Cochrane Library

Diagnostic tests for autism spectrum disorder (ASD) in preschool children (Review)

Randall M, Egberts KJ, Samtani A, Scholten RJPM, Hooft L, Livingstone N, Sterling-Levis K, Woolfenden S, Williams K

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Diagnostic tests for autism spectrum disorder (ASD) in preschool children

Melinda Randall[1,2], Kristine J Egberts[2], Aarti Samtani[3], Rob JPM Scholten[4], Lotty Hooft[4], Nuala Livingstone[5], Katy Sterling-Levis[6a], Susan Woolfenden[7,8], Katrina Williams[2,9,10]

[1]Department of Occupational Therapy, The Royal Children's Hospital, Parkville, Australia. [2]Department of Paediatrics, The University of Melbourne, Parkville, Australia. [3]UNSW Medicine, UNSW Sydney, Kensington, Australia. [4]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care / University Medical Center Utrecht, Utrecht, Netherlands. [5]Editorial & Methods Department, Cochrane, London, UK. [6]School of Women's and Children's Health, University of New South Wales & Sydney Children's Hospital, Randwick, Australia. [7]Sydney Children's Hospital Network, Randwick, Australia. [8]School of Women's & Children's Health, UNSW Medicine, UNSW Sydney, Kensington, Australia. [9]Department of Neurodevelopment and Disability, The Royal Children's Hospital, Parkville, Australia. [10]Murdoch Children's Research Institute, Parkville, Australia

[a]Deceased

Contact address: Katrina Williams, Department of Paediatrics, The University of Melbourne, Parkville, Australia. katrina.williams@rch.org.au.

A B S T R A C T

Background

Autism spectrum disorder (ASD) is a behaviourally diagnosed condition. It is defined by impairments in social communication or the presence of restricted or repetitive behaviours, or both. Diagnosis is made according to existing classification systems. In recent years, especially following publication of the *Diagnostic and Statistical Manual of Mental Disorders - Fifth Edition* (DSM-5; APA 2013), children are given the diagnosis of ASD, rather than subclassifications of the spectrum such as autistic disorder, Asperger syndrome, or pervasive developmental disorder - not otherwise specified. Tests to diagnose ASD have been developed using parent or carer interview, child observation, or a combination of both.

Objectives

Primary objectives

1. To identify which diagnostic tools, including updated versions, most accurately diagnose ASD in preschool children when compared with multi-disciplinary team clinical judgement.

2. To identify how the best of the interview tools compare with CARS, then how CARS compares with ADOS.

a. Which ASD diagnostic tool - among ADOS, ADI-R, CARS, DISCO, GARS, and 3di - has the best diagnostic test accuracy?

b. Is the diagnostic test accuracy of any one test sufficient for that test to be suitable as a sole assessment tool for preschool children?

c. Is there any combination of tests that, if offered in sequence, would provide suitable diagnostic test accuracy and enhance test efficiency?

d. If data are available, does the combination of an interview tool with a structured observation test have better diagnostic test accuracy (i.e. fewer false-positives and fewer false-negatives) than either test alone?

As only one interview tool was identified, we modified the first three aims to a single aim (Differences between protocol and review): This Review evaluated diagnostic tests in terms of sensitivity and specificity. Specificity is the most important factor for diagnosis; however, both sensitivity and specificity are of interest in this Review because there is an inherent trade-off between these two factors.

Secondary objectives

1. To determine whether any diagnostic test has greater diagnostic test accuracy for age-specific subgroups within the preschool age range.

Search methods

In July 2016, we searched CENTRAL, MEDLINE, Embase, PsycINFO, 10 other databases, and the reference lists of all included publications.

Selection criteria

Publications had to:
1. report diagnostic test accuracy for any of the following six included diagnostic tools: Autism Diagnostic Interview - Revised (ADI-R), Gilliam Autism Rating Scale (GARS), Diagnostic Interview for Social and Communication Disorder (DISCO), Developmental, Dimensional, and Diagnostic Interview (3di), Autism Diagnostic Observation Schedule - Generic (ADOS), and Childhood Autism Rating Scale (CARS);
2. include children of preschool age (under six years of age) suspected of having an ASD; and
3. have a multi-disciplinary assessment, or similar, as the reference standard.

Eligible studies included cohort, cross-sectional, randomised test accuracy, and case-control studies. The target condition was ASD.

Data collection and analysis

Two review authors independently assessed all studies for inclusion and extracted data using standardised forms. A third review author settled disagreements. We assessed methodological quality using the QUADAS-2 instrument (Quality Assessment of Studies of Diagnostic Accuracy - Revised). We conducted separate univariate random-effects logistical regressions for sensitivity and specificity for CARS and ADI-R. We conducted meta-analyses of pairs of sensitivity and specificity using bivariate random-effects methods for ADOS.

Main results

In this Review, we included 21 sets of analyses reporting different tools or cohorts of children from 13 publications, many with high risk of bias or potential conflicts of interest or a combination of both. Overall, the prevalence of ASD for children in the included analyses was 74%.

For versions and modules of ADOS, there were 12 analyses with 1625 children. Sensitivity of ADOS ranged from 0.76 to 0.98, and specificity ranged from 0.20 to 1.00. The summary sensitivity was 0.94 (95% confidence interval (CI) 0.89 to 0.97), and the summary specificity was 0.80 (95% CI 0.68 to 0.88).

For CARS, there were four analyses with 641 children. Sensitivity of CARS ranged from 0.66 to 0.89, and specificity ranged from 0.21 to 1.00. The summary sensitivity for CARS was 0.80 (95% CI 0.61 to 0.91), and the summary specificity was 0.88 (95% CI 0.64 to 0.96).

For ADI-R, there were five analyses with 634 children. Sensitivity for ADI-R ranged from 0.19 to 0.75, and specificity ranged from 0.63 to 1.00. The summary sensitivity for the ADI-R was 0.52 (95% CI 0.32 to 0.71), and the summary specificity was 0.84 (95% CI 0.61 to 0.95).

Studies that compared tests were few and too small to allow clear conclusions.

In two studies that included analyses for both ADI-R and ADOS, tests scored similarly for sensitivity, but ADOS scored higher for specificity. In two studies that included analyses for ADI-R, ADOS, and CARS, ADOS had the highest sensitivity and CARS the highest specificity.

In one study that explored individual and additive sensitivity and specificity of ADOS and ADI-R, combining the two tests did not increase the sensitivity nor the specificity of ADOS used alone.

Performance for all tests was lower when we excluded studies at high risk of bias.

Authors' conclusions

We observed substantial variation in sensitivity and specificity of all tests, which was likely attributable to methodological differences and variations in the clinical characteristics of populations recruited.

When we compared summary statistics for ADOS, CARS, and ADI-R, we found that ADOS was most sensitive. All tools performed similarly for specificity. In lower prevalence populations, the risk of falsely identifying children who do not have ASD would be higher.

Now available are new versions of tools that require diagnostic test accuracy assessment, ideally in clinically relevant situations, with methods at low risk of bias and in children of varying abilities.

PLAIN LANGUAGE SUMMARY

How accurate are diagnostic tools for autism spectrum disorder in preschool children?

Review question

How accurate are tools for diagnosing autism spectrum disorder (ASD) in preschool children?

Why is accurate ASD diagnosis important?

Not diagnosing ASD in children when it is present (false-negative result) means children with ASD may miss receiving early intervention and families may miss receiving timely support and education. An incorrect diagnosis of ASD (false-positive result) may cause family stress, lead to unnecessary investigations and treatments, and place greater strain on already limited service resources.

What is the aim of this Review?

To find out which of the commonly used tools is most accurate for diagnosing ASD in preschool children. Cochrane researchers reviewed 13 published articles to answer this question.

What was studied in the Review?

Six tests were reviewed: Four gathered information about children's behaviours from interviews with parents or carers (Autism Diagnostic Interview-Revised (ADI-R), Gilliam Autism Rating Scale (GARS), Diagnostic Interview for Social and Communication Disorder (DISCO), and Developmental, Dimensional, and Diagnostic Interview (3di)); one required that a trained professional observe a child's behaviour on specific tasks (Autism Diagnostic Observation Schedule (ADOS)); and one combined observation of the child with interview of parents or carers (Childhood Autism Rating Scale (CARS)).

What are the main results of the Review?

The Review included 21 relevant sets of analyses conducted on a total of 2900 children. Results were available for only three tools: ADOS (Modules 1 and 2), CARS, and ADI-R. If instruments were applied to 1000 children, 740 of whom had ASD, then 696, 592, and 385 children would be correctly identified by ADOS, CARS, and ADI-R, respectively, whereas 52, 31, and 42 children without ASD would be incorrectly classified as having ASD. Of 260 children without ASD, 208, 229, and 218 would be correctly classified by ADOS, CARS, and ADI-R, respectively, whereas 44, 148, and 355 children with ASD would be incorrectly classified as not having ASD.

See Figure 1.

One publication looked at using ADI-R together with ADOS and found that use of both tools together was no more accurate than use of ADOS alone.

How reliable are the results of analyses in this Review?

Using a variety of best-estimate clinical approaches led to diagnosis in children. This method is commonly used in research but does not always replicate the multi-disciplinary assessment recommended for clinical diagnosis.

Problems with how some studies were conducted and the presence of conflicts of interest in some publications may result in ADOS, CARS, and ADI-R appearing more accurate than they really are. Also, if these tools are used in populations with a lower prevalence of ASD, a higher proportion of children who do not have ASD are likely to receive an ASD diagnosis.

The numbers shown above represent average values across analyses. However, as individual estimates varied, we cannot be sure that ADOS will always produce these results. Numbers of children included in studies conducted to date, including studies comparing the accuracy of different tools, are insufficient to evoke confidence in these results.

Who do results of the Review apply to?

Studies included were carried out in Australia, Canada, India, the Netherlands, United Kingdom, and United States. Studies included children younger than six years of age, or children with a mean age less than six years, with language difficulties, developmental delay, intellectual disability, or a mental health problem, presenting to a clinical service or enrolling in a research study.

What are the implications of this Review?

Current findings suggest that ADOS is best for not missing children who have ASD and is similar to CARS and ADI-R in not falsely diagnosing ASD in a child who does not have ASD. ADOS has acceptable accuracy in populations with a high prevalence of ASD. However, overdiagnosis is likely if the tool is used in populations with a lower prevalence of ASD. This finding supports current recommended practice for ASD diagnostic tools to be used as part of a multi-disciplinary assessment, rather than as stand-alone diagnostic instruments.

How up-to-date is this Review?

This Review was up-to-date as of July 2016.

SUMMARY OF FINDINGS

Summary of findings 1.   Diagnostic accuracy of Autism Diagnostic Observation Schedule (ADOS), Childhood Autism Rating Scale (CARS), and Autism Diagnosis Interview - Revised (ADI-R) for diagnosing autism spectrum disorder in preschool children

Should ADOS, CARS, or ADI-R be used to diagnose ASD in children younger than 6 years of age?

Participants: children younger than 6 years of age

Settings: Included studies involved children from the following range of settings: hospitals and university-based clinics screening for early diagnosis of ASD; hospital-based developmental evaluation clinics; research studies; university-based child psychiatry centres (median prevalence of ASD across all studies: 74%)
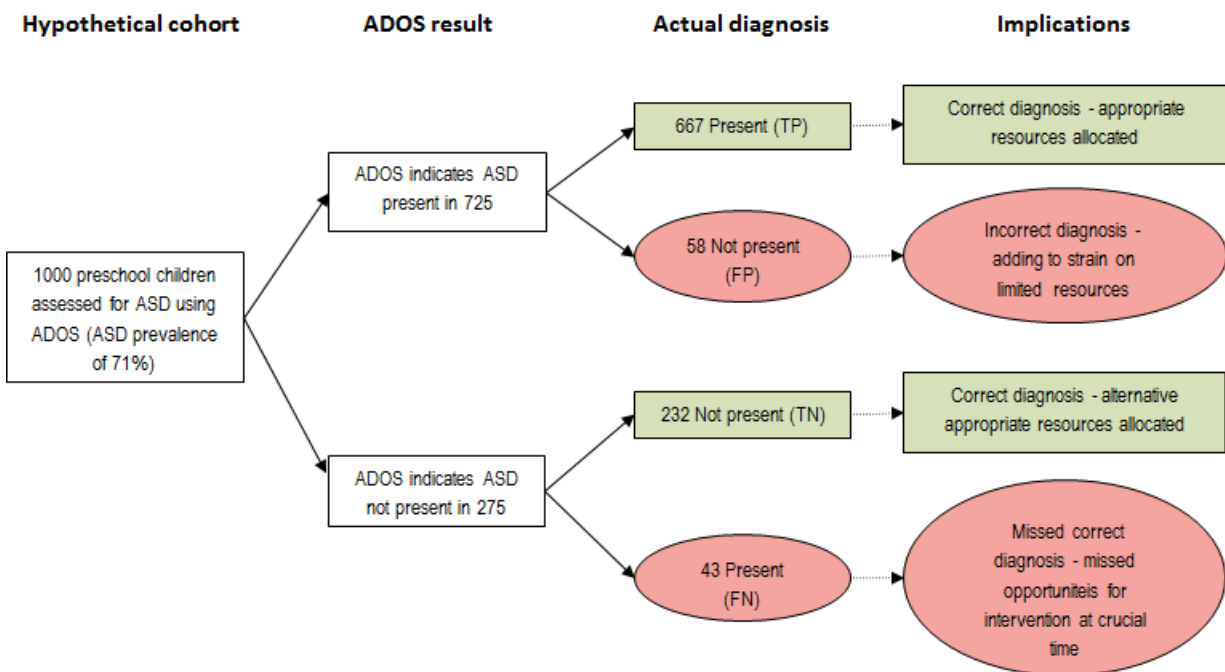
Reference standards: Assessments were administered by 1 or more professionals trained in tool administration. Best-estimate clinical diagnosis was made after review of all assessment results by 1 or more professionals experienced in the diagnosis of ASD

Study designs: cross-sectional or case-control studies

Test	Number of studies (number of participants)	Risk of bias (number of studies)	Combined sensitivity (95% CI) / Range of sensitivities	Combined specificity (95% CI) / Range of specificities	Number of true-positives per 1000 tested (95% CI)	Number of false-positives per 1000 tested (95% CI)	Number of true-negatives per 1000 tested (95% CI)	Number of false-negatives per 1000 tested (95% CI)	Interpretation in 1000 children tested with a pre-test probability of ASD of 74%
ADOS	12 (1625)	Low (0) High (8) Unclear (4)	**0.94** (0.89 to 0.97) Range = 0.76 to 0.98	**0.80** (0.68 to 0.88) Range = 0.20 to 1.00	**696** (659 to 718)	**52** (31 to 83)	**208** (177 to 229)	**44** (22 to 88)	The diagnosis will be missed in 44 children with ASD, and 52 children without ASD will be incorrectly classified as having ASD. See Figure 1
CARS	4 (641)	Low (1) High (2) Unclear (1)	**0.80** (0.61 to 0.91) Range = 0.66 to 0.89	**0.88** (0.64 to 0.96) Range = 0.21 to 1.00	**592** (451 to 673)	**31** (10 to 94)	**229** (166 to 250)	**148** (67 to 289)	The diagnosis will be missed in 148 children with ASD, and 31 children without ASD will be incorrectly classified as having ASD
ADI-R	5 (634)	High (4) Unclear (1)	**0.52** (0.32 to 0.71) Range = 0.19 to 0.75	**0.84** (0.61 to 0.95) Range = 0.63 to 1.00	**385** (237 to 525)	**42** (13 to 101)	**218** (159 to 247)	**355** (215 to 503)	The diagnosis will be missed in 355 children with ASD, and 42 children without ASD will be incorrectly classified as having ASD

ADI-R: Autism Diagnostic Interview - Revised; **ADOS:** Autism Diagnostic Observation Schedule; **ASD:** autism spectrum disorder; **CARS:** Childhood Autism Rating Scale; **CI:** confidence interval.

Figure 1. Clinical pathway.

Figure 1. Clinical pathway.

Hypothetical cohort | ADOS result | Actual diagnosis | Implications

1000 preschool children assessed for ASD using ADOS (ASD prevalence of 71%)

ADOS indicates ASD present in 725

ADOS indicates ASD not present in 275

667 Present (TP) → Correct diagnosis - appropriate resources allocated

58 Not present (FP) → Incorrect diagnosis - adding to strain on limited resources

232 Not present (TN) → Correct diagnosis - alternative appropriate resources allocated

43 Present (FN) → Missed correct diagnosis - missed opportuniteis for intervention at crucial time

ADOS: Autism Diagnostic Observation Schedule
ASD: autism spectrum disorder
TP: true positive – test is positive (indicates ASD) and patient has ASD
FP: false positive – test is positive (indicates ASD) but patient does not have ASD
TN: true negative – test is negative (indicates ASD not present) and patient does not have ASD
FN: false negative – test is negative (indicates ASD not present) but patient has ASD

BACKGROUND

Autism is a behaviourally diagnosed condition. For this diagnosis, criteria of currently accepted classification systems must be fulfilled. Recommended diagnostic evaluation includes assessment of social behaviour, language and non-verbal communication, adaptive behaviour, atypical behaviours, and cognitive status by an experienced multi-disciplinary team (Akshoomoff 2006). With regard to specific diagnostic information, it is recommended that the diagnostic process should include information from parents/carers and child observation and interaction, along with use of clinical judgement (Missouri Autism Guidelines Initiative 2010; SIGN 2007; Zwaigenbaum 2009), permitting exclusion of other diagnoses that could present in a similar way. Current diagnostic criteria in the *Diagnostic and Statistical Manual of Mental Disorders - Fifth Edition* (DSM-5) also require consistency of atypical behaviours in more than one setting (APA 2013).

Target condition being diagnosed

Autism spectrum disorder (ASD) became an official diagnostic classification with the launch of DSM-5 in 2013 (APA 2013). Although the term 'ASD' was in common usage over a decade ago, before publication of DSM-5 separate diagnostic classifications of 'childhood autism' or 'autistic disorder', 'pervasive developmental disorder - not otherwise specified' (PDD-NOS), 'other pervasive developmental disorders', 'pervasive developmental disorder, unspecified', 'Asperger syndrome' or 'Asperger disorder', and 'atypical autism' were the official possible diagnoses as defined in *DSM - Fourth Edition* (DSM-IV; APA 1994), *DSM-IV - Text Revision* (DSM-IV-TR; APA 2000), and the *International Classification of Diseases and Related Health Problems - Tenth Revision* (ICD-10; WHO 2007). For these diagnoses, impairment has been judged in three core domains - (1) communication, (2) social interaction, and (3) presence of restricted, repetitive behaviours and interests - rather than the two now used in DSM-5: (1) social communication and (2) restricted, repetitive behaviours and interests. Inconsistent use of ASD-related diagnostic classification terms has caused confusion in clinical care and service access and has complicated both the conduct of research studies and the application of research findings.

Estimates of the incidence of ASD vary (Atladottir 2015; Elsabbagh 2012; Williams 2013). In the United States, the prevalence of ASD is reported as 1 in 68 children (CDP 2016). Males are affected about four times more frequently than females (Fombonne 2009; Watkins 2014). Problems usually present in early childhood and continue throughout life. Follow-up studies have found that only 3% to 27% of people with ASD are able to live independently as adults, with variations for different diagnostic groups within the autism spectrum (Cederlund 2008; Howlin 2004). As the prevalence of ASD is growing, services are receiving increasing referrals to decide whether ASD is the appropriate diagnosis. A recent study from a regional ASD diagnostic clinic in the United States reported that 39% of children referred for ASD diagnostic assessments were not given a diagnosis of ASD following assessment (Monteiro 2015). This points to the need for accurate and appropriate assessment methods, so that a limited resource for comprehensive neurodevelopmental assessment is used most appropriately.

The reference (also known as gold) standard assessment for diagnosis involves multiple professionals and multiple assessment mechanisms, is time intensive, and requires clinical judgement. Clinical experience suggests that there would not be complete agreement between teams, and that agreement would be highest for autistic disorder or childhood autism diagnoses and lowest for diagnoses of atypical autism and PDD-NOS. We have found no published studies that compared reference standard assessments made by different multi-disciplinary teams. Emerging evidence suggests that there is low agreement between individual clinician and transdisciplinary team diagnoses (Stewart 2014), with both underdiagnosis and overdiagnosis of ASD. Nevertheless, multi-disciplinary team assessment is accepted as best practice for diagnosis of all developmental disabilities; therefore, these services are provided in many countries (Academy of Medicine Singapore 2010; Filipek 2000; Ministry of Health New Zealand 2008; SIGN 2007).

Accurate diagnosis is a critical first step in deciding which further assessments or medical investigations are needed (NICE 2011; Volkmar 2014), what interventions are likely to be needed and likely to be effective (AHRQ 2011; NICE 2013), and what services may be required in future years. It is also a critical first step for parents to gain an understanding of their child and what lies ahead and to enable them to make decisions and plan for the future (Filipek 1999).

Index test(s)

A variety of tests are used in both research and clinical settings for diagnosis of ASD. Some rely on parent or carer report, and others use observation and interview. Many of these tests are used to standardise aspects of history-taking and physical examination; others are used to reduce the length of diagnostic interviews and to reduce costs, especially in research studies. Most include additive scales and subscales and rely on diagnostic cutoffs, which have been based on the classification systems in use at the time of their development. Given the varying rates of developmental spurts in children aged from birth to three years compared with those aged from three to six years, the utility of these various diagnostic tests is likely to change with different ability levels, as well as with chronological age (Matson 2008).

Authors of this Review assessed the six diagnostic tests recommended in national guidelines, published from 1995 up to the time this Review commenced (Table 1). Since publication of the protocol for this Review (Samtani 2011), revised versions of four of these tests have been developed and published (Autism Diagnostic Observation Schedule (ADOS), Childhood Autism Rating Scale (CARS), Diagnostic Interview for Social and Communication Disorders (DISCO), and Gilliam Autism Rating Scale (GARS)) and are included in this Review if used in eligible analyses.

Parent or carer interview tests

The Autism Diagnostic Interview™ Revised (ADI-R) provides a diagnostic algorithm for ASD that is consistent with both the DSM-IV (APA 1994) and the ICD-10 (WHO 2007). Two recent studies mapped DSM-5 criteria using items from the ADI-R (Huerta 2012; Mazefsky 2013). The ADI-R is a standardised, semi-structured interview during which parents or carers report information about an individual suspected of having an ASD. It assesses behaviour across three domains: (1) reciprocal social interaction; (2) communication and language; and (3) restricted and repetitive, stereotyped interests and behaviours. For an individual to receive

a diagnosis of ASD, scores on all three domains must be elevated beyond cutoff levels. This interview is appropriate for adults and children with a mental age of 18 months and above, and it takes two hours or longer to administer and score (Lord 1994a; Mazefsky 2006a; Rutter 2003).

The tenth revision of the DISCO (DISCO-10) is a detailed, semi-structured interview, which should be used with someone who knows well the person who is being evaluated, preferably from infancy. It uses a dimensional approach to facilitate an understanding of patterns of behaviour that have developed over time. It takes three hours to administer (Wing 2002). DISCO-11 is now available (Wing 2006).

The Developmental, Dimensional, and Diagnostic Interview (3di) is a computerised parental interview that measures intensity of symptoms and co-morbidities across the autism spectrum. It takes two hours to administer (Skuse 2004a).

The GARS is a parent or teacher questionnaire based on DSM-IV (APA 1994); it focuses on four content areas: (1) stereotyped behaviours; (2) communication; (3) social interaction; and (4) developmental disturbances. GARS is an effective test for discriminating patients with ASD from those with behavioural disorders (Gilliam 1995; Mazefsky 2006a). This questionnaire consists of 56 items divided among four scales: (1) social Interaction; (2) communication; (3) stereotyped behaviours; and (4) developmental disturbances. In 2005, GARS-2 was published (Gilliam 2006); this questionnaire contains 42 items grouped into three subscales for use in people from 3 to 22 years of age. It takes 5 to 10 minutes to administer. GARS-3 was published in 2013 (Gilliam 2013). It contains 56 items based on DSM-5 criteria (APA 2013); GARS-3 is suitable for the same age group and takes the same length of time to administer.

Combination of interview and observations of unstructured activity

The CARS is an older test (its use began in 1966) that rates children on a scale of one to four across 15 criteria, to yield a composite score that is used to assign a diagnosis of non-autistic, mildly autistic, moderately autistic, or severely autistic (Schopler 1986). In 2010 CARS-2 was published (Schopler 2010), following revision of the original test. CARS-2 is reported to be useful for distinguishing between children with ASD and those with other cognitive deficits, and for distinguishing between mild-to-moderate and severe autism. It can be completed by clinicians, parents, or teachers and is often used in research studies. It takes about 20 to 30 minutes to administer (New York State Department of Health 2005; Schopler 1980).

Semi-structured observational assessment

The ADOS™-Generic (ADOS-G; Lord 2000a), also known as ADOS, is a semi-structured assessment of communication, social interaction, and play. It can be used to assess children or adults with limited or no language, as well as those who are verbally fluent. It consists of four modules that are administered according to the verbal capacity of the child or adult. Each module contains standard activities that allow the examiner to observe behaviours consistent with a diagnosis of ASD or other pervasive developmental disorders. Revision of the test resulted in the publication of ADOS-2 in 2012 (Lord 2012a). ADOS-2 contains updated protocols; revised algorithms for Modules 1, 2, and 3; and a fifth module for toddlers 12 to 30 months of age who are not yet using phrased language. This fifth module was called ADOS-T (for toddlers) during its development but is not available as a separate test. In both versions of the test; cutoff scores are provided for disorders across the autism spectrum, including classical autism and ASD. Usually one module is administered per assessment, but more may be administered if the child or adult displays unexpected abilities that require further assessment (Lord 1999). Two recent studies mapped DSM-5 criteria using items from ADOS (Huerta 2012; Mazefsky 2013).

Clinical pathway

In diagnostic practice, assessment may occur in primary or tertiary settings and is undertaken by multi-disciplinary teams comprising variable combinations of health professionals such as paediatricians, psychologists, speech pathologists, and psychiatrists. The multi-disciplinary team takes a comprehensive history and then undertakes standardised developmental or cognitive tests, behavioural assessments, speech and language assessments, and observation in clinical and usual settings (e.g. child care, home, school). For clinical history-taking or observations (or both) of children in this diagnostic process, it is best practice to use one or more standardised tests for the diagnosis of autism. Results of these tests are combined with information from other sources along with clinical judgement to develop an overall diagnosis based on the current diagnostic classification system for autism.

Prior test(s)

Children undergoing an autism diagnostic test have often completed developmental surveillance or an autism screening test, or both, as described in Alternative test(s). They also may have completed a standardised assessment of development or cognition, behavioural assessments, and speech and language assessments, as described under Clinical pathway.

Role of index test(s)

In clinical care, index tests usually are used as an adjunct to diagnosis, as described for the Clinical pathway. In research, index tests are often used in isolation or in combination to confirm a diagnosis from a clinically recruited or population-recruited sample.

Alternative test(s)

We evaluated neither tests used to screen populations for ASD nor child health surveillance tests used to assess clinical populations but not to provide a diagnosis (SIGN 2007).

Asperger syndrome (or Asperger disorder) is not a common diagnosis in this preschool age group, so we did not include diagnostic tests that have been developed specifically to diagnose this disorder.

Rationale

Accurate diagnosis of ASD is important. Current methods of diagnosis require multi-disciplinary teams and lengthy assessments. Standardised parent or carer interviews and observation instruments have been developed; these are used in clinical assessments and in the research setting. In the clinical pathway, these tests may be used in isolation or in conjunction with other tests as part of a multi-disciplinary team assessment, depending on geographical location and available services.

Clinicians need to know which of these tests has the best diagnostic accuracy and whether the tests can be used on their own to diagnose autism or only as part of a multi-disciplinary team assessment. We do not know whether these tests should be used in combination in the assessment to improve diagnostic accuracy.

For a test to be used in isolation, it would need to perform well with regard to both sensitivity and specificity because a false-positive result has implications in terms of labelling, selection of correct interventions, and resource implications of those interventions, and a false-negative result can lead to a missed opportunity for timely intervention and for family adjustment and planning and, as such, also has service implications. False-negatives are of greater concern if the result of a test inhibits future access to services; they are of less concern if review and follow-up are available if a child continues to have problems that are of concern to parents and carers or other education, health, and community-based professionals.

Instruments that are currently recommended as diagnostic tests for ASD use different assessment approaches (interview vs observation vs mixed methods); therefore, it is possible that these assessments when combined or conducted in series may offer opportunities to enhance diagnostic test accuracy or improve efficiency. Assessment of whether there are potentially suitable sequences for offering testing could save time for both families and services and could use fewer resources.

A systematic review of available diagnostic tests is required to determine which test is most accurate, and whether combinations of tests are suitable for the clinical diagnosis of ASD.

OBJECTIVES

Primary objectives

1. To identify which diagnostic tools, including updated versions, most accurately diagnose ASD in preschool children when compared with multi-disciplinary team clinical judgement.
2. To identify how the best of the interview tools compare with CARS, then how CARS compares with ADOS.
   a. Which ASD diagnostic tool - among ADOS, ADI-R, CARS, DISCO, GARS, and 3di - has the best diagnostic test accuracy?
   b. Is the diagnostic test accuracy of any one test sufficient for that test to be suitable as a sole assessment tool for preschool children?
   c. Is there any combination of tests that, if offered in sequence, would provide suitable diagnostic test accuracy and enhance test efficiency?
   d. If data are available, does the combination of an interview tool with a structured observation test have better diagnostic test accuracy (i.e. fewer false-positives and fewer false-negatives) than either test alone?

As only one interview tool was identified, we modified the first three aims to a single aim (Differences between protocol and review): This Review evaluated diagnostic tests in terms of sensitivity and specificity. Specificity is the most important factor for diagnosis; however, both sensitivity and specificity are of interest in this Review because there is an inherent trade-off between these two factors.

Secondary objectives

1. To determine whether any diagnostic test has greater diagnostic test accuracy for age-specific subgroups within the preschool age range.

METHODS

Criteria for considering studies for this review

Types of studies

Eligible studies were:

1. cohort studies or cross-sectional studies;
2. randomised studies of test accuracy - participants had been randomised to different index tests and all participants had been verified by the same gold standard; and
3. case-control studies - participants had been selected on the outcome side (i.e. a sample of patients with ASD (e.g. selected from an existing cohort) and a sample of children without ASD from a different source).

