Early diagnosis of autism spectrum disorder: Stability and change in clinical diagnosis and symptom presentation

Whitney Guthrie, **Lauren B. Swineford**, **Charly Nottke**, and **Amy M. Wetherby**
Autism Institute, Florida State University, USA

Abstract

Background—Although a diagnosis of autism spectrum disorder (ASD) appears to be stable in children as young as age three, few studies have explored stability of a diagnosis in younger children. Predictive value of diagnostic tools for toddlers and patterns of symptom change are important considerations for clinicians making early diagnoses. Most findings come from high-risk samples, but reports on children screened in community settings are also needed.

Methods—Stability of diagnosis and Autism Diagnostic Observation Schedule–Toddler Module (ADOS-T) classifications and scores was examined across two time points in a sample of 82 children identified through the FIRST WORDS[R] Project. Children received two comprehensive diagnostic evaluations at average ages of 19.39 (*SD*=2.12) and 36.89 (*SD*=3.85) months.

Results—Stability was 100% when confirming and ruling out a diagnosis of ASD based on a comprehensive diagnostic evaluation that included clinic and home observations, although diagnosis was initially deferred for 17% of the sample. Receiver Operating Characteristic curves revealed excellent sensitivity and acceptable specificity for the ADOS-T compared to concurrent diagnosis. Logistic regressions indicated good predictive value of initial ADOS-T scores for follow-up diagnosis. Finally, both ASD and Non-ASD children demonstrated a decrease in Social Affect scores (i.e., improvement), while children with ASD demonstrated an increase in Restricted and Repetitive Behavior scores (i.e., worsening), changes that were accounted for by nonverbal developmental level in mixed model analyses.

Conclusions—Short-term stability was documented for children diagnosed at 19 months on average, although a minority of children initially showed unclear diagnostic presentations. Findings highlight utility of the ADOS-T in making early diagnoses and predicting follow-up diagnoses. Children with ASD demonstrated improvement in social communication behaviors and unfolding of repetitive behaviors, suggesting that certain early patterns of change in symptoms may be characteristic of ASD.

Keywords

autism spectrum disorder; Developmental delay; Diagnosis; Development; Assessment

---

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impairments in social communication and interaction, and the presence of restricted and repetitive behaviors and interests. Although evidence suggests that ASD has genetic causes (O'Roak & State, 2008), diagnosis relies on observations of behavioral manifestations. The

---

Correspondence to: Whitney Guthrie, Florida State University Autism Institute, 1940 North Monroe Street, Suite 72, Tallahassee, FL, 32303; Phone: 850-644-4367; Fax: 850-644-3644; whitney.guthrie@med.fsu.edu.

average age of diagnosis remains well over three (Mandell, Novak, & Zubritsky, 2006), although the American Academy of Pediatrics recommends that all children be screened for ASD much earlier–at 18 and 24 months (Johnson & Myers, 2007). There is a clear need for diagnostic tools and practices for children who screen positive at these young ages. However, because diagnosis in toddlers is relatively new, clinicians face a number of important challenges that warrant further research. The stability of early diagnoses, the utility of diagnostic tools for toddlers, and patterns of symptom change in the first few years of life are among a number of questions critical to professionals making early diagnoses.

Stability of Early Clinical Diagnoses of ASD

High rates of stability of the broader diagnosis of ASD (rather than specific diagnoses within the spectrum) have been demonstrated in children first diagnosed by age three or older, with estimates ranging from 80 to 100% (see Woolfenden, Sarkozy, Ridley, & Williams, 2012 for a review). However, some estimates for children diagnosed under age three are lower and findings are more variable, ranging from 54 to 100%. While 4 of the 11 studies of children diagnosed with Autistic Disorder or ASD under age three reviewed by Woolfenden and colleagues (2011) reported a 100% stability rate, two studies reported rates under 70% (Stone et al., 1999; Turner & Stone, 2007).

In addition to being disparate, these findings may not generalize well to the larger population of toddlers with ASD, as they have consisted of high-risk children (i.e., those with an older sibling with ASD or who are referred because of parental or professional concern), or included only relatively lower-functioning children. Thus, there is a need to examine diagnostic stability in samples screened in the community that yield participants diverse in symptoms and developmental functioning who may not garner parent or professional concern. Findings also need to be extended to younger children, because although children are being screened at increasingly younger ages, just two studies of diagnostic stability have included children under age two. Encouragingly, both studies reported 100% stability for initial diagnoses of ASD (Chawarska, Klin, Paul, Macari, & Volkmar, 2009; Cox et al., 1999). However, neither study reported on stability of children with unclear diagnostic presentations at these young ages.

Accuracy of Diagnostic Tools for Young Children

Most studies utilize clinical best-estimate diagnoses made by experienced clinicians, a practice that remains the gold standard for diagnosing ASD (Volkmar, Chawarska, & Klin, 2005). However, a standardized observation is crucial to inform clinical diagnosis (Lord & Bishop, 2010). Among the most widely used and validated tools is the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999). However, it has limited utility with toddlers, as it has unacceptable specificity in children with nonverbal mental ages below 16 months (Gotham, Risi, Pickles, & Lord, 2007).

