



Published in final edited form as:

Child Adolesc Ment Health. 2013 February 1; 18(1): 52–63. doi:10.1111/j.1475-3588.2012.00664.x.

Measurement Issues: Screening and diagnostic instruments for autism spectrum disorders – lessons from research and practice

Tony Charman¹ and Katherine Gotham²

Tony Charman: t.charman@ioe.ac.uk

¹Centre for Research in Autism and Education (CRAE), Institute of Education, University of London, UK

²University of Michigan Autism and Communication Disorders Center (UMACC), Ann Arbor, Michigan, USA

Abstract

Background and Scope—Significant progress has been made over the past two decades in the development of screening and diagnostic instruments for autism spectrum disorders (ASD). This article reviews this progress, including recent innovations, focussing on those instruments for which the strongest research data on validity exists, and then turns to addressing issues arising from their use in clinical settings.

Findings—Research studies have evaluated the ability of screens to prospectively identify cases of ASD in population-based and clinically-referred samples, as well as the accuracy of diagnostic instruments to map onto ‘gold standard’ clinical best estimate diagnosis. However, extension of the findings to clinical services must be done with caution, with a full understanding that instrument properties are sample-specific. Furthermore, we are limited by the lack of a true test for ASD, which remains a behaviourally-defined disorder. In addition screening and diagnostic instruments help clinicians least in the cases where they are most in want of direction, since their accuracy will always be lower for marginal cases.

Conclusion—Instruments help clinicians to collect detailed, structured information and increase accuracy and reliability of referral for in-depth assessment and recommendations for support, but further research is needed to refine their effective use in clinical settings.

Keywords

autism spectrum disorder (ASD); screening; diagnosis; sensitivity; specificity; predictive value

Introduction

Autism spectrum disorders (ASD) affect approximately 1% of children (Baird et al., 2006; CDC, 2009). The term ASD is used to describe a range of neurodevelopmental conditions that demonstrate considerable phenotypic heterogeneity, both in terms of presentation at any one age and across development, and which are likely to differ in underlying aetiology (Levy, Mandell & Schultz, 2009); this has led to leading biologists adopting the term ‘the autisms’ (Geschwind & Levitt, 2007). However, individuals with an ASD share common impairments in social relatedness and reciprocity, the use of language for communication and rigid and repetitive ways of thinking and behaving (DSM-IV-TR; APA, 2000; ICD-10;

The authors have declared that they have no competing or potential conflicts of interest arising from the publication of this article. KG will be an author on the as-yet-unpublished Autism Diagnostic Observation Schedule, 2nd edition (ADOS-2): royalties for this will be donated to charity

WHO, 1993). Proposed revision for DSM-5 combines the social and communication impairments into one domain and includes sensory with repetitive behaviours (www.dsm5.org).

Notwithstanding progress in our understanding of the genetic and neurodevelopmental processes that lead to ASD, clinical diagnosis is reliant on the developmental and behavioural presentation. Until the 1990s it was rare for children to receive a diagnosis of autism until the age of 3 or 4 years. Today many children are now first identified in toddlerhood, although others, in particular those with average or above average language and cognitive abilities, are not diagnosed until school age or older (Charman & Baird, 2002; Mandell, Novak & Zubritsky, 2005; Manning et al., 2011).

With the recognition that ASDs are relatively common, and in many but not all cases can be identified early, there has been both burgeoning clinical interest and research activity in the development of ASD-specific screening and diagnostic instruments. In the present review we will focus on those instruments for which the strongest research data on validity exists; this applies to many more screeners than diagnostic instruments. One critical comment is that it would be better for the clinical and research fields if we knew *more* about *fewer* screening instruments, rather than the somewhat inadequate information on validity that we do have available for this ever-increasing list of ASD screens (see Al-Qabandi, Gorter & Rosenbaum, 2011; for a critical review on early screening).

1. Screening instruments

Issues in screening and surveillance—Screening and surveillance are different but related activities involving the detection of impairments with a view to prevention or amelioration of disability. Screening is the prospective identification of unrecognised disorder by the application of specific tests or examinations. Surveillance refers to the ongoing and systematic collection of data relevant to the identification of a disorder over time by an integrated health system.

Several parameters of screening instruments are important in assessing their efficacy and utility:

- i. Sensitivity is the proportion of individuals with a disorder who have a positive screen result,
- ii. Specificity is the proportion of individuals with a disorder who have a negative screen result,
- iii. Positive predictive value (PPV) is the proportion of individuals with a positive screen result who have the disorder

Sensitivity is required to be high in order that the screen misses few cases of the disorder (avoiding falsely reassuring parents and professionals). Specificity is required to be high in order that few cases without the disorder are screen positive (avoiding falsely alarming parents and costly referral for in-depth). When the sensitivity and specificity of a screen remain constant, the PPV is lower the rarer a disorder is within the population (Clark & Harrington, 1999). Hence, PPV will be lower in population than in referred samples. Glascoe (1996) has estimated that acceptable sensitivity and specificity for developmental screening tests are 70% to 80%, reflecting the nature and complexity of measuring the continuous process of child development (American Academy of Pediatrics, 2001; NICE, 2011).

Surveillance involves a parent-professional partnership that combines the observations of parents with the knowledge of the professional and the deployment of specific tests. There is

evidence that the use of screening instruments in combination with asking parents about their concerns improves the efficiency of an instrument (Glascoe, 1999). In this context, the use of specific screens where there is some concern on the part of the parent or professional can be a useful adjunct to clinical judgement.

Population prospective screening studies—Less than a handful of population screening studies have been conducted and even fewer have undertaken the long-term follow-up required in order to fully ascertain sensitivity – i.e., identifying cases missed by systematically re-visiting the sample at a later age point. Baron-Cohen and colleagues developed the CHAT (CHecklist for Autism in Toddlers) that assesses simple joint attention and pretend play behaviours by parental report and health practitioner observation through direct testing. In a sample of forty-one 18-month-old younger siblings of children diagnosed with autism the four children in the at-risk group who were subsequently diagnosed with autism were identified by the CHAT (Baron-Cohen et al., 1992). Baird et al. (2000) went on to test the effectiveness of the CHAT in a population of 16,235 18-month-olds as part of routine health surveillance. Children who failed all five joint attention and pretend play key items were the ‘high risk for autism’ group; children who failed both items measuring protodeclarative pointing (pointing for interest) were the ‘medium risk for autism’ group. In order to minimise false positives, a two-stage screening procedure was adopted. Children who were initially screen positive (at the high or medium risk threshold) received a second administration of the screen one-month later, via a telephone follow-up. Used in this 2-stage way, the PPV for all ASDs of the high and medium risk thresholds combined was good (59%). However, sensitivity was poor (21%), indicating that four-fifths of children subsequently identified as having ASD in the study population were missed by the screen. If a one-stage screening procedure only had been adopted, the sensitivity increased to 35%, although in clinical use this would have entailed the assessment of more screen false positives.