Participants

Participants were children suspected of having an ASD who were being seen prospectively because of concerns with social, communication, and/or behavioural problems of the type seen in autism. Age was restricted to the preschool years; however, if study cohorts included children beyond six years of age, we included analyses if the mean age of participants was less than six years. We placed no restrictions on setting.

Index tests

We assessed the following index tests for ASD.

1. Parent or carer interviews: Autism Diagnosis Interview - Revised (ADI-R); Diagnostic Interview for Social and Communication Disorders (DISCO) - Tenth Revision (DISCO-10) - or DISCO - Eleventh Revision (DISCO-11); Gilliam Autism Rating Scale (GARS) - Second Edition (GARS-2) - or Third Edition (GARS-3); and the Developmental, Dimensional, and Diagnostic Interview (3di).
2. Combination of interview and observations of unstructured activity: Childhood Autism Rating Scale (CARS) or CARS - Second Edition (CARS-2).
3. Semi-structured observational assessment: Autism Diagnostic Observation Schedule (ADOS), ADOS-Generic (ADOS-G), or ADOS - Second Edition (ADOS-2).

Target conditions

The target condition was ASD in preschool children. ASD can be diagnosed according to DSM-5 (APA 2013). Diagnostic subgroups of autism (childhood autism (ICD-10) or autistic disorder (DSM-IV)); pervasive developmental disorder (atypical autism (ICD-10), pervasive developmental disorder, unspecified (ICD-10), or pervasive developmental disorder - not otherwise specified (PDD-NOS) (DSM-IV)); and Asperger syndrome or Asperger disorder were grouped together as ASD (APA 1994; APA 2000; WHO 2007).

Reference standards

The reference standard was a clinical diagnosis of ASD, as defined above, based on a classification system that was accepted at the

time of the Review (DSM - Third Edition (DSM-III; APA 1980); DSM-III- Revised (DSM-III-R; APA 1987); DSM-IV (APA 1994); DSM-IV-TR (APA 2000); DSM-5 (APA 2013); ICD-9 (WHO 1992); or ICD-10 (WHO 2007)) and as assigned by an experienced multi-disciplinary team. Assessment by the multi-disciplinary team included evaluation of social behaviour, language and non-verbal communication, adaptive behaviour, atypical behaviour, and cognitive status or intellectual function. This assessment was based on information from a clinical assessment, from health professionals involved in the child's care, and from those caring for the child in community settings such as preschool or child care settings.

It is known that diagnosis of specific ASD varies over time; therefore, the reference standard assessment and the index test must have been performed within six months of each other.

Search methods for identification of studies

We developed a sensitive search strategy that combined just two concepts: population (see Participants) and the index tests that are the focus of this Review (see Index tests). We used free-text search terms for each named test, including its abbreviated form, and, when possible, indexing terms to describe the type of assessment (e.g. interview, observation). We began the searches in February 2011; these were followed by three sets of top-up searches in March 2012, May 2013, and, most recently, July 2016.

Electronic searches

We searched the following databases.

1. Cochrane Central Register of Controlled Trials (CENTRAL; 2016, Issue 6) in the Cochrane Library, which includes the Cochrane Developmental, Psychosocial and Learning Problems Specialised Register (searched 20 July 2016).
2. MEDLINE Ovid (1948 to July week 1 2016).
3. Embase Ovid (1980 to 2016 week 29).
4. PsycINFO Ovid (1887 to July week 2 2016).
5. CINAHL Plus EBSCOhost (Cumulative Index to Nursing and Allied Health Literature; 1937 to 20 July 2016).
6. Science Citation Index and Social Sciences Citation Index Web of Science (SCI and SSCI; 1970 to 21 July 2016).
7. Conference Proceedings Citation Index - Science and Conference Proceedings Citation Index - Social Science & Humanities Web of Science (CPCI-S and CPCI-SSH; 1990 to 21 July 2016).
8. ASSIA (Applied Social Sciences Index & Abstracts; 1987 to 11 February 2011). ASSIA was no longer available to the Review team after 2011.
9. Social Services Abstracts Proquest (1979 to 21 July 2016).
10. ERIC EBSCOhost (Education Resources Information Center; 1966 to 21 July 2016).
11. Database of Abstracts of Reviews of Effect (DARE; 2015, Issue 2), part of the Cochrane Library (searched 20 July 2016). DARE ceased publication after this issue.
12. National Autistic Society – Library Catalogue (www.autism.org.uk/autismdata; searched 21 July 2016). Previously known as Autism Data.

We reported the search strategy used for each database in Appendix 1. We included the strategy for each platform when databases changed supplier during the writing of this Review.

Searching other resources

We searched the reference lists of all included publications.

Data collection and analysis

Selection of studies

Two pairs of review authors (AS & KS-L or MR & NL & KE) independently assessed all publications for inclusion. We resolved disagreements by discussion or, when necessary, by consultation with a third review author (KW or SW). We made first selection by screening the titles and abstracts of identified publications. We made final decisions about inclusion by reading the full papers. We recorded our decision process in a PRISMA diagram (Moher 2009).

Data extraction and management

Two pairs of review authors (AS & KS-L or MR & NL) independently extracted data using standardised data extraction forms. We resolved disagreements by discussion and in consultation with a third review author (KW or SW). If data from publications were insufficient, we contacted study investigators for clarification.

We extracted the following data, which we used to complete the 'Characteristics of studies' tables and to conduct subgroup analyses.

1. Characteristics of participants: age; intellectual function; diagnoses for inclusion; setting for recruitment.
2. Index tests: types of tests; cutoffs for diagnostic categories.
3. Reference standards: type; diagnostic categories used; adequacy of assessment, including disciplines represented by members of the multi-disciplinary team, assessments completed, and sources of material used to inform the diagnostic assessment.
4. Study type: cross-sectional study; cohort study; randomised test accuracy study; case-control study.
5. Results: numbers of true-positives, false-positives, false-negatives, and true negatives.

Assessment of methodological quality

Two independent review authors (AS & KS-L and/or MR & NL) assessed methodological quality using the QUADAS-2 instrument (Quality Assessment of Studies of Diagnostic Accuracy - Revised) (Whiting 2011). QUADAS-2 consists of items that assess risk of bias (e.g. blind assessment of index, reference test) and concerns about applicability (e.g. whether the index test is used in the same way as it would be in clinical situations). Further information is available from www.bris.ac.uk/quadas/quadas-2. We developed criteria to aid assessment of key issues (Table 2). We resolved disagreements by discussion and, when necessary, in consultation with a third review author (KW or SW). We also gathered information about study authors' potential conflicts of interests.

Statistical analysis and data synthesis

The index tests assessed in this systematic review have different diagnostic outcome categories. To allow primary analyses, we considered all diagnoses relevant to the ASD category as ASD diagnoses and compared them with diagnoses that were not ASD.

We describe here expected diagnostic outcomes of the index tests.

1. **ADI-R.** Diagnostic categories are autistic disorder and Asperger syndrome, which we combined as ASD (Lord 1994a; Rutter 2003).

2. **ADOS.** Diagnostic categories are autism and ASD, which we combined as ASD (Lord 1999; Lord 2000a; Lord 2012b). We found no studies using ADOS-2. The appropriate ADOS module is selected for administration based on a child's expressive language skills and chronological age. Owing to the age group of interest, participants in this Review completed Module 1 (pre-verbal/single words) or Module 2 (phrase speech). Thresholds for diagnosing autism and ASD showed minimal variation between the two modules.

3. **CARS.** A score of 30 to 36 indicates mild autism, and a score of 37 or more indicates moderate or severe autism (Schopler 1980; Schopler 2010). A cutoff of < 30 is classified as not ASD, and scores ≥ 30 are classified as ASD (Schopler 1986). For the CARS-2, different cutoffs apply for different ages and abilities. We found no studies using CARS-2.

4. **DISCO-10.** The diagnostic categories based on DISCO-10 algorithms that are relevant to the ICD-10 classification system include childhood autism, atypical autism, and Asperger syndrome (Wing 2002; Wing 2006). In addition, there are diagnostic algorithms for "early infantile autism" according to Kanner 1957; "Asperger syndrome" based on the definition provided in Gillberg 1989; and "criteria for autistic spectrum disorder" according to Wing 1979. Any of these diagnostic categories would be classified as ASD. Other diagnostic categories, such as childhood disintegrative disorders and failure to fulfil ASD categories, would be classified as not ASD. We found no studies using DISCO-11.

5. **3di.** Responses on the 3di are generally coded on a three-point scale. This assessment includes 266 questions that are directly or indirectly concerned with disorders on the autism spectrum and 291 questions that relate to current mental states as relevant to other diagnoses (Skuse 2004a). For a diagnosis of ASD, cutoff scores must be achieved for the following five categories: (1) ≥ 10 for reciprocal social interaction skills; (2) ≥ 1 for social expressiveness; (3) ≥ 8 for use of language and other social communication skills; (4) ≥ 7 for use of gesture and non-verbal play; and (5) ≥ 3 for repetitive and stereotyped behaviours.

6. **GARS.** An overall autism quotient is established and then is broken down into seven ordinal categories ranging from a very low to a very high probability of autism. A diagnostic cutoff score ≥ 90 specifies that the child is probably autistic and will be classified as ASD (Gilliam 1995; Gilliam 2006; Gilliam 2013; South 2002). We found no studies using GARS-2 or GARS-3.

Test results were treated as positive or negative for the cutoff values of the index tests described above. When analyses were reported differently from required cutoff values, we generated sensitivity and specificity values for the cutoffs that were relevant to this Review, provided data were available. For example, in Risi 2006, for both eligible cohorts of children (i.e. children < 36 months (Risi 2006 Study 1 ADOS Cohort A) and children with mental retardation with mean age of 62.5 months (Risi 2006 Study 1 ADOS Cohort B)), study authors reported values for children classified with 'autism' versus children classified with 'non-autism ASD'. We calculated revised values for the diagnostic groupings of 'autism and non-autism ASD' versus 'non-spectrum' as reported in Table 3 and included these in the meta-analysis.

If analyses included participants who were not relevant to the objectives of this Review, such as children with typical development (TD), we calculated revised values for sensitivity and specificity values if data were available. For example, Cox 1999 included a small number of children with TD (n = 15) in reported sensitivity and specificity values for ADI-R. We recalculated these values while excluding TD children as reported in Table 4.

We constructed forest plots showing pairs of sensitivity and specificity values with 95% confidence intervals (CIs) for each analysis with appropriate available data. We conducted meta-analyses of pairs of sensitivity and specificity values using bivariate random-effects methods (Reitsma 2005). This enabled calculation of summary estimates while accounting for variation within and between studies and any potential correlation between sensitivity and specificity. We used Stata software for these analyses (StataCorp 2007). For tests with a small number of studies, we pooled results by performing separate meta-analyses for sensitivity and specificity using univariate random-effects logistical regressions (Takwoingi 2017), which we performed in R (module glmer) (Bates 2015).

In our protocol, we described that we would have performed the aforementioned analyses for subgroups of tests with similar cutoff points had different cutoff values for tests been applied (Samtani 2011). However, we found that cutoff values were consistent for tests in all studies with the exception of one (Oosterling 2010b ADI-R). See Differences between protocol and review.

Investigations of heterogeneity

Potential sources of heterogeneity include age of study participants; severity and type of diagnosis (autistic disorder or childhood autism vs PDD-NOS); presence or absence of language delay; presence or absence of intellectual disability or developmental delay; diagnostic mix of population included; prospectively made versus existing diagnosis for study recruitment; study type; and duration between diagnosis and diagnostic test accuracy analyses being performed. Of these, the only source of heterogeneity that was available and was sufficiently different between studies to be explored was age of study participants for two tests: ADOS and CARS (see Differences between protocol and review).

Sensitivity analyses

We performed sensitivity analyses to assess the impact of risk of bias for all tests. We considered studies to have high risk of bias if they had one or more domains with high risk of bias. We also performed sensitivity analyses by including only studies with low risk of bias for the reference standard.

RESULTS

Results of the search

We conducted our electronic literature searches in February 2011, April 2012, May 2013, and July 2016, which respectively yielded 17,393, 1513, 2146, and 5378 records once duplicates were removed. Following our initial review of titles and abstracts, we retrieved 53, 5, 21, and 3 full-text papers from our respective searches, which we assessed for eligibility against our inclusion criteria (Criteria for considering studies for this review). Of these, we excluded 69 publications as irrelevant (43 from searches in 2011; 3

from 2012; 20 from 2013; and 3 from 2016), largely because articles did not report findings from studies that included the index tests of interest, were not DTA studies, included participants outside the age range of interest, or did not include the identified reference standard. We included a total of 13 publications in this Review (10 from searches in 2011; 2 from 2012; 1 from 2013; and 0 from 2016). See Figure 2.

Figure 2. Study flow diagram.
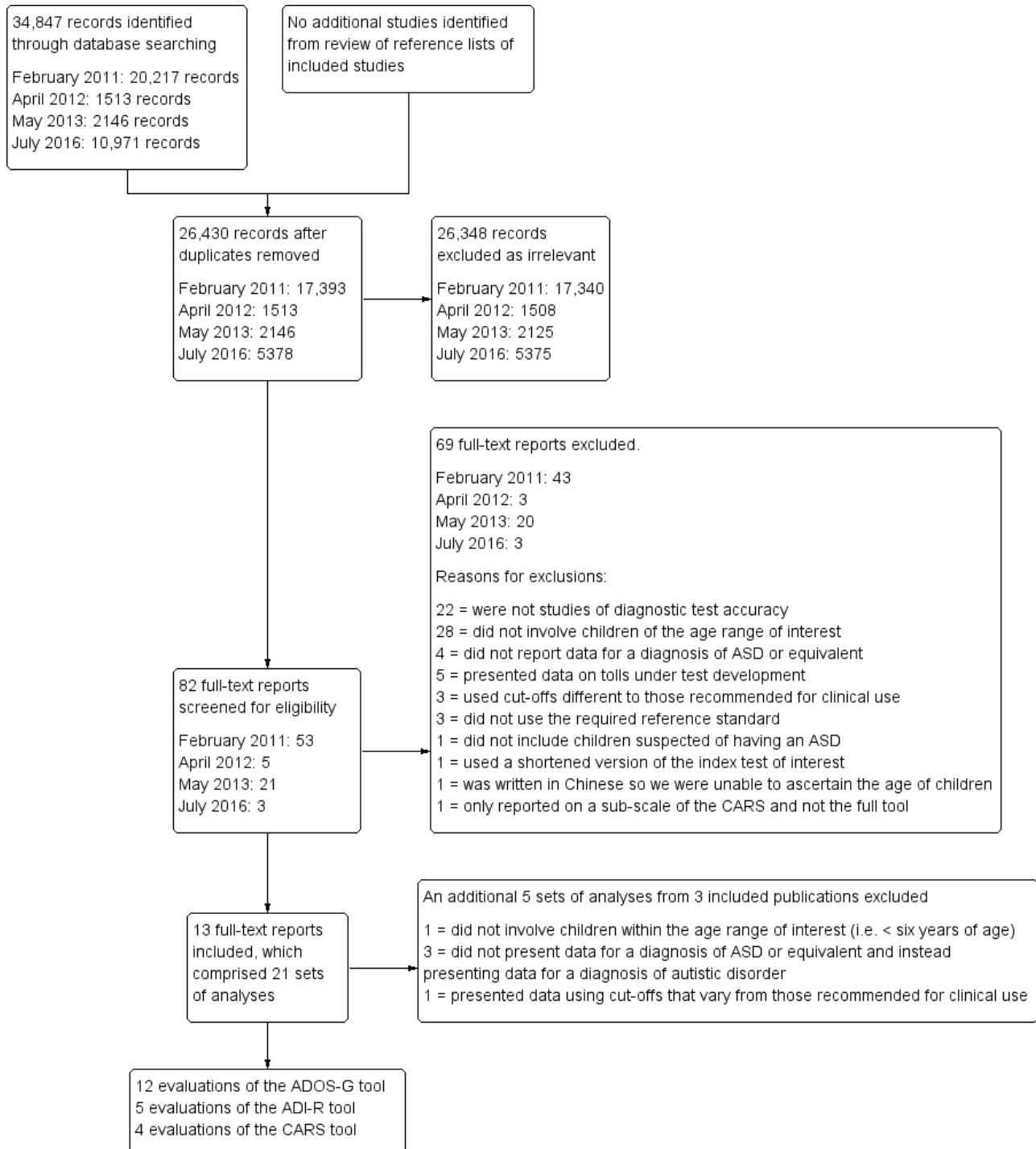
We split the 13 included publications into included 'analyses' because a number of publications described more than one study, investigated more than one tool, or reported results for more than one participant cohort. In addition, during the 'Risk of bias' and

applicability assessment and data extraction, it became apparent that not all included publications, or in some instances not all of the studies within the publications, reported results for tests in a format suitable for inclusion. Some used different cutoff criteria than those

used clinically; others used tests that are not available for clinical use. Some studies moreover did not present data in a way that allowed extraction of data for identification of children with ASD, but rather only identified children with autistic disorder. Further information is available in the Excluded studies section below.

For the purposes of this Review, we focused on 21 sets of analyses reported in 13 publications that fulfilled all of the inclusion criteria (Criteria for considering studies for this review), and we presented findings that were clinically applicable (Chlebowski 2010; Corsello 2013; Cox 1999; Gray 2008 ADI-R; Gray 2008 ADOS; Kim 2012b ADOS Cohort A; Kim 2012b ADOS Cohort B; Le Couteur 2008 ADOS; Lord 2000; Mazefsky 2006 ADOS; Oosterling 2010b ADI-R; Oosterling 2010b ADOS; Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B; Russell 2010; Ventola 2006 ADI-R; Ventola 2006 ADOS; Ventola 2006 CARS; Wiggins 2008 ADI-R; Wiggins 2008 ADOS; Wiggins 2008 CARS); see Table 3, Table 4, and Table 5. Four analyses were presented in two publications, with each publication reporting two sets of diagnostic test accuracy data for clinically different cohorts: Kim 2012b ADOS Cohort A; Kim 2012b ADOS Cohort B; Risi 2006 Study 1 ADOS Cohort A; and Risi 2006 Study 1 ADOS Cohort B. For clarity, we designated analyses by both publication information and the test being assessed if the publication included data for other tests, even if the other tests were not included in our results. For example, the Gray 2008 publication included data for both ADI-R and ADOS, which are included in our results (Gray 2008 ADI-R; Gray 2008 ADOS), whereas LeCouteur 2008 published data for both ADI-R and ADOS, and only ADOS data are included in this Review (Le Couteur 2008 ADOS). For Oosterling 2010b ADOS, study authors published only sensitivity and specificity values, so we had to obtain directly from study authors raw data for inclusion in the meta-analysis (Oosterling, I (2015)). For Risi 2006 Study 1 ADOS Cohort A and Risi 2006 Study 1 ADOS Cohort B, we used data reported in the paper to calculate values reported in this Review (i.e. by adding raw data for autism and PDD-NOS cases); these are not the values reported in the paper. For CARS in Chlebowski 2010, we also calculated sensitivity and specificity values from raw data in the paper. For ADI-R in Cox 1999, we calculated reported values with TD cases removed; these are not the values reported in the paper.

Of the included publications reporting results that compared the diagnostic test accuracy of two or more tests, only one assessed the accuracy of the combined use of tests, as well as the accuracy of each single test (Oosterling 2010b ADI-R; Oosterling 2010b ADOS).

Included studies

Types of studies

This Review includes 21 sets of analyses reported in 13 publications. Fifteen analyses were reported from prospective cohort studies of children receiving clinical assessments for developmental concerns (Chlebowski 2010; Corsello 2013; Cox 1999; Gray 2008 ADI-R; Gray 2008 ADOS; Le Couteur 2008 ADOS; Mazefsky 2006 ADOS; Oosterling 2010b ADI-R; Oosterling 2010b ADOS; Ventola 2006 ADI-R; Ventola 2006 ADOS; Ventola 2006 CARS; Wiggins 2008 ADI-R; Wiggins 2008 ADOS; Wiggins 2008 CARS); five were from studies involving secondary analyses of test scores collected from children participating in early diagnosis and intervention research projects (Kim 2012b ADOS Cohort A; Kim 2012b ADOS Cohort B; Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B; Russell 2010); one was a case-control study

that included children identified with autism, PDD-NOS, and non-spectrum disorders who were matched for verbal mental age (Lord 2000).

See Characteristics of included studies tables.

Locations of studies

Of the 21 included analyses, 12 were from studies carried out in the USA (Chlebowski 2010; Corsello 2013; Kim 2012b ADOS Cohort A; Kim 2012b ADOS Cohort B; Lord 2000; Mazefsky 2006 ADOS; Ventola 2006 ADI-R; Ventola 2006 ADOS; Ventola 2006 CARS; Wiggins 2008 ADI-R; Wiggins 2008 ADOS; Wiggins 2008 CARS); two used combined sets of data collected from the USA and Canada (Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B); two apiece were from studies conducted in the Netherlands (Oosterling 2010b ADI-R; Oosterling 2010b ADOS), the United Kingdom (Cox 1999; Le Couteur 2008 ADOS), and Australia (Gray 2008 ADI-R; Gray 2008 ADOS); and one was conducted in India (Russell 2010).

Participants

Participants were children between 12 months and 8 years of age, although we included analyses only when the mean age of participants was less than 6 years. Overall, 2900 children were included in this Review, of whom 1625 were tested via ADOS, 641 by CARS, and 634 with ADI-R. Studies usually involved children suspected of having an ASD. All but two analyses - Chlebowski 2010 and Cox 1999 - excluded TD children when calculating sensitivity and specificity values for the index test of interest. See further information below in the Methodological quality of included studies section titled 'Applicability concerns'.

In 19 included analyses, children were six years of age or younger (i.e. preschool age) (Chlebowski 2010; Corsello 2013; Cox 1999; Gray 2008 ADOS; Gray 2008 ADOS; Kim 2012b ADOS Cohort A; Kim 2012b ADOS Cohort B; Le Couteur 2008 ADOS; Lord 2000; Oosterling 2010b ADI-R; Oosterling 2010b ADOS; Risi 2006 Study 1 ADOS Cohort A; Russell 2010; Ventola 2006 ADI-R; Ventola 2006 ADOS; Ventola 2006 CARS; Wiggins 2008 ADI-R; Wiggins 2008 ADOS; Wiggins 2008 CARS). Of the remaining analyses, one included children over six years of age but with a mean age less than six years (Mazefsky 2006 ADOS), and one comprised children with mental retardation older than six years but again the mean age of the cohort was less than six years (Risi 2006 Study 1 ADOS Cohort B).

In all 21 analyses, children presented with coexisting language or developmental delay, or a combination of both. In addition, in a total of nine analyses, some children presented with intellectual disability (Kim 2012b ADOS Cohort B; Lord 2000; Oosterling 2010b ADI-R; Oosterling 2010b ADOS; Risi 2006 Study 1 ADOS Cohort A; Russell 2010), or mental health problems, including attention deficit hyperactivity disorder, anxiety, or attachment disorders (Corsello 2013; Lord 2000; Mazefsky 2006 ADOS; Oosterling 2010b ADI-R; Oosterling 2010b ADOS; Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B).

Index test

ADOS was used in a total of 12 included analyses (Corsello 2013; Gray 2008 ADOS; Kim 2012b ADOS Cohort A; Kim 2012b ADOS Cohort B; Le Couteur 2008 ADOS; Lord 2000; Mazefsky 2006 ADOS; Oosterling 2010b ADOS; Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B; Ventola 2006 ADOS; Wiggins 2008

![Cochrane Library logo]
Cochrane
Library

Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

ADOS); the ADI-R in five included analyses (Cox 1999 (20 months and 42 months); Gray 2008 ADI-R; Oosterling 2010b ADI-R; Ventola 2006 ADI-R; Wiggins 2008 ADI-R); and the CARS in four included analyses (Chlebowski 2010 (two-year-old sample); Russell 2010; Ventola 2006 CARS; Wiggins 2008 CARS). There were no suitable studies or analyses for 3di, DISCO, or GARS.

CARS was reported alone in two included analyses (Chlebowski 2010 (both two-year-old and four-year-old samples); Russell 2010) but was reported alongside ADI-R and ADOS in another two analyses (Ventola 2006 CARS; Wiggins 2008 CARS). ADOS was reported alone in two included analyses (Corsello 2013; Lord 2000) but with ADI-R in another two analyses (Gray 2008 ADI-R; Oosterling 2010b ADI-R) and with CARS and ADI-R as mentioned above in two included analyses (Ventola 2006 ADOS; Wiggins 2008 ADOS). ADI-R was reported alone in one analysis (Cox 1999 (20 months and 42 months)), alongside ADOS in the two aforementioned analyses (Gray 2008 ADI-R; Oosterling 2010b ADI-R), and with ADOS and CARS in the two previously listed analyses (Ventola 2006 ADI-R; Wiggins 2008 ADI-R).

Target conditions

Twenty-one diagnostic accuracy results were reported or could be calculated for the target condition of ASD (including subgroups of children with autism, Asperger syndrome, and PDD-NOS) for one index test. One set of results were reported in each of the following 16 analyses: Chlebowski 2010 (two-year-old sample); Corsello 2013; Cox 1999; Gray 2008 ADI-R; Gray 2008 ADOS; Le Couteur 2008 ADOS; Mazefsky 2006 ADOS; Oosterling 2010b ADI-R; Oosterling 2010b ADOS; Russell 2010; Ventola 2006 ADI-R; Ventola 2006 ADOS; Ventola 2006 CARS; Wiggins 2008 ADI-R; Wiggins 2008 ADOS; Wiggins 2008 CARS); results from two sets of analyses were reported in Kim 2012b for cohorts A and B (Kim 2012b ADOS Cohort A; Kim 2012b ADOS Cohort B) and in Risi 2006 for cohorts A and B (Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B). Lord 2000 reported separate analyses for Modules 1 and 2 of ADOS and undertook analyses on the combined data set (see Table 3).

Reference standards

Different assessments were used as the reference standard across the studies reviewed. Most studies reported using a best-estimate clinical diagnosis as the reference standard assessment. One study, Corsello 2013, applied a records-based method whereby clinicians reviewed children's records against DSM-IV-TR criteria to make a clinical diagnosis. For four included analyses (Chlebowski 2010; Wiggins 2008 ADI-R; Wiggins 2008 ADOS; Wiggins 2008 CARS), study authors reported that a clinical diagnosis was made by one professional alone. For three included analyses (Ventola 2006 ADI-R; Ventola 2006 ADOS; Ventola 2006 CARS), study authors did not specify the number or discipline of the professionals making the clinical diagnosis. Two or more clinicians or a multi-disciplinary team assessment was used for diagnosis in publications reporting on procedures implemented for 11 of the included analyses (Cox 1999; Gray 2008 ADI-R; Gray 2008 ADOS; Le Couteur 2008 ADOS; Lord 2000; Mazefsky 2006 ADOS; Oosterling 2010b ADI-R; Oosterling 2010b ADOS; Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B; Russell 2010); however, multi-disciplinary teams ranged in composition from a psychologist and a psychiatrist to potentially containing any of the following professionals: psychologist, psychiatrist, paediatrician, consultant, speech pathologist, special educator, psychiatric nurse,

or occupational therapist. Within the same study, the clinical diagnosis could also be made by a different combination and number of these professionals, and for two analyses - Kim 2012b ADOS Cohort A and Kim 2012b ADOS Cohort B - study authors reported that a clinical diagnosis was made by an experienced clinical researcher or a psychiatrist 'and/or' psychologist.

All studies reported using DSM-III (APA 1980), DSM-III-R (APA 1987), DSM-IV (APA 1994), DSM-IV-TR (APA 2000), and/or ICD-10 (WHO 2007) criteria to make a best-estimate clinical diagnosis. Assessment information and the number and/or combination of domains assessed and tests used varied between studies. Information possibly collected included formal evaluation or clinical observations of social behaviour, language and non-verbal communication, adaptive behaviour, cognitive status/intellectual function, and/or atypical behaviours. In some instances, observations or results from psychiatric evaluations were included. Variation was also present in the range of assessment results included when a best-estimate clinical judgement was made. For example, eight studies accounting for 15 of the included analyses reported including a range of standardised clinical assessment results in addition to information from interviews with families and video footage of child interactions and play (Gray 2008 ADI-R; Gray 2008 ADOS; Le Couteur 2008 ADOS; Lord 2000; Mazefsky 2006 ADOS; Oosterling 2010b ADI-R; Oosterling 2010b ADOS; Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B; Ventola 2006 ADI-R; Ventola 2006 ADOS; Ventola 2006 CARS; Wiggins 2008 ADI-R; Wiggins 2008 CARS; Wiggins 2008 ADOS).

Flow and timing

For four analyses, study authors reported that the index test and the reference standard were administered within a six-month time interval, as detailed in the study protocol (Cox 1999; Mazefsky 2006 ADOS; Oosterling 2010b ADI-R; Oosterling 2010b ADOS). Study authors for the remaining 17 analyses did not explicitly state the length of intervening time between assessment events but did report that assessments occurred at only one time point.

Conflicts of interest

For studies reporting on 13 of the included analyses, there was no direct conflict of interest evident (Chlebowski 2010; Corsello 2013; Cox 1999; Mazefsky 2006 ADOS; Oosterling 2010b ADI-R; Oosterling 2010b ADOS; Russell 2010; Ventola 2006 ADI-R; Ventola 2006 ADOS; Ventola 2006 CARS; Wiggins 2008 CARS; Wiggins 2008 ADI-R; Wiggins 2008 ADOS). For two analyses - Gray 2008 ADI-R and Gray 2008 ADOS - study authors are known to conduct training for ADI-R, ADOS-2, and ADOS-G, which raises potential conflicts of interest.

In reporting of the remaining six analyses, we could not exclude conflicts of interest because study authors were the developers of the index tools being evaluated (Kim 2012b ADOS Cohort A; Kim 2012b ADOS Cohort B; Le Couteur 2008 ADOS; Lord 2000; Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B). For analyses conducted by Risi 2006 (Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B) and for Kim 2012b (Kim 2012b ADOS Cohort A; Kim 2012b ADOS Cohort B), study authors reported conflicts of interest (see Appendix 3 for more information).