The ADOS – Toddler Module (ADOS-T) was developed to address these limitations (Lord, Luyster, Gotham, & Guthrie, 2012) and demonstrated excellent sensitivity and specificity in the validation study (Luyster et al., 2009). There is a need to examine its utility in other samples as the validation sample included only high-risk children. Of additional importance is the predictive validity of the "ranges of concern," which its authors recommend to index risk for ASD. While these concern ranges were shown to be largely consistent with concurrent clinical diagnosis (Luyster et al., 2009), their utility in predicting later diagnosis has yet to be examined.

Symptom Severity Change in Young Children

Another issue critical to the evaluation of toddlers is symptom change within the first years of life, as symptom severity may be more variable than diagnostic status. Clinicians may observe significant changes in the frequency and severity of symptoms as the clinical presentation of ASD unfolds, further complicating early diagnosis. However, existing evidence is still emerging and findings are relatively inconsistent.

Improvement of social communication skills, such as joint attention, response to name, and verbal communication, has been reported (Nadig et al., 2007; Sullivan et al., 2007; Yoder, Stone, Walden, & Malesa, 2009), although stability in more global measures of social symptom severity has also been found (Chawarska, Klin, Paul, & Volkmar, 2007). Evidence exists for other trajectories, including plateauing (i.e., developmental slowing) or even worsening (i.e., loss) of these skills in a subgroup of children (Landa, Holman, & Garrett-Mayer, 2007; Lord, Luyster, Guthrie, & Pickles, 2012; Ozonoff, Heung, Byrd, Hansen, & Hertz-Picciotto, 2008). Greater understanding of changes in symptom severity in toddlers would inform studies of diagnostic stability, as changes in symptoms are likely to accompany movement on or off the autism spectrum, but may also be observed in children with stable diagnostic presentations.

The purpose of this study was to examine stability of clinical diagnosis and symptom presentation, and the utility of a diagnostic tool in making a clinical diagnosis. The study utilized a prospectively-identified community sample that received a diagnostic evaluation at 15–24 months of age and a follow-up evaluation at least one year later. The specific research aims were to examine the (1) short-term stability of clinical diagnoses, (2) concurrent and predictive utility of ADOS-T classifications and scores, and (3) short-term change in symptom severity.

Methods

Participants

There were 82 participants identified from 5,419 children screened by the FIRST WORDS[R] Project, a longitudinal, prospective study of a general population sample that screens for communication delays and ASD in pediatric primary care settings (see Wetherby, Brosnan-Maddox, Peace, & Newton, 2008). While younger siblings were not specifically recruited, nine children (11%) had an older sibling with ASD according to parent report.

All children completed a two-step screening process that used two measures of the Communication and Symbolic Behavior Scales Developmental Profile (CSBS; Wetherby & Prizant, 2002). First the Infant-Toddler Checklist based on parent report was completed at an average of 14.23 (SD=4.02) months. Children who scored in the bottom 10th percentile (n=74 of 82) or whose parent responded "yes" to the question "Do you have any concerns about your child's development?" (n=41 of 82) were invited for a communication evaluation. All children completed the CSBS Behavior Sample and clinicians rated "red flags" for ASD using the Systematic Observation of Red Flags of ASD (Wetherby et al., 2004) to screen for ASD. Children with red flags were invited for a diagnostic evaluation to confirm or rule out ASD and for a follow-up diagnostic evaluation one to two years later. Demographic characteristics for the 82 children who met these inclusion criteria are reported in Table 1. Informed consent was obtained from all parents and the study was conducted in accordance with the Institutional Review Board at Florida State University.

Diagnostic Evaluation Procedures and Measures

Children received a diagnostic evaluation at Time 1 (T1) between 15 and 24 months of age (*M*=19.39, *SD*=2.12) that included the ADOS-T. A follow-up evaluation was completed at Time 2 (T2) at least 12 months later (*M* T2-T1=16.85, *SD*=3.72). T2 evaluations were conducted between 30 and 46 months (*M*=36.89, *SD*=3.85) and included the ADOS Module 1, 2 or 3, depending on language level.

Autism Diagnostic Observation Schedule—The ADOS is a standardized, semi-structured observation of communication, social interaction, and repetitive behaviors and interests. Symptoms relevant to a diagnosis of ASD are scored from 0–3 on the ADOS/ADOS-T, with higher numbers indicating more abnormality. Diagnostic algorithms yield domain scores for Social Affect (SA) and Restricted and Repetitive Behaviors (RRB) and a total algorithm score (Gotham et al., 2007; Luyster et al., 2009). Based on cutoffs applied to total scores, ADOS Modules 1–4 yield three diagnostic classifications: Nonspectrum, ASD, and Autism; the ADOS-T yields just two classifications: ASD and Nonspectrum. The ADOS-T also yields three ranges of concern: "little-to-no concern," "mild-to-moderate concern," and "moderate-to-severe concern."

The ADOS has demonstrated strong psychometric properties, with inter-rater reliability measured by mean exact agreement shown to be 84% for the ADOS-T and over 88% for ADOS Modules 1–4 (Luyster et al. 2009; Lord et al., 2000). ADOS inter-rater reliability was not calculated in the present study due to lack of resources to double-score a subsample of the administrations. However, all clinicians achieved research reliability with certified trainers and avoided "drift" by attending weekly meetings to calibrate on ADOS scoring and clinical diagnosis.

