



Published in final edited form as:

J Child Psychol Psychiatry. 2013 February ; 54(2): 178–185. doi:10.1111/j.1469-7610.2012.02607.x.

Diagnosis of autism spectrum disorders in two year-olds: A study of community practice

Christina M. Corsello,

Rady Children's Hospital San Diego

Natacha Akshoomoff, and

University of California, San Diego and Rady Children's Hospital San Diego

Aubyn C. Stahmer

Rady Children's Hospital San Diego and University of California, San Diego

Abstract

Background—Longitudinal research studies have demonstrated that experienced clinicians using standardized assessment measures can make a reliable diagnosis of Autism Spectrum Disorders (ASDs) in children under age 3. Limited data are available regarding the sensitivity and specificity of these measures in community settings. The aims of this study were to determine how well a standardized diagnostic observational measure (Autism Diagnostic Observation Schedule; ADOS) functions alone, and with a brief parent measure within a community setting when administered by community clinicians.

Methods—Clinical records for 138 children between the ages of 24 and 36 months of age who were evaluated for possible ASD or social/language concerns at a hospital based developmental evaluation clinic were examined. Evaluations were conducted by community based clinical psychologists. Classification results obtained from standardized diagnostic measures were compared with case reviewer diagnosis, by reviewers blind to scores on diagnostic measures, using The Records-based Methodology for ASD Case Definition that was developed by the Metropolitan Atlanta Developmental Disabilities Surveillance Program (MADDSP).

Results—When compared with case review diagnosis, the ADOS demonstrated strong sensitivity and specificity for both Autism vs. Not Autism and ASD vs. Non-spectrum (NS) diagnoses in this young sample. The Social Communication Questionnaire (SCQ), using the lower cut-off of 12, had adequate sensitivity when differentiating Autism from Not Autism, but weak sensitivity when differentiating ASD from NS, missing about 80% of the children with Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS). Using either the Modified Checklist for Autism in Toddlers (M-CHAT) or the SCQ in combination with the ADOS did not result in improved specificity over the ADOS alone and led to a drop in sensitivity when differentiating ASD from NS disorders.

Conclusions—These results demonstrate that following best practice guidelines, the ADOS can be successfully incorporated into clinical practice with relatively good sensitivity and specificity, and worked well with a referred sample of two-year-olds. A parent questionnaire did not lead to any improvement in diagnostic classification above the ADOS used in isolation.

As awareness of the early signs of autism spectrum disorders (ASD) increases, more physicians routinely screen for the presence of the disorder and children are referred to

community diagnostic clinics with suspicion of the disorder at younger ages. The American Academy of Pediatrics has suggested routine screening for ASD at the 18- and 24-month well child visits (Johnson, Myers, & American Academy of Pediatrics Council on Children With Disabilities, 2007) which can lead to an increase in referrals for diagnosis at early ages. Several studies have found that ASD can be reliably diagnosed in children under 3 years of age by experienced, highly trained clinicians in specialty clinic and research settings (Charman et al., 2005; Lord et al., 2006; Turner & Stone, 2007) and that the greatest accuracy in diagnosis of young children is achieved when using a standardized parent interview and a standardized observational measure in combination with clinical judgment (Corsello et al., 2007; Lord et al., 2006; Risi et al., 2006). Given the emphasis on appropriate early intervention and the expense of intensive early intervention services, diagnostic accuracy is crucial.

There are relatively few standardized diagnostic measures that have demonstrated strong discriminative ability in young children and predictive validity over time (Lord et al., 2006). Two of the most well validated and commonly used measures in research studies are the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 2001) and the Autism Diagnostic Interview – Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003). These measures are considered the gold standard in research protocols and have become widely known among clinical professionals. Both are available through a U.S. Publisher, but only the ADOS has become widely used by clinicians because the length of the ADI-R makes it difficult to incorporate into clinical practice. Additionally, studies using the ADI-R with children under 3 years of age have reported mixed results, with several reporting an under-identification of children with ASD primarily due to less children meeting criteria on the restricted behavior area (Ventola et al., 2006; Wiggins & Robins, 2008).

While the ADOS is widely available and recommended as one of the appropriate standardized measures recommended in several Best Practice Guidelines as an appropriate standardized diagnostic observation tool (California Department of Developmental Services, 2002; Filipek et al., 2000; Filipek et al., 1999; National Research Council, 2001), little is known about the effectiveness of this measure when administered by community clinicians who have not met the research training requirements and are not using the ADOS in combination with ADI-R. The results from studies of the ADOS conducted by community based clinicians that have achieved reliability are variable depending on clinic type and population evaluated. In an ASD specific clinic, the ADOS demonstrated good sensitivity and specificity in a sample of 2 to 8 year old children (Mazefsky & Oswald, 2006). On the other hand, the ADOS was not as specific and resulted in more false-positive ASD classifications in a complicated community based general developmental clinic sample of 2 to 16 year old children (Molloy, Murray, Akers, Mitchell, & Manning-Courtney, 2011).