Buitelaar and colleagues in the Netherlands developed a screening instrument (ESAT: Early Screening of Autistic Traits) to identify ASD (Willemsen-Swinkels, Dietz, & van Daalen, 2006). Dietz et al. (2006) completed screening of 31,724 children at 14 months of age. Health practitioners at a well-baby clinic appointment administered an initial screen of 4 items (measuring varied play with toys, readability of emotional expression and sensory abnormalities). If a child failed 1 or more of the 4 items they were offered a follow-up home visit where a longer version of the ESAT (14 items that included many social communication items such as eye contact, response to name, etc.) was administered. Children who failed 3 or more items of the 14-item ESAT were invited for a diagnostic evaluation at a mean age of 23 months. The ESAT did correctly identify children with ASD ($n = 18$) but also identified children with language disorder (LD; $n = 18$) and intellectual disability (ID; $n = 13$). The items that discriminated best between children with and without ASD were items assessing early social communication impairments, including ‘shows interest in people’, ‘smiles directly’ and ‘reacts when spoken to’. However, at this earlier age compared to the CHAT study (14 vs. 18 months) these early signs appear to be less specific to ASD, identifying almost as many children who went on to have LD and ID. This and other studies cannot report on the sensitivity of these signs to all cases of ASD until a population follow up of the whole sample has been conducted to identify missed cases.

Screening in referred samples

Robins et al. (2001) developed a modified version of the CHAT (the M-CHAT) that included additional items measuring other aspects of early social communication impairments characteristic of autism (e.g. response to name, imitation) as well as repetitive behaviours (e.g. unusual fingers mannerisms) and sensory abnormalities (e.g. oversensitivity

to noise). The M-CHAT is a parent report instrument and the health practitioner does no direct testing. In their initial report, Robins et al. (2001) had tested 1,122 unselected children (initially at 18 months but subsequently at 24 months of age) and 171 children referred for early intervention services (considered to be at high risk of having an ASD or other developmental disability). Following analysis of the first 600 returns, a cut-off was set as failing 2 from 6 'critical items' or any 3 items from the total of 23 items (Robins et al., 2001). Once a child failed the M-CHAT the research team re-administered the screen by telephone, and if a child still scored above cut-off the family was invited for an assessment. Of the 58 children who failed on both administrations of the M-CHAT, 39 subsequently received an ASD diagnosis and the remaining 19 were found to have language or global developmental delay. Only 3 of the 39 children with ASD were from the unselected population, with the majority being identified from the sample referred for early intervention services. The items that best discriminated between children with ASD and children with other developmental problems were those that measured joint attention behaviours (pointing, bringing things to show), social relatedness (interest in other children, imitation) and communication (response to name). Robins et al. (2001) calculated sensitivity, specificity and PPV for various combinations of M-CHAT items and demonstrated that in this largely referred sample its instrument parameters were reasonably strong (PPV for 2-stage 68%).

Two recent studies (Kleinman et al., 2008; Robins, 2008) have reported on the M-CHAT with new samples of 3,793 children and 4,797 children, respectively, aged 14 to 30 months. In a sample combining an unselected and a high risk (referred) sample, Kleinman and colleagues (2008) found a PPV of 0.36 for the initial screening, which improved to 0.74 for the screening plus the follow-up telephone interview. Again, most cases were identified from the 'high risk' sample of children referred for early intervention services or due to a developmental concern. Robins (2008) found a much lower (0.06) PPV in an unselected sample attending well-child visits, but following the telephone interview the PPV increased to 0.57. Follow-up studies will allow us to estimate the instrument's parameters when used on an unselected population, in particular its sensitivity in detecting cases of ASDs in children about whom there had not been previous developmental concerns. Canal-Bedia et al. (2011) report on a Spanish translation of the M-CHAT and found screening plus follow-up PPVs of 0.35 and 0.19 in a combined sample of low and high risk (referred) (N=2,480) and low risk (N=2,055) children, respectively.

Wetherby and colleagues (Wetherby, Brosnan-Maddox, Peace, & Newton, 2008; Wetherby et al., 2004) developed an early screening tool that can be used from 6 to 24 months of age – the Infant Toddler Checklist (ITC). The ITC is a broader developmental screen that successfully identifies children with developmental delay as well as children with an ASD. The ITC is a 24-item parent completed screen that asks about early social responsiveness, gestures, babble and early language and motor development. In their most recent work, 5,385 families with infants were recruited between 6 and 24 months of age and then re-screened every 3 months (Wetherby et al., 2008). They have shown that it is possible in some cases to prospectively identify children who will go on to have a diagnosis of an ASD towards the end of the first year of life, although most cases were identified during the second year. PPV was low at 6–8 months (43%) but increased to over 70% in the second year of life (Wetherby et al., 2008).

Overall these prospective screening studies (summarised in Table 1) have shown that it is possible to prospectively identify ASD, including in children about whom parents and professionals did not have pre-existing concerns, from the age of 18 months or even earlier in the second year of life. The most common early signs captured by the screen are impairments or delays in early emerging social communication behaviours such as response to name, joint attention and play behaviours, although sensory abnormalities or a restricted

repertoire of play activities might also be early indicators of later ASD. However, these early signs are neither universal nor specific to ASD as opposed to other neurodevelopmental disorders. In addition, one-stage screening has been shown to have low PPV with the attendant risk of over-referral. These research findings suggest caution about recommending universal population screening. Despite this, the American Academy of Pediatrics recently published guidelines that recommended routine use of screens to help community paediatricians to identify at-risk cases at 18 month and 24 month health checks (Johnson et al., 2007; see Al Abandi et al., 2011). It remains to be determined whether screening instrument parameters warrant this clinical guidance. However, screens can be a useful adjunct to ongoing parent-practitioner surveillance.

Two recent, innovative studies remind us that screening instruments do not act in isolation to improve early detection. When introducing screening instruments to practitioners one inevitably is involved in training the practitioners to use the screen and about the early signs of ASD. In a controlled study, Oosterling et al. (2010a) showed that *in combination*, training health professionals in the early signs of autism and introducing use of the ESAT screen and a clear referral and diagnostic pathway led to a reduction in age of diagnosis from 83 to 64 months. The proportion of cases diagnosed before the age of 24 months increased from 12.7% to 24.3%. Dereu et al. (2010) also provided training and introduced a new screening instrument the CESDD (Checklist for Early Signs of Developmental Disorders) to daycare workers caring for 6,808 children between 3 and 39 months of age. Dereu and colleagues report a moderate to high sensitivity of 0.68 but a low PPV of 0.10, although the sample has yet to be followed up to determine cases missed. By being trained in the use of screens and the early signs of ASD, health and education practitioners will develop their expertise to identify possible ASD cases and come to understand the benefits but also the potential limitations (e.g. false positives) of using the screening instruments themselves.

One question is whether the ultimate goal should be the development of a universal population screen to identify undetected cases of ASD or the development of instruments that can be used, in combination with parental and professional expressions of concern regarding a child's development, in ongoing health surveillance. Another consideration is whether screens should target ASD specifically or whether they should also attempt to identify children with language and general developmental delays or other neurodevelopmental conditions.