Home observation—In order to gather information across settings, a systematic, video-recorded home observation of interactions between parent and child was obtained for 78 children at T1 and 75 children at T2. Parents were instructed to interact with their child for one hour during everyday routines.

Parent report measures—At T1, parents completed the Early Screening for Autism and Communication Disorders (ESAC; Wetherby et al., 2009). At T2, parents completed the Repetitive Behavior Scale (Bodfish, Symons, & Lewis, 1999) and either the ESAC if the child was under 36 months or the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003) if over 36 months.

Developmental functioning—Children were administered the Mullen Scales of Early Learning (MSEL; Mullen, 1995) to measure developmental abilities. Because of the substantial proportion of children who received a T-score of 20 (i.e., the floor) on one or more subtests (39% at T1), developmental quotients (DQ) were calculated. Nonverbal DQ scores were derived from the average of the Visual Reception and Fine Motor age equivalents divided by chronological age, multiplied by 100. Calculation of verbal DQ scores followed the same procedure using the Receptive and Expressive Language subscales. The average nonverbal DQ at T1 was 92.66 (*SD*=15.79), with a normal distribution. See Table 2 for sample characteristics by diagnosis.

Adaptive behavior—Adaptive functioning was measured with the Vineland Adaptive Behavior Scales, 2nd edition (VABS; Sparrow, Cicchetti, & Balla, 2005).

Clinical diagnosis—An initial clinical diagnosis to confirm or rule out ASD was made by the experienced clinician who evaluated the child. Next, diagnoses were verified by another

clinician who did not complete testing, based on review of all available information (i.e., evaluating clinician's report, ADOS/ADOS-T scores and classification, video-recordings of evaluations and home observations, parent-report measures, and MSEL and VABS scores). While diagnoses were made and verified by experienced clinicians, diagnoses at T2 were not independent of diagnoses made at T1 because of lack of resources to ensure a blind clinician at every time point. Thus, the decision was made to provide information on diagnostic history for all T2 evaluations to ensure that identical procedures were followed for all children, regardless of whether they had a new clinician at T2.

Clinicians made one of three diagnostic decisions: confirm ASD, rule out ASD, or defer diagnosis. Children were diagnosed with ASD (n=56 at T1) if they demonstrated impairments consistent with DSM-IV criteria in all three domains across settings according to expert clinical judgment. Diagnostic distinctions within the autism spectrum were not made given the young age of these children. Scores above the cutoff for ASD on the ADOS/ADOS-T were not required for diagnosis of ASD because clinical judgment at age two has been shown to better predict outcome diagnosis than ADOS classifications (Lord et al., 2006), and the ADOS-T is a new instrument with little research to date on agreement with clinical diagnosis. However, only 6 of the 56 children who met clinical judgment for ASD at T1 and 2 of the 59 at T2 had ADOS scores under the ASD cutoff.

When ASD was ruled out according to diagnostic criteria and clinical judgment, children were classified as developmental delay (DD; n=11) or typically developing (TD; n=1) based on MSEL scores. Children were classified as DD if they had at least one MSEL T-score below 38, and classified as TD if all MSEL T-scores were at or above 38.

Finally, diagnosis was deferred (n=14) if ASD could not be confirmed or ruled out according to diagnostic criteria and clinical judgment, regardless of developmental functioning. Reasons for deferred diagnosis included inconsistent observation of symptoms across settings (i.e., symptoms observed during the ADOS but not during the home observation), lack of symptoms in all three domains (e.g., no RRBs), or symptoms were present in all domains but they were judged to be less severe or insufficient to meet criteria for a diagnosis.

Participation in treatment—Parents indicated whether their child was currently receiving intervention and the type(s) of treatment at T1 and T2. Twenty-three percent of children received an average of 1.74 (*SD*=.95) hours per week of speech therapy (ST), physical therapy (PT), occupational therapy (OT), and/or general early intervention therapy (EI) at T1 At T2, 46% of children received an average of 1.67 hours (*SD*=1.09) of ST, PT, OT, EI, and/or Applied Behavior Analysis (ABA).

In addition, 30% of children participated in the Early Social Interaction project, an 18-month treatment study that provided nine months of a weekly play-group and nine months of parent-implemented intervention that included three home sessions per week. Six percent participated in intervention projects targeting minority, low-income, and low-resource families that provided 3 or 6 months of parent-implemented intervention in two home sessions per week.

Results

Stability of Clinical Diagnosis

Clinical diagnosis of confirming ASD (n=56 at T1) and ruling out ASD (i.e., DD or TD; n=12 at T1) was stable for all children between T1 and T2, yielding a stability rate of 100% when ASD was initially confirmed and ruled out (see Figure 1). In contrast, stability of the

deferred diagnosis group (i.e., unclear diagnostic presentation so that ASD could neither be confirmed nor ruled out) was very poor (7%), as was expected given the unstable nature of this group. Almost all children with a deferred diagnosis at T1 were later diagnosed at T2, as ASD was later ruled out for 10 of the 14 (DD n=4, TD n=6) and confirmed for three children at T2. The remaining child, a younger sibling of a child with ASD, continued to have diagnosis deferred at T2. Overall stability for the sample was 84%, with instability driven by children with deferred diagnoses at T1. For purposes of all additional analyses, the Non-ASD group was comprised of children with TD, DD, and deferred diagnosis (i.e., all children not diagnosed with ASD).