Excluded studies

We excluded 69 publications after full-text review (see Figure 2). Reasons for exclusion were as follows: 22 publications did not

report on diagnostic test accuracy; 28 did not involve children within the age range of interest (i.e. < six years of age); four did not present data for a diagnosis of ASD or equivalent and instead presented data for a diagnosis of autistic disorder (Lord 1993; Lord 1994; Perry 2005; Shin 1998); five presented data on test development with varying cutoffs or for tests that are not in clinical use (Gotham 2007; Gotham 2008; Guthrie 2013; Lord 2006; Luyster 2009); three for presenting data using cutoffs that vary from those recommended for clinical use (Kim 2012a; Kim 2013; Oosterling 2010a); three did not use the required reference standard (Lecavalier 2006; Moss 2008; Saemundsen 2003); one did not include children suspected of having an ASD (Soke 2011); one used only a shortened version of the index test of interest (the 3di) rather than the complete tool (Chuthapisith 2012); one was written in Chinese and we were unable to ascertain the age of the included children (Li 2005); and one reported sensitivity and specificity values for the social impairment scale of the CARS - not for the full test (DiLalla 1994).

In addition, three publications already included in this Review (Le Couteur 2008 ADOS; Mazefsky 2006 ADOS; Risi 2006 Study 1 ADOS Cohort A and Risi 2006 Study 1 ADOS Cohort B), which also contained irrelevant analyses, were excluded (Le Couteur 2008 ADI-R; Mazefsky 2006 ADI-R; Mazefsky 2006 GARS; Risi 2006 study 1 ADI-R; Risi 2006 study 2). One publication - Risi 2006 study 2 - did not involve children within the age range of interest (i.e. < six years of age); three did not present data for a diagnosis of ASD or equivalent and instead presented data for a diagnosis of autistic disorder (Mazefsky 2006 ADI-R; Mazefsky 2006 GARS; Risi 2006 study 1 ADI-R); and one presented data using cutoffs that vary from those recommended for clinical use (Le Couteur 2008 ADI-R).

See Characteristics of excluded studies tables.

Methodological quality of included studies

Risk of bias

We assessed all studies accounting for the 21 analyses for risk of bias. We considered only one study reporting on one CARS analysis - Russell 2010 - to be at low risk of bias across all domains: patient selection, index test, reference standard, and flow and timing (see Figure 3). We judged a further study reporting on one ADOS analysis to be at low risk of bias for three domains (patient selection, index test, and reference standard) and at uncertain risk of bias for flow and timing (Corsello 2013).

Figure 3. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.

Major concerns for risk of bias were known lack of blinding between the index text and the reference standard, both at the time of assessment using the index test and in development of the reference standard diagnosis. Only the two studies named above - Corsello 2013 and Russell 2010 - included a description of blinding for both the index test and the reference standard diagnosis.

For studies in which the index test was completed blinded to diagnosis, we considered two studies reporting analyses for ADOS to be at low risk of bias for index test assessment (Corsello 2013; Mazefsky 2006 ADOS). We judged risk of bias for the index test assessment as unknown for studies reporting on six ADOS analyses (Kim 2012b ADOS Cohort A; Kim 2012b ADOS Cohort B; Le Couteur 2008 ADOS; Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B; Wiggins 2008 ADOS) but high for studies reporting on the remaining four ADOS analyses (Gray 2008 ADOS; Lord 2000; Oosterling 2010b ADOS; Ventola 2006 ADOS). For analyses reported on CARS, we rated one study - Russell 2010 - as having low risk of bias, another unknown risk of bias (Wiggins 2008 CARS), and two high risk of bias (Chlebowski 2010; Ventola 2006 CARS). For analyses reported on ADI-R, we considered no studies to be at low risk of bias but judged four to be at high risk of bias (Cox 1999; Gray 2008 ADI-R; Oosterling 2010b ADI-R; Ventola 2006 ADI-R) and one to be at unknown risk of bias (Wiggins 2008 ADI-R).

We rated three studies reporting ADOS analyses as introducing low risk of bias for the manner in which the reference standard was conducted to reach a diagnosis (Corsello 2013; Gray 2008 ADOS; Lord 2000), seven unknown risk of bias (Kim 2012b ADOS Cohort A; Kim 2012b ADOS Cohort B; Le Couteur 2008 ADOS; Oosterling 2010b ADOS; Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B; Wiggins 2008 ADOS), and two high risk of bias (Mazefsky 2006 ADOS; Ventola 2006 ADOS). For analyses reported on CARS, we judged one study as having low risk of bias (Russell 2010), one unknown risk of bias (Wiggins 2008 CARS), and two high risk of bias (Chlebowski 2010; Ventola 2006 CARS). For analyses reported on ADI-R, we rated two studies as having low risk of bias (Cox 1999; Gray 2008 ADI-R), two unclear risk of bias (Oosterling 2010b ADI-R; Wiggins 2008 ADI-R), and one high risk of bias (Ventola 2006 ADI-R).

Applicability concerns

Using the QUADAS-2, we assessed studies reporting on 10 analyses as applicable. These 10 studies included seven of the 12 ADOS analyses (Corsello 2013; Cox 1999; Gray 2008 ADOS; Le Couteur 2008 ADOS; Lord 2000; Mazefsky 2006 ADOS; Oosterling 2010b ADOS), one of the four CARS analyses (Russell 2010), and two of the five ADI-R analyses (Gray 2008 ADI-R; Oosterling 2010b ADI-R).

Most studies were applicable for patient selection, with the exception being the cohorts reported in Risi 2006 Study 1 ADOS Cohort A, in which children were taken from a longitudinal study, with most receiving a diagnosis of ASD, and Risi 2006 Study 1 ADOS Cohort B, which comprised only children with profound mental retardation. Although some children with normal development were included in two analyses (Chlebowski 2010; Cox 1999), all included children had failed an autism screening test. As such, patient selection is similar to selection of children for referral to services for developmental assessment.

Findings

Twenty-one included analyses provided data eligible for inclusion in meta-analyses (Data table 1; Data table 2; Data table 3). As reported earlier, four analyses were presented in two publications, with each publication including two sets of diagnostic test accuracy data for clinically different cohorts (Kim 2012b ADOS Cohort A; Kim 2012b ADOS Cohort B; and Risi 2006 Study 1 ADOS Cohort A; Risi 2006 Study 1 ADOS Cohort B), so we included data from these four analyses in the ADOS meta-analysis. In Lord 2000, we included data from combined (i.e. Modules 1 and 2) analyses only to prevent duplication.

The prevalence of ASD across all studies ranged from 51% to 86% (median 74%).

Individual tool accuracy

ADOS

For ADOS, we combined the diagnostic categories of autism and ASD as ASD, for analysis and reporting purposes.

There were 12 analyses (1625 children) of sensitivity and specificity reported for all versions and modules of ADOS, with 74% of children in the ADOS analyses receiving a diagnosis of ASD. Prevalence of ASD across these analyses ranged from 51% to 85% (median 75%). Sensitivity ranged from 0.76 to 0.98, and specificity from 0.20 to 1.00 (see Data table 1). The summary sensitivity (bivariate method) was 0.94 (95% CI 0.89 to 0.97), and specificity was 0.80 (95% CI 0.68 to 0.88). See Figure 4.

![Cochrane Library logo] Cochrane
Library

Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

Figure 4. Summary ROC Plot of tests: ADOS, CARS, and ADI-R.
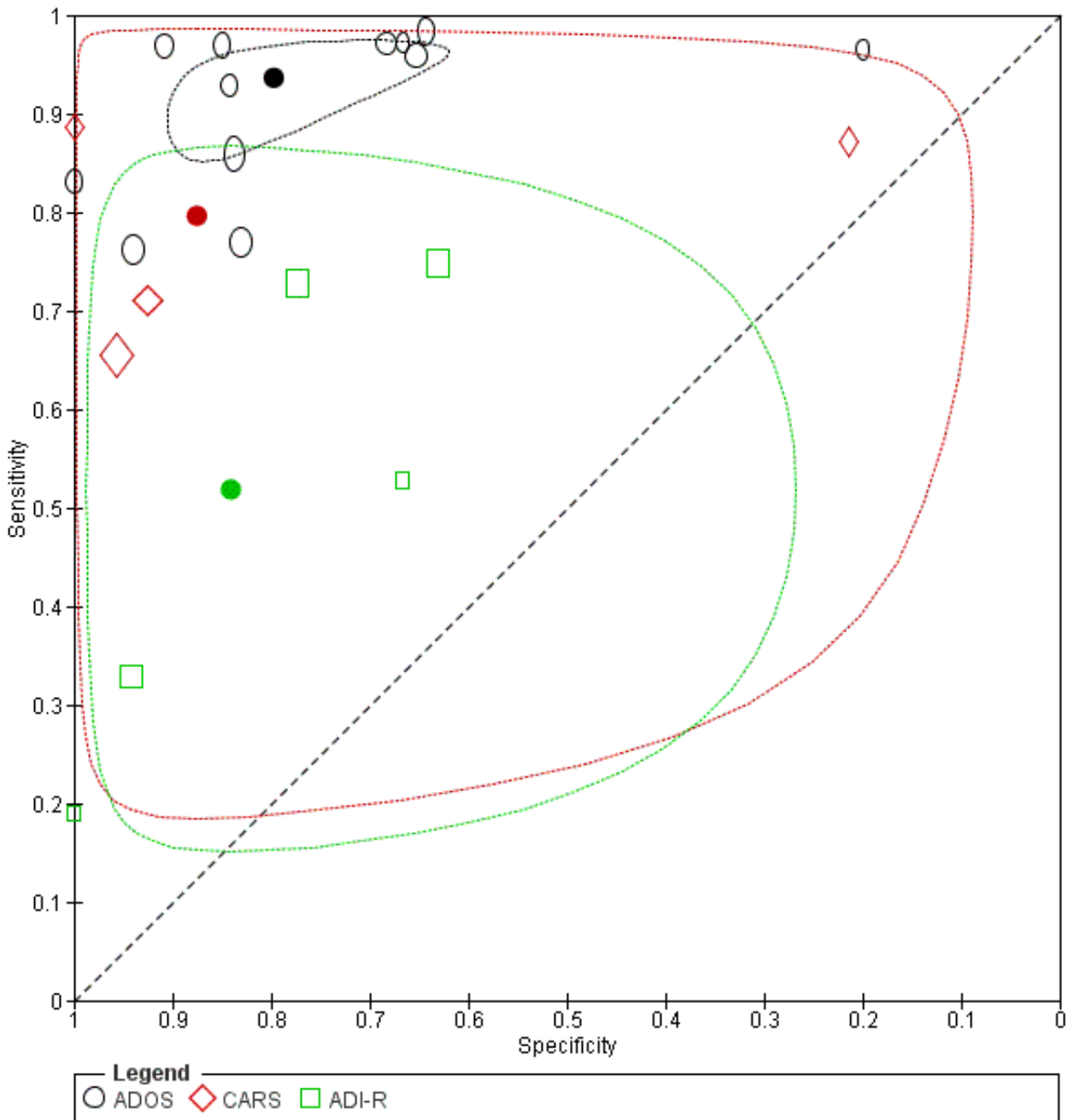
In Lord 2000, in addition to sensitivity and specificity values reported for the overall test, study authors calculated separate sensitivity and specificity results for subgroups of children according to their verbal ability level. Children of different verbal abilities were administered Module 1 or Module 2 of ADOS. Sensitivity and specificity values of 0.98 and 0.94 were reported for Module 1, and 0.95 and 0.88 for Module 2, respectively.

One analysis included only children with an intellectual disability (Risi 2006 Study 1 ADOS Cohort B). Specificity was considerably

lower (0.20, 95% CI 0.03 to 0.56) than reported specificity from other studies.

In a meta-regression analysis, mean age range (26 months to 62.5 months) was not a significant modifier of sensitivity (P = 0.56) nor of specificity (P = 0.41).

With inclusion of only data from three analyses calculated from three studies that were not at high risk of bias (Corsello 2013; Kim 2012b ADOS Cohort A; Wiggins 2008 ADOS), summary sensitivity changed from 0.94 (95% CI 0.89 to 0.98) to 0.97 (95% CI 0.94 to 0.98),

and summary specificity from 0.80 (95% CI 0.70 to 0.88) to 0.68 (95% CI 0.60 to 0.75).

The summary sensitivity did not change when only analyses from the two studies at low risk of bias were included for the reference standard (Corsello 2013; Gray 2008 ADOS); however, the summary specificity increased from 0.80 (95% CI 0.68 to 0.88) to 0.91 (95% CI 0.84 to 0.95).

CARS

For CARS, we classified children with a cutoff score ≥ 30 as having ASD, for analysis and reporting purposes.

Four analyses involving 641 children suspected of having ASD, aged 16 months to 6 years 8 months, were reported for CARS (Chlebowski 2010; Russell 2010; Ventola 2006 CARS; Wiggins 2008 CARS). Sixty-seven per cent of children in analyses undertaken on CARS received the diagnosis of ASD. Prevalence of ASD across these analyses ranged from 51% to 86% (median 73%). We included data from analyses undertaken on the two-year-old cohort in Chlebowski 2010.

Analyses reported sensitivity for CARS ranging from 0.66 to 0.89 and specificity ranging from 0.21 to 1.00 (Data table 2). We could not perform a bivariate meta-analysis owing to too few analyses for CARS. In separate random-effects logistical regression meta-analyses for sensitivity and specificity, the summary sensitivity for CARS was 0.80 (95% CI 0.61 to 0.91) and the summary specificity was 0.88 (95% CI 0.64 to 0.96). See Figure 4.

In a meta-regression analysis, mean age (three studies with a mean age of 26 months; and one study with a mean age of 61 months) increased sensitivity (P = 0.06) and decreased specificity (P < 0.001).

With exclusion of analyses calculated from the two studies deemed at high risk of bias (Chlebowski 2010; Ventola 2006 CARS), the summary sensitivity changed from 0.78 (95% CI 0.65 to 0.88) to 0.88 (95% CI 0.81 to 0.92), and the summary specificity from 0.85 (95% CI 0.43 to 0.98) to 0.65 (95% CI 0.03 to 0.99).

Analyses calculated for the only study at low risk of bias for the reference standard - Russell 2010 - found a similar estimate for sensitivity and an extremely low value for specificity (0.21, 95% CI 0.05 to 0.51).

ADI-R

For ADI-R, we combined the diagnostic categories of autistic disorder and Asperger syndrome as ASD, for analysis and reporting purposes.

Five analyses involving 634 children reported the diagnostic accuracy of ADI-R (Cox 1999; Gray 2008 ADI-R; Oosterling 2010b ADI-R; Ventola 2006 ADI-R; Wiggins 2008 ADI-R). Sixty-six per cent of children in the ADI-R analyses received the diagnosis of ASD. Prevalence of ASD in these analyses ranged from 51% to 80% (median 69%). We included data from the younger cohort in Cox 1999.

Published sensitivity and specificity values for ASD versus non-ASD for ADI-R ranged from 0.19 to 0.75 for sensitivity and from 0.63 to 1.00 for specificity (Data table 3). Lower sensitivity levels were noted in studies of children screened for ASD (Cox 1999; Ventola 2006 ADI-R; Wiggins 2008 ADI-R) compared with clinical samples. We could

not perform a bivariate meta-analysis owing to too few analyses for ADI-R. In separate random-effects logistical regression meta-analyses for sensitivity and specificity, the summary sensitivity for ADI-R was 0.52 (95% CI 0.32 to 0.71) and the summary specificity was 0.84 (95% CI 0.61 to 0.95). See Figure 4.

In a meta-regression analysis, mean age (range 20 to 38.5 months) increased sensitivity (P < 0.001) and decreased specificity - but not significantly (P = 0.12).

We considered four of the five studies reporting analyses for ADI-R to be at high risk of bias for one or more criteria (Cox 1999; Gray 2008 ADI-R; Oosterling 2010b ADI-R; Ventola 2006 ADI-R), and the remaining study to be at unclear risk of bias for two criteria (Wiggins 2008 ADI-R). In analyses from this study, sensitivity was 0.33 (95% CI 0.22 to 0.45) and specificity was 0.94 (95% CI 0.86 to 0.98), compared with 0.52 (95% CI 0.32 to 0.72) and 0.80 (95% CI 0.63 to 0.91) for the summary estimates, respectively.

We found no major change in the summary estimates of sensitivity and specificity when only analyses from the two studies at low risk of bias were included for the reference standard (Cox 1999; Gray 2008 ADI-R).

3di, DISCO, and GARS

We found no studies reporting relevant data for 3di, DISCO-10 or DISCO-11, or GARS that met the inclusion criteria for this Review.

Comparison of ADOS, CARS, and ADI-R

The sensitivities of CARS (0.80, 95% CI 0.61 to 0.91) and ADI-R (0.52, 95 % CI 0.32 to 0.71) in the random-effects logistical regression analysis were significantly lower than those of ADOS (0.94, 95% CI 0.89 to 0.97) (P = 0.019 and P < 0.001, respectively). For specificities, CARS (0.88, 95% CI 0.64 to 0.96) and ADI-R (0.84, 95% CI 0.61 to 0.95) were not significantly different from ADOS (0.80, 95% CI 0.68 to 0.88) (P = 0.52 and P = 0.75, respectively).

Studies reporting between-test comparisons within the same study

Table 6 provides sensitivity and specificity data for the studies reporting analyses for two or more tests in the same cohort.

ADOS was always as or more sensitive than ADI-R in four studies (Gray 2008 ADI-R and Gray 2008 ADOS; Oosterling 2010b ADI-R and Oosterling 2010b ADOS; Ventola 2006 ADI-R, Ventola 2006 ADOS, and Ventola 2006 CARS; Wiggins 2008 ADI-R, Wiggins 2008 ADOS, and Wiggins 2008 CARS) and was as or more specific than ADI-R in three of these four studies (Gray 2008 ADI-R and Gray 2008 ADOS; Oosterling 2010b ADI-R and Oosterling 2010b ADOS; Ventola 2006 ADI-R, Ventola 2006 ADOS, and Ventola 2006 CARS). In two studies (Ventola 2006 ADI-R, Ventola 2006 ADOS, and Ventola 2006 CARS; Wiggins 2008 ADI-R, Wiggins 2008 ADOS, and Wiggins 2008 CARS), ADOS was more sensitive than CARS but was less specific (Table 6). CARS was more sensitive than ADI-R in two studies (Ventola 2006 ADI-R, Ventola 2006 ADOS, and Ventola 2006 CARS; Wiggins 2008 ADI-R, Wiggins 2008 ADOS, and Wiggins 2008 CARS), with similar or higher specificity (Table 6). Overlap of the CI indicates lack of statistically significant differences between most of the reported within-study findings.

Studies reporting combined tool accuracy

Only one of the included publications compared the accuracy of combined use of ADI-R and ADOS against the use of each single test (Oosterling 2010b ADI-R; Oosterling 2010b ADOS). This publication reported that although the combination of ADI-R and ADOS improved specificity (0.94, 95% CI 0.85 to 0.98) by 11% compared with using ADOS alone (0.83, 95% CI 0.72 to 0.91), this came at a cost of a 14% reduction in sensitivity (i.e. sensitivity for ADOS alone was 0.77 (95% CI 0.69 to 0.84) compared with 0.63 (95% CI 0.54 to 0.71) when the tools were used in combination). However, because of the respective 95% CI overlap (especially for specificity), differences between the two approaches could not be demonstrated or refuted.

DISCUSSION

Diagnosing autism spectrum disorder (ASD) is not straightforward owing to the wide spectrum of the condition and reliance on behavioural symptoms and signs. Current recommended diagnostic practice requires that information from clinical assessment, child care, or educational settings as well as standardised instruments (especially for developmental or intellectual ability) should be included, with diagnostic assessment tests for autism as optional additions (AACAP 2014; NICE 2011), rather than use of diagnostic tests alone. This assessment requires involvement of a multi-disciplinary team consisting of several health professionals and often is time-consuming with limited availability of resources. However, accurate diagnosis is critical. If diagnosis is inaccurate, young children who have ASD and who are not given the diagnosis will fail to receive tailored early interventions that may provide them and their families with valuable strategies to facilitate their development and manage their behaviours. In addition, inaccurate diagnosis may result in children who do not have ASD receiving an ASD diagnosis, which could have a detrimental effect for the child and the family and may result in misallocation of limited service resources.

We conducted a systematic review to compare the accuracy of the Developmental, Dimensional, and Diagnostic Interview (3di), the Autism Diagnostic Interview™ Revised (ADI-R), the Autism Diagnostic Observation Schedule (ADOS), the Childhood Autism Rating Scale (CARS), the Diagnostic Interview for Social and Communication Disorders (DISCO), and the Gilliam Autism Rating Scale (GARS) against a reference standard for diagnosis that involved a best-estimate clinical diagnosis made by more than one professional, using available information, to decide whether criteria are met for an acceptable diagnostic classification system such as the *Diagnostic and Statistical Manual of Mental Disorders - Third Edition* (DSM-III), *DSM - Fourth Edition* (DSM-IV), *DSM - Fourth Edition - Text Revision* (DSM-IV-TR), *International Classification of Diseases and Related Health Problems - Ninth Revision* (ICD-9), or *ICD - Tenth Revision* (ICD-10). We assessed sensitivity and specificity of these tests, recognising that both sensitivity and specificity should be high for a diagnostic test.

In all included studies, the prevalence of ASD was high for all tests and might have been greater than in clinical settings in which the test is used. Prevalence in practice will vary depending on the nature of the service. If a service specialises in ASD assessment and management, likely prevalence at the time of ASD assessment would be high. However, a diagnostic and management service that is not autism specific will detect a lower prevalence of ASD at assessment presentation. Practitioners conducting ASD assessments should estimate the prevalence of ASD diagnosed in their service and should take that into account when making decisions about diagnostic test performance.

Summary of main results

Given the widespread use of these tests, diagnostic test accuracy data are relatively limited. Overall, only one study was at low risk of bias for all criteria and four studies were known to be at low risk of bias for three or more factors, as assessed by QUADAS-2 (Quality Assessment of Studies of Diagnostic Accuracy - Revised). Blinding of the reference standard diagnosis was often ignored at the time of index testing, and the index test was sometimes incorporated into the reference standard. These two issues introduce high risk of bias to studies, as they potentially inflate the sensitivity and specificity of the index test. In addition, 29% (6 of 21) of included analyses were published by study authors with a potential conflict of interest. Sensitivity analyses for ADOS and CARS indicate that the specificity of tools was susceptible to risk of bias, with lower calculated specificity in studies assessed at low risk of bias. Although studies were, for the most part, applicable to practice in relation to patient selection, for nearly half of the included studies, the reference standard was not, or it was not clear if it was, in keeping with current recommendations for diagnostic practice.

We included in this Review 21 sets of analyses found in 13 publications. An overview of results of meta-analyses can be found in Summary of findings 1. Published sensitivity and specificity values for ASD versus non-ASD for clinically available tools ranged between 0.19 and 0.98 for sensitivity and between 0.20 and 1.00 for specificity, with ADOS reporting the highest summary sensitivity and similar specificity to CARS and ADI-R. Lowest sensitivities were reported in studies involving children who had ASD with associated intellectual disability. Lower sensitivity levels were also noted in studies using cohorts of children being screened for ASD versus studies using clinical samples. No articles reported sensitivity and specificity values for GARS, DISCO, or 3di. New versions of included tests have emerged since the publication of our protocol (Samtani 2011), but no diagnostic test accuracy data were available at the time of writing this Review.

Four studies compared the diagnostic test accuracy of two (ADOS and ADI-R) or three (ADOS, ADI-R, and CARS) index tests in the same cohort of children. The magnitude of sensitivity and specificity for each test varied between studies. Within studies, few significant between-test differences were noted for sensitivity and specificity. In one study, we found a difference between ADOS (with higher sensitivity) and ADI-R and CARS (both with higher specificity), demonstrating the well-known trade-off between sensitivity and specificity. One of these studies also conducted analyses to assess the accuracy of two tests - ADOS and ADI-R - used alone or in combination and found no conclusive difference between between these two approaches.

Assessment of whether diagnostic test accuracy varied for important clinical factors was limited. In one set of analyses of ADOS calculated on a cohort that included only children with intellectual disability, specificity was lower than in other analyses. Also, specificity was higher in the meta-analysis of data from ADOS when used with older children. This was not replicated for CARS, with persistent heterogeneity and few analyses limiting interpretation of findings.

Strengths and weaknesses of the review

A range of study methods and different approaches to analysis meant that data extraction and synthesis were not straightforward. For example, different modules of ADOS were used in different studies and were reported separately in some studies but combined in other studies. We omitted some studies because they used updated algorithms for ADOS or updated modules (ADOS-T (toddler)). We decided to omit them from this version of the Review because the algorithms are not yet used in clinical diagnosis and assessment, and because the modules are not available separately for clinical use. However, these components are now included as part of the revised ADOS-2, which will be included in future updates of this Review.

Review authors also encountered difficulties when reviewing studies for age of inclusion, with several studies recruiting a wide age range of participants and sometimes reporting results for subgroups of children within the larger sample. Despite the extensive search strategy, lack of available data for GARS, 3di, and DISCO has meant that it is not currently possible to assess the diagnostic test accuracy of all tests as intended.

A potential limitation of this Review is that diagnoses for ASD were grouped from DSM-IV, ICD-10, and earlier classification systems, and this does not directly match *DSM - Fifth Edition* (DSM-5) ASD. This decision was made to ensure that the Review reflected current practice as much as possible. We remain confident that this approach would not have differentially influenced differences in results between studies for the same test or between tests.

Applicability of findings to the review question

Most available data were about single-tool diagnostic test accuracy, allowing indirect comparisons between different types of tests (interview, observation, combined). Limited data were available for direct tool comparisons and combined use of tests. The latter deficiency makes it difficult to definitively answer questions posed by this Review. Another concern is that many studies did not recruit in a way that would reflect the current clinical context in which these tests are used. Nor did the reference diagnosis always meet current recommended standards for making a diagnosis of ASD. Of particular relevance is the high proportion of children with a diagnosis of ASD in the included analyses. Although sensitivity and specificity of these tests will hold in different prevalence populations, the utility of the tests will change. Specifically, there will be risk of overdiagnosis, with a higher proportion of those testing positive not having an ASD when it is used in low prevalence settings.

AUTHORS' CONCLUSIONS

Implications for practice

It is important for a diagnostic test for ASD to have high sensitivity and specificity. A diagnostic test with high sensitivity and low specificity would result in overdiagnosis, consequently placing further strain on already limited resources. Conversely, a diagnostic test with low sensitivity and high specificity could result in missed opportunities for intervention at a crucial period.

From current data, among the three tests with available diagnostic test accuracy data, ADOS has the highest summary sensitivity and similar specificity to CARS and ADI-R. However, there are important caveats to be noted in interpretation of all of these findings, with few high-quality studies reported and most studies having incomplete or uncertain applicability to usual clinical use. It is also important to be aware that the diagnostic test performance of ADOS is acceptable in high prevalence populations; however, if the test is used in low prevalence settings or in settings where children have an associated intellectual disability, there is risk of overdiagnosis. It is not known whether combining tests increases diagnostic test accuracy because only one study investigating this was found and the results were inconclusive.

Each of the reviewed tests recommends that it is not to be used in isolation to make a diagnosis of ASD. Diagnostic test accuracy requirements for tests that are to be used as part of a multi-disciplinary team assessment will be fewer than requirements for those to be used in isolation, as multi-disciplinary team assessment activity will provide opportunities to improve sensitivity and specificity, even though, to our knowledge, there are no reports of this to date. Accepted best practice for this preschool age group is to use a combination of multi-disciplinary assessment (including a paediatrician, a speech pathologist, and a psychologist, with other disciplines included depending on identified abilities and needs) and DSM-5 or ICD-10 criteria when making a diagnosis, and to include information from clinical assessment and from child care or educational settings, as well as results of standardised instruments especially for developmental or intellectual ability (AACAP 2014; NICE 2011). Findings of this Review support currently recommended clinical diagnostic practice, in which addition of a diagnostic test is optional, but could add value given its use in a setting that is likely to have a high prevalence of ASD.

Implications for research

Some studies included in this Review were at high risk of bias, were of uncertain application to clinical care, and did not report findings in a way that is in keeping with best practice for diagnostic test accuracy studies. All future studies should aim to minimise risk of bias, maximise application to clinical care, and provide data in a way that is readily interpretable. In particular, attention should be paid to the reference standard, so that it is consistent with current best practice recommendations.

There is a need for studies in populations that are usually seen by clinicians diagnosing ASD (e.g. consecutive patients suspected of ASD with a mixture of concomitant conditions that might mimic ASD), so that diagnostic test accuracy and clinical utility can be assessed simultaneously. In particular, children with intellectual disability should be included. New versions of tests should be assessed against current diagnostic classification systems applied using best clinical practice.

We also suggest that future studies should work towards diagnostic test accuracy protocols that reflect the stepwise approach to ASD diagnosis that is used clinically, and that if tests are combined, the sequence of administration is reported.

Research is needed if we are to better understand the utility, including added value, of diagnostic accuracy and identification of specific needs that will assist future intervention and management planning of autism diagnostic tests amongst other aspects of multi-disciplinary assessment.