Clinicians working outside an ASD specialty clinic or research setting often have less rigorous training on standardized diagnostic measures, see a lower percentage of children with ASD, and are more likely to evaluate children who are diagnostically complex (Molloy et al., 2011). Clinic-referred samples may be more likely to include children from minority backgrounds, parents with a lower level of education, and families with more disadvantaged circumstances, all factors that have been associated with increased age of diagnosis (Mandell, Ittenbach, Levy, & Pinto-Martin, 2007; Mandell, Novak, & Zubritsky, 2005; Shattuck et al., 2009). Complicated diagnostic presentations, combined with the often limited training of community clinicians on research based measures, may affect measurement accuracy when used in community programs. Additionally, most studies on community based samples compare the classification on the measure with clinical diagnosis that included the scores and observations from the measures. This decreases the objectivity and independence of the clinical diagnosis. The CDC has developed a method for giving a

diagnosis based on a record review (Centers for Disease Control, 2007). This method allows for an outside reviewer to provide a diagnosis without scores or classifications on measures.

The ADI-R is also available but is less commonly used in clinical practice given the difficulty of obtaining insurance funding for the lengthy interview. Aside from the ADI-R, there are very few diagnostic measures based on parent report. For this reason, many clinicians constrained by funding have resorted to using screening measures as well as parent questionnaires to obtain standardized parent information to augment their clinical interview. One such screening measure, the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003), is based on the ADI-R and is more widely utilized given its brevity. The SCQ has been demonstrated to discriminate ASD from Non-Spectrum (NS) diagnoses in children under age 5 with a reduced cut-off (Corsello et al., 2007) but little is known about how accurate it is with children under age 3. Another screening measure, the Modified Checklist for Autism in Toddlers (M-CHAT; Robins, Fein, Barton, & Green, 2001) is widely used in children under the age of 30 months (Kleinman et al., 2008; Pandey et al., 2008; Robins et al., 2001). While the M-CHAT can accurately identify many children with ASD, some children without an ASD screen positive for the disorder (Pandey et al., 2008), especially when a follow-up interview is not employed which is often the case in community clinics (Pandey, et. al., 2008; Robins & Dumont-Mathieu, 2006). Despite this weakness, without other options clinicians may turn to this screener when evaluating very young children referred for an ASD.

Most of the evidence for making an accurate diagnosis in young children is based on studies conducted with experienced and highly trained staff in research settings. We investigated whether these results can be replicated within a community based sample in a general developmental evaluation clinic with very young children when the clinician, in this case a clinical psychologist, has not received intensive training on standardized measures. The present study was designed to address the following questions: 1) How well does the ADOS perform when used by community clinicians in a young community based sample? 2) Does using a brief parent measure (M-CHAT or SCQ) in combination with the ADOS improve diagnostic accuracy in this young community sample?

Method

Subjects

The sample consisted of 138 consecutive children between the ages of 24 and 36 months evaluated for an ASD at a children's hospital developmental evaluation clinic, between October 2005 and August 2007. IQ was available for all but 3 of the children in the autism group. SCQ scores were available for 67 children, M-CHAT scores were available for 38 children (7 of these children also had the SCQ), and ADOS algorithm scores (communication + social + restricted and repetitive + play) were available for 94 children (see Table 1). There was not a significant difference between diagnostic groups in terms of age, $F(2, 135) = 1.02, p = .36$. There was a significant difference in IQ scores between diagnostic groups, $F(2, 132) = 6.72, p < .01$, with the Autism (AUT) group scoring significantly lower than both the Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS) and Nonspectrum (NS) groups.

There was not a significant difference in gender between diagnostic groups ($\chi^2(2) = .53$). Consistent with the 4:1 male to female ratio typical of children with ASD, males comprised 86% of the AUT group and 90% of the PDD-NOS group. Interestingly, there were many more males (81%) than females in the NS group as well. There was not a significant difference between diagnostic groups in terms of race, ($\chi^2(8) = .58$) with Caucasian children comprising 41% of the sample, and a much smaller percentage of African

American (2%) and Asian (3%) children. Fifteen percent of the sample identified themselves as 'Other' and race was unknown for 39% of the sample. There were a large percentage of Hispanic children included in this sample (28%), which is consistent with the Hispanic population in San Diego County. The majority of the sample was English speaking (72%). The remainder of the sample was bilingual (15%), Spanish speaking only (9%) or spoke another language (4%). Twenty three percent of the children in this study were enrolled in Medicaid.