Screening with older children

The Social Communication Questionnaire (SCQ; Rutter et al., 2003) is a screening tool for ASD based on the Autism Diagnostic Interview-Revised (ADI-R; Lord et al., 1994) and is increasingly widely-used in research and clinical practice. It consists of 40 items answered yes/no and is suitable for both verbal and non-verbal children. In the initial validation study that included children and adults, the SCQ discriminated well between ASD and non-ASD cases with a sensitivity of 0.85 and a specificity of 0.75 (Berument et al., 1999). Chandler et al. (2007) found the SCQ to work well with at-risk 9-to-10 year old children (sensitivity 0.88, specificity 0.72). Chandler et al. (2007) also reported that in 2 small general population samples between 4% and 5% of children scored above the ASD cutoff, including 1.5% who scored above the autism cut-off. Many of these high-scoring children had an ASD diagnosis and almost all (~90%) had a diagnosed neurodevelopmental disorder. Corsello et al. (2007) reported reduced sensitivity (0.68) for detecting ASD in children below the age of 5 years compared to a sensitivity of 0.80 for children 11 years and older. Recently the first cross-cultural validation study also found strong instrument properties in a German sample (Bölte, Holtmann & Poustka, 2008). The ability of the SCQ to discriminate between ASD and non-ASD cases in two samples of 3-to-6-year-old children has recently been reported. In 2 studies Eaves and colleagues found moderate sensitivity ~0.70 and specificity (0.79; Eaves

et al., 2006a) but reduced specificity (0.54) in the second study (Eaves et al., 2006b). One recent study has reported the ability of the SCQ to detect ASD in a younger (20–40 month old) referred/high-risk sample and found a sensitivity of 0.66 and a PPV of 0.79 (Oosterling et al., 2010).

The Social Responsiveness Scale (SRS; Constantino & Gruber, 2005; Constantino et al., 2000) is a parent and teacher completed questionnaire with 65 items rated on a 4-point Likert scale (from 'not true' to 'almost always true'). Charman et al. (2007) found that the SRS had a specificity of 0.78, a sensitivity of 0.67 and a PPV of 0.63 in a clinical sample (N=119) of 9 to 13 year olds, with higher PPV (0.78) in the high IQ (>70) compared to the low IQ subsample (0.52). Constantino et al. (2007) found that in a large sample (N=442) of children with ASD and non-ASD, a combined parent and teacher T-score cut point of >60 yielded a sensitivity of 0.75 and a specificity of 0.96 (PPV not reported). A German version of the SRS also reported strong instrument properties (Bölte, Poustka & Constantino, 2008).

We have chosen to focus on the two most widely-used and best-validated ASD screening instruments for children and adolescents. Other ASD screens for children and young people have been developed over the past decade, including the Childhood Autism Screening Test (CAST; Williams et al., 2005, 2008) and the Autism Spectrum Screening Questionnaire (ASSQ; Elhers, Gillberg & Wing, 1999; Posserud, Lundervold & Gillberg, 2006, 2009); summaries of their instrument properties are available in other recent reviews (Norris & Lecavalier, 2010). Only a few studies have directly compared different screens against each other, providing insufficient data at this stage to make clear recommendations regarding which screens 'work best' at a particular age or ability level. Charman et al. (2007) found that the SRS and the Children's Communication Checklist (CCC; Bishop, 1998) had lower precision than the SCQ in the subsample with low IQ (<70). Oosterling et al. (2009) reported on the ESAT, SCQ and CHAT screens in a high risk sample of children screened around 30 months of age (8 to 44 months, N = 238). The instrument with the highest PPV was the CHAT (0.97 for 5 high risk key items) but this was associated with the lowest sensitivity (0.18), as was found in the original CHAT population study. The ESAT and the SCQ both had moderate PPV (0.68 and 0.79, respectively) and moderate sensitivity (0.88 and 0.66). The decision of which screen to use will be sample specific in both clinical and research use, and some understanding of how sample and respondent characteristics might systematically affect instrument properties is required to make recommendations for one screen over another.

Clinical issues in screening and surveillance

When screening for undetected cases of ASD, some parents' first recognition that something might be wrong may follow 'failure' of a screen and consequent discussion about their child's development with the professional involved. For a parent to make use of information about their child it first has to make sense and they have to be ready to agree on it. Recognition, belief and acceptance can be particularly difficult when the professional is giving completely unexpected information. One of the benefits of active surveillance is the opportunity to discuss 'risk status' with parents and what it means when a particular child fails a screen. In practice, being screen positive does not constitute a diagnosis, even when tests have very high positive predictive value; rather, the initial screening process should be seen as the beginning of a dialogue between the parent and professional about the child's development, with additional assessments being couched as helpful checks to make sure things are going OK.

Another caution is that screening results are sample-specific. The prevalence of ASD cases within a sample, the characteristics (e.g. clinical diagnosis, IQ, age) of the ASD and non-ASD cases, family factors (e.g. parental education, parental knowledge about autism), and

methodological factors, including whether the screen was completed prior to or after a diagnostic assessment, can all affect how a screen performs. However, the utility of any particular screening instrument and the application of any particular cut-point for identifying ‘screen positives’ depend both on the sample characteristics and on the intended purpose of screening. Charman et al. (2007) outline different hypothetical clinical and research scenarios that illustrate how three different screens would perform on different tasks. The choice of which screen to use, and for which particular purpose, critically depends on the relative costs of false positives and false negatives. It is also important to remember that these costs tend to fall on different parties. False positives involve costly further investigation and parental anxiety. False negatives may deprive children of clinical and education resources or place the burden of provision entirely on parents.

Diagnostic Instruments

ASD screening instruments function to identify children in need of further monitoring or diagnostic evaluation. At that point, standardised autism diagnostic instruments are often employed to structure the information-gathering from both parents and identified children within a diagnostic assessment. The existence of and ongoing improvements to such measures are associated with more accurate diagnosis of ASD, including the ability to reliably describe milder and younger cases, as well as increased comparability of research findings, based on better agreement as to “caseness” across research teams. However, as with screening tools, diagnostic instruments are often limited by inadequate power to correctly identify individuals with and without ASD. Further, the estimates of such performance validity for each particular measure are necessarily limited by the absence of an absolute test for ASD, and as such are influenced by clinical experience in diagnosing ASD, training and experience in using the diagnostic measure, and evolution within the field in terms of what is recognised and labeled “ASD.” Four commonly used autism diagnostic instruments are reviewed briefly below in terms of intended purpose, administration and scoring, and psychometric properties. Instrument parameters of these measures based on the largest available samples are briefly summarized in Table 3. For more detailed psychometric information, see the National Institute of Clinical Excellence guidelines (NICE, 2011).

The Autism Diagnostic Interview-Revised (ADI-R)

The Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) is a standardised, semi-structured interview that is administered by a trained clinician to a parent or caretaker familiar with the developmental history and current behaviour of the individual being evaluated. Scoring is based on the interviewer’s judgment of the behavioural reports obtained, rather than on the informant’s judgment. Administration and scoring of this interview takes approximately 1.5–3 hours in a face-to-face setting. The ADI-R version published by Western Psychological Services (WPS) can be used to assess those with mental ages above 24 months (Rutter, Le Couteur, & Lord, 2003); newly-created algorithms extend the use to children aged 12–47 months and down to nonverbal mental ages of 10 months (Kim & Lord, in press).

The ADI-R is comprised of 93 items focusing on Early Development, Language/Communication, Reciprocal Social Interactions, and Restricted, Repetitive Behaviours and Interests. Most items includes distinct current and historical scores, the latter based either on the period between the individual’s 4th and 5th birthdays (“Most Abnormal 4–5”) or the point in the individual’s lifetime at which the behaviour in question was regarded as most atypical (“Ever”). Scores are assigned on a 0–3 scale; higher numbers indicate more definite presence or greater severity of symptoms. Diagnostic algorithms for children aged 2:0 to 3:11, or 4:0 and older are based on sums of specific item scores across the domains noted above. Algorithm cut-offs for all domains must be met or exceeded to achieve the

instrument classification of “autism.” Current Behaviour algorithms exist but do not yield classifications.