ACKNOWLEDGEMENTS

REFERENCES

References to studies included in this review

Chlebowski 2010 *{published data only}*

Chlebowski C, Green JA, Barton ML, Fein D. Using the Childhood Autism Rating Scale to diagnose autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2010;**40**(7):787-99. [DOI: 10.1007/s10803-009-0926-x; PMC3612531; PUBMED: 20054630]

Corsello 2013 *{published data only}*

Corsello CM, Akshoomoff N, Stahmer AC. Diagnosis of autism spectrum disorders in 2-year-olds: a study of community practice. *Journal of Child Psychology and Psychiatry* 2013;**54**(2):178-85. [DOI: 10.1111/j.1469-7610.2012.02607.x; PMC3505251; PUBMED: 22905987]

Cox 1999 *{published data only}*

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**(5):719-32. [PUBMED: 10433406]

Gray 2008 ADI-R *{published data only}*

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**(4):657-67. [DOI: 10.1007/s10803-007-0432-y]

Gray 2008 ADOS *{published data only}*

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**(4):657-67. [DOI: 10.1007/s10803-007-0432-y]

Kim 2012b ADOS Cohort A *{published data only}*

Kim SH, Lord C. Combining information from multiple sources for the diagnosis of autism spectrum disorders for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Child Psychology and Psychiatry* 2012;**53**(2):143-51. [DOI: 10.1111/j.1469-7610.2011.02458.x; PMC3235227; PUBMED: 21883205]

Kim 2012b ADOS Cohort B *{published data only}*

Kim SH, Lord C. Combining information from multiple sources for the diagnosis of autism spectrum disorders for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Child Psychology and Psychiatry* 2012;**53**(2):143-51. [DOI: 10.1111/j.1469-7610.2011.02458.x; PMC3235227; PUBMED: 21883205]

Le Couteur 2008 ADOS *{published data only}*

Le Couteur A, Haden G, Hammal D, McConachie H. Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: the ADI-R and the ADOS. *Journal of Autism and Developmental Disorders* 2008;**38**(2):362-72. [DOI: 10.1007/s10803-007-0403-3; PUBMED: 17605097]

Lord 2000 *{published data only}*

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**(3):205-23. [PUBMED: 11055457]

Mazefsky 2006 ADOS *{published data only}*

Mazefsky CA, Oswald DP. The discriminative ability and diagnostic utility of the ADOS-G, ADI-R, and GARS for children in a clinical setting. *Autism* 2006;**10**(6):533-49. [DOI: 10.1177/1362361306068505; PUBMED: 17088271]

Oosterling 2010b ADI-R *{published data only}*

Oosterling I, Rommelse N, de Jonge M, van der Gaag RJ, Swinkels S, Roos S, et al. How useful is the Social Communication Questionnaire in toddlers at risk of autism spectrum disorder?. *Journal of Child Psychology and Psychiatry* 2010;**51**(11):1260-8. [DOI: 10.1111/j.1469-7610.2010.02246.x; PUBMED: 20626528]

Oosterling 2010b ADOS *{published data only}*

Oosterling I, Rommelse N, de Jonge M, van der Gaag RJ, Swinkels S, Roos S, et al. How useful is the Social Communication Questionnaire in toddlers at risk of autism spectrum disorder?. *Journal of Child Psychology and Psychiatry* 2010;**51**(11):1260-8. [DOI: 10.1111/j.1469-7610.2010.02246.x; PUBMED: 20626528]

Risi 2006 Study 1 ADOS Cohort A *{published data only}*

Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry* 2006;**45**(9):1094-103.

Risi 2006 Study 1 ADOS Cohort B *{published data only}*

* Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry* 2006;**45**(9):1094-103. [DOI: 10.1097/01.chi.0000227880.42780.0e; PUBMED: 16926617]

Russell 2010 *{published data only}*

Russell PS, Daniel A, Russell S, Mammen P, Abel JS, Raj LE, et al. Diagnostic accuracy, reliability and validity of Childhood Autism Rating Scale in India. *World Journal of Pediatrics* 2010;**6**(2):141-7. [DOI: 10.1007/s12519-010-0029-y; PUBMED: 20490769]

Ventola 2006 ADI-R *{published data only}*

Ventola PE, Kleiman J, Pandey J, Barton M, Allen S, Green J, et al. Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and*

Developmental Disorders 2006;**36**(7):839-47. [DOI: 10.1007/s10803-006-0128-8; PUBMED: 16897398]

Ventola 2006 ADOS *{published data only}*

Ventola PE, Kleiman J, Pandey J, Barton M, Allen S, Green J, et al. Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders* 2006;**36**(7):839-47. [DOI: 10.1007/s10803-006-0128-8; PUBMED: 16897398]

Ventola 2006 CARS *{published data only}*

Ventola PE, Kleiman J, Pandey J, Barton M, Allen S, Green J, et al. Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders* 2006;**36**(7):839-47. [DOI: 10.1007/s10803-006-0128-8; PUBMED: 16897398]

Wiggins 2008 ADI-R *{published data only}*

Wiggins LD, Robins DL. Brief report: excluding the ADI-R behavioral domain improves diagnostic agreement in toddlers. *Journal of Autism and Developmental Disorders* 2008;**38**(5):972-6. [DOI: 10.1007/s10803-007-0456-3; PUBMED: 17879150]

Wiggins 2008 ADOS *{published data only}*

Wiggins LD, Robins DL. Brief report: excluding the ADI-R behavioral domain improves diagnostic agreement in toddlers. *Journal of Autism and Developmental Disorders* 2008;**38**(5):972-6. [DOI: 10.1007/s10803-007-0456-3; PUBMED: 17879150]

Wiggins 2008 CARS *{published data only}*

Wiggins LD, Robins DL. Brief report: excluding the ADI-R behavioral domain improves diagnostic agreement in toddlers. *Journal of Autism and Developmental Disorders* 2008;**38**(5):972-6. [DOI: 10.1007/s10803-007-0456-3; PUBMED: 17879150]

References to studies excluded from this review

Becker 2012 *{published data only}*

Becker MM, Wagner MB, Bosa CA, Schmidt C, Longo D, Papaleo C, et al. Translation and validation of Autism Diagnostic Interview-Revised (ADI-R) for autism diagnosis in Brazil [Tradução e validação da ADI-R (Autism Diagnostic Interview-Revised) para diagnóstico de autismo no Brasil]. *Arquivos de Neuro-Psiquiatria* 2012;**70**(3):185-90. [DOI: 10.1590/S0004-282X2012000300006]

Bölte 2001 *{published data only}*

Bölte S, Poustka F. The factor structure of the Autism Diagnostic Interview-Revised (ADI-R): a study on the dimensional versus the categorical classification of autistic disorders [Die Faktorenstruktur des Autismus Diagnostischen Interviews-Revision (ADI-R): Eine Untersuchung zur dimensionalen versus kategorialen klassifikation autistischer störungen originalarbeit]. *Zeitschrift fur Kinder-und Jugendpsychiatrie und Psychotherapie* 2001;**29**(3):221-9. [DOI: 10.1024//1422-4917.29.3.221; PUBMED: 11524898]

Bölte 2004 *{published data only}*

Bölte S, Poustka F. Diagnostic Observation Scale for autistic disorders: initial results of reliability and validity [Diagnostische Oeobachtungsskala für Autistische Störungen (ADOS): erste ergebnisse zur zuverlässigkeit und gültigkeit]. *Zeitschrift fur Kinder-und Jugendpsychiatrie und Psychotherapie* 2004;**32**(1):45-50. [DOI: 10.1024/1422-4917.32.1.45; PUBMED: 14992047]

Charman 2005 *{published data only}*

Charman T, Taylor E, Drew A, Cockerill H, Brown JA, Baird G. Outcomes at 7 years of children diagnosed with autism at age 2: predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry* 2005;**46**(5):500-13. [DOI: 10.1111/j.1469-7610.2004.00377.x; PUBMED: 15845130]

Charman 2013 *{published data only}*

Charman T, Gotham K. Measurement issues: screening and diagnostic instruments for autism spectrum disorders - lessons from research and practice. *Child and Adolescent Mental Health* 2013;**18**(1):52-63. [DOI: 10.1111/j.1475-3588.2012.00664.x; PMC3607539; PUBMED: 23539140]

Chawarska 2007 *{published data only}*

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* 2007;**48**(2):128-38. [DOI: 10.1111/j.1469-7610.2006.01685.x; PUBMED: 17300551]

Chuthapisith 2012 *{published data only}*

Chuthapisith J, Taycharpipranai P, Ruangdaraganon N, Warrington R, Skuse D. Translation and validation of the Developmental, Dimensional and Diagnostic Interview (3di) for diagnosis of autism spectrum disorder in Thai children. *Autism* 2012;**16**(4):350-6. [DOI: 10.1177/1362361311433770; PUBMED: 22399447]

Clancy 1969 *{published data only}*

Clancy H, Dugdalei A, Rendle-Shortt J. The diagnosis of infantile autism. *Developmental Medicine and Child Neurology* 1969;**11**(4):432-42. [PUBMED: 5805347]

Constantino 2012 *{published data only}*

Constantino JN, Zhang Y, Abbacchi AM, Calhoun A, Scofield F, Grafeman SJ. Rapid phenotyping of autism spectrum disorders: inclusion of direct observation in feasible paradigms for clinical assessment. *Neuropsychiatry* 2012;**2**(3):203-12. [DOI: 10.2217/NPY.12.28]

de Bildt 2004 *{published data only}*

de Bildt A, Sytema S, Ketelaars C, Kraijer D, Mulder E, Volkmar F, et al. 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**(2):129-37. [PUBMED: 15162932]

de Bildt 2009 *{published data only}*

de Bildt A, Sytema S, van Lang NDJ, Minderaa RB, van Engeland H, de Jonge MV. Evaluation of the ADOS revised algorithm: the applicability in 558 Dutch children and adolescents. *Journal of Autism and Developmental Disorders* 2009;**39**(9):1350-8. [DOI: 10.1007/s10803-009-0749-9; PMC2727366 ; PUBMED: 19452268]

de Bildt 2013 *{published data only}*

de Bildt A, Oosterling IJ, van Lang NDJ, Kuijper S, Dekker V, Sytema S, et al. How to use the ADI-R for classifying autism spectrum disorders? Psychometric properties of criteria from the literature in 1204 Dutch children. *Journal of Autism and Developmental Disorders* 2013;**43**(10):2280–94. [DOI: 10.1007/s10803-013-1783-1; PUBMED: 23397166]

Diken 2012a *{published data only}*

Diken IH, Ardiç A, Diken Ö, Gilliam JE. Exploring the validity and reliability of Turkish version of Gilliam Autism Rating Scale-2: Turkish standardization study. *Eğitim ve Bilim (Education and Science)* 2012;**37**(166):318-28.

Diken 2012b *{published data only}*

Diken IH, Diken O, Gilliam JE, Ardic A, Sweeney D. Validity and reliability of Turkish version of Gilliam Autism Rating Scale-2: results of preliminary study. *International Journal of Special Education* 2012;**27**(2):207-15. [files.eric.ed.gov/fulltext/EJ982874.pdf]

DiLalla 1994 *{published data only}*

DiLalla DL, Rogers SJ. Domains of the Childhood Autism Rating Scale: relevance for diagnosis and treatment. *Journal of Autism and Developmental Disorders* 1994;**24**(2):115-28. [PUBMED: 8040157]

Eaves 2006 *{published data only}*

Eaves RC, Wood-Groves S, Williams TO Jr, Fall A-M. Reliability and validity of the Pervasive Developmental Disorders Rating Scale and the Gilliam Autism Rating Scale. *Education and Training in Developmental Disabilities* 2006;**41**(3):300-9. [www.jstor.org/stable/23880203]

Falkmer 2013 *{published data only}*

Falkmer T, Anderson K, Falkmer M, Horlin C. Diagnostic procedures in autism spectrum disorders: a systematic literature review. *European Child & Adolescent Psychiatry* 2013;**22**(6):329–40. [DOI: 10.1007/s00787-013-0375-0; PUBMED: 23322184]

Fisch 2012 *{published data only}*

Fisch GS. Autism and epistemology III: child development, behavioral stability, and reliability of measurement. *American Journal of Medical Genetics Part A* 2012;**158a**(5):969-79. [DOI: 10.1002/ajmg.a.35269; PUBMED: 22488915]

Goldfischer 2001 *{unpublished data only}*

Goldfischer HM. Improving the Diagnositc Utility of the Childhood Autism Rating Scale [PhD Thesis]. Oxford (OH): Miami University, 2001.

Gotham 2007 *{published data only}*

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**(4):613-27. [DOI: 10.1007/s10803-006-0280-1; PUBMED: 17180459]

Gotham 2008 *{published data only}*

Gotham K, Risi S, Dawson G, Tager-Flusberg H, Joseph R, Carter A, et al. A replication of the Autism Diagnostic Observation Schedule (ADOS) revised algorithms. *Journal of the American Academy of Child and Adolescent Psychiatry* 2008;**47**(6):642-51. [DOI: 10.1097/CHI.0b013e31816bffb7; PMC3057666; PUBMED: 18434924]

Guthrie 2013 *{published data only}*

Guthrie W, Swineford LB, Nottke C, Wetherby AM. Early diagnosis of autism spectrum disorder: stability and change in clinical diagnosis and symptom presentation. *Journal of Child Psychology and Psychiatry* 2013;**54**(5):582-90. [DOI: 10.1111/jcpp.12008; PMC3556369; PUBMED: 23078094]

Huerta 2012 *{published data only}*

Huerta M, Lord C. Diagnostic evaluation of autism spectrum disorders. *Pediatric Clinics of North America* 2012;**59**(1):103-11. [DOI: 10.1016/j.pcl.2011.10.018; NIHMS331663; PMC3269006; PUBMED: 22284796]

Jackson 2012 *{unpublished data only}*

Jackson LS. Translation, Validation, and Norm-Referencing of the Spanish Adaptation of the Gilliam Autism Rating Scale-2 [PhD thesis]. Minneapolis (MN): Walden University, 2012.

Kamp-Becker 2013 *{published data only}*

Kamp-Becker I, Ghahreman M, Heinzel-Gutenbrunner M, Peters M, Remschmidt H, Becker K. Evaluation of the revised algorithm of Autism Diagnostic Observation Schedule (ADOS) in the diagnostic investigation of high-functioning children and adolescents with autism spectrum disorders. *Autism* 2013;**17**(1):87-102. [DOI: 10.1177/1362361311408932; PUBMED: 21610187]

Kim 2012a *{published data only}*

Kim SH, Lord C. New Autism Diagnostic Interview - Revised algorithms for toddlers and young preschoolers from 12-47 months of age. *Journal of Autism and Developmental Disorders* 2012;**42**(1):82-93. [DOI: 10.1007/s10803-011-1213-1]

Kim 2013 *{published data only}*

Kim SH, Thurm A, Shumway S, Lord C. Multisite study of New Autism Diagnositc Interview-Revised (ADI-R) algorithms for toddlers and young preschoolers. *Journal of Autism and Developmental Disorders* 2013;**43**(7):1527-38. [DOI: 10.1007/s10803-012-1696-4; NIHMS419066; PMC3594108; PUBMED: 23114567]

Klose 2012 *{published data only}*

Klose L, Plotts C, Kozeneski N, Skinner-Foster J. A review of assessment tools for diagnosis of autism spectrum disorders: implications for school practice. *Assessment for Effective Intervention* 2012;**37**:236-42.

Lecavalier 2006 *{published data only}*

Lecavalier L, Aman MG, Scahill L, McDougle CJ, McCracken JT, Vitiello B, et al. Validity of the Autism Diagnostic Interview-Revised. *American Journal of Mental Retardation* 2006;**111**(3):199-215. [DOI: 10.1352/0895-8017(2006)111[199:VOTADI]2.0.CO;2; PUBMED: 16597187]

Le Couteur 1989 *{published data only}*

Le Couteur A, Rutter M, Lord C, Rios P, Robertson S, Holdgrafer M, et al. Autism Diagnostic Interview: a standardized investigator-based instrument. *Journal of Autism and Developmental Disorders* 1989;**19**(3):363-87. [PUBMED: 2793783]

Le Couteur 2008 ADI-R *{published data only}*

Le Couteur A, Haden G, Hammal D, McConachie H. Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: the ADI-R and the ADOS. *Journal of Autism and Developmental Disorders* 2008;**38**(2):362-72. [DOI: 10.1007/s10803-007-0403-3; PUBMED: 17605097]

Leekam 2002 *{published data only}*

Leekam R, 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* 2002;**43**(3):327-42. [PUBMED: 11944875]

Li 2005 *{published data only}*

Li J-H, Zhong J-M, Cai L-Y, Chen Y, Zhou M-Z. Comparison of clinical application of three autism rating scales. *Chinese Journal of Contemporary Pediatrics* 2005;**7**(1):59-62.

Lord 1993 *{published data only}*

Lord C, Storoschuk S, Rutter M, Pickles A. Using the ADI-R to diagnose autism in preschool children. *Infant Mental Health Journal* 1993;**14**(3):234-52. [DOI: 10.1002/1097-0355(199323)14:3<234::AID-IMHJ2280140308>3.0.CO;2-F]

Lord 1994 *{published data only}*

Lord C, Rutter M, Le Couteur A. 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**(5):659-85. [PUBMED: 7814313]

Lord 2006 *{published data only}*

Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Pickles A. Autism from 2 to 9 years of age. *Archives of General Psychiatry* 2006;**63**(6):694-701. [DOI: 10.1001/archpsyc.63.6.694; PUBMED: 16754843]

Lord 2012a *{published data only}*

Lord C, Petkova E, Hus V, Gan W, Lu F, Martin DM, et al. A multisite study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry* 2012;**69**(3):306-13. [DOI: 10.1001/archgenpsychiatry.2011.148; PMC3626112; PUBMED: 22065253]

Lozowski-Sullivan 2011 *{unpublished data only}*

Lozowski-Sullivan S. Psychometric Properties of Diagnostic Assessment Instruments for Autism Spectrum Disorders in a Community Sample Aged 2 Through 17 Years [PhD thesis]. Kalamazoo (MI): Western Michigan University, 2011. [scholarworks.wmich.edu/cgi/viewcontent.cgi?article=1436&context=dissertations]

Luyster 2009 *{published data only}*

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 autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2009;**39**(9):1305-20.

Maljaars 2012 *{published data only}*

Maljaars JN, Scholte I, Van Berckelaer-Onnes E. Evaluation of the criterion and convergent validity of the Diagnostic Interview for Social and Communication Disorders in young and low-functioning children. *Autism* 2012;**16**(5):487-97.

Matson 2010 *{published data only}*

Matson J, Hess J, Mahan S, Fodstad JC. Convergent validity of the Autism Spectrum Disorder-Diagnostic for Children (ASD-DC) and Autism Diagnostic Interview-Revised (ADI-R). *Research in Autism Spectrum Disorders* 2010;**4**(4):741-5.

Mayes 2009 *{published data only}*

Mayes SD, Calhoun SL, Murray MJ, Morrow JD, Yurich KKL, Mahr F, et al. Comparison of scores on the Checklist for Autism Spectrum Disorder, Childhood Autism Rating Scale, and Gilliam Asperger's Disorder Scale for children with low functioning autism, high functioning autism, Asperger's disorder, ADHD, and typical development. *Journal of Autism and Developmental Disorders* 2009;**39**(12):1682-93.

Mayes 2012 *{published data only}*

Mayes SD, Calhoun SL, Murray MJ, Morrow JD, Yurich KKL, Cothren S, et al. Use of the Childhood Autism Rating Scale (CARS) for children with high functioning autism or Asperger syndrome. *Focus on Autism and Other Developmental Disabilities* 2012;**27**(1):31-8.

Mazefsky 2006 ADI-R *{published data only}*

Mazefsky CA, Oswald DP. The discriminative ability and diagnostic utility of the ADOS-G, ADI-R, and GARS for children in a clinical setting. *Autism* 2006;**10**(6):533-49.

Mazefsky 2006 GARS *{published data only}*

Mazefsky CA, Oswald DP. The discriminative ability and diagnostic utility of the ADOS-G, ADI-R, and GARS for children in a clinical setting. *Autism* 2006;**10**(6):533-49.

McGarry Klose 2012 *{published data only}*

McGarry Klose L, Plotts C, Kozeneski N, Skinner Foster J. A review of assessment tools for diagnosis of autism spectrum disorders: implications for school practice. *Assessment for Effective Intervention* 2012;**37**(4):236-42.

Mick 2007 *{published data only}*

Mick K. Diagnosing autism: comparison of the Childhood Autism Rating Scale (CARS) and the Autism Diagnostic Observation Schedule (ADOS). ProQuest Information and Learning 2007.

Molloy 2011 *{published data only}*

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):143-62.

Moss 2008 *{published data only}*

Moss J, Magiati I, Charman T, Howlin P. Stability of the Autism Diagnostic Interview – Revised from pre-school to elementary school age in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 2008;**38**:1081-91.

Nordin 1996 *{published data only}*

Nordin V, Gillberg C. Autism spectrum disorders in children with physical or mental disability or both. II: screening aspects. *Developmental Medicine and Child Neurology* 1996;**38**(4):314-24. [PUBMED: 8641536]

Noterdaeme 1999 *{published data only}*

Noterdaeme M, Kurz U, Mildenberger K, Sitter S, Amorosa H. Exclusion of receptive speech disorders with the ADOS (Autism Diagnostic Observation Schedule) [Ausschluß rezeptiver sprachstörungen mittels des ADOS (Autism Diagnostic Observation Schedule)]. *Zeitschrift fur Kinder-und Jugendpsychiatrie und Psychotherapie* 1999;**27**(4):251-7. [DOI: 10.1024//1422-4917.27.4.251; PUBMED: 10637975]

Nygren 2009 *{published data only}*

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**(5):730–41. [DOI: 10.1007/s10803-008-0678-z; PUBMED: 19148741]

Oosterling 2010a *{published data only}*

Oosterling I, Roos S, de Bildt A, Rommelse N, de Jonge M, Visser J, et al. Improved diagnostic validity of the ADOS revised algorithms: a replication study in an independent sample. *Journal of Autism and Developmental Disorders* 2010;**40**(6):689-703. [DOI: 10.1007/s10803-009-0915-0; PMC2864898; PUBMED: 20148299]

Papanikolaou 2009 *{published data only}*

Papanikolaou K, Paliokosta E, Houliaras G, Vgenopoulou S, Giouroukou E, Pehlivanidis A, et al. Using the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule-Generic for the diagnosis of autism spectrum disorders in a Greek sample with a wide range of intellectual abilities. *Journal of Autism and Developmental Disorders* 2009;**39**(3):414-20. [DOI: 10.1007/s10803-008-0639-6; PUBMED: 18752062]

Perry 2005 *{published data only}*

Perry A, Condillac RA, Freeman NL, Dunn-Geier J, Belair J. Multi-site study of the Childhood Autism Rating Scale (CARS) in five clinical groups of young children. *Journal of Autism and Developmental Disorders* 2005;**35**(5):625-34. [DOI: 10.1007/s10803-005-0006-9; PUBMED: 16172810]

Rellini 2004 *{published data only}*

Rellini E, Tortolani D, Trillo S, Carbone S, Montecchi F. Childhood Autism Rating Scale (CARS) and Autism Behavior Checklist (ABC) correspondence and conflicts with DSM-IV criteria in diagnosis of autism. *Journal of Autism and Developmental Disorders* 2004;**34**(6):703-8. [PUBMED: 15679189]

Risi 2006 study 1 ADI-R *{published data only}*

Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry* 2006;**45**(9):1094-103. [DOI: 10.1097/01.chi.0000227880.42780.0e; PUBMED: 16926617]

Risi 2006 study 2 *{published data only}*

Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry.* 2006;**45**(9):1094-103. [DOI: 10.1097/01.chi.0000227880.42780.0e; PUBMED: 16926617]

Rutter 1988a *{published data only}*

Rutter M, Schopler E. Autism and pervasive developmental disorders: concepts and diagnostic issues. In: Schopler E, Mesibov GB editor(s). Diagnosis and Assessment in Autism. New York (NY): Springer Science + Business Media, 1988:15-36. [ISBN 978-1-4899-0794-3]

Rutter 1988b *{published data only}*

Rutter M, LeCouteur A, Lord C, MacDonald H, Rios P, Folstein S. Diagnosis and subclassification of autism: concepts and instrument development. In: Schopler E, Mesibov GB editor(s). Diagnosis and Assessment in Autism. New York (NY): Springer Science + Business Media, 1988:239-59. [ISBN 978-1-4899-0794-3]

Saemundsen 2003 *{published data only}*

Saemundsen E, Magnússon P, Smári J, Sigurdardóttir S. Autism Diagnostic Interview-Revised and the Childhood Autism Rating Scale: convergence and discrepancy in diagnosing autism. *Journal of Autism and Developmental Disorders* 2003;**33**(3):319-28. [PUBMED: 12908834]

Schopler 1988 *{published data only}*

Schopler E, Mesibov GB. Introduction to diagnosis and assessment of autism. In: Schopler E, Mesibov GB editor(s). Diagnosis and Assessment in Autism. New York (NY): Springer Science + Business Media, 1988:3-14. [ISBN 978-1-4899-0794-3]

Sevin 1991 *{published data only}*

Sevin J, Matson JL, Coe DA, Fee VE, Sevin BM. A comparison and evaluation of three commonly used autism scales. *Journal of Autism and Developmental Disorders* 1991;**21**(4):417-32. [PUBMED: 1778958]

Shin 1998 *{published data only}*

Shin MS, Kim Y-H. Standardization study for the Korean version of the Childhood Autism Rating Scale: reliability,

validity and cut-off score. *Korean Journal of Clinical Psychology* 1998;**17**(1):1-15.

Sikora 2008 *{published data only}*

Sikora D, Hall TA, Hartley SL, Gerrard-Morris AE, Cagle S. Does parent report of behavior differ across ADOS-G classifications: analysis of scores from the CBCL and GARS. *Journal of Autism and Developmental Disorders* 2008;**38**(3):440-8. [DOI: 10.1007/s10803-007-0407-z; PUBMED: 17619131]

Skuse 2004 *{published data only}*

Skuse D, Warrington R, Bishop D, Chowdhury U, Lau J, Mandy W, et al. 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):548-58. [DOI: 10.1097/00004583-200405000-00008; PUBMED: 15100561]

Soke 2011 *{published data only}*

Soke GN, Philofsky A, Diguiseeppi C, Lezotte D, Rogers S, Hepburn S. Longitudinal changes in scores on the Autism Diagnostic Interview - Revised (ADI-R) in preschool children with autism: implications for diagnostic classification and symptom stability. *Autism* 2011;**15**(5):545-62. [DOI: 10.1177/1362361309358332; PMC4426200; PUBMED: 21586639]

Stella 2002 *{published data only}*

Stella JL. Predictive Validity of the Factor Structure of the Childhood Autism Rating Scale [PhD thesis]. Coral Gables (FL): University of Miami, 2002.

Tachimori 2003 *{published data only}*

Tachimori H, Osada H, Kurita H. Childhood Autism Rating Scale-Tokyo version for screening pervasive developmental disorders. *Psychiatry and Clinical Neurosciences* 2003;**57**(1):113-8. [DOI: 10.1046/j.1440-1819.2003.01087.x; PUBMED: 12519463]

Teal 1982 *{unpublished data only}*

Teal MLB. A Validity Analysis of Selected Instruments Used to Assess Autism [PhD thesis]. Denton (TX): Texas Woman's University, 1981.

Tsuchiya 2013 *{published data only}*

Tsuchiya KJ, Matsumoto K, Yagi A, Inada N, Kuroda M, Inokuchi E, et al. Reliability and validity of Autism Diagnostic Interview-Revised, Japanese Version. *Journal of Autism and Developmental Disorders* 2013;**43**(3):643-62. [DOI: 10.1007/s10803-012-1606-9; PUBMED: 22806002]

Vaughan 2011 *{published data only}*

Vaughan CA. Test review: E Schopler, ME Van Bourgondien, GJ Wellman and SR Love Childhood Autism Rating Scale (2nd edition). Los Angeles (CA): Western Psychological Services, 2010. *Journal of Psychoeducational Assessment* 2011;**29**(5):489-93. [DOI: 10.1177/0734282911400873]

Ventola 2007 *{published data only}*

Ventola P, Kleinman J, Pandey J, Wilson L, Esser E, Boorstein H, et al. Differentiating between autism spectrum disorders and other developmental disabilities in children who failed a screening instrument for ASD. *Journal of Autism and Developmental Disorders* 2007;**37**(3):425-36. [DOI: 10.1007/s10803-006-0177-z; PUBMED: 16897377]

Zwaigenbaum 2011 *{published data only}*

Zwaigenbaum L. Assessment and diagnosis of autism spectrum disorders. *BMJ* 2011;**343**:d6628. [DOI: 10.1136/bmj.d6628; PUBMED: 22021469]

Additional references

AACAP 2014

Volkmar F, Siegel M, Woodbury-Smith M, King B, McCracken J, State M, et al. Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* 2014;**53**(2):237–57. [DOI: 10.1016/j.jaac.2013.10.013; PUBMED: 24472258]

Academy of Medicine Singapore 2010

Academy of Medicine Singapore - Ministry of Health Clinical Practice Guidelines Workgroup on Autism Spectrum Disorders. Academy of Medicine Singapore - Ministry of Health Clinical Practice Guidelines: autism spectrum disorders in pre-school children. *Singapore Medical Journal* 2010;**51**(3):255-63. [PUBMED: 20428749]

AHRQ 2011

Weitlauf AS, McPheeters ML, Peters B, Sathe N, Travis R, Aiello R, et al. Therapies for children with autism spectrum disorders: behavioral interventions update [internet]. www.ncbi.nlm.nih.gov/books/NBK241433/ (accessed before 26 April 2018).

Akshoomoff 2006

Akshoomoff N. Use of the Mullen Scales of Early Learning for the assessment of young children with autism spectrum disorders. *Child Neuropsychology* 2006;**12**(4-5):269-77. [DOI: 10.1080/09297040500473714; NIHMS8885; PMC1550495; PUBMED: 16911972]

APA 1980

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-III). 3rd Edition. Washington (DC): American Psychiatric Association, 1980.

APA 1987

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM- III-R). 3rd Edition. Washington (DC): American Psychiatric Association, 1987.

APA 1994

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). 4th Edition. Washington (DC): American Psychiatric Association, 1994.

APA 2000

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR). 4th Edition. Washington (DC): American Psychiatric Association, 2000.

APA 2013

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th Edition. Washington (DC): American Psychiatric Association, 2013.

Atladottir 2015

Atladottir HO, Gyllenberg D, Langridge A, Sandin S, Hansen SN, Leonard H, et al. The increasing prevalence of reported diagnoses of childhood psychiatric disorders: a descriptive multinational comparison. *European Child & Adolescent Psychiatry* 2015;**24**(2):173-83. [DOI: 10.1007/s00787-014-0553-8; PUBMED: 24796725]

Bates 2015

Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 2015;**67**(1):1-48. [DOI: 10.18637/jss.v067.i01; cran.r-project.org/web/packages/lme4/citation.html]

CDP 2016

Centers for Disease Control and Prevention. Prevalence of autism spectrum disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. www.cdc.gov/mmwr/pdf/ss/ss6302.pdf (accessed before 16 April 2018).