Characteristics of Children with a Deferred Diagnosis at Time 1

A series of analyses of variance modeled differences between children with deferred diagnosis, ASD, and DD/TD on evaluation and demographic characteristics. Analyses indicated some distinct patterns of ADOS scores (SA: F=34.47, p<.001; RRB: F=4.70, p=.01). Planned comparisons revealed that children with a deferred diagnosis demonstrated lower ADOS SA scores than children diagnosed with ASD (p<.001), but marginally higher scores than children diagnosed with DD/TD (p=.09). Differences in RRB scores were not found in the planned comparisons, indicating that the deferred group demonstrated comparable scores to the ASD (p=.17) and DD/TD groups (p=.55). At the omnibus level, differences were found for verbal DQ (F=4.33, p=.02), but not for nonverbal DQ (F=1.57, p=.21). However, planned comparisons indicated that the deferred diagnosis group showed comparable verbal DQ scores to the ASD (p=.15) and DD/TD groups (p=1.00). The small sample size of the deferred diagnosis group (n=14) limited power to detect statistically significant planned comparisons within significant omnibus tests (i.e., for RRB and verbal DQ). No differences were found for age at initial evaluation or any demographic variables.

Concurrent Utility of Time 1 ADOS-T Classifications

Receiver Operating Characteristic (ROC) curves were run to estimate the discrimination, sensitivity, and specificity of the ADOS-T research classifications (i.e., Nonspectrum or ASD). However, sample size was only sufficient to compute ROC curves for the 12–20/21–30 nonverbal algorithm (n=51 ASD, n=24 Non-ASD), so children receiving the 21–30 verbal algorithm (n=5 ASD, n=2 Non-ASD) were excluded for these analyses. Results indicated excellent discrimination (area under the curve =.91), as well as excellent sensitivity (.90) and acceptable specificity (.71) of the established cutoff (12).

Predictive Value of Time 1 ADOS-T Classifications

Logistic regressions were run to examine the predictive value of T1 ADOS-T scores rather than ROC curves, as they provide an index of calibration (rather than discrimination), a critical component of prognostic models (see Cook, 2008). Thus, clinical diagnosis at T2 was regressed on T1 ADOS-T and nonverbal DQ scores (^{2}(2) 43.54, p<.001, *pseudo R^2*=. 59). Results revealed that ADOS-T scores significantly predicted T2 clinical (B=.49, p<. 001, *odds ratio*=1.63), but nonverbal DQ scores did not (p=.91). Thus, nonverbal DQ was dropped from the model when evaluating calibration using the Hosmer Lemeshow goodness-of-fit test, which compares expected and predicted event rates and is commonly used in prediction models. This metric indicated good calibration and fit of the model predicting follow-up clinical diagnosis from initial ADOS-T scores (^{2}(8)=4.99, p=.76).

Percent agreements between ADOS-T classifications at T1 (i.e., ranges of concern and research classifications) and clinical consensus diagnosis at T2 are also presented, in order to examine predictive value within risk ranges and research classifications (see Table 3). Of the 12 children who scored in the "little-to-no concern" range on the ADOS-T, 83% were given a clinical diagnosis of Non-ASD (n=2 DD, n=8 TD) at T2; of the 26 children in the

"mild-to-moderate concern" range, 58% were ASD and 42% were Non-ASD (n=8 DD, n=3 TD) at T2; and of the 44 children in the "moderate-to-severe concern" range, 95% were ASD and 5% were Non-ASD (n=1 TD, n=1 diagnosis deferred) at T2. A similar pattern of agreement was found between ADOS-T research classifications and T2 follow-up clinical diagnosis (see Table 3).

Change in Symptom Severity by Domain between Time 1 and Time 2

A series of repeated-measures mixed-model analyses of covariance (ANCOVA) was conducted to examine changes in SA and RRB scores between T1 and T2 (see Table 4). Although ADOS-T algorithm totals were used for all previous analyses, they are less appropriate for analyses across modules as the 21–30 verbal ADOS-T algorithm is comprised of a different number of SA and RRB items than the 12–20/21–30 nonverbal and Module 1–3 algorithms. Thus, Module 1 algorithms were calculated using ADOS-T items for T1 data, and are used in the following analyses to increase comparability across modules. Every item that appears on the Module 1 algorithms appears on the ADOS-T and each was closely inspected to ensure that it shares very similar phrasing and coding structure across module.

