domain – Modified [†]	26 (92.9)	28 (100.0)	26 (92.9)	0 (0.0)	2 (7.1)	0 (0.0)

Note. ADI-R = Autism Diagnostic Interview–Revised.

^aIncorporates change in the cut-off for the restricted/repetitive behavior domain from 3 to 2.

^bIncludes the broader ASD criteria for children who did not meet the current published ADI-R criteria.