Cederlund 2008

Cederlund M, Hagberg B, Billstedt E, Gillberg C, Gillberg C. Asperger syndrome and autism: a comparative longitudinal follow-up study more than 5 years after original diagnosis. *Journal of Autism and Developmental Disorders* 2008;**38**(1):72–85. [DOI: 10.1007/s10803-007-0364-6; PUBMED: 17340200]

Elsabbagh 2012

Elsabbagh M, Divan G, Koh YJ, Kim YS, Kauchali S, Marcín C, et al. Global prevalence of autism and other pervasive developmental disorders. *Autism Research* 2012;**5**(3):160-79. [DOI: 10.1002/aur.239; PMC3763210; PUBMED: 22495912]

Filipek 1999

Filipek PA, Accardo PJ, Baranek GT, Cook EH Jr, Dawson G, Gordon B, et al. The screening and diagnosis of autistic spectrum disorders. *Journal of Autism and Developmental Disorders* 1999;**29**(6):439-84. [PUBMED: 10638459]

Filipek 2000

Filipek PA, Accardo PJ, Ashwal S, Baranek GT, Cook EH Jr, Dawson G, et al. Practice parameter: screening and diagnosis of autism: report of the Quality Standards Subcommittee of the American Academy of Neurology and Child Neurology Society. *Neurology* 2000;**55**(4):468-79. [PUBMED: 10953176]

Fombonne 2009

Fombonne E. Epidemiology of pervasive developmental disorders. *Pediatric Research* 2009;**65**(6):591-8. [DOI: 10.1203/PDR.0b013e31819e7203; PUBMED: 19218885]

Gillberg 1989

Gillberg IC, Gillberg C. Asperger syndrome - some epidemiological considerations: a research note. *Journal of Child Psychology and Psychiatry* 1989;**30**(4):631-8. [PUBMED: 2670981]

Gilliam 1995

Gilliam JE. The Gilliam Autism Rating Scale: GARS. Austin (TX): Pro-Ed, 1995.

Gilliam 2006

Gilliam JE. Gilliam Autism Rating Scale (2nd Edition): GARS 2. Austin (TX): Pro-Ed, 2006.

Gilliam 2013

Gilliam JE. Gilliam Autism Rating Scale (3rd Edition): GARS 3. Austin (TX): Pro-Ed, 2013.

Howlin 2004

Howlin P, Goode S, Hutton J, Rutter M. Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry* 2004;**45**(2):212-29. [PUBMED: 14982237]

Johnson 2007

Johnson CP, Myers SM, Council on Children With Disabilities. Identification and evaluation of children with autism spectrum disorders. *Pediatrics* 2007;**120**(5):1183-215. [DOI: 10.1542/peds.2007-2361]

Kanner 1957

Kanner L, Eisenberg L. Early infantile autism, 1943-1955. *Psychiatric Research Reports of the American Psychiatric Association* 1957;**26**(7):55-65. [PUBMED: 13432078]

Lord 1994a

Lord C, Rutter M, Le Couteur A. 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**(5):659-85. [PUBMED: 7814313]

Lord 1999

Lord C, Rutter M, DiLavore PC, Risi S. Autism Diagnostic Observation Schedule (ADOS): Manual. Los Angeles (CA): WPS, 1999.

Lord 2000a

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**(3):205-23. [PUBMED: 11055457]

Lord 2012b

Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Bishop SL, et al. Autism Diagnostic Observation Schedule: ADOS-2. 2nd Edition. Los Angeles (CA): Western Psychological Services, 2012.

Matson 2008

Matson JL, Wilkins J, González M. Early identification and diagnosis in autism spectrum disorders in young children and infants: how early is too early?. *Research in Autism Spectrum Disorders* 2008;**2**(1):75-84. [DOI: 10.1016/j.rasd.2007.03.002]

Mazefsky 2006a

Mazefsky CA, Oswald DP. The discriminative ability and diagnostic utility of ADOS-G, ADI-R and GARS for children in a clinical setting. *Autism* 2006;**10**(6):533-49. [DOI: 10.1177/1362361306068505; PUBMED: 17088271]

Mazefsky 2013

Mazefsky CA, McPartland JC, Gastgeb HZ, Minshew NJ. Brief report: comparability of DSM-IV and DSM-5 ASD research samples. *Journal of Autism and Developmental Disorders* 2013;**43**(5):1236-42. [DOI: 10.1007/s10803-012-1665-y; PMC3635090; PUBMED: 23011251]

McConachie 2005

McConachie H, Randle V, Hammal D, Le Couteur A. A controlled trial of a training course for parents of children with suspected autism spectrum disorder. *Journal of Pediatrics* 2005;**147**(3):335-40. [DOI: 10.1016/j.jpeds.2005.03.056; PUBMED: 16182672]

Ministry of Health New Zealand 2008

Ministry of Health New Zealand. New Zealand autism spectrum disorder guideline. www.health.govt.nz/publication/new-zealand-autism-spectrum-disorder-guideline (accessed 14 October 2014).

Ministry of Health Singapore 2010

Ministry of Health Singapore. Autism Spectrum Disorders in Pre-School Children. www.moh.gov.sg/content/moh_web/home/Publications/guidelines/cpg/2010/autism_spectrum_disorders_in_pre-school_children.html (accessed 26 April 2018).

Missouri Autism Guidelines Initiative 2010

Missouri Autism Guidelines Initiative. Autism spectrum disorders: Missouri best practice guidelines for screening, diagnosis and assessment. www.autismguidelines.dmh.mo.gov/pdf/Guidelines.pdf (accessed 14 October 2014).

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine* 2009;**6**(7):e1000097. [DOI: 10.1371/journal.pmed.1000097; PMC2707599; PUBMED: 19621072]

Monteiro 2015

Monteiro SA, Spinks-Franklin A, Treadwell-Deering D, Berry L, Sellers-Vinson S, Smith E, et al. Prevalence of autism spectrum disorder in children referred for diagnostic autism evaluation. *Clinical Pediatrics* 2015;**54**(14):1322-7. [DOI: 10.1177/0009922815592607; PUBMED: 26130396]

New York State Department of Health 2005

New York State Department of Health. Report of the recommendations - autism / pervasive developmental disorders: assessment and intervention for young children (age 0-3 years). www.health.state.ny.us/community/infants_children/early_intervention/disorders/autism/index.htm#Table_of_Contents (accessed 14 October 2014).

NICE 2011

National Institute for Health and Care Excellence. Autism spectrum disorder in under 19s: recognition, referral and diagnosis. www.nice.org.uk/guidance/cg128 (accessed before 15 May 2017).

NICE 2013

National Institute for Health and Care Excellence. Autism spectrum disorder in under 19s: support and management. www.nice.org.uk/guidance/cg170 (accessed before 15 May 2017).

Ohio Developmental Disabilities Council 2010

Ohio Developmental Disabilities Council 2010. Autism: reaching for a brighter future. Service guidelines for individuals with autism spectrum disorder through the lifespan. www.ocali.org/up_doc/Autism_Service_Guidelines.pdf (accessed before 9 May 2018).

Oosterling, I (2015)

Oosterling, i. Personal communication 2015.

Reitsma 2005

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005;**58**(10):982-90. [DOI: 10.1016/j.jclinepi.2005.02.022; PUBMED: 16168343]

Robins 2001

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**(2):131-44.

Rutter 2003

Rutter M, Le Couteur A, Lord C. Autism Diagnostic Interview - Revised: Manual. Los Angeles (CA): Western Psychological Services, 2003.

Schopler 1980

Schopler E, Reichler RJ, DeVellis RF, Daly K. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders* 1980;**10**(1):91-103. [PUBMED: 6927682]

Schopler 1986

Schopler E, Reichler RJ, Renner BR. The Childhood Autism Rating Scale (CARS): For Diagnostic Screening and Classification of Autism. New York (NY): Irvington, 1986.

Schopler 2010

Schopler E, Van Bourgondien ME, Wellman GJ, Love SR, Western Psychological Services. The Childhood Autism Rating Scale, Second Edition (CARS-2). Torrance (CA): WPS, 2010.

Shearer 2001

Shearer H. Executive Function and Autistic Symptomatology in Very Young Children [PhD thesis]. Durham (UK): University of Durham, 2001.

SIGN 2007

Scottish Inercollegiate Guidelines Network. Assessment, diagnosis and clinical interventions for children and young people with autism spectrum disorders: a national clinical guideline. www.autismeurope.org/wp-content/uploads/2017/08/Assessment-diagnosis-and-clinical-interventions-for-children-and-young-people-with-ASD.pdf (accessed 14 October 2014).

Skuse 2004a

Skuse D, Warrington R, Bishop D, Chowdhury U, Lau J, Mandy W, et al. 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):548-58. [DOI: 10.1097/00004583-200405000-00008; PUBMED: 15100561]

South 2002

South M, Williams BJ, McMahon WM, Owley T, Filipek PA, Shernoff E, et al. Utility of the Gilliam Autism Rating Scale in research and clinical populations. *Journal of Autism and Developmental Disorders* 2002;**32**(6):593-9. [PUBMED: 12553595]

StataCorp 2007 [Computer program]

StatCorp. Stata Statistical Software: Release 10. College Station (TX): StataCorp LP, 2007.

Stewart 2014

Stewart JR, Vigil DC, Ryst E, Yang W. Refining best practices for the diagnosis of autism: a comparison between individual healthcare practitioner diagnosis and transdisciplinary assessment. *Nevada Journal of Public Health* 2014;**11**(1):1-13. [pdfs.semanticscholar.org/8ebe/16b8c745c85d8f04791d96a3b37174b770b1.pdf]

Takwoingi 2017

Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Statistical Methods in Medical Research* 2017;**26**(4):1896-911. [DOI: 10.1177/0962280215592269; PMC5564999; PUBMED: 26116616]

Volkmar 2014

Volkmar F, Siegel M, Woodbury-Smith M, King B, McCracken J, State M, et al. Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* 2014;**53**(2):237-57. [DOI: 10.1016/j.jaac.2013.10.013; PUBMED: 24472258]

Watkins 2014

Watkins E. The Gender of Participants in Published Research Involving People with Autism Spectrum Disorders [Masters thesis]. Michigan (MI): Western Michigan University, 2014. [scholarworks.wmich.edu/masters_theses/482]

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies.

Annals of Internal Medicine 2011;**155**(8):529-36. [DOI: 10.7326/0003-4819-155-8-201110180-00009; PUBMED: 22007046]

WHO 1992

World Health Organization. The ICD-9 Classification of Mental and Behavioural Disorders. Geneva (CH): World Health Organization, 1992.

WHO 2007

World Health Organization. International Statistical Classification of Diseases and Related Health Problems - 10th Revision. apps.who.int/classifications/apps/icd/icd10online (accessed 14 October 2014).

Williams 2013

Williams KJ. Prevalence of autism spectrum disorders by the age of eight in one region, 2007-2011. International Autism-Europe Congress: New Dimensions for Autism; 2013 Sept 26-28; Budapest, Hungary. 2013:33. [www.autismeurope.org/wp-content/uploads/2017/08/autism-europe-congress-2013-programme.pdf]

Wing 1979

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 2002

Wing L, Leekam SR, Libby SJ, Gould J, Larcombe M. The Diagnostic Interview for Social and Communication Disorders: background, inter-rater reliability and clinical use. *Journal of Child Psychology and Psychiatry* 2002;**43**(3):307-25. [PUBMED: 11944874]

Wing 2006

Wing L. Diagnostic Interview for Social and Communication Disorders. 11th Edition. Bromley (UK): Centre for Social and Communication Disorders, 2006.

Zwaigenbaum 2009

Zwaigenbaum L, Bryson S, Lord C, Rogers S, Carter A, Carver L, et al. Clinical assessment and management of toddlers with suspected autism spectrum disorder: Insights from studies of high-risk infants. *Pediatrics* 2009;**123**(5):1383-91. [DOI: 10.1542/peds.2008-1606; PMC2833286; PUBMED: 19403506]

References to other published versions of this review

Samtani 2011

Samtani A, Sterling-Levis K, Scholten RJPM, Woolfenden S, Hooft L, Williams K. Diagnostic tests for autism spectrum disorders (ASD) in preschool children. *Cochrane Database of Systematic Reviews* 2011, Issue 3. [DOI: 10.1002/14651858.CD009044]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Chlebowski 2010

Study characteristics	
Patient sampling	Prospective cohort study of children who failed a screening evaluation and telephone follow-up about developmental concerns. All children were invited for a developmental evaluation at 2 years of age and a re-evaluation of development at 4 years of age. Although some children with normal development were identified, this method of sampling is consistent with a clinic providing a service to children with developmental concerns
Patient characteristics and setting	**Number of study groups:** 2 (group 1 = 2 years of age; group 2 = children re-evaluated at 4 years of age, with subset of 173 children being assessed at both time points) **Number of participants:** 606 (group 1 = 376 children; group 2 = 230 children) **Diagnosis:** group 1 = autistic disorder (n = 142); pervasive developmental disorder not otherwise specified (n = 101); non-autism spectrum disorder (n = 95); no diagnosis (n = 38); group 2 = autistic disorder (n = 104); pervasive developmental disorder not otherwise specified (n = 44); non-autism spectrum disorder (n = 34); no diagnosis (n = 48) **Comorbidity:** NS **Age:** group 1: range = 21-30 months (equivalent to 1.7-2.5 years); group 2: 42-66 months (equivalent to 3.5-5.5 years) **Sex:** group 1: 296 males, 80 females; group 2: 186 males, 44 females **Ethnicity:** NS **Inclusion criteria:** children who failed the Modified Checklist for Autism in Toddlers **Exclusion criteria:** NS **Location:** University of Connecticut **Setting:** university psychological service clinic **Training assessor:** NS **Method of participant selection:** Participants were part of a large screening study, and all were children who failed the M-CHAT and received a follow-up telephone call and a developmental evaluation. Some children included in the analysis were known to be developing normally
Index tests	CARS, which was administered by a licensed psychologist. Cutoff of 32 in the 2-year-old sample and 30 in the 4-year-old sample; hence, 2-year-old cutoff not consistent with clinical use of the tool
Target condition and reference standard(s)	**Target condition:** ASD (AD, PDD-NOS). Non-ASD diagnosis consisting of children with diagnoses of intellectual disability, global developmental delay, developmental language disorder, or other DSM-IV-TR diagnoses. No diagnosis consisting of children who did not meet criteria for any DSM-IV-TR diagnoses, nor children judged to be typically developing by clinicians in the study **Reference standard (type):** DSM-IV-TR **Procedure for diagnosis:** made by 1 licensed clinician (psychologist or developmental paediatrician). The clinician who completed CARS also made the clinical diagnosis for some children, which likely inflated the relationship between CARS scores and clinical diagnoses. CARS, ADOS, ADI-R, and MSEL (without the gross motor subscale) assessments completed for diagnosis, and parent interview and direct observation of the child used to inform diagnosis
Flow and timing	**Administration:** All children underwent diagnosis using the reference standard; CARS was then administered to all children

Chlebowski 2010 *(Continued)*

Duration: NS

Timing of assessment: It is not stated how long apart the DSM-IV-TR was administered before CARS at each time point

Missing data/withdrawals: It is not clear that all children who failed the M-CHAT and were subsequently offered a free evaluation chose to have the evaluation and therefore were included in the study. However, all children who did receive a free evaluation were accounted for and no withdrawals were reported

Uninterpretable results: NS

Comparative	
Notes	**Relevant clinical data available:** unclear, as these children had an M-CHAT, which is not usual practice
	Conflicts of interest: none listed or apparent, as study authors are not the developers of the tool being reported
	Funding: nil
	Study start and end dates: NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
		High	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to	Yes		

Chlebowski 2010 *(Continued)*

correctly classify the target condition?

Were the reference standard results interpreted without knowledge of the results of the index tests?	No	
	High	**Unclear**

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes	
	Unclear	

Corsello 2013

Study characteristics	
Patient sampling	Prospective cohort study of a consecutive sample of children attending a children's hospital developmental evaluation clinic for evaluation for ASD
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 138
	Diagnosis: autism (n = 56); PDD-NOS (n = 50); non-spectrum (n = 32)
	Comorbidity: NS
	Age: 24-36 months (mean age: autism = 29.77 (SD 3.16) months; PDD-NOS = 30.58 (SD 3.39) months; non-spectrum = 30.50 (SD 2.82) months)
	Sex: autism = 86% male; PDD-NOS = 90% male; non-spectrum = 81% male
	Ethnicity: Caucasian (41%); African American (2%); Asian (3%); other (15%); unknown (39%)
	Inclusion criteria: consecutive children between ages 24 and 36 months referred for ASD assessment
	Exclusion criteria: NS
	Location: children's hospital developmental evaluation clinic
	Training assessor: clinical psychologists who had attended 2-day training on ADOS and had had a consultation with one of the authors of the tool
	Method of participant selection: consecutive attendances
Index tests	ADOS, which was administered to each child by 1 of the 8 clinical psychologists at the clinic
Target condition and reference standard(s)	**Target condition:** autism or ASD
	Reference standard: DSM-IV-TR

Corsello 2013 *(Continued)*

	Procedure for diagnosis: Records of children with possible ASD were reviewed and coded by clinicians, and a determination of autism or ASD 'caseness' was made using the records-based method for ASD case definition, as developed by the Metropolitan Atlanta Developmental Disabilites Surveillance Program (MADDSP). This method is used to determine if a case meets DSM-IV-TR criteria for a specific ASD diagnosis. The record reviewers (CC and NA) were blind to scores on the measures and to final clinical diagnosis when coding reports. Most children received ADOS, with either M-CHAT or SCQ, along with a developmental assessment (BSID-II, Bayley-III, Mullen, or WPPSI-II)
Flow and timing	**Administration:** As the reference standard was applied using review of child records, timing of case review and index test administration was not a concern. However, it is unclear whether all assessments that were reviewed in the case file were administered within a suitable time frame **Duration:** NS **Timing of assessment:** NS **Delay between tests:** NS **Missing data/withdrawals:** NS **Uninterpretable results:** NS
Comparative	
Notes	**Relevant clinical data available:** yes **Quote:** "Each child was given a diagnosis by the psychologist conducting the evaluation, who then wrote a clinical report that included a summary of the assessment with developmental scores and diagnostic classifications on the standardized measures" **Conflicts of interest:** none declared **Funding:** grants from the National Institutes of Health (K23MH071796 and K01MH065325) **Study start and end dates:** October 2005 and August 2007

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the	Yes		

![Cochrane Library logo] Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

Corsello 2013 *(Continued)*

results of the reference standard?			
If a threshold was used, was it pre-specified?	Yes		
		Low	Low

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	Low

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Unclear		
	Unclear		

Cox 1999

Study characteristics

Patient sampling	Prospective cohort study
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 45 - although results for 15 typically developing children were removed for the analyses reported
	Diagnosis: autism (n = 8); pervasive developmental disorder not otherwise specified (n = 13); language disorder (n = 9); typical (n = 15)
	Comorbidity: NS
	Age: 20 months and 42 months (equivalent to 1.6 and 3.5 years)
	Sex: autism = 8 (100%) males; pervasive developmental disorder not otherwise specified = 11 (85%) males; language disorder = 4 (44%) males; typical = 12 (80%) males
	Ethnicity: NS
	Inclusion criteria: children failing 2 to 5 key items on CHAT upon 2 assessments

Cox 1999 *(Continued)*

	Exclusion criteria: children with profound developmental delay, gross physical disability, or intellectual impairment
	Location: UK
	Setting: South-East Thames health region
	Training assessors: experienced clinicians
	Method of participant selection: All children identified as being at high risk of developing autism and a random selection of children identified as being at moderate or low risk of developing autism (after being screened on 2 occasions using the CHAT) were recruited to the study
Index tests	ADI-R, but it is not stated whether raters of ADI-R were blinded as to the child's clinical diagnosis at the 20- or 42-month assessment point
Target condition and reference standard(s)	**Target condition:** childhood autism
	Reference standard: ICD-10 scored by clinicians
	Procedure for diagnosis: final diagnosis made with ICD-10 at 20 months wherein clinical diagnosis was independent of developmental history gained from parents through ADI-R. However, incorporation not avoided at 42 months when clinical diagnosis was not independent of the clinical diagnosis at 20 months, nor from information gained from the ADI-R interview. Reference standard results blinded at 20 months when clinical diagnosis based on ICD-10 criteria was independent of ADI-R findings, but reference standard results not blinded at 42 months when clinical diagnosis was not independent of information gained from the ADI-R interview
Flow and timing	**Administration:** ICD-10 administered at both time points before evaluation with ADI-R. All children were subjected to ICD-10
	Duration: follow-up = 2 years
	Timing of assessment: NS
	Delay between tests: NS
	Missing data/withdrawals: Study authors explain that data were excluded owing to 1 child lost to follow-up; 2 with incomplete ADI-R; and 1 with cerebral palsy
	Uninterpretable results: none identified
Comparative	
Notes	**Relevant clinical data available:** no, as all children had been screened on 2 occasions using CHAT, which is not usual practice
	Conflicts of interest: none reported or identified
	Funding: Reserach was supported by 2 MRC project grants to authors of the publication (SBC, AC, and GB; 1992-1996)
	Study start and end dates: NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			

Cox 1999 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low

DOMAIN 2: Index Test All tests

Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
		High	Low

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
		Low	Low

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes	
	Low	

Gray 2008 ADI-R

Study characteristics

Patient sampling	Cohort study

Gray 2008 ADI-R *(Continued)*

Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 209
	Diagnosis: autism (n = 120); pervasive developmental disorder not otherwise specified (n = 23); developmental delay ± language delay (n = 66)
	Comorbidity: developmental delay = 171 (82%); delayed language = 200 (96%)
	Age: 20-55 months (equivalent to 1.6-4.6 years)
	Sex: 174 (17%) males; 35 (83%) females
	Ethnicity: broad range of social class and ethnic mix
	Inclusion criteria: children referred to an assessment clinic for children with developmental problems or suspected of having autism, or both
	Exclusion criteria: NS
	Location: Melbourne, Australia
	Setting: assessment clinic
	Training assessors: NS
	Method of participant selection: children referred to an assessment clinic for children with developmental problems or suspected of having autism, or both
Index tests	ADI-R and ADOS-G, preceding the reference standard
Target condition and reference standard(s)	**Target condition:** autistic disorder, ASD, no ASD
	Reference standard: best-estimate clinical DSM-IV diagnosis, taking into account all information obtained during assessment
	Procedure for diagnosis: Diagnoses were made according to DSM-IV diagnostic criteria for autistic disorder (APA 2000). One of two clinicians (KG and DS) gave ADI-R, and the second clinician gave ADOS-G whilst blind to the results of ADI-R. Clinicians then arrived at a consensus best-estimate clinical DSM-IV diagnosis, taking into account all information obtained during the assessment. Clinicians were blind to total scores on ADI-R and ADOS-G assessments during this case conferencing process. Index tests were administered and scored before the clinical diagnosis was made; then results of the index test were used to formulate the reference standard
Flow and timing	**Administration:** Individuals in the whole sample were given a diagnosis according to DSM-IV, which was the planned reference assessment. All participants underwent the same composite reference standard
	Duration: assessments completed across 3 sessions
	Timing of assessment: NS
	Delay between tests: not reported (probably low risk of bias)
	Missing data/withdrawals: numbers in all analyses not reported. Not stated whether all those attending the clinic participated
	Uninterpretable results: NS
Comparative	
Notes	**Relevant clinical data available:** yes, including assessments of behaviour, cognitive function, and language

Gray 2008 ADI-R *(Continued)*

Conflicts of interest: unclear if avoided, as study authors were involved in training others to use the diagnostic tools examined

Funding: National Health and Medical Research Council Project Grant (236834)

Study start and end dates: March 2002 and November 2005

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Unclear		
		High	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		

Gray 2008 ADI-R *(Continued)*

Unclear

Gray 2008 ADOS

Study characteristics	
Patient sampling	Cohort study
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 209
	Diagnosis: autism (n = 120); pervasive developmental disorder not otherwise specified (n = 23); developmental delay ± language delay (n = 66)
	Comorbidity: developmental delay = 171 (82%); delayed language = 200 (96%)
	Age: 20-55 months (equivalent to 1.6-4.6 years)
	Sex: 174 (17%) males; 35 (83%) females
	Ethnicity: broad range of social classes and broad ethnic mix
	Inclusion criteria: children referred to an assessment clinic for children with developmental problems or suspected of having autism, or both (from March 2002 to November 2005)
	Exclusion criteria: NS
	Location: Melbourne, Australia
	Setting: assessment clinic
	Training assessors: NS
	Method of participant selection: children referred to an assessment clinic for children with developmental problems or suspected of having autism, or both
Index tests	ADI-R and ADOS-G, preceding the reference standard
Target condition and reference standard(s)	**Target condition:** autistic disorder, ASD, no ASD
	Reference standard: best-estimate clinical DSM-IV diagnosis, taking into account all information obtained during assessment
	Procedure for diagnosis: Diagnoses were made according to DSM-IV diagnostic criteria for autistic disorder (APA 2000). One of two clinicians (KG and DS) gave ADI-R, and the second clinician gave ADOS-G whilst blind to results of the ADI-R. Clinicians then arrived at a consensus best-estimate clinical DSM-IV diagnosis, taking into account all information obtained during the assessment. Clinicians were blind to total scores on ADI-R and ADOS-G assessments during this case conferencing process. Index tests were administered and scored before the clinical diagnosis was made; then results of the index test were used to formulate the reference standard
Flow and timing	**Administration:** Individuals in the whole sample were given a diagnosis according to DSM-IV, which was the planned reference assessment. All participants underwent the same composite reference standard
	Duration: assessments completed across 3 sessions
	Timing of assessment: NS
	Delay between tests: not reported (probably low risk of bias)

Gray 2008 ADOS *(Continued)*

	Missing data/withdrawals: numbers in all analyses not reported. Not stated whether all those attending the clinic participated

Comparative	

Notes	**Relevant clinical data available:** yes, including assessments of behaviour, cognitive function, and language
	Conflicts of interest: unclear if avoided, as study authors were involved in training others to use the diagnostic tools examined
	Funding: National Health and Medical Research Council Project Grant (236834)
	Study start and end dates: March 2002 and November 2005

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Unclear		
		High	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Gray 2008 ADOS *(Continued)*

		Low	Low

DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
		Unclear	

Kim 2012b ADOS Cohort A

Study characteristics	
Patient sampling	Convenient sampling of children from (1) prior 'first words and toddlers' study OR (2) university clinic
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 151
	Diagnosis: autism spectrum = 123; non-spectrum = 28
	Comorbidity: NS
	Age: 21-47 months (mean age 34 months)
	Sex: 103 (68%) males
	Ethnicity: 74% Caucasian
	Inclusion criteria: Children participated in the study if they (1) had complete ADOS, ADI-R, and nonverbal IQ scores and best-estimate clinical diagnosis (collected from participating in a prior 'first words and toddlers' study) OR (2) were patients at a university clinic
	Exclusion criteria: NS
	Location: Michigan, USA
	Setting: university clinic
	Training assessors: Assessors had completed research training and had obtained research reliability
	Method of participant selection: meeting inclusion criteria stated above
Index tests	ADOS
Target condition and reference standard(s)	**Target condition:** autism, PDD-NOS, NS
	Reference standard: best-estimate clinical DSM-IV diagnosis, taking into account all information obtained during assessment
	Procedure for diagnosis: Clinicians arrived at a consensus best-estimate clinical DSM-IV diagnosis, taking into account all information obtained during the assessment. Clinicians performed one assessment using ADI-R, then they or another clinician did an assessment using ADOS - possibly introducing bias if the same clinician did both assessments
Flow and timing	**Administration:** within a few days

Kim 2012b ADOS Cohort A *(Continued)*

Duration: NS

Timing of assessment: unclear when best-estimate clinical diagnosis was made in relation to assessment by ADOS

Delay between tests: few days

Missing data/withdrawals: NS

Comparative	
Notes	**Relevant clinical data available:** Standard hierarchy of cognitive and IQ assessment tools were also administered
	Conflicts of interest: declared that 1 study author receives royalties for the ADOS tool. Profits from this research were donated to charity
	Funding: funded by NIMH (RO1 MH066469, MH57167, and HD 35482-01)
	Study start and end dates: NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
		Unclear	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without	No		

Kim 2012b ADOS Cohort A *(Continued)*

knowledge of the results of
the index tests?