Time (T1 and T2) and T1 diagnosis (ASD and Non-ASD) were entered into the 2×2 models as the within and between subjects factors, respectively. First, models without covariates were run to determine whether the groups demonstrated symptom change over time. For SA, significant main effects of diagnosis and time were observed with large and medium effect sizes respectively. However, no interaction effect was observed, indicating that although the ASD group showed significantly higher SA scores than the Non-ASD group at T1, the groups demonstrated similar rates of decrease in SA scores (i.e., improvement) between T1 and T2. For RRBs, a significant interaction was observed between time and diagnosis, indicating that the ASD and Non-ASD groups demonstrated different patterns of change in RRBs. Follow-up analyses revealed a significant effect of time for the ASD group with a medium effect size, but not for the Non-ASD group, indicating that the ASD group demonstrated an increase in RRB scores (i.e., worsening) while the Non-ASD group demonstrated stability.

Next, nonverbal DQ was entered into the model to determine whether developmental level could explain the changes over time observed across diagnostic groups. For SA, the effect of diagnosis remained significant with a large effect size, while the effect of time was no longer significant. For RRB, the interaction between time and diagnosis remained significant. However, follow-up analyses indicated that the effect of time for the ASD group was no longer significant. Thus, the improvement in SA and worsening in RRB scores demonstrated in children with ASD was accounted for by initial nonverbal developmental level.

Discussion

Stability of Clinical Diagnosis

All children given a clinical diagnosis that confirmed or ruled out ASD (i.e., diagnosis was not deferred) demonstrated diagnostic stability when re-evaluated one to two years later. Findings were consistent with the other studies of stability of ASD diagnoses made under two (Chawarska et al., 2009; Cox et al., 1999) and together support the growing practice of making early clinical diagnoses when diagnostic presentations are clear. The present study also extends findings in high-risk samples, by reporting on children recruited from a community-screening program who did not necessarily garner parent or professional concern at pediatric visits. In fact, although 90% of children screened positive on the CSBS

Infant-Toddler Checklist, only 50% of parents reported having concern about their child's development at the time of screening. Parental concern appeared nonspecific, as the percentage of children who received a clinical diagnosis of ASD at T1 was identical in children with and without parental concern at screening (i.e., 68% in both groups).

Importantly, diagnostic stability was found for a group of children who participated in a two-tiered process that included broadband and autism screening followed by a comprehensive diagnostic evaluation. Stability of confirming or ruling out ASD was likely facilitated by the use of experienced clinicians, who have been shown to generate more reliable diagnoses than inexperienced clinicians (Klin, Lang, Cicchetti, & Volkmar, 2000; Stone et al., 1999). The use of observation across clinic and home settings also likely contributed to stability, as home observation may be particularly valuable for toddlers whose behavior can be highly influenced by the testing environment. This model of diagnostic evaluation is presented as a "best-practices" model, though it is recognized that community settings may not have sufficient resources to implement diagnostic practices as outlined here.

Stability of confirming and ruling out ASD was also impacted by clinicians' ability to defer diagnosis, as an initial diagnosis was made for only 83% of the sample. A small proportion of children was initially classified into the deferred diagnosis group, suggesting that symptom presentation is not always clear at these young ages. These findings suggest that 100% stability of clinical diagnosis of ASD is possible when diagnostic practices allow sufficient flexibility to not make an official diagnosis until diagnostic presentation is clear. In contrast to stability of confirmed and ruled out ASD diagnoses, stability of an unclear diagnostic presentation was quite low, indicating that almost all of these children could be given a diagnosis one to two years later.

The deferred subgroup was distinguishable from children given a clinical diagnosis of ASD, DD, or TD based on social communication deficits. Many of these children (*n*=9 of 14) fell in the "mild-to-moderate concern" range of the ADOS-T, indicating the importance of monitoring children with this risk classification. Finally, the majority of the deferred diagnosis group (*n*=10 of 14) was later diagnosed with DD or TD, suggesting that many of these young children with subthreshold symptoms show improvement and no longer warrant clinical concern for ASD at follow-up. However, this finding should be interpreted cautiously given the small number of children with a deferred diagnosis. Clinicians should carefully monitor all children with a deferred diagnosis, as some children did go on to receive a diagnosis of ASD.

Concurrent and Predictive Value of ADOS-T Classifications

The excellent discrimination and sensitivity in the present sample were comparable to findings from the ADOS-T validation study by Luyster and colleagues (2009), while the acceptable specificity was lower. This lower accuracy in excluding children without ASD may be due to the wider range of developmental level in this low-risk sample, which may better reflect the performance of the ADOS-T in community settings. On the other hand, the ADOS-T was able to detect almost all children with ASD, supporting its use as a diagnostic tool as well as the established cutoffs and diagnostic classifications. Results also highlight the importance of using the ADOS-T in the context of a comprehensive evaluation of symptoms across settings rather than in isolation.

Results from the predictive analyses indicate that initial ADOS-T scores were also highly predictive of clinical diagnosis, with each additional algorithm point corresponding to 1.63 times greater likelihood of having an ASD diagnosis at follow-up. Examination of risk ranges suggested greatest accuracy of the lowest and highest risk ranges, as 83% of "little-

to-no" concern range at initial evaluation were diagnosed as Non-ASD at follow-up and 95% of "moderate-to-severe" concern range were diagnosed with ASD. On the other hand, the "mild-to-moderate" concern group at initial evaluation was comprised almost equally of children with and without ASD at follow-up. Thus, while ROC and logistic regression analyses suggest excellent sensitivity and acceptable specificity of the research classifications and substantial predictive value of total scores, use of the recommended risk ranges appears to provide slightly distinct and clinically useful information.