Measures

The M-CHAT is a 23-item, parent completed, screening questionnaire (Robins et al., 2001), including items assessing early communication and social behaviors as well as atypical behaviors suggestive of an ASD. Each item is checked as "yes" or "no," indicating abnormality depending on the question. Six of these items are identified as critical items. A child's score is suggestive of a possible ASD if it includes endorsement of abnormality on any 2 of the 6 critical items or any 3 of the 23 total items. The M-Chat follow-up interview was not used as part of the evaluations because data was collected prior to the publication of studies indicating that specificity was improved with the interview. The M-CHAT was used as a part of the evaluation and clinical interviews were conducted with the family.

The SCQ is a 40-item, parent completed, screening questionnaire, based on the initial mandatory probes from the original ADI (Le Couteur et al., 1989). It covers the areas of communication, reciprocal social interactions, and restricted and repetitive behaviors and interests (Rutter, Bailey, et al., 2003). Each item is checked as "yes" or "no," and is assigned a point rating of "1" (presence of abnormal behavior) or "0" (absence of abnormal behavior). The first item is not included in the scoring, as it indicates if the child has sufficient verbal skills for language items to be scored. If the child is not scored as verbal, the six language items are skipped. The points are summed and result in a total possible score of 0 – 33 for nonverbal children or 0 – 39 for verbal children. Totals are compared to a cut off of 15 for ASD. A lower cut-off score of 12 has been suggested for children under the age of 5 years (Corsello et al., 2007). There are two different versions of the SCQ: 1) a "current" version designed for children under the age of 5 years and 2) a "lifetime" version designed for children 5 years of age or older, with all questions based on lifetime or past behavior. Given their young age, the current version was used for the children in this study.

The ADOS (Lord et al., 2001) is a standardized observational diagnostic measure. It is organized into four separate modules, based on the age and expressive language level of the child, ranging from pre-verbal toddlers to verbally fluent adults. A child meets criteria for a classification of autism if the scores in the social and communication domains and the total on the algorithm meet or exceed cut-off scores. All children in this study were administered either module 1 or 2 of the ADOS. In the present study, the majority of examiners had not achieved inter-rater reliability, although many had attended a clinical two-day workshop on the measure. This was considered to be representative of the use of this measure in a community clinic. ADOS total algorithm scores (communication + social + restricted and repetitive + play) were used for data analyses.

One of four developmental/cognitive measures was also administered: the Mullen Scales of Early Learning (Mullen, 1995), the Bayley Scales of Infant Development-Second Edition (BSID-II; (Bayley, 1993) or Bayley Scales of Infant and Toddler Development -Third Edition (Bayley-III; (Bayley, 2005) or the Wechsler Preschool and Primary Scales of Intelligence-Third Edition (WPPSI-III; Wechsler, 2002). In this study, Full Scale IQ scores consisted of the Early Learning Composite from the Mullen, the Full Scale IQ score from the WPPSI-III, the Mental Development Index from the BSID-II, or the mean of the Language and Cognitive standard scores from the Bayley-III. If the child's scores were

below the lowest standard or scaled scores available, ratio IQs were calculated by taking the mean of the age equivalents on the measure and dividing by the child's chronological age.

The Records-based Methodology for ASD Case Definition was developed by the Metropolitan Atlanta Developmental Disabilities Surveillance Program (MADDSP) of the Centers for Disease Control and Prevention. Records of children with possible ASD were reviewed and coded by clinicians and a determination of autism or ASD "caseness" was made (Centers for Disease Control, 2007). The MADDSP ASD coding scheme includes operational definitions of each diagnostic criterion related to autism and each criterion is rated as present or not present in a report. These are used to determine if a case meets Diagnostic and Statistical Manual, 4th Ed., Text Revision (DSM-IV-TR) criteria for a specific ASD diagnosis. The reviewer determines a classification of ASD or not based on the information included in the report.