		Unclear	Unclear

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes

	Unclear

Kim 2012b ADOS Cohort B

Study characteristics	
Patient sampling	Convenient sampling from (1) prior 'first words and toddlers' study or (2) university clinic
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 110
	Diagnosis: ASD = 69; non-spectrum = 41
	Comorbidity: NS
	Age: 21-47 months (mean age 40 months)
	Sex: 89 (81%) males
	Ethniticity: 74% Caucasian
	Inclusion criteria: Children participated in the study if they (1) had complete ADOS, ADI-R, or non-verbal IQ scores and best-estimate clinical diagnosis (collected from participating in a prior 'first words and toddlers' study) OR (2) were patients at a university clinic
	Exclusion criteria: NS
	Location: Michigan, USA
	Setting: university clinic
	Training assessors: Assessors had completed research training and had obtained research reliability
	Method of participant selection: meeting inclusion criteria stated above
Index tests	ADOS
Target condition and reference standard(s)	**Target condition:** autism, PDD-NOS, NS
	Reference standard: best-estimate clinical DSM-IV diagnosis, taking into account all information obtained during assessment
	Procedure for diagnosis: Clinicians arrived at a consensus best-estimate clinical DSM-IV diagnosis, taking into account all information obtained during the assessment. Clinicians performed one assessment using ADI-R, then they or another clinician did an assessment using ADOS - possibly introducing bias if the same clinician did both assessments
Flow and timing	**Administration:** within a few days

Kim 2012b ADOS Cohort B *(Continued)*

	Duration: NS
	Timing of assessment: unclear when best-estimate clinical diagnosis was made in relation to assessment by ADOS
	Delay between tests: few days
	Missing data/withdrawals: NS

Comparative	
Notes	**Relevant clinical data available:** Standard hierarchy of cognitive and IQ assessment tools were also administered
	Conflicts of interest: declared that 1 study author receives royalties for the ADOS tool. Profits from this research were donated to charity
	Funding: funded by NIMH (RO1 MH066469, MH57167, and HD 35482-01)

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
		Unclear	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		

Kim 2012b ADOS Cohort B *(Continued)*

	Unclear	Unclear

DOMAIN 4: Flow and Timing		
Did all patients receive the same reference standard?	Yes	

	Unclear

Le Couteur 2008 ADOS

Study characteristics	
Patient sampling	Prospective cohort study in which children were assessed with ADOS
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 101
	Diagnosis: autism = 49; ASD = 28; other = 24
	Comorbidity: NS
	Age: 24-49 months (equivalent to 2-4 years)
	Sex: 81 (80%) males
	Ethnicity: NS
	Inclusion criteria: Quote: "All children were initially identified from within the North East of England by local speech and language therapists and paediatricians as having speech or communication difficulties, or suspected ASD. At the time of recruitment, not all had a firm clinical diagnosis"
	Exclusion criteria: NS
	Location: North-East England
	Setting: children's own homes
	Training assessors: A local speech and language therapist and a paediatrician diagnosed suspected ASD or communication difficulties. Study authors diagnosed PDD. ADOS was executed by trained associates
	Method of participant selection: Children were recruited from 2 previous studies: the first, an evaluation of a group parent training intervention, and the second, a study of the relationship between executive function and autistic symptoms in very young children
Index tests	ADOS. Not blinded and a best-estimate clinical diagnosis (BECD) was made on the basis of all available clinical information, including ADOS results
Target condition and reference standard(s)	**Target condition:** autism, ASD, other
	Reference standard: ICD-10 and 2 clinicians
	Procedure for diagnosis: Quote: "A best estimate clinical diagnosis (BECD) was made by the senior authors (ALC, HM) based on all available clinical information across settings, along with the ADI-R, ADOS and all other research assessment information. This procedure is in line with accepted best practice for research assessments"

Le Couteur 2008 ADOS *(Continued)*

The BECD did include information from ADOS, although the BECD was made by different researchers than those who administered ADOS. Not clearly stated, but it is reported that ADOS was administered before the BECD was made and results of the index test were used to formulate the reference standard

Flow and timing	**Administration:** NS
	Duration: NS
	Timing of assessment: at 1 time point only
	Delay between tests: delay between ADI-R, ADOS, and best-estimate clinical diagnosis not stated
	Missing data/withdrawals: none; all 101 participants accounted for
	Uninterpretable results: no; all results were interpreted
Comparative	
Notes	**Relevant clinical data available:** yes, as senior authors are clinicians in the Regional Children's PDD service and thus had access to additional clinical information and reports about many of the children
	Conflicts of interest: not avoided, as Rutter and Lord are authors of ADOS
	Funding: nil reported
	Study start and end dates: NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
		High	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Unclear	Low

Le Couteur 2008 ADOS *(Continued)*

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	No

	Unclear	Low

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes

	Low

Lord 2000

Study characteristics

Patient sampling	Case-control study in which children identified with autism, PDD-NOS, and non-spectrum were matched for verbal mental age
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 129 (Module 1 = 74; Module 2 = 55)
	Diagnosis: Module 1: autism = 40; PDD-NOS = 17; non-spectrum = 17; Module 2: autism = 21; PDD-NOS = 18; non-spectrum = 16
	Age: Module 1 = 3.51-4.92 years; Module 2 = 3.78-4.56 years
	Sex: Module 1: 57 males (77%); Module 2: 31 males (71%)
	Ethnicity: 80% Caucasian, 11% African American, 4% Hispanic, 2% Asian American, 2% other
	Inclusion criteria: English speakers, developmental disorder clinic diagnosis, ASD and non-spectrum in close verbal and IQ age range
	Exclusion criteria: diagnosis of ASD in non-spectrum cohort. Williams syndrome and mild cerebral palsy in non-spectrum group. For Module 1, if children with autism were initially recruited but could not be matched to other diagnostic groups on language level, children with autism were excluded from the sample
	Location: Chicago, USA
	Setting: developmental disorders clinic
	Training assessors: Clinical research staff administered ADOS-G and a best-estimate diagnosis was made by a clinical psychologist and a child psychiatrist
	Method of participant selection: total of 381 consecutive referrals to the Developmental Disorders Clinic at the University of Chicago. Group was split into AS, PDD-NOS, and NS participants (which included mental re-

---

Lord 2000 *(Continued)*

	tardation, receptive-expressive language disorder, attention-deficit hyperactivity disorder and/or oppositional defiant disorder, anxiety disorder, major depression, obsessive-compulsive disorder) and typically developing children and adults
Index tests	ADOS-G. Direct observations of the individual participant occurred during ADOS-G, physical examination, psychological testing, and free time with the parents. Direct observation was also utilised for a best-estimate diagnosis
Target condition and reference standard(s)	**Target condition:** autism, PDD, PDD-NOS, non-spectrum **Reference standard:** clinical psychologist/child psychiatrist **Procedure for diagnosis:** Clinical diagnoses were assigned according to clinical impressions of a clinical psychologist and a child psychiatrist, each of whom interviewed the parents and observed the child separately and discussed discrepant impressions until they reached a "best-estimate" diagnosis. Clinicians had access to history, results of a physical examination, and scores on ADI-R. Direct observations of the individual participant occurred during ADOS-G, physical examination, psychological testing, and free time with the parents. Index tests were performed before a best-estimate diagnosis was made
Flow and timing	**Administration:** NS **Duration:** NS **Timing of assessment:** assessment at 1 time point only **Delay between tests:** not stated **Missing data/withdrawals:** none; all participants accounted for **Uninterpretable results:** NS
Comparative	
Notes	**Relevant clinical data available:** yes, as clinicians had access to history, results of a physical examination, and scores on ADI-R, along with direct observations of the individual participant during ADOS-G, physical examination, psychological testing, and free time with the parents **Conflicts of interest:** not avoided, as authors Lord, Risi, and DiLavore were involved in development of the ADOS-G **Funding:** nil reported **Study start and end dates:** NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		

Lord 2000 *(Continued)*

	Low	Low

DOMAIN 2: Index Test All tests

Were the index test results interpreted without knowledge of the results of the reference standard?	No	
If a threshold was used, was it pre-specified?	Yes	

	High	Low

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition?	Yes	
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes	

	Low	Low

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	No	

	Low

Mazefsky 2006 ADOS

Study characteristics	

Patient sampling	Prospective cohort study
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 78 but data reported for only 75
	Diagnosis: autism = 32 (40.7%); other PDDs = 27 (33.7%) (e.g. Asperger disorder = 6; PDD-NOS = 21); 'non-PDD' (diagnoses outside the autism spectrum such as language disorders) = 19 (24.4%)
	Co-morbidity: 9% children with psychotropic medication
	Age: 22 months (equivalent to 1.8 years) to 8 years; mean age 4 years (SD = 1.4)

Mazefsky 2006 ADOS *(Continued)*

Sex: 72% male, 28% female

Ethnicity: 69% Caucasian, 10% African American, 21% mixed race or other ethnicity

Inclusion criteria: Children had received diagnostic assessments

Exclusion criteria: NS

Location: Virginia, USA

Setting: specialised clinic for assessment of PDD in a medical care setting. Multi-disciplinary diagnostic clinic

Training assessors: licensed clinical psychologist, child psychiatrist, education specialist, speech/language pathologist, occupational therapist

Method of participant selection: NS

Index tests	ADOS-G. At the visit, an examiner administered ADOS-G while observed by the evaluation team, then scored ADOS–G following its completion. Scores and ADOS–G diagnostic recommendations generally were not discussed with the evaluation team
Target condition and reference standard(s)	**Target condition:** autism, PDD **Reference standard:** DSM-IV and multi-disciplinary team within 3 years **Procedure for diagnosis:** All participants took part in a multi-disciplinary diagnostic clinic aimed at clarifying the child's disability and generating recommendations to guide intervention. At a minimum, the evaluation team consisted of a licensed clinical psychologist, a child psychiatrist, an education specialist, a speech/language pathologist, and an occupational therapist, all of whom had extensive experience with children with autism (ranging from 5 to over 20 years of expertise). Decision regarding diagnosis consistent with criteria from DSM-IV. An experienced licensed clinical psychologist, certified to both administer and train others on ADOS–G, administered ADOS–G, while all other team members observed from behind a one-way mirror. Reference standard diagnosis occurred after completion of the index test
Flow and timing	**Administration:** Each clinic evaluation consisted of structured assessments, observations, and team discussion. All participants took part in a multi-disciplinary team diagnosis, which utilised the index test **Duration:** approximately 4 hours **Timing of assessment:** at 1 time point only **Delay between tests:** All assessments occurred on the same day **Missing data/withdrawal:** Data from 75 children are reported in the results' tables, but 78 children participated in the study and study authors provided no explanation regarding the 3 sets of missing scores. No withdrawals, as there was no follow-up **Uninterpretable results:** NS
Comparative	
Notes	**Relevant clinical data available:** unclear, possibly because participants were already receiving some form of special education or early intervention services before the clinic assessment, but this was not explicitly stated **Conflicts of interest:** avoided, as published by SAGE on behalf of the Autistic National Society **Funding:** nil reported **Study start and end dates:** NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Mazefsky 2006 ADOS *(Continued)*

DOMAIN 1: Patient Selection

Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**

DOMAIN 2: Index Test All tests

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
		High	**Low**

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes		
		Unclear	

Oosterling 2010b ADI-R

Study characteristics

Patient sampling	Cohort study
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 208
	Diagnosis: ASD = 143 (92 = autism, 49 = PDD-NOS, 2 = Asperger syndrome). Non-ASD = 65 (10 = mental retardation without ASD, 21 = language disorders, 17 = externalising disorders (attention deficit/hyperactivity disorder or oppositional defiant disorder), 3 = internalising disorders (mood or anxiety disorder), 13 = other developmental disorders; 1 = functioning normally)
	Comorbidity: NS
	Age: 20-40 months (equivalent to 1.6-3.3 years)
	Sex: NS for the 208 participants studied; however, of the larger sample of 426 children, 78% were male
	Ethnicity: NS for the 208 participants studied; however, of the larger sample of 426 children from which these participants were drawn, 95% were Dutch Caucasian
	Inclusion criteria: children with ADI-R/ADOS data for Oosterling 2009 sample (Oosterling 2010b ADI-R)
	Exclusion criteria: children with no ADI-R/ADOS data for Oosterling 2009 study (Oosterling 2010b ADI-R)
	Location: Netherlands
	Settings: university centres for child and adolescent psychiatry (Nijmegen, Utrecht, and Groningen)
	Training assessors: Clinical psychologist administered ADOS and ADI-R; child psychiatrist and psychologist with extensive multi-disciplinary programme made best-estimate clinical diagnosis
	Method of participant selection: children referred for clinical assessment to Nijmegen, Utrecht, or Groningen University Centres in the Netherlands
Index tests	ADI-R, results of which were incorporated from a best-estimate diagnosis
Target condition and reference standard(s)	**Target condition:** autism, non-autism ASD, non-ASD
	Reference standard: multi-disciplinary best-estimate clinical diagnosis
	Procedure for diagnosis: A consensus best-estimate clinical diagnosis was established by at least 2 experienced professionals - a child psychiatrist and a psychologist - based on DSM-IV-TR (APA 2000) criteria and using all available information, except for SCQ and CSBS-DP Infant-Toddler Checklist. Clinical psychologists who met standard requirements for research reliability administered ADOS and ADI-R. Not known if clinical psychologists were aware of the other results before administering ADI-R and ADOS
Flow and timing	**Administration:** All children were evaluated by the same diagnostic evaluation programme
	Duration: 6 weeks
	Timing of assessment: at 1 time point only
	Delay between tests: occurred within 6-week assessment period
	Missing data/withdrawals: NS
	Uninterpretable results: NS
Comparative	

Oosterling 2010b ADI-R *(Continued)*

Notes	**Relevant clinical data available:** yes, as results from an unstructured developmental interview, a psychiatric evaluation, a parent–child play observation, and, for research purposes, psychometric testing (cognition and language) were made available
	Conflicts of interest: avoided, as study authors were not involved in development of the tool
	Funding: supported by a grant from the Korczak Foundation
	Study start and end dates: NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**

Oosterling 2010b ADI-R *(Continued)*

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes

Unclear

Oosterling 2010b ADOS

Study characteristics

Patient sampling	Cohort study
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 208
	Diagnosis: ASD = 143 (92 = autism, 49 = PDD-NOS, 2 = Asperger syndrome). Non-ASD = 65 (10 = mental retardation without ASD, 21 = language disorders, 17 = externalising disorders (attention deficit/hyperactivity disorder or oppositional defiant disorder), 3 = internalising disorders (mood or anxiety disorder), 13 = other developmental disorders; 1 = functioning normally)
	Comorbidity: NS
	Age: 20-40 months (equivalent to 1.6-3.3 years)
	Sex: NS for the 208 participants studied; however, of the larger sample of 426 children from which these participants were drawn, 78% were male
	Ethnicity: NS for the 208 participants studied; however, of the larger sample of 426 children, 95% were Dutch Caucasian
	Inclusion criteria: children with ADI-R/ADOS data for Oosterling 2009 sample (Oosterling 2010b ADI-R)
	Exclusion criteria: children with no ADI-R/ADOS data for Oosterling 2009 study (Oosterling 2010b ADI-R)
	Location: Netherlands
	Settings: university centres for child and adolescent psychiatry (Nijmegen, Utrecht, and Groningen)
	Training assessors: Clinical psychologist administered ADOS and ADI-R, and child psychiatrist and psychologist with extensive multi-disciplinary programme made best-estimate clinical diagnosis
	Method of participant selection: children referred for clinical assessment to Nijmegen, Utrecht, or Groningen university centres in the Netherlands
Index tests	ADOS, results of which were incorporated from a best-estimate diagnosis
Target condition and reference standard(s)	**Target condition:** autism, non-autism ASD, non-ASD
	Reference standard: multi-disciplinary best-estimate clinical diagnosis
	Procedure for diagnosis: A consensus best-estimate clinical diagnosis was established by at least 2 experienced professionals - a child psychiatrist and a psychologist - according to DSM-IV-TR (APA 2000) criteria and using all available information, except for SCQ and CSBS-DP Infant-Toddler Checklist. Clinical psychologists who met standard requirements for research reliability administered ADOS and ADI-R. Not known if the clinical psychologists were aware of the other results before administering ADI-R and ADOS
Flow and timing	**Administration:** All children were evaluated by the same diagnostic evaluation programme

Oosterling 2010b ADOS *(Continued)*

Duration: 6 weeks

Timing of assessment: at 1 time point only

Delay between tests: occurred within 6-week assessment period

Missing data/withdrawals: NS

Uninterpretable results: NS

Comparative	
Notes	**Relevant clinical data available:** yes, as results from an unstructured developmental interview, a psychiatric evaluation, a parent–child play observation, and, for research purposes, psychometric testing (cognition and language) were made available
	Conflicts of interest: avoided, as study authors were not involved in development of the tool
	Funding: supported by a grant from the Korczak Foundation
	Study start and end dates: NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
		High	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to cor-	Yes		

Oosterling 2010b ADOS *(Continued)*

rectly classify the tar-
get condition?

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear	
	Unclear	**Low**

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes
	Unclear

Risi 2006 Study 1 ADOS Cohort A

Study characteristics	
Patient sampling	Secondary analyses of a large cohort of children from 5 clinical and research centres
Patient characteristics and setting	**Number of study groups:** 1 (cohort A)

Number of participants: 270 (data were collected from participants who were recruited from 2 different research projects)

Diagnosis: ASD = 227 (84%); non-spectrum = 43 (16%) (including non-specific mental retardation; language disorder; oppositional defiant disorder or attention-deficit/hyperactivity disorder, or both; Down syndrome; mood or anxiety disorder, or both; Tourette syndrome)

Comorbidity: Included children had a known developmental, cognitive, or behavioural diagnosis

Age: < 3 years (mean age 28.5 months)

Sex: 215 (80%) males

Ethnicity: NS for the 270 participants studied; however, of the larger sample of 1039 children from which these participants were drawn, 82% were white, 13% African American, 4% Asian American, and 1% other or multi-racial

Inclusion criteria: Children had completed a diagnostic evaluation at the University of Michigan Autism and Communication Disorders Clinic or the University of Chicago Clinic Developmental Disorders Clinic or were part of a longitudinal study conducted through centres at the University of North Carolina, Chapel Hill, and the University of Chicago

Exclusion criteria: participants with visual, hearing, or motor impairments that precluded standard administration of an instrument

Location: USA

Setting: university clinic, autism clinic, and 2 research teams at university hospitals

Training assessors: All examiners had completed research training for ADOS and met standard requirements for reliability |

Risi 2006 Study 1 ADOS Cohort A *(Continued)*

	Method of participant selection: participant of existing cohort
Index tests	ADOS, which was probably blinded, but this was not explicitly mentioned
Target condition and reference standard(s)	**Target condition:** autism, non-autism ASD, non-spectrum **Reference standard:** consensus best-estimate diagnosis **Procedure for diagnosis:** procedure not completely clear, with reference to physicians conducting a review and then "all of the clinicians" agreeing on the diagnosis but no mention of the disciplines and the number of clinicians for each child. Index tests were part of the reference standard
Flow and timing	**Administration:** All participants underwent the same composite reference standard. The sample was probably selected on the basis of having undergone the reference standard; however, it is not clear whether the reference standard was applied to a selection of participants who were part of a bigger group who and scored positive on the index tests **Duration:** NS **Timing of assessment:** at 1 time point only **Delay between tests:** NS **Missing data/withdrawals:** NS **Uninterpretable results:** NS
Comparative	
Notes	**Relevant clinical data available:** unclear, as not explicitly mentioned **Conflicts of interest:** not avoided, as study authors are developers of the ADOS. "Disclosure: Drs. Risi, Lord, Corsello, and Pickles receive royalties for the ADOS; profits accrued from this study were donated to charity. The other authors have no financial relationships to disclose" **Funding:** NS **Study start and end dates:** NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
		High	High
DOMAIN 2: Index Test All tests			

Risi 2006 Study 1 ADOS Cohort A *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
		Unclear	**Low**

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
		Unclear	**Low**

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes		
		High	

Risi 2006 Study 1 ADOS Cohort B

Study characteristics	
Patient sampling	Secondary analyses of a large cohort of children from 5 clinical and research centres
Patient characteristics and setting	**Number of study groups:** 1 (cohort B)
	Number of participants: 67
	Diagnosis: ASD = 57 (85%); non-spectrum = 10 (15%)
	Co- morbidity: non-specific mental retardation
	Age: 36-112 months with profound mental retardation (mean age 62.5 months)
	Sex: 48 (72%) males
	Ethnicity: NS for the 67 participants studied; however, of the larger sample of 1039 children from which these participants were drawn, 82% were white, 13% African American, 4% Asian American, and 1% other or multi-racial

Risi 2006 Study 1 ADOS Cohort B *(Continued)*

	Inclusion criteria: Children had completed a diagnostic evaluation at the University of Michigan Autism and Communication Disorders Clinic or the University of Chicago Clinic Developmental Disorders Clinic or were part of a longitudinal study conducted through centres at the University of North Carolina, Chapel Hill, and the University of Chicago
	Exclusion criteria: participants with visual, hearing, or motor impairments that precluded standard administration of an instrument
	Location: USA
	Setting: university clinic, autism clinic, and 2 research teams at university hospitals
	Training assessors: All examiners had completed research training and met standard requirements for reliability
	Method of participant selection: participant of existing cohort
Index tests	ADOS. Probably blinded but not explicitly stated
Target condition and reference standard(s)	**Target condition:** autism, non-autism ASD, non-spectrum
	Reference standard: consensus best-estimate diagnosis
	Procedure for diagnosis: not completely clear, with reference to physicians conducting a review and then "all of the clinicians" agreeing on the diagnosis but no mention of the disciplines and the number of clinicians for each child. Index tests were part of the reference standard
Flow and timing	**Administration:** All participants underwent the same composite reference standard. The sample was probably selected on the basis of having undergone the reference standard; however, it is not clear whether the reference standard was applied to a selection of participants who were part of a bigger group and who scored positive on the index tests
	Duration: NS
	Timing of assessment: at 1 time point only
	Delay between tests: NS
	Missing data/withdrawals: NS
	Uninterpretable results: NS
Comparative	
Notes	**Relevant clinical data available:** unclear, as not explicitly mentioned
	Conflicts of interest: not avoided, as study authors are developers of ADOS
	"Disclosure: Drs. Risi, Lord, Corsello, and Pickles receive royalties for the ADOS; profits accrued from this study were donated to charity. The other authors have no financial relationships to disclose"
	Funding: NS
	Study start and end dates: NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			

Risi 2006 Study 1 ADOS Cohort B *(Continued)*

Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
		High	**High**

DOMAIN 2: Index Test All tests

Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
		Unclear	**Low**

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
		Unclear	**Low**

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes		
		High	

Russell 2010

Study characteristics

Russell 2010 *(Continued)*

Patient sampling	Cohort study involving a secondary analysis of CARS scores from charts of children assessed for suspected autism
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 103
	Diagnosis: autism = 86 (28 = childhood autism, 54 = atypical autism, 3 = Asperger syndrome, 1 = Rett syndrome); non-ASD = 14 (7 = average intelligence, 6 = compromised intelligence, 1 = compromised intelligence with selective mutism); no available data = 3
	Comorbidity: 72 children with severe/profound intellectual disability; 21 with autism also had unspecified intellectual disabilities. Seizures, cerebral palsy in non-ASD cohort. The mix of children seen was similar to the mix of children presenting for autism diagnostic services
	Age: mean age 5.10 years (SD = 2.20)
	Sex: NS, but study authors report higher representation of boys in the study group
	Ethnicity: expected high percentage of Indian children due to location and setting (see details below)
	Inclusion criteria: children suspected to have autism
	Exclusion criteria: diagnosis of overactive disorder associated with mental retardation and stereotyped movements
	Location: Southern India
	Setting: autism clinic, child and adolescent psychiatry unit of tertiary care, teaching hospital
	Training assessors: NS
	Method of participant selection: by audit of clinic charts from 2001-2007
Index tests	CARS, which was rated independently by clinical psychologists or rehabilitation psychologists and speech therapists
Target condition and reference standard(s)	**Target condition:** ASD, non-ASD
	Reference standard: ICD-10 and multi-disciplinary team within 6 years (2001-2007)
	Procedure for diagnosis: ICD-10-based clinical diagnosis of autism (pervasive developmental disorders) (childhood autism (F84.0), atypical autism (F84.1), Rett's syndrome (F84.2), other childhood disintegrative disorder (F84.3), and Asperger syndrome (F84.5)) made by consultant psychiatrists and later endorsed by the multi-disciplinary team consisting of special educators, occupational therapists, speech therapists, and psychiatric nurses, was used as the reference standard in this study. CARS was assessed after autism was clinically diagnosed by the psychiatrists on the team
Flow and timing	**Administration:** All participants were subjected to the same reference standard diagnosis
	Duration: NS
	Timing of assessment: at 1 time point only
	Delay between tests: NS
	Missing data/withdrawals: Absence of data was reported for 3 cases
	Uninterpretable results: nil reported
Comparative	

Russell 2010 *(Continued)*

Notes	**Relevant clinical data available:** yes; children's skills in 7 areas - memory, language, conceptual thinking, reasoning, numerical reasoning, visuo-motor coordination, and social intelligence - were assessed via the Bi-net Kamat Scale of Intelligence, which is the Indian adaptation of the Stanford-Binet Scale of Intelligence. Developmental skills in the 4 areas of motor behaviour, adaptive behaviour, language, and personal and social behaviour were assessed on Gesell's Developmental Schedule (1940)
	Conflicts of interests: avoided, as no benefits, in any form, have been received directly from any extramural grants or indirectly through funds
	Funding: none
	Study start and end dates: 2001 and 2007

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the	Yes		

Russell 2010 *(Continued)*
results of the index
tests?

		Low	Low

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes		

	Low		

Ventola 2006 ADI-R

Study characteristics		

Patient sampling	Cohort study
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 45
	Diagnosis: AD = 27 (60%); PDD-NOS = 9 (20%); non-spectrum = 9 (20%)
	Comorbidity: NS
	Age: mean chronological age at screening = 22 months (range 16–30 months); mean chronological age at diagnosis = 26 months (range 16-31 months)
	Sex: male = 37 (82%); female = 8 (18%)
	Ethnicity: 89% Caucasian; 9% Latino; 2% other
	Inclusion criteria: All children failed the Modified Checklist for Autism in Toddlers (Robins 2001)
	Exclusion criteria: NS
	Location: USA
	Setting: psychological services clinic at the University of Connecticut (n = 41); child's home (n = 1); child's early intervention office (n = 3)
	Training assessors: licensed psychologist
	Method of participant selection: Participants were part of a larger screening study
Index tests	ADI-R
Target condition and reference standard(s)	**Target condition:** autism, non-autism
	Reference standard: consensus best-estimate diagnosis based on DSM-IV criteria
	Procedure for diagnosis: NS
Flow and timing	**Administration:** All children were evaluated by the same diagnostic evaluation programme
	Duration: NS
	Timing of assessment: occurred at same point in time

Ventola 2006 ADI-R *(Continued)*

	Delay between tests: nil
	Missing data/withdrawals: nil
	Uninterpretable results: nil

Comparative	

Notes	**Relevant clinical data available:** yes; other clinical data collected included Mullen Scales of Early Learning and Vineland Adaptive Behaviour Scales
	Conflicts of interest: nil reported
	Funding: supported by the University of Connecticut's research foundation faculty grant
	Study start and end dates: NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	No		
If a threshold was used, was it pre-specified?	Yes		
		High	**Unclear**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
		High	**Low**
DOMAIN 4: Flow and Timing			

Ventola 2006 ADI-R *(Continued)*

Did all patients receive the same reference standard?	Yes

Low

Ventola 2006 ADOS

Study characteristics	
Patient sampling	Cohort study
Patient characteristics and setting	**Number of study groups**: 1
	Number of participants: 45
	Diagnosis: AD = 27 (60%); PDD-NOS = 9 (20%); non-spectrum = 9 (20%)
	Comorbidity: NS
	Age: mean chronological age at screening = 22 months (range 16–30 months); mean chronological age at diagnosis = 26 months (range 16-31 months)
	Sex: male = 37 (82%); female =8 (18%)
	Ethnicity: 89% Caucasian; 9% Latino; 2% other
	Inclusion criteria: All children failed the Modified Checklist for Autism in Toddlers (Robins 2001)
	Exclusion criteria: NS
	Location: USA
	Setting: psychological services clinic at the University of Connecticut (n = 41); child's home (n = 1); child's early intervention office (n = 3)
	Training assessors: NS; ADOS was completed by a doctoral student
	Method of participant selection: Participants were part of a larger screening study
Index tests	ADOS
Target condition and reference standard(s)	**Target condition:** autism, ASD, non-ASD
	Reference standard: consensus best-estimate diagnosis based on DSM-IV criteria
	Procedure for diagnosis: NS
Flow and timing	**Administration:** All children were evaluated by the same diagnostic evaluation programme
	Duration: NS
	Timing of assessment: occurred at same point in time
	Delay between tests: nil
	Missing data/withdrawals: nil
	Uninterpretable results: nil
Comparative	

Ventola 2006 ADOS *(Continued)*

Notes	**Relevant clinical data available:** yes; other clinical data collected included Mullen Scales of Early Learning and Vineland Adaptive Behaviour Scales
	Conflicts of interest: nil reported
	Funding: supported by the University of Connecticut's Resarch Foundation Faculty Grant
	Study start and end dates: NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
		High	Unclear
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
		High	Low
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
		Low	

Ventola 2006 CARS

Study characteristics

Patient sampling	Cohort study
Patient characteristics and setting	**Number of study groups**: 1
	Number of participants: 45
	Diagnosis: AD = 27 (60%); PDD-NOS = 9 (20%); non-spectrum = 9 (20%)
	Comorbidity: NS
	Age: mean chronological age at screening = 22 months (range 16–30 months); mean chronological age at diagnosis = 26 months (range 16-31 months)
	Sex: male = 37 (82%); female =8 (18%)
	Ethnicity: 89% Caucasian; 9% Latino; 2% other
	Inclusion criteria: All children failed the Modified Checklist for Autism in Toddlers (Robins 2001)
	Exclusion criteria: NS
	Location: USA
	Setting: psychological services clinic at the University of Connecticut (n = 41); child's home (n = 1); child's early intervention office (n = 3)
	Training assessors: CARS completed by both the licensed psychologist and a doctoral student
	Method of participant selection: Participants were part of a larger screening study
Index tests	CARS
Target condition and reference standard(s)	**Target condition:** autism, non-autism
	Reference standard: best-estimate clinical diagnosis based on DSM- IV criteria
	Procedure for diagnosis: NS
Flow and timing	**Administration:** All children were evaluated by the same diagnostic evaluation programme
	Duration: NS
	Timing of assessment: occurred at same point in time
	Delay between tests: nil
	Missing data/withdrawals: nil
	Uninterpretable results: nil
Comparative	
Notes	**Relevant clinical data available:** yes; other clinical data collected included Mullen Scales of Early Learning and Vineland Adaptive Behaviour Scales
	Conflicts of interest: nil reported
	Funding: supported by the University of Connecticut's research foundation faculty grant
	Study start and end dates: NS

Ventola 2006 CARS *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
		High	**Unclear**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		
		High	**Low**
DOMAIN 4: Flow and Timing			
Did all patients receive the same reference standard?	Yes		
		Low	

Wiggins 2008 ADI-R

Study characteristics

Patient sampling	Cohort study
Patient characteristics and setting	**Number of study groups:** 1 **Number of participants:** 142

Wiggins 2008 ADI-R *(Continued)*

Diagnosis: AD = 43 (30%); PDD-NOS = 29 (20%); Asperger disorder = 1 (< 1%); non-ASD = 69 (49%)

Comorbidity: NS

Age: 16-37 months (equivalent to 1.3-3.08 years)

Sex: male = 112 (79%); female = 30 (21%)

Ethnicity: NS

Inclusion criteria: Children who failed M-CHAT were considered at-risk for ASD, along with those who had completed ADI-R, ADOS, and CARS and had received a clinical diagnosis

Exclusion criteria: NS

Location: USA

Setting: University of Connecticut and Georgia State University

Training assessor: Trained clinicians executed ADI-R and ADOS

Method of participant selection: Children were part of a large-scale screening study and were identified by their primary care physician or their early intervention provider

Index tests	ADI-R, but not stated whether results were blinded
Target condition and reference standard(s)	**Target condition:** AD, non-AD **Reference standard:** DSM-IV and clinical judgement **Procedure for diagnosis:** Clinical judgement was determined by a clinician who applied DSM-IV criteria for autism and PDD-NOS to guide the clinical decision. Not stated whether reference standard results were blinded
Flow and timing	**Administration:** All children received diagnosis by clinical judgement based on DSM-IV criteria **Duration:** NS **Timing of assessment:** assessment at 1 time point only **Delay between tests:** NS **Missing data/withdrawals:** NS, but it does not appear that any were missing **Uninterpretable results:** none reported or apparent
Comparative	
Notes	**Relevant clinical data available:** unclear, as whilst relevant clinical information was probably available from which sound clinical judgements could be made, this was not clearly stated **Conflicts of interests:** avoided, as study authors were not involved in development of the tools **Funding:** supported in part by the University of Connecticut's research foundation faculty grant **Study start and end dates:** NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			