In two large independent samples aged 3 and older (N=960 American; N=232 Canadian), the ADI-R correctly identified 89–95% of children with ASD, however it yielded a non-spectrum classification for only 56–59% of children with non-ASD disorders (Risi et al., 2006). When used in combination with the Autism Diagnostic Observation Schedule (ADOS), this specificity improved to 77% and 75% by sample. This supports the recommendation by de Bildt and colleagues (2004) that the ADI-R and ADOS are most valuable in combination. WPS ADI-R diagnostic algorithms have been found to be over-inclusive for individuals with nonverbal mental ages below 18 months and those with severe to profound intellectual disability (Lord, Storoschuk, Rutter, & Pickles, 1993; Nordin & Gillberg, 1998; Risi et al., 2006); see Kim & Lord, in press, for algorithm revisions to address those issues.

Interrater reliability was found to be very good in a sample of 20 children with autism or intellectual or language impairments (Lord et al., 1994), as well as in larger sample of 94 preschoolers (Lord et al., 1993). A later study of seven reliable examiners rating one administration reported good to perfect agreement on 87% of items (Cicchetti, Lord, Koenig, Klin, & Volkmar, 2008). Original test-retest data were available for six children: with blind interviewers administering the measure 2–3 months later, exact agreement exceeded 83% for all but 6 items (Lord et al., 1994).

The Diagnostic Interview for Social and Communication Disorders (DISCO)

The Diagnostic Interview for Social and Communication Disorders (DISCO; Wing, Leekam, Libby, Gould, & Larcombe, 2002), now in its 11th version, is a semi-structured, standardized interview used in ASD assessment for diagnosis and educational and/or treatment planning. Like the ADI-R, the DISCO is administered face-to-face by a trained examiner interviewing a parent or close caregiver of an individual with suspected ASD. However, ratings on the DISCO can be based on any available information, including direct observation of the individual or reports from teachers or other caregivers. The instrument is intended for individuals of all chronological and mental ages, although published data on the diagnostic validity of the measure in special populations (such as very young children or those with above average or impaired intellectual ability) are limited.

The DISCO takes approximately 2–4 hours to administer (including scoring). It encompasses 362 items covering domains of social interaction, communication, imagination, and repetitive behaviours, as well as domains assessing developmental levels and daily living skills, and non-ASD-specific behaviours, such problems with attention, overactivity, sexual or psychiatric difficulties, and other challenging behaviours. Developmental items are rated on a 3-point scale of “delay,” “minor delay,” or “no problem”; atypical behaviour items receive separate ratings based on current behaviour and most atypical behaviour ever, with a scale of “severe,” “minor,” or “not present.” The DISCO was originally intended to assist in clinical assessment and treatment planning for an individual rather than to yield a categorical diagnosis, however computerized diagnostic algorithms for research use have been created (Leekam, Libby, Wing, Gould, & Gillberg, 2000; Leekam, Libby, Wing, Gould, & Taylor, 2002). These include two algorithms operationalizing the ICD-10/DSM-IV criteria for autistic disorder and Asperger’s Disorder, a 5-item algorithm based on Wing and Gould (1979) criteria for autism spectrum disorders, and an algorithm based on Ehlers and Gillberg (1993) criteria for Asperger’s Disorder.

In a DISCO-9 validity study of children aged 3–11 years (Leekam et al., 2002), all 36 children with ASD were correctly identified by either Wing & Gould or ICD-10 autism

algorithms, however 10 out of 17 with intellectual disability and 4 out of 14 with language disorders were also identified as “ASD.” In a study using the Swedish translation of the DISCO-10, the ICD-10 autism algorithm identified 42 out of 51 individuals with an ASD diagnosis and incorrectly identified 1 individual out of 6 without ASD (Nygren et al., 2009). Another Swedish study reported that 89 adult longitudinal participants with an ASD were correctly identified by the DISCO as having an autism spectrum condition by either ICD-10 autism or atypical autism algorithms (Billstedt, Gillberg, & Gillberg, 2007). Inter-rater reliability of the DISCO-10 was assessed in 33 children (aged 2.5–15) and 7 adults (20–38 years), with over 90% of items showing good or excellent inter-rater reliability and the majority of algorithm items showing at least moderate reliability (Nygren et al., 2009).

The Developmental, Dimensional, and Diagnostic Interview (3Di)

The Developmental, Dimensional, and Diagnostic Interview (3Di; Skuse et al., 2004) is a parent/caregiver interview administered face-to-face by a trained examiner using a laptop computer. Prior to the interview, parent/caregivers complete questionnaires that are entered into the software and used to tailor the order and wording of questions asked in the in-person interview. Parent responses are entered directly into the software in the moment, and immediately following the 90–180 minute interview, a computer-generated report, including algorithm scores and classification, is available to the clinician. In addition, a spreadsheet of up to 1300 3Di variables is automatically created. The 3Di was developed primarily to assess ASD symptoms in children with average-range abilities and to differentiate ASD and non-spectrum conditions in a general population; the authors suggest it may also be used in populations with moderate to severe intellectual disability (Skuse et al., 2004).

The 3Di is comprised of mandatory modules covering autism spectrum symptoms (266 items), optional modules on comorbid symptoms (291 items), and 183 items concerning patient demographics, family background, developmental history, and motor skills. A short 53-item form of the 3Di has also been piloted (3Di-sv; Santosh et al., 2009). On both the original and short forms, response options vary within 3-point scales. Computer-generated algorithms weight and sum responses within domains, though published reports are not clear as to which domains are represented in the algorithm, how items were chosen for algorithm inclusion, and which classifications are produced.

The majority of data on the reliability and validity of the 3Di comes from the authors' original 2004 paper (Skuse et al., 2004) and the published report of the short form (Santosh et al., 2009). In the former report, the measure discriminated between 27 children with ASD, 60 typically developing children, and 33 with unreported diagnoses with 100% sensitivity and 98% specificity. In the latter report, sensitivity of the short form ranged from .90 to .96 and specificity ranged from .85 to .96 by domain in an independent sample of 439 children, 58% of whom had ASD (Santosh et al., 2009).

The Autism Diagnostic Observation Schedule

The Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000) is a semi-structured, standardized observation of children and adults referred for ASD. With the recent addition of a Toddler module (Luyster et al., 2009), the ADOS has five development- and language-dependent modules, which are 30–60 minute protocols of activities that are based on talking and/or play-based interaction. Trained examiners choose the appropriate module for the individual's age (from 12 months through adulthood) and verbal level (from no words to fluent speech), and follow that protocol of activities using standardized materials (e.g., books, toys), for a semi-structured social interaction. An adapted ADOS module to address diagnostic needs in older individuals with minimal language is currently undergoing validity testing (Hus & Lord, in preparation).

Like its companion measure, the ADI-R, the ADOS was created by operationalizing DSM-IV criteria for autism. Item scores on a 0–3 scale, with higher scores indicating greater symptom severity, are assigned immediately after completing the administration. Specific items from Communication, Reciprocal Social Interaction, and Restricted and Repetitive Behaviour domains comprise algorithms, which yield a classification of “autism,” “autism spectrum disorder,” or “non-spectrum.” In 2007, revised algorithms were created with the same number of items and of similar content across modules 1–3 (Gotham, Risi, Pickles, & Lord, 2007), and in 2009, two algorithms were created to correspond to the Toddler module (Luyster et al., 2009). In these seven new algorithms, raw scores of algorithm items from a “Social Affect” and a “Restricted Repetitive Behaviour” domain are summed and applied to one set of cut-offs to yield the instrument classification. In addition, Toddler algorithm totals can be located within three “ranges of concern,” in order to discuss the scores dimensionally rather than applying a cut-off score. For the original algorithm still in use for Module 4 (for older adolescents and adults with fluent speech), separate cut-offs exist for Communication, Social Interaction, and the combination of those two totals; all three sets of cut-offs must be met or exceeded to achieve an “autism” or “autism spectrum” classification.