Procedure

Reports were reviewed for 138 children. Permission to review these records was obtained under the approval of the University of California, San Diego and Rady Children's Hospital Human Research Protections Programs. Each child was evaluated by one of the eight clinical psychologists in the clinic. The majority of children received the ADOS, either the M-CHAT or the SCQ, and a developmental assessment using either the BSID-II (N=42), Bayley-III (N=67), Mullen (N=14) or WPPSI-III (N=3). The clinicians had all attended a two-day training on the ADOS and had consultation from one of the authors (CC or NA) who was research reliable on the measure. The clinicians did not attend a research workshop or achieve inter-rater reliability on the measure. For 11 children, a developmental test score was not available and the Adaptive Behavior Composite of the Vineland Adaptive Behavior Scales, 2nd Edition (Vineland; (Sparrow, Cicchetti, & Balla, 2005) was used instead. Each child was given a diagnosis by the psychologist conducting the evaluation who then wrote a clinical report that included a summary of the assessment with developmental scores and diagnostic classifications on the standardized measures.

Written clinical reports were coded using the MADDSP ASD coding system. The record reviewers (CC and NA) were blind to scores on the measures and final clinical diagnosis when coding reports, with the goal of listing diagnostic criteria met and providing a second diagnosis that was not influenced by scores on the measures. Report reviewers first reviewed reports provided by the CDC along with coding keys and established reliability at an 80% agreement for symptoms and at a 90% level for overall diagnosis. Thirty percent of the reports were double coded and reliability was maintained at 80% or better for diagnostic criteria and at 90% or better for overall diagnosis.

Design

We evaluated the diagnostic validity of the measures by comparing the classification of the measures alone and in combination to reviewer diagnosis. The discriminative validity of the measures was compared alone and in combination by examining the sensitivity (the ability to accurately classify children with an autism or ASD diagnosis: True Positives/(True Positives + False Negatives) and specificity (the ability to accurately exclude children without an autism or ASD diagnosis: True Negatives/(True Negatives + False Positives). Comparisons were made between children with autism (AUT) and those without autism (Not AUT: PDD-NOS and NS disorders), as well as between children with ASD (including Autism and PDD-NOS) and NS. Throughout the analyses, the groups will be referred to in these ways. Clinician and reviewer diagnoses were compared and diagnostic stability was reviewed for the subset of children who returned for a re-evaluation as another check of the accuracy of the initial diagnosis.

Results

Diagnostic Validity of Standardized Measures

In order to determine the validity of the ADOS and SCQ in this community sample, the classifications on the measures were compared with reviewer diagnosis, blinded to scores and classifications on the measures as well as clinician diagnosis, based on a review and coding of the written report. When compared with reviewer diagnosis, the ADOS demonstrated strong sensitivity and specificity for both AUT vs. Not AUT and ASD vs. NS diagnoses (see Table 2). The SCQ, using the lower cut-off of 12 had adequate sensitivity when differentiating AUT from Not AUT, but weak sensitivity when differentiating ASD from NS, missing about 80% of the children with PDD-NOS. Consistent with previous studies, the M-CHAT captured the majority of the children with PDD-NOS and AUT in this 2 year old sample, but also included 30% of the NS children. Using either the M-CHAT or the SCQ in combination with the ADOS did not result in improved specificity over the ADOS alone and led to a dramatic drop in sensitivity for the ASD vs. NS group primarily because the SCQ missed so many of the children with PDD-NOS (See Figure 1). Figure 1 depicts the percentage of children meeting the cut-off scores on each measure compared with their reviewer-identified diagnosis.

To determine if age may have affected the diagnostic discrimination of these two measures, the M-CHAT was evaluated for the younger age group for which it is recommended (30 months and under) and the SCQ evaluated for the children older than 30 months. When discriminating between ASD and NS in this younger group, the sensitivity (.95) and specificity (.75) improved for the M-CHAT. However, the NS sample was extremely small (N=4). For the SCQ, neither sensitivity (.63) nor specificity (.33) improved when discriminating ASD from NS diagnoses in children over 30 months of age.

Clinician and Reviewer Diagnostic Agreement

Clinicians and reviewers agreed 80% of the time. Disagreements were primarily between PDD-NOS and AUT with reviewers more likely to classify the diagnosis AUT than clinicians (See Table 3). Encouragingly, there were no disagreements between AUT and NS disorders. That is, a reviewer never classified a child as NS when a clinician classified a child with AUT or vice versa. Of the five children that the reviewer classified as NS and the clinician classified as PDD-NOS, three had received a nonspecific ASD diagnoses from the clinician ("Features of PDD-NOS", "PDD-NOS provisional" and "Rule out ASD"). Of the 19 children that were diagnosed PDD-NOS by the clinician and AUT by the reviewer, all had a sufficient symptom count to meet criteria for a diagnosis of autism and 10 (53%) met criteria for autism on the ADOS.