Change in Symptom Severity by Domain

Results also indicated that significant changes in symptom severity were observed even when diagnosis was stable, suggesting that the toddler years are marked by changes in global symptom severity. Although children with and without ASD differed on symptom severity at initial evaluation, improvement in social communication and interaction was observed at similar rates across diagnostic groups. In contrast, restricted and repetitive behaviors increased (i.e., worsened) in children with ASD, while they remained stable in children without ASD. Results further support the finding that these behaviors are observable in toddlers and distinguish young children with ASD. The significant increase in number and/or severity of restricted and repetitive behaviors in the ASD group suggests that an unfolding of these behaviors over the second and third years of life may also be an indicator of ASD.

Nonverbal developmental level at initial diagnosis accounted for the improvement in social communication and interaction observed for all children and worsening in restricted and repetitive behaviors observed for the ASD group. Such findings suggest that initial developmental level and subsequent change in autism symptoms may covary within children, thereby highlighting the importance of measuring both abilities in toddlers to facilitate understanding of symptom changes over time.

Limitations and Future Directions

While this study had a number of unique strengths, limitations should also be considered. Perhaps most importantly, follow-up diagnoses were not independent of initial diagnosis, as some children saw the same clinician for both evaluations and all clinicians utilized the child's complete history to make clinical diagnoses. Lack of blind clinicians represents a common solution to the tradeoff between external and internal validity, as community clinics are unlikely to have the resources to provide blind clinicians. In order to partially address this potential bias, all diagnoses were reviewed by another clinician who did not complete testing. However, there is a clear need for additional studies of diagnostic stability that utilize a blinded, independent diagnostic process.

The present study examined changes within the narrow window of the toddler years, and additional follow-up is planned in order to examine the stability into the preschool and school-aged years. Finally, changes in symptom severity were examined at the group level, thereby failing to model individual change. Future studies utilizing more frequent measurement and growth modeling have the capacity to elucidate individual symptom trajectories and reveal differential patterns of change within the first few years of life.

Conclusions

The present study addressed several important diagnostic issues faced by clinicians evaluating toddlers suspected of ASD. Results demonstrated that toddlers can be diagnosed with ASD at 15–24 months through a community-based screening program and that diagnoses of ASD and Non-ASD can be stable one to two years later, though a number of

important factors likely contributed to this stability. Clinicians making diagnoses were experienced with young children with and without ASD, used multiple sources of information to make clinical decisions, and had flexibility to defer diagnosis according to their clinical judgment. This study also demonstrated that a small proportion of children with a positive autism screen may have unclear diagnostic presentations before 24 months, show distinct social behaviors, and require additional follow-up. These findings also highlight the utility of the ADOS-T in making early diagnoses and predicting follow-up diagnoses, thereby extending current evidence of the psychometric properties of this diagnostic tool. Finally, toddlers with ASD demonstrated improvement in social communication and interaction and unfolding of restricted and repetitive behaviors during the toddler years, suggesting that early patterns of change in symptoms may be characteristic of ASD.

Acknowledgments

References

Bodfish, JW.; Symons, F.; Lewis, M. The Repetitive Behavior Scales: A test manual. Morganton, NC: Western Carolina Center Research Reports; 1999.

Chawarska K, Klin A, Paul R, Macari S, Volkmar F. A prospective study of toddlers with ASD: short-term diagnostic and cognitive outcomes. Journal of Child Psychology and Psychiatry. 2009; 37:89–126.

Chawarska K, Klin A, Paul R, Volkmar F. Autism spectrum disorder in the second year: stability and change in syndrome expression. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2007; 48:128–138.

Cook N. Statistical Evaluation of Prognostic versus Diagnostic Models: Beyond the ROC Curve. Clinical Chemistry. 2008; 54:17–23. [PubMed: 18024533]

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]

Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. Journal of Autism and Developmental Disorders. 2007; 37:613–627. [PubMed: 17180459]

Johnson CP, Myers SM. Identification and evaluation of children with autism spectrum disorders. Pediatrics. 2007; 120:1183–1215. [PubMed: 17967920]

Klin A, Lang J, Cicchetti DV, Volkmar FR. Brief report: Interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: Results of the DSM-IV autism field trial. Journal of Autism and Developmental Disorders. 2000; 30:163–167. [PubMed: 10832781]

Landa RJ, Holman KC, Garrett-Mayer E. Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. Archives of General Psychiatry. 2007; 64:853–864. [PubMed: 17606819]

Lord C, Bishop SL. Autism spectrum disorder: Diagnosis, prevalence, and services for children and families. Society for Research in Child Development. 2010; 24:1–21.

Lord, C.; Luyster, R.; Gotham, K.; Guthrie, W. Autism Diagnostic Observation Schedule–Toddler Module manual. Los Angeles, CA: Western Psychological Services; 2012.