Both original and revised diagnostic algorithms have strong predictive validity against best estimate clinical diagnoses, with the revised set of algorithms showing minimal association between ADOS totals and chronological age, generally decreased association between ADOS total and verbal IQ when compared to the original algorithms, and improved sensitivity in lower-functioning developmental groups (Gotham et al., 2007; Gotham et al., 2008; Luyster et al., 2009). In 2009, de Bildt and colleagues offered a caveat that the inclusion of the RRB domain toward the classification cut-off may lead to over-inclusion of children with cognitive impairments on the revised algorithms (de Bildt et al., 2009). Of note, the ADOS performs better in autism clinic samples in which the information gained from the measure is directly used in clinicians’ diagnostic decisions; in samples in which the ADOS examiner is not the primary diagnostician (such as the Canadian sample in Risi et al., 2006, or the American medical center sample in Molloy, Murray, Akers, Mitchell, & Manning-Courtney, 2011), the predictive validity of the measure tends to be significantly lower.

Inter-rater reliability, internal consistency, and test-retest reliability of the measure were found to be good to excellent in the original ADOS reliability sample of 98 individuals and 12 reliable examiners, as well as in updated data on the revised algorithms (see the ADOS-2 manual, Lord et al., in press) and the toddler module (Luyster et al., 2009).

Issues in the Use of Autism Diagnostic Measures

Because of their strong discriminant validity, the ADI-R and ADOS have been translated into over 18 languages and are used worldwide to establish caseness and aid in treatment planning; they have also been linked to diverse genetic and neuroimaging findings. Scores from both instruments also have been used to measure severity of autism symptoms and changes over time, however it is important to keep in mind that these measures were developed with the goal of differentiating individuals with and without ASD. Recent updates to the ADOS have resulted in the creation of a 10-point calibrated severity scale proposed as an alternative method of quantifying ASD severity with greater independence from participant characteristics such as chronological age and IQ (Gotham, Pickles, & Lord, 2009). Calibrated severity scores do not measure functional impairment but rather provide an alternative for comparing ADOS scores across modules and time. Though initial replications provide some evidence for the utility of the metric (de Bildt et al., 2011), it has not yet been widely studied.

A strength of the DISCO is that it provides an overall profile of skills and abilities, challenges and disabilities, and areas of needed remediation. Therefore, for clinical purposes, it allows simultaneous information-gathering for diagnosis and treatment planning. The 3Di offers a dimensional measurement of symptoms, assessment of comorbid symptoms, a less time-consuming training protocol than comparable measures, and quickly and easily accessible results. However, the base of empirical research on the 3Di is extremely small, and predictive validity estimates have been generated largely in comparison to typically-developing children, a circumstance which inflates predictive validity scores and does not reflect the populations in which ASD diagnostic decision-making usually takes place. In light of the strengths of this measure, it will be worthwhile to explore the utility of the 3Di in larger, more diverse ASD populations, as well as its ability to differentiate between children with ASD and those with clearly defined and reported cognitive, language, and behavioural disorders.

For most autism diagnostic measures, training is time-consuming and can be both expensive and difficult to procure. One avenue to achieving research reliability on the ADI-R and ADOS includes achieving three consecutive reliable administrations with a research-reliable examiner (at 90% item exact agreement on both the protocol and algorithm items for the ADI-R, and 80% for the ADOS). A second means of achieving reliability is to attend training workshops, score video-recorded administrations, and submit recordings of one's own administrations that are scored by a reliable examiner for acceptable reliability. Training on the DISCO involves a 3-day introductory workshop, an additional 2-day workshop, and subsequently meeting quality standards for accreditation. 3Di training is completed through DVDs and internet-based training modules, although in-person training courses are also available.

Another drawback of commonly used parent interviews is the length of administration. For some samples and purposes, the SCQ may be a viable substitute for the ADI-R, in particular when used in combination with the ADOS (see Chandler et al., 2007; Corsello et al., 2007). Additionally, a shortened telephone screener, the Autism Screening Interview, has been created based in part on the ADI-R and currently is undergoing validation (S. Bishop and C. Lord, personal communication).

In general, an ASD evaluation should include, at minimum, a caregiver-based developmental history, a direct observation of the referred individual using a semi-structured observational measure, and measures of cognitive, language, and adaptive skill (see also Gotham, Bishop, & Lord, 2011, for a more detailed discussion of autism diagnostic assessment). Despite the strong predictive validity of some of the assessment tools described above, an individual's diagnosis of ASD should never depend on the diagnostic classification of a single measure or combination of measures. In addition, professionals must be realistic about the limitations of diagnostic instruments: ultimately these measures cannot "solve" a difficult diagnostic decision, and they may not be universally necessary (e.g., for clear-cut cases in which accurate diagnosis is the sole aim). Experienced clinical judgment is essential for accurate diagnosis, regardless of how carefully a clinician weighs the benefits and disadvantages of commonly used instruments. Nevertheless, choosing the best performing instruments for a particular clinical need can provide invaluable standardisation and structure to support clinical judgement and to aid in treatment planning and recommendations.

Conclusions

Over the past two decades both the research and clinical fields have benefited from the development of ASD screening and diagnostic instruments. Both provide valuable sources of information about a child or young person which can help clinicians make more informed

judgements about onward referral and diagnosis. However, they do not ‘do the job’ for the clinical team in that no instrument score equates to a diagnosis. In addition, screening and diagnostic instruments help clinicians least in the cases where they are most in want of direction since their accuracy will always be lower for marginal cases. Depending on service-configuration, timely and expensive ASD-specific diagnostic instruments will not always be feasible or appropriate, but in our experience, clinical teams benefit from practitioners being trained in these methods. Whilst further research hopefully will refine our knowledge of both screening and diagnostic instruments and where they can prove useful, to date many of them remain inadequately studied and their instrument parameters, whilst promising, insufficient to recommend them as universal clinical tools.

Acknowledgments

This review article was invited by CAMH Editors for publication in the Measurement Issues section of the journal.

The work of TC on screening is supported by the COST Action BM1004. Research at the Centre for Research in Autism and Education (CRAE) is supported by the Clothworkers’ Foundation and Pears Foundation. The work of K. on diagnostic instruments has been supported by United States National Institute of Health grants R01HD065277 (PI: Somer Bishop), R01 MH066469 (PI: Catherine Lord), and an Autism Speaks Predoctoral Training Fellowship.