Diagnostic stability

Diagnosis often becomes clearer as children grow older (Lord, et. al., 2006). Diagnosis at follow up appointments for the small number of children who were re-evaluated was reviewed and compared with initial clinical diagnosis, rather than reviewer diagnosis, as another gauge of diagnostic accuracy. Of the 138 children included in this study, 23% or 32 children were seen for a re-evaluation when they were between 35 and 101 months of age, with 72% of the children falling between 4 – 6 years of age. Seventeen (53%) of the children remained in the same diagnostic category. As predicted, the AUT and NS diagnoses remained the most stable, with only 2 shifting from autism to PDD-NOS and 2 shifting from NS to ASD (1 to autism and 1 to PDD-NOS). The PDD-NOS diagnoses had the most change, with 4 remaining in the PDD-NOS group, 5 moving to autism and 5 moving to NS.

All 11 of the children who were initially diagnosed with AUT and later re-evaluated met criteria for AUT on the ADOS when evaluated at the age of 2 years. This included the 2 children who moved from a clinical diagnosis of autism to PDD-NOS. All but 2 of the 14 children who were initially diagnosed with PDD-NOS fell within the ASD classification ($N = 8$) or autism classification ($N = 4$) on the ADOS when they were initially evaluated of the 4 children who remained within the PDD-NOS group, the 3 who were given the ADOS fell within the ASD classification on the measure. Of the 5 children whose clinical diagnosis moved from PDD-NOS to Autism, 3 received an autism classification on the ADOS. All 5 of the children who moved from PDD-NOS to NS had an ASD score on the ADOS with 1 falling within the classification of autism. This child was later diagnosed with selective mutism. Of the 6 children initially diagnosed as NS and later re-evaluated, only 2 were given the ADOS. One met criteria for ASD on the measure and was later diagnosed with autism while the other was NS on the measure and later was diagnosed with an anxiety disorder.

Discussion

The ADOS worked encouragingly well in this group of young children diagnosed by experienced psychologists in a general developmental clinic. Although the clinicians had not received intensive training on the measure or achieved research reliability, the ADOS demonstrated very good diagnostic discrimination when compared to a diagnosis based on a report review by an expert on ASD who was blinded to clinician diagnosis and scores on diagnostic measures. The group of clinicians included in this study had received atwo-day training on the measure and worked closely with clinicians who had achieved research reliability on the ADOS and were available to provide consultation and support as needed. Although this level of consultation may not be typical for community clinicians using the ADOS, this suggests that the ADOS can be an effective diagnostic tool even with very young children in a community-based clinic with experienced clinicians who receive consultative support.

In contrast to previous studies, classification on the parent measure did not improve diagnostic accuracy in this young sample and the combination of the ADOS and a parent measure led to a decrease in sensitivity when differentiating ASD from NS disorders. The parent questionnaires used by the clinicians in the current study were designed for screening rather than diagnostic purposes. These screening questionnaires were completed by parents of young children who may not have the experience or expertise to accurately rate the diagnostic behaviors these measures are targeting. This may have contributed to lack of improvement in diagnostic accuracy over the ADOS alone. Most of the longitudinal research studies used the lengthier standardized diagnostic interview, the ADI-R, rather than a parent questionnaire. The ADI-R is administered and scored by experienced clinicians who have a solid understanding of how to interpret the behaviors that they are coding. Another potential reason for the ineffectiveness of the M-CHAT and SCQs diagnostic measures in this study is that neither one was designed to cover the entire age span between 24 and 36 months. Brief diagnostic measures that could be easily incorporated into a clinical diagnostic assessment do not exist for this age group.

Although these measures did not improve the accuracy of an ASD diagnosis, the M-CHAT appeared to be effective as a screener in this sample, as the majority of ASD children were captured by this measure. This was interesting given that the majority of the children were older than the recommended age. It is important to note that the evaluations included in this study were conducted prior to the wide-spread introduction of the follow up interview for the M-CHAT which has led to improved specificity (Pandey, et. al., 2008). The SCQ, which was designed and tested on an older sample of children initially, missed many children with PDD-NOS, indicating that it is not the best screener or diagnostic measure in this young age

group. This is similar to studies that have found that the ADI-R, on which the SCQ is based, tends to miss children in this age group (Corsello et al., 2007). More research is needed on the SCQ in younger ages, including an evaluation of subscales and items to determine if the pattern of symptoms missed parallels findings in studies of the ADI-R. Further research may lead to an effective diagnostic and screening measure for children between 2 and 3 years of age, which is important given that this is the group most often initially identified and referred and that two measures in combination with clinical judgment leads to the most accurate diagnosis in 2 year old children suspected of an ASD.