Lord C, Luyster R, Guthrie W, Pickles A. Patterns of developmental trajectories in toddlers with autism spectrum disorder. Journal of Consulting and Clinical Psychology. 2012; 80:477–489. [PubMed: 22506796]

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, Leventhal BL, DiLavore PC, Pickles A, Rutter M. The Autism Diagnostic Observation Schedule Generic: A standard measure of social and communication de cits associated with the spectrum of autism. Journal of Autism and Developmental Disorders. 2000; 30:205–223. [PubMed: 11055457]

Lord, C.; Rutter, M.; DiLavore, P.; Risi, S. Autism Diagnostic Observation Schedule- Generic. Los Angeles, CA: Western Psychological Services; 1999.

Luyster R, Gotham K, Guthrie W, Coffing M, Petrak R, DiLavore P, Pierce K, Bishop SL, Esler A, Hus V, Oti R, Richler J, Risi S, Lord CE. The Autism Diagnostic Observation Schedule–Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. Journal of Autism and Developmental Disorders. 2009; 39:1305–1320. [PubMed: 19415479]

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]

Mullen, E. The Mullen Scales of Early Learning. Circle Pines, MN: American Guidance; 1995.

Nadig AS, Ozonoff S, Young GS, Rozga A, Sigman M, Rogers SJ. A prospective study of response to name in infants at risk for autism. Archives of Pediatrics Adolescent Medicine. 2007; 161:378–383. [PubMed: 17404135]

O'Roak BJ, State MW. Autism genetics: Strategies, challenges, and opportunities. Autism Research. 2008; 1:4–17. [PubMed: 19360646]

Ozonoff S, Heung K, Byrd R, Hansen R, Hertz-Picciotto I. The onset of autism: patterns of symptom emergence in the first years of life. Autism Research. 2008; 1:320–328. [PubMed: 19360687]

Rutter, M.; Bailey, A.; Lord, C. The Social Communication Questionnaire. Los Angeles, CA: Western Psychological Services; 2003.

Sparrow, S.; Cicchetti, DV.; Balla, DA. Vineland Adaptive Behavior Scales. 2. Circle Pines, MN: AGS; 2005.

Stone WL, Lee EB, Ashford L, Brissie J, Hepburn SL, Coonrod EE, et al. Can autism be diagnosed accurately in children under three years? Journal of Child Psychology and Psychiatry. 1999; 40:219–226. [PubMed: 10188704]

Sullivan M, Finelli J, Marvin A, Garrett-Mayer E, Bauman M, Landa R. Response to joint attention in toddlers at risk for autism spectrum disorder: a prospective study. Journal of Autism and Developmental Disorders. 2007; 37:37–48. [PubMed: 17216332]

Turner L, Stone W. Variability in outcome for children with an ASD diagnosis at age 2. Journal of Child Psychology and Psychiatry. 2007; 48:793–802. [PubMed: 17683451]

Volkmar F, Chawarska K, Klin A. Autism in infancy and early childhood. Annual Review of Psychology. 2005; 56:315–336.

Wetherby A, Brosnan-Maddox S, Peace V, Newton L. Validation of the Infant-Toddler Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. Autism. 2008; 12(5):455–479.

Wetherby, AM.; Lord, C.; Wood, J.; Pierce, K.; Shumway, S.; Thurm, A.; Ozonoff, S. The Early Screening for Autism and Communication Disorders: Preliminary Field-Testing of an Autism-Specific Screening Tool for Children 12 to 36 Months of Age. Paper presented at the International Meeting for Autism Research (IMFAR); Chicago, IL. 2009 May.

Wetherby, A.; Prizant, B. Communication and symbolic behavior scales developmental profile. Baltimore, MD: Paul H. Brookes; 2002. First Normed edition

Wetherby A, Woods J, Allen L, Cleary J, Dickinson H, Lord C. Early indicators of autism spectrum disorders in the second year of life. Journal of Autism and Developmental Disorders. 2004; 34:473–493. [PubMed: 15628603]

Woolfenden S, Sarkozy V, Ridley G, Williams K. A systematic review of the diagnostic stability of autism spectrum disorder. Research in Autism Spectrum Disorders. 2012; 6:345–354.

Yoder P, Stone WL, Walden T, Malesa E. Predicting social impairment and ASD diagnosis in younger
    siblings of children with autism spectrum disorder. Journal of Autism and Developmental
    Disorders. 2009; 39:1381–1391. [PubMed: 19449096]

Key points

- Stability of ASD is well established over age three, but has not yet been extensively researched under 24 months. While the toddler years are marked by growth and development, patterns of change in autism symptoms are not well-understood.

- A diagnosis of ASD made between 15 and 24 months was stable one to two years later when made as part of a comprehensive evaluation that utilized multiple sources. Despite this, some toddlers initially showed unclear diagnostic presentations requiring follow-up.

- The ADOS-T was a useful component of initial evaluations and predicted follow-up diagnosis.

- Even though clinical diagnosis was stable, toddlers showed improvement in social communication/interaction, and worsening of restricted, repetitive behaviors.