References

- Al-Qabandi M, Gorter JW, Rosenbaum P. Early autism detection: are we ready for routine screening? *Pediatrics*. 2011; 128:e211–217. [PubMed: 21669896]
- American Academy of Pediatrics (Committee on Children with Disabilities). Developmental surveillance and screening of infants and young children. *Pediatrics*. 2001; 108:192–196. [PubMed: 11433077]
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4. Washington, DC: American Psychiatric Association; 2000. Text Revision (DSM-IV-TR)
- Baird G, Charman T, Baron-Cohen S, Cox A, Swettenham J, Wheelwright S, Drew A. A screening instrument for autism at 18 month of age: A six-year follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2000; 39:694–702. [PubMed: 10846303]
- Baird G, Simonoff E, Pickles A, Chandler S, Loucas T, Meldrum D, Charman T. Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet*. 2006; 368:210–215. [PubMed: 16844490]
- Baron-Cohen S, Allen J, Gillberg C. Can autism be detected at 18 months? The needle, the haystack and the CHAT. *British Journal of Psychiatry*. 1992; 138:839–843. [PubMed: 1483172]
- Berument SK, Rutter M, Lord C, Pickles A, Bailey A. Autism screening questionnaire: Diagnostic validity. *British Journal of Psychiatry*. 1999; 175:444–451. [PubMed: 10789276]
- Billstedt E, Gillberg C, Gillberg C. Autism in adults: Symptom patterns and early childhood predictors. Use of the DISCO in a community sample followed from childhood. *Journal of Child Psychology and Psychiatry*. 2007; 48(11):1102–1110. [PubMed: 17995486]
- Bishop DVM. Development of the Children’s Communication Checklist (CCC): A method for assessing qualitative aspects of communicative impairment in children. *Journal of Child Psychology and Psychiatry*. 1998; 39:879–891. [PubMed: 9758196]
- Canal-Bedia R, García-Primo P, Martín-Cilleros MV, Santos-Borbujo J, Guisuraga-Fernández Z, Herráez-García L, et al. Modified Checklist for Autism in Toddlers: Cross-Cultural Adaptation and Validation in Spain. *Journal of Autism and Developmental Disorders*. 2011 online early.
- Center for Disease Control. Prevalence of autism spectrum disorders - Autism and Developmental Disabilities Monitoring Network, United States, 2006. *MMWR Surveillance*. 2009; 18:1–20.
- Chandler S, Charman T, Baird G, Simonoff E, Loucas T, Meldrum D, Scott M, Pickles A. Validation of the Social Communication Questionnaire in a population cohort of children with autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2007; 46(10):1324–1332. [PubMed: 17885574]

- Charman T, Baird G. Practitioner review: Diagnosis of autism spectrum disorder in 2- and 3-year-old children. *Journal of Child Psychology and Psychiatry*. 2002; 43:289–305. [PubMed: 11944873]
- Charman T, Baird G, Simonoff E, Loucas T, Chandler S, Meldrum D, Pickles A. Efficacy of three screening instruments in the identification of autism spectrum disorder. *British Journal of Psychiatry*. 2007; 191:554–559. [PubMed: 18055961]
- Cicchetti DV, Lord C, Koenig K, Klin A, Volkmar FR. Reliability of the ADI-R: Multiple examiners evaluate a single case. *Journal of Autism and Developmental Disorders*. 2008; 38:764–770. [PubMed: 18058216]
- Clark A, Harrington R. On diagnosing rare disorders rarely: Appropriate use of screening instruments. *Journal of Child Psychology and Psychiatry*. 1999; 40:287–290. [PubMed: 10188711]
- Corsello C, Hus V, Pickles A, Risi S, Cook EH, Leventhal BL, 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; 48:932–940. [PubMed: 17714378]
- Constantino, JN.; Gruber, CP. *Social Responsiveness Scale (SRS)*. Los Angeles, LA: Western Psychological Services; 2005.
- Constantino JN, Lavesser PD, Zhang Y, Abbacchi AM, Gray T, Todd RD. Rapid quantitative assessment of autistic social impairment by classroom teachers. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2007; 46:1668–1676. [PubMed: 18030089]
- de Bildt A, Mulder EJ, Hoekstra PJ, van Lang ND, Minderaa RB, Hartman CA. Validity of the Children's Social Behaviour Questionnaire (CSBQ) in children with intellectual disability: comparing the CSBQ with ADI-R, ADOS, and clinical DSM-IV-TR classification. *Journal of Autism and Developmental Disorders*. 2009; 39(10):1464–70. [PubMed: 19495951]
- de Bildt A, Oosterling I, van Lang N, Sytema S, Minderaa R, van Engeland H, Roos S, Buitelaar JK, van der Gaag RJ, de Jonge M. Standardized ADOS scores: Measuring severity of autism spectrum disorders in a Dutch sample. *Journal of Autism and Developmental Disorders*. 2011; 41:311–319. [PubMed: 20617374]
- de Bildt A, Sytema S, Ketelaars C, Kraijer D, Mulder E, Volkmar F, Minderaa R. Interrelationship between Autism Diagnostic Observation Schedule-Generic (ADOS-G), Autism Diagnostic Interview-Revised (ADI-R), and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders*. 2004; 34:129–137. [PubMed: 15162932]
- Dereu M, Warreyn P, Raymaekers R, Meirsschaut M, Pattyn G, Schietecatte I, Roeyers H. Screening for Autism Spectrum Disorders in Flemish Day-Care Centres with the Checklist for Early Signs of Developmental Disorders. *Journal of Autism and Developmental Disorders*. 2010; 40:1247–1258. [PubMed: 20198413]
- Dietz C, Willemsen-Swinkels SHN, van Daalen E, van Engeland H, Buitelaar JK. Screening for Autistic Spectrum Disorder in children aged 14 to 15 Months. II: Population screening with the Early Screening of Autistic Traits (ESAT). Design and general findings. *Journal of Autism and Developmental Disorders*. 2006; 36:713–722. [PubMed: 16633887]
- Eaves LC, Wingert HD, Ho HH, Mickelson ECR. Screening for autism spectrum disorders with the social communication questionnaire. *Journal of Developmental and Behavioural Pediatrics*. 2006a; 27:S95–S103.
- Eaves LC, Wingert H, Ho HH. Screening for autism - Agreement with diagnosis. *Autism*. 2006b; 10:229–242. [PubMed: 16682396]
- Ehlers S, Gillberg C. The epidemiology of Asperger syndrome: A total population study. *Journal of Child Psychology and Psychiatry*. 1993; 34:1327–1350. [PubMed: 8294522]
- Ehlers S, Gillberg C, Wing L. A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children. *Journal of Autism and Developmental Disorders*. 1999; 29:129–141. [PubMed: 10382133]
- Geschwind DH, Levitt P. Autism spectrum disorders: developmental disconnection syndromes. *Current Opinion in Neurobiology*. 2007; 17:103–111. [PubMed: 17275283]
- Glascocoe, FP. *Developmental screening*. In: Wolraich, M., editor. *Disorders of development and learning: A practical guide to 20*. 2. St. Lois, MO: Mosby; 1996.