Of the children who were initially evaluated, 22% were re-evaluated providing an additional way of assessing diagnostic accuracy of the ADOS in this 2 year old sample. The ADOS scores at initial evaluation were fairly good predictors of later diagnosis. Only two children who initially were given a clinical diagnosis of autism and met criteria for autism on the measure, moved into the PDD-NOS group at follow up. The remainder continued to meet criteria for autism. The PDD-NOS group had more variability in diagnosis at follow up, which is consistent with previous studies. While all but two of the children within this group had an ASD classification on the ADOS at the time of the initial evaluation, five no longer met criteria for an ASD at follow up. The only misclassification of autism on the measure at initial evaluation was for a child who was later diagnosed with selective mutism. The remaining children who initially met criteria for PDD-NOS continued to meet for PDD-NOS or moved into a diagnosis of autism.

Surprisingly, there was fairly good agreement between reviewer diagnoses, blind to scores on measures and clinician diagnosis, based on report review and coding and the clinician's clinical judgment. This is encouraging as the MADDSP ASD coding system has been used on a broader scale in an older sample of children. The strong agreement between reviewer and clinician suggests that this surveillance methodology also shows promise in a younger sample. However, reports were all from one clinic and were carefully written, included the same measures, and generally the same level of detail leading to greater ease for the reviewer. The use of the ADOS may influence the observations of the clinician as well as the data that they include in their report ultimately also influencing the diagnosis of the reviewer even though they were blind to measure scores and classification. Diagnostic agreement may not be as easy to achieve in other situations given the variability of measures used and content of reports in the larger community.

Limitations

This study was based on the practices in a community clinic and report review. The stringent controls that are in place for a prospective research study were not in place in this study. Each child did not receive all three standardized diagnostic measures, limiting the conclusions that we can draw about the relationship between these measures. An attempt was made to provide a reviewer diagnosis independent of the clinician generated scores and classification on diagnostic measures by using a standardized record review strategy by a research clinician blind to the scores on the diagnostic measures and clinician diagnosis. However, the reviewer diagnosis is based on a report written by a clinician who used specific diagnostic measures to gather information and therefore, subject to the biases of the clinician who wrote the report. Additionally, follow up evaluations were only available for a small subset of the sample, again limiting the conclusions that can be made in terms of predictive value of the measure.

Summary

Within this community based clinic, the ADOS evidenced good diagnostic discrimination in a 2 year old sample when administered by clinicians who had received a two day clinical

training followed by consultative support. More research is needed to identify or develop an appropriate efficient parent report tool to accompany a standardized diagnostic measure, as neither the SCQ nor the M-CHAT added to the diagnostic accuracy of the ADOS in this sample.

Acknowledgments

This work was supported by grants from the National Institutes of Health: K23MH071796 (to N.A.) and K01MH065325 (to A.C.S.). We would also like to thank our research assistants, Rebecca Piñon Ruiz, Jasper Estabillo, and Therese Gadomski.