- Findings support the practice of conducting diagnostic evaluations for children under 24 months and elucidate the importance of using standardized diagnostic tools.
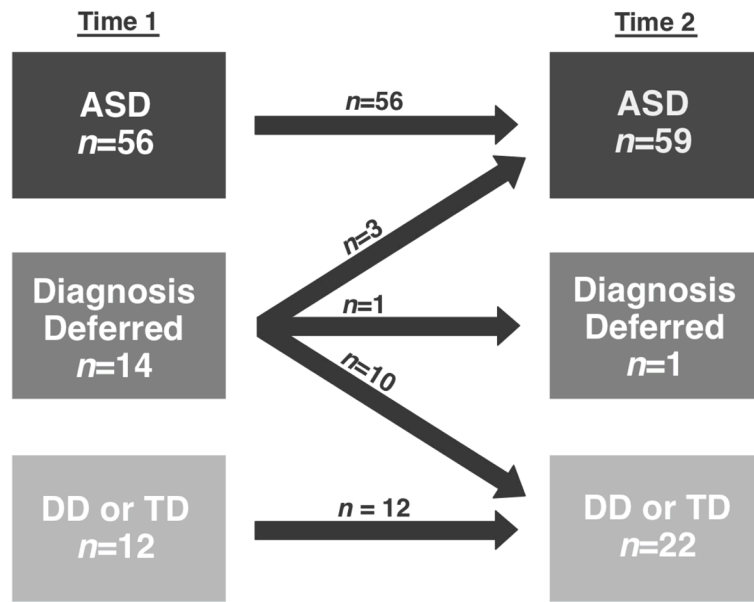
Figure 1.
Stability of Clinical Diagnosis

Table 1

Demographic Characteristics

Characteristic	*N*=82
Gender *n* (%)	
Male	64 (78.0%)
Race *n* (%)	
White	60 (73.2%)
Black	12 (14.6%)
Asian	2 (2.4%)
Biracial	8 (9.8%)
Ethnicity *n* (%)	
Hispanic	8 (9.8%)
Maternal education in years *M* (*SD*)	14.63 (2.45)

Table 2

Diagnostic Evaluation Characteristics

Characteristic *M (SD)*	Time 1			Time 2	
	ASD	Diagnosis Deferred	DD/TD	ASD	DD/TD*
N	56	14	12	59	23
Age at ADOS	19.07 (2.01)	19.71 (2.30)	20.50 (2.15)	37.39 (3.24)	35.57 (4.95)
ADOS Social Affect Total	14.36 (3.73)	8.93 (2.37)	6.08 (3.66)	10.44 (3.78)	4.22 (3.66)
ADOS Restricted/Repetitive Behavior Total	2.93 (1.83)	2.00 (1.52)	1.42 (1.31)	4.68 (2.11)	1.30 (1.36)
Age at Mullen	18.88 (2.22)	19.36 (2.53)	20.58 (2.23)	37.10 (3.41)	35.22 (5.00)
Mullen Nonverbal DQ	90.60 (15.32)	96.23 (17.39)	98.12 (15.32)	81.91 (21.54)	93.18 (23.47)
Mullen Verbal DQ	61.72 (18.02)	75.95 (24.28)	74.61 (20.48)	74.51 (27.31)	90.31 (28.77)

*
Note. Diagnosis was deferred for just one child at Time 2, so this child is included in the DD/TD group.

Table 3

Agreement Between Time 1 Classification and Time 2 Diagnosis/Classification

		T2 Clinical Diagnosis		T2 ADOS Diagnostic Classification		
		ASD *n* (% within classification)	Non- ASD	Autism *n* (% within classification)	ASD	Non- spectrum
T1 ADOS-T Ranges of Concern	Little to no (*n*=12)	2 (17%)	10 (83%)	1 (8%)	1 (8%)	10 (83%)
	Mild to moderate (*n*=26)	15 (58%)	11 (42%)	13 (50%)	4 (15%)	9 (35%)
	Moderate to severe (*n*=44)	42 (95%)	2 (5%)	37 (84%)	5 (11%)	2 (5%)
		$\chi^2(2)$=32.84, *p*<.001		$\chi^2(4)$=32.84, *p*<.001		
T1 ADOS-T Diagnostic Classification	Nonspectrum (*n*=25)	7 (28%)	18 (72%)	7 (28%)	2 (8%)	16 (64%)
	ASD (*n*=57)	52 (91%)	5 (9%)	44 (77%)	8 (14%)	5 (9%)
		$\chi^2(1)$=34.43, *p*<.001		$\chi^2(2)$=27.98, *p*<.001		

Table 4

Mixed-Model ANCOVA: Change Over Time By Time 1 Diagnosis

	Model without covariate			Model with NVQ		
	F	p	η^2		F	p	η^2
ADOS Social Affect				**ADOS Social Affect**			
Diagnosis	58.22	<.001	.41	Diagnosis	52.95	<.001	.39
Time	41.13	<.001	.13	Time	2.14	.15	.01
Time*Diagnosis	1.79	.19	.00	Time*Diagnosis	1.53	.22	.00
ADOS Repetitive/Restricted Behavior				**ADOS Repetitive/Restricted Behavior**			
Diagnosis	37.21	<.001	.31	Diagnosis	32.72	<.001	.31
Time	1.93	.17	.01	Time	1.85	.19	.01
Time*Diagnosis	8.12	.006	.04	Time*Diagnosis	6.67	.01	.04
Time*ASD	12.15	.001	.17	Time*ASD	1.24	.27	.00
Time*Non-ASD	1.23	.28	.01	Time*Non-ASD	1.51	.23	.02