Wiggins 2008 ADI-R *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**

DOMAIN 2: Index Test All tests

Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
		Unclear	**Low**

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Unclear**

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes		
		Low	

Wiggins 2008 ADOS

Study characteristics

Patient sampling	Cohort study
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 142
	Diagnosis: AD = 43 (30%); PDD-NOS = 29 (20%); Asperger disorder = 1 (< 1%); non-ASD = 69 (49%)
	Comorbidity: NS

Wiggins 2008 ADOS *(Continued)*

Age: 16-37 months (equivalent to 1.3-3.08 years)

Sex: male = 112 (79%); female = 30 (21%)

Ethnicity: NS

Inclusion criteria: children who failed M-CHAT and were considered at-risk for ASD and those who had completed ADI-R, ADOS, or CARS and had received a clinical diagnosis

Exclusion criteria: NS

Location: USA

Setting: University of Connecticut and Georgia State University

Training assessor: Trained clinicians executed ADI-R and ADOS

Method of participant selection: Children were part of a large-scale screening study and were identified by their primary care physician or their early intervention provider

Index tests	ADOS-G, but not stated whether results were blinded
Target condition and reference standard(s)	**Target condition:** ASD (children with AD and other spectrum diagnoses were combined into the ASD diagnostic category); non-ASD **Reference standard:** DSM-IV and clinical judgement **Procedure for diagnosis:** Clinical judgement was determined by a clinician who applied DSM-IV criteria for autism and PDD-NOS to guide the clinical decision. Not stated whether reference standard results were blinded
Flow and timing	**Administration:** All children received diagnosis by clinical judgement based on DSM-IV criteria **Duration:** NS **Timing of assessment:** assessment at 1 time point only **Delay between tests:** NS **Missing data/withdrawals:** NS, but it does not appear that any were missing **Uninterpretable results:** none reported or apparent
Comparative	
Notes	**Relevant clinical information available:** unclear, as whilst relevant clinical information was probably available from which sound clinical judgements could be made, this was not clearly stated **Conflicts of interests:** avoided, as study authors were not involved in development of the tools **Funding:** supported in part by the University of Connecticut's research foundation faculty grant **Study start and end dates:** NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		

Wiggins 2008 ADOS *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**

DOMAIN 2: Index Test All tests

Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
		Unclear	**Low**

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Unclear**

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes	
		Low

Wiggins 2008 CARS

Study characteristics

Patient sampling	Cohort study
Patient characteristics and setting	**Number of study groups:** 1
	Number of participants: 142
	Diagnosis: AD = 43 (30%); PDD-NOS = 29 (20%); Asperger disorder = 1 (< 1%); non-ASD = 69 (49%)
	Comorbidity: NS
	Age: 16-37 months (equivalent to 1.3-3.08 years)
	Sex: male = 112 (79%); female = 30 (21%)

Wiggins 2008 CARS *(Continued)*

	Ethnicity: NS
	Inclusion criteria: children who failed M-CHAT and were considered at-risk for ASD and those who had completed ADI-R, ADOS, or CARS and had received a clinical diagnosis
	Exclusion criteria: NS
	Location: USA
	Setting: University of Connecticut and Georgia State University
	Training assessor: Trained clinicians executed ADI-R and ADOS
	Method of participant selection: Children were part of a large-scale screening study and were identified by their primary care physician or their early intervention provider
Index tests	CARS, but not stated whether results were blinded
Target condition and reference standard(s)	**Target condition:** AD, non-AD
	Reference standard: DSM-IV and clinical judgement
	Procedure for diagnosis: Clinical judgement was determined by a clinician who applied DSM-IV criteria for autism and PDD-NOS to guide the clinical decision. Not stated whether reference standard results were blinded
Flow and timing	**Administration:** All children received diagnosis by clinical judgement based on DSM-IV criteria
	Duration: NS
	Timing of assessment: assessment at 1 time point only
	Delay between tests: NS
	Missing data/withdrawals: NS, but it does not appear that any were missing
	Uninterpretable results: none reported or apparent
Comparative	
Notes	**Relevant clinical information available:** unclear, as whilst relevant clinical information was probably available from which sound clinical judgements could be made, this was not clearly stated
	Conflicts of interest: avoided, as study authors were not involved in development of the tools
	Funding: supported in part by the University of Connecticut's research foundation faculty grant
	Study start and end dates: NS

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		

Wiggins 2008 CARS *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
		Low	Low

DOMAIN 2: Index Test All tests

Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
		Unclear	Low

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	Unclear

DOMAIN 4: Flow and Timing

Did all patients receive the same reference standard?	Yes	
		Low

AD: autistic disorder; **ADI-R:** Autism Diagnosis Interview - Revised; **ADOS:** Autism Diagnostic Observation Schedule; **ADOS-G:** Autism Diagnostic Observation Schedule - Generic; **AS:** autism spectrum; **ASD:** autism spectrum disorder; **AUT:** autism; **Bayley-III:** Bayley Scales of Infant and Toddler Development - Third Edition; **BECD:** best-estimate clinical diagnosis; **BSID-II:** Bayley Scales of Infant Development - Second Edition; **CARS:** Childhood Autism Rating Scale; **CHAT:** Checklist for Autism in Toddlers; **CSBS-DP:** Communication and Symbolic Behavior Scales Developmental Profile; **DSM-IV:** *Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition*; **DMS-IV-TR:** *Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition - Text Revision*; **ICD-10:** *International Classification of Diseases and Related Health Problems - Tenth Revision;* **IQ:** intelligence quotient; **MADDSP:** Metropolitan Atlanta Developmental Disabilites Surveillance Program; **M-CHAT:** Modified Checklist for Autism in Toddlers; **MSEL:** Mullen Scales of Early Learning; **NIMH:** National Institute of Mental Health; **NS:** not specified; **PDD-NOS:** pervasive developmental disorder - not otherwise specified; **SCQ:** social communication quotient; **SD:** standard deviation; **WPPSI-II:** Wechsler Preschool and Primary Scale of Intelligence - Second Edition.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Becker 2012	No DTA result
Bölte 2001	No DTA result

Study	Reason for exclusion
Bölte 2004	Age range > 6 years
Charman 2005	No DTA result
Charman 2013	No DTA data reported
Chawarska 2007	No DTA result
Chuthapisith 2012	Reports on shortened, not full, version of the Thai 3di
Clancy 1969	Mean age > 6 years
Constantino 2012	Mean age > 6 years
de Bildt 2004	Age range > 6 years
de Bildt 2009	Age range > 6 years
de Bildt 2013	Mean age > 6 years
Diken 2012a	Mean age > 6 years
Diken 2012b	Mean age > 6 years
DiLalla 1994	Different cutoff criteria used to investigate factor structure of the CARS; therefore did not meet review criteria
Eaves 2006	Mean age > 6 years
Falkmer 2013	Not a DTA study
Fisch 2012	Not a DTA study
Goldfischer 2001	Mean age > 6 years
Gotham 2007	Study to further develop the ADOS, so investigated use of different cutoff criteria from those used clinically. Seeking to establish ADOS-2
Gotham 2008	Study to further develop the ADOS, so investigated use of different cutoff criteria from those used clinically. Seeking to establish ADOS-2
Guthrie 2013	Analyses conducted on different version of index tool (i.e. on ADOS-T)
Huerta 2012	Not a DTA study
Jackson 2012	No DTA data to report
Kamp-Becker 2013	Mean age > 6 years
Kim 2012a	Analyses conducted on different cutoff criteria for ADOS
Kim 2013	Analyses conducted on different cutoff criteria for ADOS
Klose 2012	Not a DTA study

Study	Reason for exclusion
Le Couteur 1989	Participants recruited were above preschool age
Le Couteur 2008 ADI-R	Analyses conducted on different cutoff criteria for ADI-R
Lecavalier 2006	No gold standard used as reference standard
Leekam 2002	Participants recruited were above preschool age
Li 2005	Two Chinese authors could not ascertain the age group
Lord 1993	Used different diagnostic groups of Autism versus non-Autism (not ASD vs non-ASD)
Lord 1994	Used different diagnostic groups of Autism versus non-Autism (not ASD vs non-ASD)
Lord 2006	Analyses conducted on different cutoff criteria for both ADOS (using ADOS-PL cutoffs) and ADI-R
Lord 2012a	Not a DTA study
Lozowski-Sullivan 2011	Mean age > 6 years
Luyster 2009	Analyses conducted on different cutoff criteria for ADOS (using ADOS-T criteria)
Maljaars 2012	Mean age > 6 years
Matson 2010	Participants recruited were above preschool age
Mayes 2009	Mean age > 6 years
Mayes 2012	Not a DTA study
Mazefsky 2006 ADI-R	Used different diagnostic groups of Autism versus non-Autism (not ASD vs non-ASD)
Mazefsky 2006 GARS	Used different diagnostic groups of Autism versus non-Autism (not ASD vs non-ASD)
McGarry Klose 2012	No DTA raw data reported
Mick 2007	Thesis participants recruited were above preschool age
Molloy 2011	Mean age of groups studied was > 6 years. One group studied was < 5 years and was assessed on Module 2 of ADOS, but no DTA results were reported
Moss 2008	No gold standard used and no non-ASD comparison group included. Study of diagnostic stability over time
Nordin 1996	No DTA result
Noterdaeme 1999	8 children with average ages of 10.6 (SD = 2) and 10.0 (SD = 2), respectively. Excluded on the basis of age
Nygren 2009	Mean age of group is 7.8 years (i.e. > 6 years)
Oosterling 2010a	Used different cutoff criteria for ADOS
Papanikolaou 2009	Participants recruited were above preschool age

Study	Reason for exclusion
Perry 2005	Used different diagnostic groups of Autism versus non-Autism (not ASD vs non-ASD)
Rellini 2004	No DTA result
Risi 2006 study 1 ADI-R	Used different diagnostic groups of Autism versus non-Autism (not ASD vs non-ASD)
Risi 2006 study 2	Mean age > 6 years
Rutter 1988a	No DTA result
Rutter 1988b	Age range > 6 years
Saemundsen 2003	No gold standard used. Comparative study between ADI-R and CARS
Schopler 1988	No DTA result
Sevin 1991	Participants recruited were above preschool age
Shin 1998	Used different cutoff criteria for CARS (used 28, not 30)
Sikora 2008	No DTA result
Skuse 2004	Mean age of participants was > 6 years
Soke 2011	Study groups do not meet inclusion criteria (i.e. no group of children not suspected of having ASD OR no control group of non-ASD children)
Stella 2002	No DTA result. No comparative group included, and only mean group scores from CARS reported
Tachimori 2003	Participants recruited were above preschool age
Teal 1982	Age range > 6 years
Tsuchiya 2013	Mean age > 6 years. Sample included individuals aged 3-19 years
Vaughan 2011	No DTA result
Ventola 2007	No DTA result
Zwaigenbaum 2011	Not a DTA study, but editorial on NICE guidelines for assessment and diagnosis

ADI-R: Autism Diagnostic Interview - Revised; **ADOS:** Autism Diagnostic Observation Schedule;**ADOS-2:** Autism Diagnostic Observation Schedule - Second Edition; **ADOS-PL:** Autism Diagnostic Observation Schedule - Pre-Linguistic; **ADOS-T:** Autism Diagnostic Observation Schedule - Toddler Module; **ASD:** autism spectrum disorder; **CARS:** Childhood Autism Rating Scale; **DTA:** diagnostic test accuracy; **NICE:** National Institute for Health and Care Excellence; **SD:** standard deviation; **3di:** Developmental, Dimensional, and Diagnostic Interview.

D A T A

Presented below are all the data for all of the tests entered into the review.

Table Tests.   Data tables by test

Test	No. of studies	No. of participants
1 ADOS	12	1625
2 CARS	4	641
3 ADI-R	5	634

Test 1. ADOS.

Test 2. CARS.

Test 3. ADI-R.

ADDITIONAL TABLES

Table 1.   Tests, method of administration, and guidelines in which they were listed at the time of commencement of this review

Test	Administration	Guidelines that included each test					
		SIGN 2007	Ministry of Health New Zealand 2008	Ministry of Health Singapore 2010	Missouri Autism Guidelines Initiative 2010	Ohio Developmental Disabilities Council 2010	Johnson 2007
ADI-R	Parent or carer interview, face-to-face	X	X	X	X	X	-
DISCO-10	Parent or carer interview, face-to-face	X	-	X	-	-	-
3di	Parent or carer interview, face-to-face with electronic data entry	X	-	-	-	-	-
GARS-2	Parent or carer interview, questionnaire	-	X	-	-	X	-
CARS	Combination of interview and observations of unstructured activity	X	X	X	X	X	X
ADOS or ADOS-G	Semi-structured observational assessment	X	X	X	X	X	-

ADI-R: Autism Diagnostic Interview - Revised; **ADOS:** Autism Diagnostic Observation Schedule; **ADOS-G:** Autism Diagnostic Observation Schedule - Generic; **CARS:** Childhood Autism Rating Scale; **DISCO-10:** Diagnostic Interview for Social and Communication Disorders - Tenth Revision; **GARS-2:** Gilliam Autism Rating Scale; **3di:** Developmental, Dimensional, and Diagnostic Interview.

Table 2.   Operationalisation of issues relevant to 'Risk of bias' and applicability assessment

Items and guide to classification

Domain 1: patient selection

A. Risk of bias

1. Was a consecutive or random sample of patients enrolled?
   a. Classify as 'yes' if the study enrolled all consecutive, or a random sample of, eligible patients referred for further diagnosis of ASD
   b. Classify as 'no' if there was clear evidence of selective sampling
   c. Classify as 'unclear' if insufficient information was given to make a judgement
2. Was a case-control design avoided?
   a. Classify as 'yes' if the study consisted of children referred for further diagnosis of ASD
   b. Classify as 'no' if the study used only healthy controls or enrolled patients with a known diagnosis of ASD and a control group without a diagnosis
   c. Classify as 'unclear' if insufficient information was given to make a judgement
3. Did the study avoid inappropriate exclusions?
   a. Classify as 'yes' if the study consisted of children representing a mixture of conditions (including absence of any condition) that are usually present (e.g. autistic disorder; pervasive developmental disorder not otherwise specified; developmental disability that is not autism but has some characteristics in common, such as global developmental delay in association with language delay, language delay alone, attachment disorders, ADHD, anxiety disorders)
   b. Classify as 'no' if the study made inappropriate exclusions, such as excluding 'difficult to diagnose' patients
   c. Classify as 'unclear' if insufficient information was given to make a judgement

B. Concerns regarding applicability

Is there concern that the included patients do not match the review question?

1. Classify concern: low/high/unclear

Domain 2: index test(s)

A. Risk of bias

1. Were the index test results interpreted without knowledge of results of the reference standard?
   a. Classify as 'yes' if results of the index test were interpreted blind to results of the reference test
   b. Classify as 'no' if the assessor of the index test was aware of the results of the reference standard
   c. Classify as 'unclear' if insufficient information was given on independent or blind assessment of the index test
2. If a threshold was used, was it pre-specified?
   a. Classify as 'yes' if a threshold was used and pre-specified
   b. Classify as 'no' if a threshold was used but was not pre-specified
   c. Classify as 'unclear' if insufficient information was given on the use of a threshold

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or its interpretation differ from the review question?

1. Classify concern: low/high/unclear

Domain 3: reference standard

A. Risk of bias

1. Is the reference standard likely to correctly classify the target condition?
   a. Classify as 'yes' if the reference standard consists of a clinical diagnosis of autism or other ASD using a current, accepted classification system (DSM-III, DSM-III-R, DSM-IV, DSM-IV-TR, ICD-9, or ICD-10), as assigned by an experienced multi-disciplinary team

Table 2.   Operationalisation of issues relevant to 'Risk of bias' and applicability assessment *(Continued)*

(including assessment of social behaviour, language and non-verbal communication, adaptive behaviour, motor skills, atypical behaviours, and cognitive status/intellectual function), and based on information from a clinical assessment and from health professionals involved in the child's care and those caring for the child in community settings such as preschool or child care settings

   b.   Classify as 'no' if the above-mentioned methods were not used

   c.   Classify as 'unclear' if insufficient information was given on the reference standard

2. Were the reference standard results interpreted without knowledge of results of the index test?

   a.   Classify as 'yes' if results of the reference standard were interpreted blind to results of the index test

   b.   Classify as 'no' if the assessor of the reference standard was aware of results of the index test

   c.   Classify as 'unclear' if insufficient information was given on independent or blind assessment of the reference standard

---

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question?

1. Classify concern: low/high/unclear

Domain 4: flow and timing

A. Risk of bias

---

1. Was there an appropriate interval between index test(s) and reference standard?

   a.   Classify as 'yes' if the time period between the index test and the reference standard was 6 months or shorter

   b.   Classify as 'no' if the time period between the index test and the reference standard was longer than 6 months

   c.   Classify as 'unclear' if there was insufficient information on the time period between the index test and the reference standard

2. Did all patients receive a reference standard?

   a.   Classify as 'yes' if it is clear that all patients or a random selection of those who received the index test went on to receive a reference standard, even if the reference standard was not the same for all patients

   b.   Classify as 'no' if not all patients or a random selection of those who received the index test received verification by a reference standard

   c.   Classify as 'unclear' if insufficient information was provided to assess this item

3. Did patients receive the same reference standard?

   a.   Classify as 'yes' if it is clear that all patients who received the index test were subjected to the same reference standard

   b.   Classify as 'no' if different reference standards were used

   c.   Classify as 'unclear' if insufficient information was provided to assess this item

4. Were all patients included in the analysis?

   a.   Classify as 'yes' if it is clear what happened to all patients who entered the study (all patients are accounted for, preferably in a flow chart), or if study authors explicitly reported the absence of any withdrawals

   b.   Classify as 'no' if it is clear that not all patients who were entered completed the study (received both index test and reference standard), and not all patients were accounted for

   c.   Classify as 'unclear' when the paper did not clearly describe whether or not all patients completed all tests and are included in the analysis

---

Notes

---

1. Relevant clinical information: Were the same clinical data available when the index test results were interpreted as would be available when the test is used in practice?

   a.   Classify as 'yes' if only clinical data (e.g. speech and language therapy; occupational therapy; developmental or psychology reports that address general assessments that are not specific for autism assessments; information from a doctor, nurse, teacher or allied health professional that lists why autism is of concern) were available in the study that normally would be available when the test results would be interpreted

   b.   Classify as 'no' if this is not the case (e.g. if other test results are available that cannot be regarded as part of routine care)

   c.   Classify as 'unclear' if the paper did not explain what clinical information was available at the time of assessment

Table 2.   Operationalisation of issues relevant to 'Risk of bias' and applicability assessment *(Continued)*

2. Conflicts of interest avoided: Were conflicts of interest avoided or absent?

    a. Classify as 'yes' if study authors/researchers were not involved in development of the diagnostic instrument

    b. Classify as 'no' if study authors/researchers were involved in development of the diagnostic instrument

    c. Classify as 'unclear' if insufficient information was given

ADHD: attention-deficit/hyperactivity disorder;**ASD:** autism spectrum disorder; **DSM-III:** *Diagnostic and Statisticial Manual of Mental Disorders - Third Edition;* **DSM-III-R:** *Diagnostic and Statistical Manual of Mental Disorders - Third Edition - Revised*; **DSM-IV:** *Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition;***DSM-IV-TR:** *Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition - Text Revision*; **ICD-9:** *International Classification of Diseases - Ninth Revision*; **ICD-10:** *International Classification of Diseases - Tenth Revision.*

Table 3. Study results for ADOS

Study	Number of partici-pants	Age of group (mean age, if available)	Study group source	Diagnos-tic groups (number of partici-pants)	Test	Module (cutoff)	Sensitiv-ity (%) (95% CI)	Speci-ficity (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)
Corsello 2013	118	24 to 36 months	Sample was 138 consecu-tive children between the ages of 24 and 36 months evaluated for ASD at a chil-dren's hospital develop-mental evaluation clinic	ASD (98) NS (20)	ADOS	M1 or M2	**97** (0.91 to 0.99)	**85** (0.62 to 0.97)	**97** (0.91 to 0.99)	**85** (0.62 to 0.97)
Gray 2008 ADOS	209 (M1: n = 195; M2: n = 14)	20 to 55 months (38.5 months)	Assessment clinic for chil-dren with developmental concerns or ASD	ASD (139) NS (56)	ADOS	M1 and M2	**76** (0.68 to 0.83)	**94** (0.85 to 0.98)	**96** (0.91 to 0.99)	**65** (0.54 to 0.74)
Kim 2012b ADOS Co-hort A	151	21 to 47 months (34 months)	Non-verbal (NV) children from 3 data sources: 1. Early diagnosis of autism 2. First words and toddlers at University of Michigan Autism and Communica-tion Disorders Centre 3. Clinic patients at Univer-sity of Michigan Autism Clinic	ASD (123) NS (28)	ADOS	ADOS modules not spec-ified but assume M1 and M2, given age of chil-dren	**98** (0.94 to 1.00)	**64** (0.44 to 0.81)	**92** (0.86 to 0.96)	**90** (0.68 to 0.99)
Kim 2012b ADOS Co-hort B	110	21 to 47 months (40 months)	Children with phrase speech from 3 sources: 1. Early diagnosis of autism 2. First words and toddlers at University of Michigan Autism and Communica-tion Disorders Centre 3. Clinic patients at Univer-sity of Michigan Autism Clinic	ASD (69) NS (41)			**97** (0.90 to 1.00)	**68** (0.52 to 0.82)	**84** (0.74 to 0.91)	**93** (0.78 to 0.99)

Table 3. **Study results for ADOS** *(Continued)*

Study	N	Age	Setting	Group	Test	Module				
Le Couteur 2008 ADOS	101	24 to 49 months (37 months)	Recruited from 2 previous unrelated studies (McConachie 2005;Shearer 2001); children suspected of having ASD	ASD (77) NS (24)	ADOS	All M1 but 2 who received M2	**83** (0.73 to 0.91)	**100** (0.86 to 1.00)	**100** (0.94 to 1.00)	**65** (0.47 to 0.80)
Lord 2000* *There were other analyses in this publication for older cohorts that were not eligible for inclusion in this review. Overall results reported here were generated from M1 and M2 data for children who did meet the age limit for inclusion	129	51 months	University of Chicago Developmental Disorders Clinic, USA	ASD (96) NS (33)	ADOS-G	Overall	**97** (0.91 to 0.99)	**91** (0.76 to 0.98)	**97** (0.91 to 0.99)	**91** (0.76 to 0.98)
	74	15 months to 10 years (50 months)		ASD (57) NS (17)		M1	**98** (not calculated)	**94** (not calculated)	**98** (not calculated)	**94** (not calculated)
	55	2 to 7 years (51 months)		ASD (39) NS (16)		M2	**95** (not calculated)	**88** (not calculated)	**95** (not calculated)	**88** (not calculated)
Mazefsky 2006 ADOS	75	22 months to 8 years (48 months)	Specialised clinic for assessment of pervasive developmental disorders at a US university medical centre	ASD (56) NS (19)	ADOS-G	M1 and M2	**93** (0.83 to 0.98)	**84** (0.60 to 0.97)	**95** (0.85 to 0.99)	**80** (0.56 to 0.94)
Oosterling 2010b ADOS	208	20 to 40 months (32.5 months)	Karakter Child and Adolescent Psychiatry University Centre, Netherlands	ASD (143) NS (65)	ADOS	M1 (204) and M2 (4)	**77** (0.69 to 0.84)	**83** (0.72 to 0.91)	**91** (0.84 to 0.95)	**62** (0.51 to 0.72)
Risi 2006 Study 1 ADOS Cohort A	270	< 36-Month group	1. University of Michigan Autism and Communication Disorders Clinic, USA 2. TEACCH® Centers at the University of North Car-	ASD (227) NS (43)	ADOS	ADOS module not specified but assume	**86** (0.81 to 0.90)	**84** (0.69 to 0.93)	**97** (0.93 to 0.99)	**53** (0.40 to 0.65)

Table 3. Study results for ADOS *(Continued)*

		(mean age not reported); 21 to 34-month group (28 months)	olina, Chapel Hill, and the University of Chicago 3. Universiy of Chicago Developmental Disorders Clinic			M1 and M2, given age of children				
Risi 2006 Study 1 ADOS Cohort B	67	36 to 112 months (62.5 months)	Mental Retardation, USA	ASD (57) NS (10)			**96** (0.88 to 1.00)	**20** (0.03 to 0.56)	**87** (0.77 to 0.94)	**50** (0.07 to 0.93)
Ventola 2006 ADOS	45	16 to 31 months (26 months)	Screening study for toddlers who failed the Modified Checklist for Autism in Toddlers	ASD (36) NS (9)	ADOS-G	M1	**97** (0.85 to 1.00)	**67** (0.30 to 0.93)	**92** (0.79 to 0.98)	**86** (0.42 to 1.00)
Wiggins 2008 ADOS	142	16 to 37 months (26 months)	Screening study of toddlers who failed the Modified Checklist for Autism in Toddlers	ASD (73) NS (69)	ADOS-G	M1 mostly reported	**96** (0.88 to 0.99)	**65** (0.53 to 0.76)	**74** (0.64 to 0.83)	**94** (0.83 to 0.99)

ADOS: Autism Diagnostic Observation Schedule; **ADOS-G:** Autism Diagnostic Observation Schedule - Generic; **ASD:** autism spectrum disorder; **CI:** confidence interval; **M:** module; **NPV:** negative predictive value; **NS:** non-spectrum;**NV:** non-verbal; **PPV:** positive predictive value.

Table 4. Study results for Autism Diagnosis Interview - Revised

Study	Number of participants	Age of group (age range)	Study group source	Diagnostic groups (number of participants)	Test/Algorithm/Variations (i.e. variation from cutoffs met for 3 domains of social interaction, communication, and repetitive behaviours) Specific cutoffs: Social interaction = 10 Restricted and repetitive behaviours = 3	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)

					Communication = 8 verbal and 7 non-verbal				
Cox 1999	30 for calculations (as 15 TD cases removed from original 45)	20 months (range not reported)	Group 1: high AD risk Group 2: medium AD risk Group 3: no AD risk	ASD (21) NS (9)	Elevated scores in all 3 domains	**19** (0.05 to 0.42)	**100** (0.66 to 1.00)	**100** (0.40 to 1.00)	**35** (0.17 to 0.56)
Gray 2008 ADI-R	209	38.5 months (20 to 55 months)	Assessment clinic for children with developmental concerns or ASD	ASD (143) NS (66)	Not specified, but assume elevated scores in all 3 domains	**73** (0.65 to 0.80)	**77** (0.65 to 0.87)	**87** (0.80 to 0.93)	**57** (0.46 to 0.67)
Oosterling 2010b ADI-R*	208	32.5 months (20 to 40 months)	Karakter Child and Adolescent Psychiatry University Centre, Netherlands	ASD (143) NS (65)	Revised algorithms for ASD (as per Risi 2006 study 1 ADI-R). Meets criteria for: 1. social interaction and communication (not behavioural); 2. social interaction AND within 2 points for communication; 3. communication AND within 2 points for social interaction; or 4. within 1 point for both social and communication	**75** (0.67 to 0.82)	**63** (0.50 to 0.75)	**82** (0.74 to 0.88)	**53** (0.42 to 0.65)
Ventola 2006 ADI-R	45	26 months (16 to 31 months)	Screening study for toddlers who failed the Modified Checklist for Autism in Toddlers	ASD (36) NS (9)	Elevated scores in all 3 domains ADI-R (n = 35) Toddler ADI-R (n = 10)	**53** (0.35 to 0.70)	**67** (0.30 to 0.93)	**86** (0.65 to 0.97)	**26** (0.10 to 0.48)
Wiggins 2008 ADI-R	142	26 months (16 to 37 months)	Screening study for toddlers who failed the Modified Checklist for Autism in Toddlers	ASD (73) NS (69)	Elevated scores in all 3 domains	**33** (0.22 to 0.45)	**94** (0.86 to 0.98)	**86** (0.67 to 0.96)	**57** (0.47 to 0.66)

Table 5. Study results for Childhood Autism Rating Scale (cutoff < 30 not autism spectrum disorder)

Study	Number of partici-pants	Mean age of group (age range)	Diagnostic groups (num-ber of participants)	CARS cut-off values	Sensitivi-ty (%) (95% CI)	Specifici-ty (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)	+ LR	− LR
Chłebows-ki 2010	354	26 months (21 to 30 months)	ASD (236) NS (118)	30	**66** (0.59 to 0.72)	**96** (0.90 to 0.99)	**97** (0.93 to 0.99)	**58** (0.51 to 0.65)	15.5	0.4
Russell 2010	100	61 months (range not re-ported)	ASD (86), included 1 child with Rett's syndrome NS (14) Severe/profound intellec-tual disability (72) and Unspecified intellectual disability (21)	30	**87** (0.78 to 0.93)	**21** (0.05 to 0.51)	**87** (0.78 to 0.93)	**21** (0.05 to 0.51)	1.1	0.6
Ventola 2006 CARS	45	26 months (16 to 31 months)	ASD (36) NS (9)	Not clearly stated but assume 30	**89** (0.74 to 0.97)	**100** (0.66 to 1.00)	**100** (0.89 to 1.00)	**69** (0.39 to 0.91)	-	0.1
Wiggins 2008 CARS	142	26 months (16 to 37 months)	ASD (73) NS (69)	Not clearly stated but assume 30	**71** (0.59 to 0.81)	**93** (0.84 to 0.98)	**91** (0.81 to 0.97)	**75** (0.65 to 0.84)	9.8	0.3

ASD: autism spectrum disorder; **CI:** confidence interval; **+ LR:** positive likelihood ratio;**- LR:** negative likelihood ratio; **NPV:** negative predictive value; **NS:** non-spectrum;**PPV:** positive predictive value.