References

- Bayley, N. Bayley Scales of Infant Development-Second Edition (BSID-II). 2nd ed.. Psychological Corp.; San Antonio, TX: 1993.
- Bayley, N. Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III). Third Edition ed.. Harcourt Brace & Co.; San Antonio, TX: 2005.
- California Department of Developmental Services. Autistic Spectrum Disorders: Best Practice Guidelines for Screening, Diagnosis and Assessment. Sacramento: 2002.
- Centers for Disease Control. Prevalence of autism spectrum disorders--autism and developmental disabilities monitoring network, 14 sites, United States, 2002. *Mor Mortal Wkly Rep Surveill Summ.* 2007; 56:12–28.
- Charman T, Taylor E, Drew A, Cockerill H, Brown JA, Baird G. Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry.* 2005; 46:500–513. [PubMed: 15845130]
- Corsello C, Hus V, Pickles A, Risi S, Cook E, Leventhal B, Lord C. Between a ROC and a hard place: Decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry.* 2007; 48(9):932–940. [PubMed: 17714378]
- Filipek PA, Accardo PJ, Ashwal S, Baranek GT, Cook EH Jr, Dawson G, Gordon B, Gravel JS, Johnson CP, Kallen RJ, Levy SE, Minshew NJ, Ozonoff S, Prizant BM, Rapin I, Rogers SJ, Stone WL, Teplin SW, Tuchman RF, Volkmar FR. Practice parameter: screening and diagnosis of autism: report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. *Neurology.* 2000; 55:468–479. [PubMed: 10953176]
- Filipek PA, Accardo PJ, Baranek GT, Cook EH Jr, Dawson G, Gordon B, Gravel JS, Johnson CP, Kallen RJ, Levy SE, Minshew NJ, Prizant BM, Rapin I, Rogers SJ, Stone WL, Teplin S, Tuchman RF, Volkmar FR. The screening and diagnosis of autistic spectrum disorders. *Journal of Autism and Developmental Disorders.* 1999; 29:439–484. [PubMed: 10638459]
- Johnson CP, Myers SM, American Academy of Pediatrics Council on Children With Disabilities. Identification and evaluation of children with autism spectrum disorders. *Pediatrics.* 2007; 120:1183–1215. [PubMed: 17967920]
- Kleinman JM, Robins DL, Ventola PE, Pandey J, Boorstein HC, Esser EL, Wilson LB, Rosenthal MA, Sutura S, Verbalis AD, Barton M, Hodgson S, Green J, Dumont-Mathieu T, Volkmar F, Chawarska K, Klin A, Fein D. The Modified Checklist for Autism in Toddlers: A follow-up study investigating the early detection of autism spectrum disorders. *Journal of Autism and Developmental Disorders.* 2008; 38:827–839. [PubMed: 17882539]
- Le Couteur A, Rutter M, Lord C, Rios P, Robertson S, Holdgrafer M, McLennan J. Autism diagnostic interview: a standardized investigator-based instrument. *Journal of Autism and Developmental Disorders.* 1989; 19(3):363–387. [PubMed: 2793783]
- Lord C, Risi S, DiLavore PS, 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.; Rutter, M.; DiLavore, PC.; Risi, S. Autism Diagnostic Observation Schedule. Western Psychological Services; Los Angeles: 2001.
- Mandell DS, Ittenbach RF, Levy SE, Pinto-Martin JA. Disparities in diagnoses received prior to a diagnosis of autism spectrum disorder. *Journal of Autism and Developmental Disorders.* 2007

- Mandell DS, Novak MM, Zubritsky CD. Factors associated with age of diagnosis among children with autism spectrum disorders. *Pediatrics*. 2005; 116:1480–1486. [PubMed: 16322174]
- Mazefsky CA, Oswald DP. The discriminative ability and diagnostic utility of the ADOS-G, ADI-R, and GARS for children in a clinical setting. *Autism*. 2006; 10:533–549. [PubMed: 17088271]
- Molloy CA, Murray DS, Akers R, Mitchell T, Manning-Courtney P. Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting. *Autism*. 2011; 15(2):143–162. [PubMed: 21339248]
- Mullen, EM. Mullen Scales of Early Learning. AGS ed.. American Guidance Service Inc.; Circle Pines, MN: 1995.
- National Research Council. Educating children with autism. National Academy Press; Washington, DC: 2001.
- Pandey J, Verbalis A, Robins DL, Boorstein H, Klin A, Babitz T, Chawarska K, Volkmar F, Green J, Barton M, Fein D. Screening for autism in older and younger toddlers with the Modified Checklist for Autism in Toddlers. *Autism*. 2008; 2:513–535. [PubMed: 18805945]
- Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, Cook EH, Leventhal B, Pickles A. 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]
- Robins DL, Fein D, Barton ML, Green JA. The modified-Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*. 2001; 31:131–144. [PubMed: 11450812]
- Rutter, M.; Bailey, A.; Lord, C. Social Communication Questionnaire. Western Psychological Services; Los Angeles: 2003.
- Rutter, M.; Le Couteur, A.; Lord, C. ADI-R. Autism Diagnostic Interview-Revised. WPS Edition. Western Psychological Services; Los Angeles: 2003.
- Shattuck PT, Durkin M, Maenner M, Newschaffer C, Mandell DS, Wiggins L, Lee LC, Rice C, Giarelli E, Kirby R, Baio J, Pinto-Martin J, Cunniff C. Timing of identification among children with an autism spectrum disorder: Findings from a population-based surveillance study. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2009; 48(5):474–483. [PubMed: 19318992]
- Sparrow, SS.; Cicchetti, DV.; Balla, DA. Vineland Adaptive Behavior Scales. II. Survey forms manual. 2nd ed.. American Guidance Service; Circle Pines, MN: 2005.
- Turner LM, Stone WL. Variability in outcome for children with an ASD diagnosis at age 2. *Journal of Child Psychology and Psychiatry*. 2007; 48:793–802. [PubMed: 17683451]
- Ventola PE, Kleinman J, Pandey J, Barton M, Allen S, Green J, Robins D, Fein D. Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders*. 2006; 36:839–847. [PubMed: 16897398]
- Wechsler, D. Wechsler Preschool and Primary Scale of Intelligence-Third Edition (WPPSI-III). The Psychological Corporation; San Antonio, TX: 2002.
- Wiggins LD, Robins DL. Brief Report: Excluding the ADI-R behavioral domain improves diagnostic agreement in toddlers. *Journal of Autism and Developmental Disorders*. 2008; 38:972–976. [PubMed: 17879150]