- Glascie FP. The value of parents' concerns to detect and address developmental and behavioural problems. *Journal of Paediatrics and Child Health*. 1999; 35:1–8.
- Gotham, K.; Bishop, SL.; Lord, C. Diagnosis of autism spectrum disorders. In: Amaral, DG.; Dawson, G.; Geschwind, DH., editors. *Autism Spectrum Disorders*. New York: Oxford University Press; 2011. p. 30–43.
- Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2009; 39(5):693. [PubMed: 19082876]
- Gotham K, Risi S, Dawson G, Tager-Flusberg H, Joseph R, Carter A, Lord C. A replication of the Autism Diagnostic Observation Schedule (ADOS) revised algorithms. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2008; 47(6):643–651.
- Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule (ADOS): Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*. 2007; 37:400–408.
- Gray KM, Tonge BJ, Sweeney DJ. Using the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule with young children with developmental delay: Evaluating diagnostic validity. *Journal of Autism and Developmental Disorders*. 2008; 38:657–667. [PubMed: 17690967]
- Hus, V.; Lord, C. The ADOS Adapted Modules: An autism diagnostic observation measure for adolescents and adults with limited language. in preparation.
- Johnson CP, Myers SM. 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, et al. 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]
- Kim S, Lord C. New Autism Diagnostic Interview-Revised (ADI-R) algorithms for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Autism and Developmental Disorders*. (in press).
- Leekam SR, Libby SJ, Wing L, Gould J, Gillberg C. Comparison of ICD-10 and Gillberg's criteria for Asperger syndrome. *Autism*. 2000; 4:11–28.
- Leekam SR, Libby SJ, Wing L, Gould J, Taylor C. The Diagnostic Interview for Social and Communication Disorders: Algorithms for ICD-10 childhood autism and Wing and Gould autistic spectrum disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. 2002; 43(3):327–342.
- Levy SE, Mandell DS, Schultz R. Autism. *Lancet*. 2009; 374:1627–1638. [PubMed: 19819542]
- Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, 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.; DiLavore, P.; Risi, S. *Autism Diagnostic Observation Schedule: Manual*. Los Angeles: Western Psychological Services; 1999.
- Lord, C.; Rutter, M.; DiLavore, P.; Risi, S.; Gotham, K.; Luyster, R.; Bishop, S. *Autism Diagnostic Observation Schedule. 2*. Los Angeles: Western Psychological Services; (in press)(ADOS-2): Manual
- Lord C, Rutter ML, LeCouteur A. The Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*. 1994; 24:659–685. [PubMed: 7814313]
- Lord C, Storoschuk S, Rutter M, Pickles A. Using the ADI-R to diagnose autism in preschool children. *Infant Mental Health Journal*. 1993; 14:1234–1252.
- Luyster R, Gotham K, Guthrie W, Coffing M, Petrak R, Pierce K, et al. The Autism Diagnostic Observation Schedule—Toddler Module: A new module of a standardized diagnostic measure for ASD. *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]
- Manning SE, Davin CA, Barfield WD, Kotelchuck M, Clements K, et al. Early Diagnoses of Autism Spectrum Disorders in Massachusetts Birth Cohorts, 2001–2005. *Pediatrics*. 2011; 127:1043–1051. [PubMed: 21576313]
- 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]
- National Institute of Clinical Excellence. 2011. <http://guidance.nice.org.uk/CG/Wave15/78/PrepublicationCheck>
- Nordin V, Gillberg C. The long-term course of autistic disorders: Update on follow-up studies. *Acta Psychiatrica Scandinavica*. 1998; 97:99–108. [PubMed: 9517902]
- Norris M, Lecavalier L. Screening Accuracy of Level 2 Autism Spectrum Disorder Rating Scales A Review of Selected Instruments. *Autism*. 2010; 14:263–284. [PubMed: 20591956]
- Nygren G, Hagberg B, Billstedt E, Skoglund A, Gillberg C, Johansson M. The Swedish version of the Diagnostic Interview for Social and Communication Disorders (DISCO-10): Psychometric properties. *Journal of Autism and Developmental Disorders*. 2009; 39:730–741. [PubMed: 19148741]
- Oosterling IJ, Swinkels SH, van der Gaag RJ, Visser JC, Dietz C, Buitelaar JK. Comparative Analysis of Three Screening Instruments for Autism Spectrum Disorder in Toddlers at High Risk. *Journal of Autism and Developmental Disorders*. 2009; 39:897–909. [PubMed: 19205862]
- Oosterling IJ, Wensing M, Swinkels SH, van der Gaag RJ, Visser JC, et al. Advancing early detection of autism spectrum disorder by applying an integrated two-stage screening approach. *Journal of Child Psychology and Psychiatry*. 2010a; 51:250–258. [PubMed: 19843319]
- Oosterling I, Roos S, de Bildt A, Rommelse N, de Jonge M, et al. Improved diagnostic validity of the ADOS revised algorithms: A replication study in an independent sample. *Journal of Autism and Developmental Disorders*. 2010b; 40:689–703. [PubMed: 20148299]
- Oosterling I, Rommelse N, de Jonge M, van der Gaag RJ, Swinkels S, Roos S, Visser J, Buitelaar J. How useful is the Social Communication Questionnaire in toddlers at risk of autism spectrum disorder? *Journal of Child Psychology and Psychiatry*. 2010c; 51:1260–1268. [PubMed: 20626528]
- Posserud MB, Lundervold AJ, Gillberg C. Autistic features in a total population of 7–9-year-old children assessed by the ASSQ (Autism Spectrum Screening Questionnaire). *Journal of Child Psychology and Psychiatry*. 2006; 47:167–175. [PubMed: 16423148]
- Posserud MB, Lundervold AJ, Gillberg C. Validation of the Autism Spectrum Screening Questionnaire in a Total Population Sample. *Journal of Autism and Developmental Disorders*. 2009; 39:126–134. [PubMed: 18592364]
- Risi S, Lord C, Gotham K, Corsello C, Szatmari P, Cook E, 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]
- Robins DL. Screening for autism spectrum disorders in primary care settings. *Autism*. 2008; 12:537–556. [PubMed: 18805946]
- 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.; LeCouteur, A.; Lord, C. *The Autism Diagnostic Interview-Revised (ADI-R)*. Los Angeles, CA: Western Psychological Services; 2003.
- Rutter, M.; Bailey, A.; Lord, C. *Social Communication Questionnaire (SCQ)*. Los Angeles, LA: Western Psychological Services; 2003.
- Santosh PJ, Mandy WP, Puura K, Kaartinen M, Warrington R, Skuse DH. The construction and validation of a short form of the developmental, diagnostic and dimensional interview. *European Child and Adolescent Psychiatry*. 2009; 18:521–524. [PubMed: 19263189]
- Skuse D, Warrington R, Bishop D, Chowdhury U, Lau J, Mandy W, Place M. The Developmental, Dimensional, and Diagnostic Interview (3di): A novel computerized assessment for autism

- spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2004; 43(5):548–558. [PubMed: 15100561]
- Wetherby AM, 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; 12:487–511. [PubMed: 18805944]
- Wetherby AM, 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]
- Williams J, Scott F, Stott C, Allison C, Bolton P, et al. The CAST (Childhood Asperger and Syndrome Test) - Test accuracy. *Autism*. 2005; 9:45–68. [PubMed: 15618262]
- Williams JG, Allison C, Scott FJ, Bolton PF, Baron-Cohen S, et al. The Childhood Autism Spectrum Test (CAST): Sex differences. *Journal of Autism and Developmental Disorders*. 2008; 38:1731–1739. [PubMed: 18408991]
- Wing L, Gould J. Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*. 1979; 9(1):11–29. [PubMed: 155684]
- Wing L, Leekam SR, Libby SJ, Gould J, Locombe M. The Diagnostic Interview for Social and Communication Disorders: Background, inter-rater reliability and clinical use. *Journal of Child Psychology and Psychiatry*. 2002; 43:307–325. [PubMed: 11944874]
- World Health Organisation. *Mental Disorders: A Glossary and Guide to their Classification in Accordance with the 10th Revision of the International Classification of Diseases: Research Diagnostic Criteria (ICD-10)*. Geneva: WHO; 1993.

Key Practitioner Message

- Over the past two decades screening instruments (many) and diagnostic measures (few) for ASD have been developed though few are well-evaluated.
- Extension of research findings to clinical services must be done with caution, understanding that instrument properties are sample-specific.
- Screening and diagnostic instruments help clinicians least in the cases where they are most in want of direction since their accuracy will always be lower for marginal cases.
- Timely and expensive ASD-specific diagnostic instruments will not always be feasible or appropriate, but clinical teams benefit from practitioners being trained in these methods.

Table 1

Prospective screening studies

Reference	Instrument	Age (Months)	Study design	Administration	Instrument parameters
Baird et al. (2000)	CHAT	18m	2-stage population screening (N = 16,235)	Parental questionnaire and health practitioner observation	Se = 0.21 (2-stage) / PPV = 0.59 (2-stage)
Dietz et al. (2006)	ESAT	14m	2-stage population screening (N = 31,724)	Health practitioner completes items following interview with parents	PPV ASD = 0.25 (2-stage) PPV DD = 0.68 (2-stage)
Robins et al. (2001)	M-CHAT	24m	2-stage combined high-risk and low-risk (N = 1,293)	Parent questionnaire	PPV = 0.36 (1-stage) PPV = 0.68 (2-stage)
Kleinman et al. (2008)	M-CHAT	16–30m	2-stage combined high-risk and low-risk (N = 3,793)	Parent questionnaire	PPV = 0.36 (1-stage) PPV = 0.74 (2-stage)
Robins (2008)	M-CHAT	16–27m	2-stage low-risk (N = 4,797)	Parent questionnaire	PPV = 0.06 (1-stage) PPV = 0.57 (2-stage)
Wetherby et al. (2004, 2008)	ITC	6–24m	Population screening (n = 5,385); multiple stage follow-ups	Parent questionnaire	PPV DD = 0.43 (6–8m) PPV DD = 0.79 (21–24m)

¹ Date reported for the 2-stage, high and medium risk combined

CHAT (Checklist for Autism in Toddlers); ESAT (Early Screening of Autistic Traits); M-CHAT (Modified- Checklist for Autism in Toddlers)

ITC (Infant Toddler Checklist); Se: sensitivity; PPV: positive predictive value; ASD: autism spectrum disorder; DD: developmental disorder

Table 2

Studies reporting on the Social Communication Questionnaire (SCQ) and the Social Responsiveness Scale (SRS) – ASD vs. non-spectrum

Reference	Age (Years)	Sample	Sensitivity	Specificity	PPV
SCQ					
Berument et al. (1999)	4–40 yrs	160 ASD 40 non-ASD	0.85	0.75	0.93
Chandler et al. (2007)	9–10 yrs	158 ASD 97 non-ASD	0.88	0.72	0.64
Corsello et al. (2007)	2–16 yrs	439 ASD 151 non-ASD	0.71	0.71	0.88
Eaves et al. (2006a)	3–7 yrs	49 ASD 102 non-ASD	0.71	0.79	0.65
Eaves et al. (2006b)	4–6 yrs	35 ASD 59 non-ASD	0.74	0.54	0.65
SRS					
Charman et al. (2007)	9–13 yrs	70 ASD 49 non-ASD	0.78	0.67	0.63
Constantino et al. (2007)	4–18 yrs	271 ASD 171 non-ASD	0.75	0.96	NR

NR: Not reported; Se = sensitivity; PPV : positive predictive value

Table 3

Overview of autism diagnostic instruments

	Intended Patient	Administration	Content	Inter-rater Reliability	Predictive Validity
ADOS	Ages 12 m and up; any verbal level; $nvma \geq 12$ ¹	Examiner-Patient Interactive Observation (30–60 min)	Social interaction, Communication, Play & Imagination, Repetitive Behaviors	Toddler Module: $MK_w = .76$ for 30/38 items; mean exact agreement = 84%. ¹ See also Lord et al., in press, for revised algorithm reliability.	AUT vs. NS: Se=91–98, Sp=84–94 (N=1072) ² ; Se=82–94, Sp=80–100 (N=949) ³ ; Se=87–92, Sp=71–76 (N=337) ⁴ <i>Non-Autism ASD vs. NS</i> ; Se=72–84, Sp=76–83 (N=649) ² ; Se=60–95, Sp=75–100 (N=238) ³ ; Se=53–86, Sp=62–63 (N=354) ⁴ <i>Toddlers</i> : Se=88–91, Sp=91 (n=104, ASD vs. NS) ¹ See Gray et al. (2008) & Oosterling et al. (2010b) for more data from Modules 1 & 2, and Lord et al. (1999) for data on Module 4.
ADI-R	Ages 12 m and up; ⁵ any verbal level; $nvma \geq 10$ ⁵	Examiner-Parent/Caregiver Interview (90–180 min)	Social interaction, Communication, Play & Imagination, Repetitive Behaviors, Developmental Milestones, Problem Behaviors & Special Skills	$K_w = .73$ –.78, Mean Agreement = 90–93% on all items (n=10 AUT, 10 NS) ⁶ $K_w = .62$ –.89 (preschoolers; n=51 ASD, 43 NS) ⁷	Se=89, Sp=59, PPV=74, NPV=81 (N=960, ASD vs. NS) ⁸ Se=83, Sp=72, PPV=82, NPV=74 (N=270 under age 3, ASD vs. NS) ⁸ Se=72; Sp=79 (N=137, AUT+ID vs. NS +ID); Se=77, Sp=63 (N=136, non-aut ASD +ID vs. NS+ID) ⁹
DISCO	All ages, verbal levels, and mental ages	Examiner-Parent/Caregiver Interview (+ Outside Information) (120–240 min)	Social interaction, Communication, Imagination, Repetitive Behaviors, Developmental Level & Daily Living Skills, Problem Behaviors	$K = .75$ + for 80% of all items, ^{10,11} $K = .75$ + for 50–68% of items across algorithms. ¹²	Se=100, Sp=55 (N=67, ASD vs. NS), ¹¹ Se=82, Sp=83 (N=57; ASD vs. NS) ¹²
3Di	Children with normal-range abilities, unselected clinical and general population samples	Examiner-Parent/Caregiver Interview (Computerized; 45–180 min, depending on optional module inclusion)	Social interaction, Communication, Repetitive Behaviors, Demographics, Developmental History, Motor skills; Optional: Comorbid symptoms	ICC = .86+ (n=23 ASD, 27 NS) ¹³	Correctly identified 100% of 27 children with ASD and 98% of 60 typical children and 33 with unknown diagnoses. ¹³

Note: *nvma*: nonverbal mental age; MK_w : mean weighted kappa; K_w : weighted kappa; K: kappa; Se: sensitivity; Sp: specificity; ID: intellectual disability.

¹ Based on the Toddler module (see Luyster et al., 2009);

² Gotham et al., 2007;

³ Gotham et al., 2008;

⁴ de Bildt et al., 2009;

- ⁵Based on the Toddler algorithms (see Kim & Lord, in press);
- ⁶Lord et al., 1993;
- ⁷Lord et al., 1994;
- ⁸Risi et al., 2006;
- ⁹de Bildt et al., 2004;
- ¹⁰Wing et al., 2002;
- ¹¹Leekam et al., 2002;
- ¹²Nygren et al., 2009;
- ³Skuse et al., 2004