- The ADOS was effective in a two year old sample in a community clinic when administered by psychologists who had received a two day training followed by consultation
- The use of a parent screening measure did not improve diagnostic accuracy over the ADOS alone
- Diagnosis remained stable in the majority of children who were diagnosed with autism or non-spectrum disorders and less stable in toddlers diagnosed with PDD-NOS
- There are no diagnostic parent interviews or questionnaires currently available for this age group that are easily incorporated in a clinical evaluation

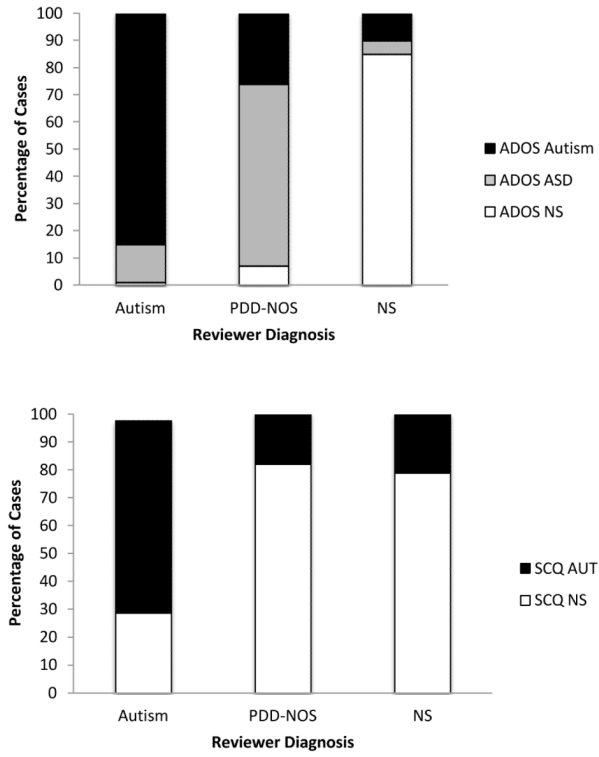


Figure 1. Reviewer diagnosis agreement with the ADOS (top) and the SCQ (bottom)

Table 1
Age, IQ score, SCQ score, and ADOS Total Score by Clinical Diagnosis

	Clinical Diagnosis		
	Autism (n=56)	PDD-NOS (n=50)	Non-Spectrum (n=32)
Age (months) <i>M</i>	29.77	30.58	30.50
<i>SD</i>	3.16	3.39	2.82
Range	24-35	24-35	24-35
<i>IQ</i> ^a <i>M</i>	63.38	73.20	76.28
<i>SD</i>	4.18	17.58	21.76
Range	24-102	25-112	13-97
<i>SCQ</i> ^b <i>M</i>	17.16	12.44	9.17
<i>SD</i>	6.16	6.35	5.38
Range	3-29	3-24	1-19
<i>ADOS</i> ^b <i>M</i>	23.74	14.08	8.25
<i>SD</i>	3.72	6.06	6.25
Range	15-30	3-27	2-22

^aSignificant difference between Autism and Non-Spectrum groups

^bSignificant difference between all three groups

Table 2
Classification of Autism (AUT) or Not Autism (NotAUT) and ASD or Non-Spectrum (NS)
on Screening and Diagnostic Measures Compared to Reviewer Diagnosis

	<i>n</i>	Sensitivity	Specificity
<i>AUT vs. Not AUT</i>			
ADOS	118	.84	.80
SCQ 12	67	.82	.64
M-CHAT	38	.88	.33
M-CHAT or SCQ & ADOS	84	.82	.65
<i>ASD vs. NS</i>			
ADOS	118	.97	.85
SCQ 12	67	.69	.56
M-CHAT	38	.86	.67
M-CHAT or SCQ & ADOS	84	.76	.85

Table 3
Agreement between Clinical Diagnosis and Reviewer Diagnosis

Clinical Diagnosis	Reviewer Diagnosis		
	Autism	PDD-NOS	Non-Spectrum
Autism	55	1	0
PDD-NOS	19	27	5
Non-Spectrum	0	4	27