Table 6. Studies addressing more than one instrument in the same study sample

Study authors	Number of partici-pants	Age range	Sensitivity				Specificity			
			Number with ASD	ADOS (95% CI)	CARS (95% CI)	ADI-R (95% CI)	Number without ASD	ADOS (95% CI)	CARS (95% CI)	ADI-R (95% CI)

Gray 2008 ADI-R; Gray 2008 ADOS	209	20 to 55 months	143	0.76 (0.68 to 0.83)	-	0.73 (0.65 to 0.80)	66	0.94 (0.85 to 0.98)	-	0.77 (0.65 to 0.87)
Oosterling 2010b ADI-R; Oosterling 2010b ADOS	208	20 to 40 months	143	0.77 (0.69 to 0.84)	-	0.75 (0.67 to 0.82)	65	0.83 (0.72 to 0.91)	-	0.63 (0.50 to 0.75)
Ventola 2006 ADI-R; Ventola 2006 ADOS; Ventola 2006 CARS	45	16 to 31 months	36	0.97 (0.85 to 1.00)	0.89 (0.74 to 0.97)	0.53 (0.35 to 0.70)	9	0.67 (0.30 to 0.93)	1.00 (0.66 to 1.00)	0.67 (0.30 to 0.93)
Wiggins 2008 ADI-R; Wiggins 2008 ADOS; Wiggins 2008 CARS	142	16 to 37 months	73	0.96 (0.88 to 0.99)	0.71 (0.59 to 0.81)	0.33 (0.22 to 0.45)	69	0.65 (0.53 to 0.76)	0.93 (0.84 to 0.98)	0.94 (0.86 to 0.98)

ADI-R: Autism Diagnostic Interview - Revised;**ADOS:** Autism Diagnostic Observation Schedule;**ASD:** autism spectrum disorder**; CARS:** Childhood Autism Rating Scale; **CI:** confidence interval.

A P P E N D I C E S

Appendix 1. Search strategies

1. Cochrane Central Register of Controlled Trials (CENTRAL), in the Cochrane Library

Searched: 11 February 2011 (275 records); 1 April 2012 (12 records); 14 May 2013 (5 records); 21 July 2016 (176 records)

#1MeSH descriptor: [Child Development Disorders, Pervasive] explode all trees
#2pervasive development* disorder*
#3(PDD or PDDs or ASD or ASDs)
#4autis*
#5asperger*
#6kanner*
#7childhood schizophrenia
#8Rett*
#9#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8
#10gilliam* near/5 autis*
#11GARS
#12(diagnos* next interview*) near/5 (communic* next disorder*)
#13DISCO
#14autis* next diagnos* next interview*
#15ADI-R
#16development* near/3 dimension* near/3 diagnos*
#173di
#18 child* next autis* next rating
#19 CARS
#20(autis* next diagnos* next observ*) or ADOS
#21MeSH descriptor: [Psychiatric Status Rating Scales] this term only
#22MeSH descriptor: [Psychometrics] this term only
#23MeSH descriptor: [Neuropsychological Tests] explode all trees
#24MeSH descriptor: [Psychological Tests] this term only
#25MeSH descriptor: [Interview, Psychological] this term only
#26MeSH descriptor: [Interviews as Topic] this term only
#27MeSH descriptor: [Personality Assessment] this term only
#28MeSH descriptor: [Observation] this term only
#29MeSH descriptor: [Questionnaires] this term only
#30rating next scale*
#31((diagnos* or screen*) near/3 (algorithm* or assess* or interview* or instrument* or observation* or questionnaire* or schedule* or tool*))
#32((parent*) near/3 (interview* or questionnaire* or report*))
#33((carer*) near/3 (interview* or questionnaire* or report*))
#34((caregiver*) near/3 (interview* or questionnaire* or report*))
#35#10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34
#36#9 and #35
#37(infant* or child* or toddler* or preschool* or pre-school*)
#38#36 and #37 in Trials [Note: final line of 2011 search]
#39#36 and #37 Publication Year from 2011 to 2012, in Trials [Note: final line of 2012 search]
#40#36 and #37 Publication Year from 2012 to 2013, in Trials [Note: final line of 2013 search]
#41#36 and #37 Publication Year from 2013 to 2016, in Trials [Note: final line of 2016 search]

2. MEDLINE Ovid

Searched: 10 February 2011 (3943 records); 31 March 2012 (535 records); 13 May 2013 (689 records); 20 July 2016 (1849 records)

1 exp child development disorders, pervasive/
2 pervasive development$ disorder$.tw. (1
3 (PDD or PDDs or ASD or ASDs).tw.
4 autis$.tw.
5 asperger$.tw.
6 kanner$.tw.
7 childhood schizophrenia.tw.

Diagnostic tests for autism spectrum disorder (ASD) in preschool children (Review)

8 Rett$.tw.
9 or/1-8
10 (gilliam$ adj5 autis$).tw.
11 GARS.tw.
12 DISCO.tw.
13 (diagnos$ interview$ adj5 communic$ disorder$).tw.
14 ADI-R.tw.
15 autis$ diagnos$ interview$.tw.
16 3di.tw.
17 (development$ adj3 dimension$ adj3 diagnos$).tw.
18 child$ autis$ rating.tw.
19 CARS.tw.
20 autis$ diagnos$ observ$.tw.
21 ADOS.tw.
22 Psychiatric Status Rating Scales/
23 Psychometrics/
24 neuropsychological tests/
25 psychological tests/
26 Interview, Psychological/
27 Interviews as Topic/
28 Personality Assessment/
29 observation/
30 Questionnaires/
31 rating scale$.tw.
32 ((diagnos$ or screen$) adj3 (algorithm$ or assess$ or interview$ or instrument$ or observation$ or questionnaire$ or schedule$ or test $ or tool$)).tw.
33 ((parent$ or carer$ or caregiver$) adj3 (interview$ or questionnaire$ or report$)).tw.
34 or/10-33
35 9 and 34
36 remove duplicates from 35
37 limit 36 to ED=20110201-20120331 [Note: final line of 2012 search]
38 limit 36 to ED=20120301-20130513 [Note: final line of 2013 search]
39 limit 36 to ED=20130501-20160707 [Note: final line of 2016 search]

3. Embase Ovid

Searched: 10 February 2011 (3628 records); 31 March 2012 (711 records); 13 May 2013 (845 records); 20 July 2016 (3253 records)

1 exp autism/
2 pervasive development$ disorder$.tw.
3 (PDD or PDDs or ASD or ASDs).tw.
4 autis$.tw.
5 asperger$.tw.
6 kanner$.tw.
7 childhood schizophrenia.tw.
8 Rett$.tw.
9 or/1-8
10 (gilliam$ adj5 autis$).tw.
11 GARS.tw.
12 (diagnos$ interview$ adj5 communic$ disorder$).tw.
13 DISCO.tw.
14 autis$ diagnos$ interview$.tw.
15 ADI-R.tw.
16 (development$ adj3 dimension$ adj3 diagnos$).tw.
17 3di.tw.
18 child$ autis$ rating.tw.
19 CARS.tw.
20 autis$ diagnos$ observ$.tw.
21 ADOS.tw.
22 Psychiatric Status Rating Scales/
23 psychological rating scale/
24 psychometry/
25 neuropsychological tests/

26 psychologic test/
27 personality test/
28 interview/
29 semi structured interview/
30 structured interview/
31 observation/
32 questionnaire/
33 developmental screening/
34 ((diagnos$ or screen$) adj3 (algorithm$ or assess$ or interview$ or instrument$ or observation$ or questionnaire$ or schedule$ or test$ or tool$)).tw.
35 ((parent$ or carer$ or caregiver$) adj3 (interview$ or questionnaire$ or report$)).tw.
36 or/10-35
37 9 and 36
38 limit 37 to yr="2011 -Current" [Note: final line of 2012 search]
39 limit 37 to yr="2012 -Current" [Note: final line of 2013 search]
39 limit 37 to yr="2013 -Current" [Note: final line of 2016 search]

4. PsycINFO

PsycINFO via Ovid

Searched: 31 March 2012 (425 records); 13 May 2013 (446 records); 20 July 2016 (1659 records)

1 exp Pervasive Developmental Disorders/
2 pervasive development$ disorder$.tw.
3 (PDD or PDDs or ASD or ASDs).tw.
4 (autis* or asperg* or kanner* or rett*).tw.
5 childhood schizophrenia.tw.
6 or/1-5
7 (gilliam$ adj5 autis$).tw.
8 GARS.tw.
9 (diagnos$ interview$ adj5 communic$ disorder$).tw.
10 DISCO.tw. (6
11 autis$ diagnos$ interview$.tw.
12 ADI-R.tw.
13 child$ autis$ rating.tw.
14 CARS.tw.
15 autis$ diagnos$ observ$.tw.
16 ADOS.tw.
17 (development$ adj3 dimension$ adj3 diagnos$).tw.
18 3di.tw.
19 Rating Scales/
20 Psychometrics/
21 Observation Methods/
22 Questionnaires/
23 Diagnostic Interview Schedule/
24 Interview Schedules/
25 Psychodiagnostic Interview/
26 Psychological Assessment/
27 Screening Tests/
28 Structured Clinical Interview/
29 Behavioral Assessment/
30 Neuropsychological Assessment/)
31 Cognitive Assessment/
32 ((diagnos$ or screen$) adj3 (algorithm$ or assess$ or interview$ or instrument$ or observation$ or questionnaire$ or schedule$ or test or tool$)).tw.
33 ((parent$ or carer$ or caregiver$) adj3 (interview$ or questionnaire$ or report$)).tw.
34 or/7-33
35 6 and 34
36 limit 35 to up=20110210-20120331 [Note: final line 2012]
37 limit 35 to up=20120326-20130513 [Note: final line 2013]
38 limit 35 to yr="2013 -Current" [Note: final line 2016]

PsycINFO via EBSCOhost

Searched: 9 February 2011 (3528 records)

S37 S6 and S36
S36 S7 or S8 or S9 or S10 or S11 or S12 or S13 or S14 or S15 or S16 or S17 or S18 or S19 or S20 or S21 or S22 or S23 or S24 or S25 or S26 or S27 or S28 or S29 or S30 or S31 or S32 or S33 or S34 or S35
S35 (parent* N3 interview*) or (parent* N3 questionnaire*) or (parent* N3 report*)
S34 (parent* N3 interview*) or (parent* N3 questionnaire*) or (parent* N3 report*)
S33 (screen* N3 algorithm*) or (screen* N3 assess*) or (screen* N3 interview*) or (screen* N3 instrument*) or (screen* N3 observation*) or (screen* N3 questionnaire*) or (screen* N3 schedule*) or (screen* N3 tool*)
S32 (diagnos* N3 algorithm*) or (diagnos* N3 assess*) or (diagnos* N3 interview*) or (diagnos* N3 instrument*) or (diagnos* N3 observation*) or (diagnos* N3 questionnaire*) or (diagnos* N3 schedule*) or (diagnos* N3 test*) or (diagnos* N3 tool*)
S31 DE "Cognitive Assessment"
S30 DE "Neuropsychological Assessment"
S29 DE "Behavioral Assessment"
S28 DE "Structured Clinical Interview"
S27 DE "Screening Tests"
S26 DE "Psychological Assessment"
S25 DE "Psychodiagnostic Interview"
S24 DE "Interview Schedules"
S23 DE "Diagnostic Interview Schedule"
S22 DE "Questionnaires"
S21 DE "Observation Methods"
S20 DE "Psychometrics"
S19 DE "Rating Scales"
S18 3di
S17 development* N3 dimension* N3 diagnos*
S16 ADOS
S15 autis* diagnos* observ*
S14 CARS
S13 child* autis* rating
S12 ADI-R
S11 autis* diagnos* interview*
S10 DISCO
S9 diagnos* interview* N5 communic* disorder*
S8 GARS
S7 gilliam* N5 autis*
S6 S1 or S2 or S3 or S4 or S5
S5 childhood schizophrenia
S4 autis* or asperg* or kanner* or rett*
S3 PDD or PDDs or ASD or ASDs
S2 pervasive development* disorder*
S1 DE "Pervasive Developmental Disorders" OR DE "Aspergers Syndrome" OR DE "Autism" OR DE "Rett Syndrome"

5. CINAHL Plus EBSCOhost (Cumulative Index to Nursing and Allied Health Literature)

Searched: 10 February 2011 (2509 records); 31 March 2012 (425 records); 14 May 2013 (717 records); 20 July 2016 (1887 records)

S42 S36 AND S41 [Note: final line 2016]
S41 EM 20130501-
S40 S36 AND S39 [Note: final line 2013]
S39 EM 20120301-
S38 S36 and S37 [Note: final line 2012]
S37 EM 20110200-
S36 S7 and S35
S35 S8 or S9 or S10 or S11 or S12 or S13 or S14 or S15 or S16 or S17 or S18 or S19 or S20 or S21 or S22 or S23 or S24 or S25 or S26 or S27 or S28 or S29 or S30 or S31 or S32 or S33 or S34
S34 (caregiver* N3 interview*) or (caregiver* N3 questionnaire*) or (caregiver* N3 report*) S33 (carer* N3 interview*) or (carer* N3 questionnaire*) or (carer* N3 report*)
S32 (parent* N3 interview*) or (parent* N3 questionnaire*) or (parent* N3 report*)
S31 (screen* N3 algorithm*) or (screen* N3 assess*) or (screen* N3 interview*) or (screen* N3 instrument*) or (screen* N3 observation*) or (screen* N3 questionnaire*) or (screen* N3 schedule*) or (screen* N3 tool*)

S30 (diagnos* N3 algorithm*) or (diagnos* N3 assess*) or (diagnos* N3 interview*) or (diagnos* N3 instrument*) or (diagnos* N3 observation*) or (diagnos* N3 questionnaire*) or (diagnos* N3 schedule*) or (diagnos* N3 test*) or (diagnos* N3 tool*)
S29 rating scale*
S28 (MH "Observational Methods")
S27 (MH "Interviews")
S26 (MH "Psychometrics")
S25 (MH "Diagnosis, Psychosocial")
S24 (MH "Personality Assessment")
S23 (MH "Clinical Assessment Tools")
S22 (MH "Questionnaires")
S21 (MH "Neuropsychological Tests")
S20 (MH "Psychological Tests")
S19 (MH "Behavior Rating Scales")
S18 (development* N3 dimension* N3 diagnos*) OR "3di"
S17 ADOS
S16 autis* diagnos* observ*
S15 CARS
S14 child* autis* rating
S13 ADI-R
S12 autis*
S11 DISCO
S10 diagnos* interview* N5 communic* disorder*
S9 GARS
S8 gilliam* N5 autis*
S7 S1 or S2 or S3 or S4 or S5 or S6
S6 childhood schizophrenia
S5 autis* or asperg* or kanner* or rett*
S4 PDD or PDDs or ASD or ASDs
S3 pervasive development* disorder*
S2 (MH "Rett Syndrome")
S1 (MH "Child Development Disorders, Pervasive+")

6 Science Citation Index and Social Science Citation Index Web of Science

Searched: 10 February 2011 (3980 records); 31 March 2012 (930 records); 14 May 2013 (803 records); 21 July 2016 (1314 records)

#28 #24 AND #5 [Note Final line 2016]
Indexes=SCI-EXPANDED, SSCI Timespan=2013-2016
#27 #24 AND #5 [Note Final line 2013]
Indexes=SCI-EXPANDED, SSCI Timespan=2012-2013
#26 #24 AND #5 [Note Final line 2012]
Indexes=SCI-EXPANDED, SSCI Timespan=2011-2012
#25 #24 AND #5 [Note Final line: 2011]
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#24 #23 OR #22 OR #21 OR #20 OR #19 OR #18 OR #17 OR #16 OR #15 OR #14 OR #13 OR #12 OR #11 OR #10 OR #9 OR #8 OR #7 OR #6
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#23 TS= ((caregiver*) NEAR/5 (interview* or questionnaire* or report*))
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#22 TS= ((carer*) NEAR/5 (interview* or questionnaire* or report*))
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#21 TS= ((parent*) NEAR/5 (interview* or questionnaire* or report*))
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#20 TS=rating scale*
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#19 TS= psychometric*
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#18 TS= ((screen*) NEAR/5 (algorithm* OR assess* OR interview* OR instrument* OR observation* OR questionnaire* OR schedule* OR tool*))
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#17 TS= ((diagnos*) NEAR/5 (algorithm* or assess* or interview* or instrument* or observation* or questionnaire* or schedule* or test* or tool*))
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#16 TS=3di

Indexes=SCI-EXPANDED, SSCI Timespan=All years
#15 TS= (development* NEAR/3 dimension* NEAR/3 diagnos*)
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#14 Ts=ADOS
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#13 TS= (autis* diagnos* observ*)
Indexes=SCI-EXPANDED, SSCI;
#12 TS=CARS
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#11 TS= "ADI-R"
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#10 TS= ("child* autis* rating")
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#9 TS= "autis* diagnos* interview* "
Indexes=SCI-EXPANDED, SSCI;
#8 TS=disco
Indexes=SCI-EXPANDED, SSCI;
#7 TS= ((diagnos* NEAR/1 interview*) NEAR/3 (communic* NEAR/1 disorder*))
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#6 TS=(gilliam NEAR/5 autis*) or TS=(GARS);
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#5 #4 OR #3 OR #2 OR #1
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#4 TS=(PDD or PDDs or ASD or ASDs)
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#3 TS=(childhood schizophrenia)
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#2 TS=(pervasive development* disorder*)
Indexes=SCI-EXPANDED, SSCI Timespan=All years
#1 TS=(autis* or asperg* or kanner* or rett*)
Indexes=SCI-EXPANDED, SSCI Timespan=All years

7. Conference Proceedings Citation Index-Science and Conference Proceedings Citation Index-Social Sciences, Arts & Humanities Citation Index (Web of Science)

Searched: 10 February 2011 (215 records); 31 March 2012 (9 records); 14 May 2013 (11 records); 21 July 2016 (13 records)

#28 #24 AND #5 [Note: Final line 2016]
Indexes=CPCI-S, CPCI-SSH Timespan=2013-2016
#27 #24 AND #5 [Note Final line: 2013]
Indexes=CPCI-S, CPCI-SSH Timespan=All years Timespan=2012-2013
#26 #24 AND #5 [Note: Final line 2012]
Indexes=CPCI-S, CPCI-SSH Timespan=All years Timespan=2011-2012
#25 #24 AND #5 [Note: Final line: 2011]
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#24 #23 OR #22 OR #21 OR #20 OR #19 OR #18 OR #17 OR #16 OR #15 OR #14 OR #13 OR #12 OR #11 OR #10 OR #9 OR #8 OR #7 OR #6
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#23 TS= ((caregiver*) NEAR/5 (interview* or questionnaire* or report*))
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#22 TS= ((carer*) NEAR/5 (interview* or questionnaire* or report*))
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#21 TS= ((parent*) NEAR/5 (interview* or questionnaire* or report*))
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#20 TS=rating scale*
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#19 TS= psychometric*
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#18 TS= ((screen*) NEAR/5 (algorithm* OR assess* OR interview* OR instrument* OR observation* OR questionnaire* OR schedule* OR tool*))
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#17 TS= ((diagnos*) NEAR/5 (algorithm* or assess* or interview* or instrument* or observation* or questionnaire* or schedule* or test* or tool*))
Indexes=CPCI-S, CPCI-SSH Timespan=All years

#16 TS=3di
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#15 TS= (development* NEAR/3 dimension* NEAR/3 diagnos*)
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#14 Ts=ADOS
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#13 TS= (autis* diagnos* observ*)
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#12 TS=CARS
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#11 TS= "ADI-R"
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#10 TS= ("child* autis* rating")
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#9 TS= "autis* diagnos* interview* "
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#8 TS=discov
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#7 TS= ((diagnos* NEAR/1 interview*) NEAR/3 (communic* NEAR/1 disorder*))
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#6 TS=(gilliam NEAR/5 autis*) or TS=(GARS)
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#5 #4 OR #3 OR #2 OR #1
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#4 TS=(PDD or PDDs or ASD or ASDs)
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#3 TS=(childhood schizophrenia)
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#2 TS=(pervasive development* disorder*)
Indexes=CPCI-S, CPCI-SSH Timespan=All years
#1 TS=(autis* or asperg* or kanner* or rett*)

Indexes=CPCI-S, CPCI-SSH Timespan=All years

8. ASSIA (Cambridge Scientific Abstracts)

Searched: 11 February 2011 (482 records). ASSIA was not available to the editorial base or the review authors after 2011.

((KW= ((diagnos* or screen*) within 3 (algorithm* or assess* or interview* or instrument* or observation* or questionnaire* or schedule* or tool*))) or(KW= ((parent* or carer* or caregiver*) within 3 (interview* or questionnaire* or report*))) or(DE=("questionnaires" or "screening" or "structured interviews" or "interviews" or "structured behavioural interviews" or "structured clinical interviews" or "diagnostic testing" or "neuropsychological tests" or "personality tests" or "psychological tests" or "psychometric tests"))) and((DE=("pervasive developmental disorders" or "aspergers syndrome" or "autistic spectrum disorders" or "rett syndrome")) or(KW=autis* or asperger* or rett* or kanner*) or(KW=childhood schizophrenia) or(KW=PDD or PDDs or ASD or ASDs))

9. Social Services Abstracts

Social Services Abstracts via ProQuest

Searched: 14 May 2013 (limited by publication year 2011-2013); 21 July 2016 (limited by publication year 2013-2016)

((SU.EXACT("Indexes (Measures)") OR SU.EXACT("Diagnosis") OR SU.EXACT("Questionnaires") OR SU.EXACT("Scales") OR SU.EXACT("Tests") OR SU.EXACT("Measures (Instruments)") OR SU.EXACT ("Interview Schedules") OR SU.EXACT("Interviews") OR ((diagnos* OR screen*) NEAR/5 (algorithm*)) OR ((diagnos* OR screen*) NEAR/5 (assess*)) OR ((diagnos* OR screen*) NEAR/5 (interview*)) OR ((diagnos* OR screen*) NEAR/5 (instrument*)) OR ((diagnos* OR screen*) NEAR/5 (observation*)) OR ((diagnos* OR screen*) NEAR/5 (questionnaire*)) OR ((diagnos* OR screen*) NEAR/5 (schedule*)) OR ((diagnos* OR screen*) NEAR/5 (tool*)) OR ((diagnos* OR screen*) NEAR/5 (test*)) OR ((parent* OR carer* OR caregiver*) NEAR/5 (interview* OR questionnaire* OR report*)) OR (gilliam NEAR/5 autis*) OR gars OR ("diagnostic interview" NEAR/5 "communication disorder*") OR DISCO OR "autis* diagnos* interview*" OR "ADI-R" OR "autis* diagnos* observ*" OR ados OR "child* autis* rating" OR CARS OR ("develop* near/5 dimension* near/5 diagnos*") OR 3di) AND (SU.EXACT("Developmental Disabilities") OR SU.EXACT("Autism") OR AUTIS* OR ASPERGER* OR RETT OR RETTS OR KANNER OR KANNERS OR PDD OR PDDs OR ASD OR ASDs OR "pervasive development* disorder*"))

Social Services Abstracts via Cambridge Scientific Abstracts

Searched: 11 February 2011 (28 records). Not available in 2012.

Query: ((DE=("developmental disabilities" or "autism")) or(KW=(autis* or asperger* or rett* or kanner*) or(KW=childhood schizophrenia) or(KW=PDD or PDDs or ASD or ASDs)) and((DE=("diagnosis" or "interview schedules" or "interviews" or "measures instruments" or "questionnaires" or "tests")) or(KW= (diagnos* or screen*)within 3 (algorithm* or assess* or interview* or instrument* or observation* or questionnaire* or schedule* or tool*)) or(KW= (parent* or carer* or caregiver*) within 3 (interview* or questionnaire* or report*)))

10. ERIC (Education Resources Information Center)

ERIC Via EBSCOhost

Searched: 14 May 2013, limited by publication year 2012-2013 (409 records); 21 July 2016, limited by publication year 2013-2016; S1 DE "Pervasive Developmental Disorders" OR DE "Asperger Syndrome" OR DE "Autism"

S2 AUTIS* OR ASPERGER* OR RETT* OR KANNER*
S3 PDD* OR ASD OR ASDs
S4 "pervasive development* disorder*"
S5 S1 OR S2 OR S3 OR S4
S6 DE "Diagnostic Tests"
S7 DE "Clinical Diagnosis"
S8 DE "Screening Tests"
S9 DE "Test Reliability"
S10 DE "Comparative Testing"
S11 DE "Screening Tests"
S12 DE "Questionnaires"
S13 DE "Rating Scales"
S14 DE "Interviews"
S15 ((parent* OR carer* OR caregiver* Or care-giver*)) N3 (interview* OR questionnaire*))
S16 gilliam N3 autis*
S17 gars
S18 "diagnostic interview" N5 "communication disorder*"
S19 DISCO
S20 "autis* diagnos* interview*"
S21 "ADI-R"
S22 "autis* diagnos* observ*"
S23 ados
S24 "child* autis* rating"
S25 CARS
S26 (development* n3 dimension*) N3 diagnos*
S27 3di
S28 S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27
S29 S5 AND S28

ERIC Via ProQuest

Searched: 1 April 2012 (232 records)

(((SU.EXACT("Pervasive Developmental Disorders") OR SU.EXACT("Autism") OR SU.EXACT("Asperger Syndrome") OR (AUTIS* OR ASPERGER* OR RETT OR RETTS OR KANNER OR KANNERS OR PDD OR PDDs OR ASD OR ASDs OR "pervasive development* disorder*")) AND (SU.EXACT("Diagnostic Tests") OR SU.EXACT("Test Interpretation") OR SU.EXACT("Test Reliability") OR SU.EXACT("Test Validity") OR SU.EXACT("Comparative Testing") OR SU.EXACT("Measures (Individuals)") OR SU.EXACT("Screening Tests") OR SU.EXACT("Clinical Diagnosis") OR SU.EXACT("Rating Scales") OR SU.EXACT("Questionnaires") OR SU.EXACT("Observation") OR SU.EXACT("Interviews"))) OR ((SU.EXACT("Pervasive Developmental Disorders") OR SU.EXACT("Autism") OR SU.EXACT("Asperger Syndrome") OR (AUTIS* OR ASPERGER* OR RETT OR RETTS OR KANNER OR KANNERS OR PDD OR PDDs OR ASD OR ASDs OR "pervasive development* disorder*")) AND ("rating scale*" OR ((diagnos* OR screen*) NEAR/5 (algorithm* OR assess* OR interview* or instrument* OR observation* OR questionnaire* OR schedule* OR tool*)))) OR ((SU.EXACT("Pervasive Developmental Disorders") OR SU.EXACT("Autism") OR SU.EXACT("Asperger Syndrome") OR (AUTIS* OR ASPERGER* OR RETT OR RETTS OR KANNER OR KANNERS OR PDD OR PDDs OR ASD OR ASDs OR "pervasive development* disorder*")) AND (((parent* OR carer* OR caregiver*) NEAR/5 (interview* OR questionnaire* OR report*)) OR ((gilliam NEAR/5 autis*) OR gars) OR ("diagnostic interview" NEAR/5 ("communication disorder*") OR DISCO) OR (("autis* diagnos* interview*") OR "ADI-R") OR (("autis* diagnos* observ*") OR ados) OR ("child* autis* rating" OR CARS) OR (("child* near/5 dimension* near/5 diagnos*") OR 3di)))) AND pd(20110101-20121231)

ERIC Via DataStar

Searched: 10 February 2011 (1767 records)

"((Pervasive-Developmental-Disorders#.DE.) OR (rett$1 ADJ syndrome) OR (autis$) OR (pervasive ADJ development$2 ADJ disorder$1) OR (asperg$3) OR (kannerS1) OR (PDD OR PDDs OR ASD OR ASDs)) AND ((gilliam$ NEAR autis$ OR GARS) OR (diagnos$ ADJ interview$ NEAR communic$ ADJ disorder$ OR DISCO) OR (autis$ ADJ diagnos$ ADJ interview$ OR ADI-R) OR (autis$ ADJ diagnos$ ADJ observ$ OR ADOS) OR (child$ ADJ autis$ ADJ rating OR CARS) OR (child$ NEAR dimension$ NEAR diagnos$ OR 3di) OR (PSYCHOLOGICAL-TESTING.DE.) OR (TEST-VALIDITY.DE.) OR (DIAGNOSTIC-TESTS.DE.) OR (TEST-RELIABILITY.DE.) OR (TEST-INTERPRETATION.DE.) OR (COMPARATIVE-TESTING.DE.) OR (MEASURES-INDIVIDUALS.DE.) OR (CLINICAL-DIAGNOSIS.DE.) OR (SCREENING-TESTS.DE.) OR (QUESTIONNAIRES.W..DE.) OR (rating ADJ scale$1) OR (RATING-SCALES.DE.) OR (Observation.W..DE.) OR (Interviews.W..DE.) OR ((diagnos$3 OR screen$3) NEAR (algorithm$2 OR assess$4 OR interview$1 OR instrument$1 OR observation$1 OR questionnaire$1 OR schedule$1 OR tool$1)) OR ((parent$1 OR carer$1 OR caregiver$1) NEAR (interview$1 OR questionnaire$1 OR report$1)))"

11 Database of Abstracts of Reviews of Effect (DARE), part of the Cochrane Library

Searched: 11 February 2011 (32 records); 14 May 2013, limited by publication year 2012-2013 (0 records); 21 July 2016, limited by publication year 2013 -2016 (0 records)

#1 MeSH descriptor: [Child Development Disorders, Pervasive] explode all trees
#2 pervasive development* disorder*
#3 (PDD or PDDs or ASD or ASDs)
#4 autis*
#5 asperger*
#6 kanner*
#7 childhood schizophrenia
#8 Rett*
#9 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8
#10 gilliam* near/5 autis*
#11 GARS
#12 (diagnos* next interview*) near/5 (communic* next disorder*)
#13 DISCO
#14 autis* next diagnos* next interview*
#15 ADI-R
#16 development* near/3 dimension* near/3 diagnos*
#17 3di
#18 child* next autis* next rating
#19 CARS
#20 (autis* next diagnos* next observ*) or ADOS
#21 MeSH descriptor: [Psychiatric Status Rating Scales] this term only
#22 MeSH descriptor: [Psychometrics] this term only
#23 MeSH descriptor: [Neuropsychological Tests] explode all trees
#24 MeSH descriptor: [Psychological Tests] this term only
#25 MeSH descriptor: [Interview, Psychological] this term only
#26 MeSH descriptor: [Interviews as Topic] this term only
#27 MeSH descriptor: [Personality Assessment] this term only
#28 MeSH descriptor: [Observation] this term only
#29 MeSH descriptor: [Questionnaires] this term only
#30 rating next scale*
#31 ((diagnos* or screen*) near/3 (algorithm* or assess* or interview* or instrument* or observation* or questionnaire* or schedule* or tool*))
#32 ((parent*) near/3 (interview* or questionnaire* or report*))
#33 ((carer*) near/3 (interview* or questionnaire* or report*))
#34 ((caregiver*) near/3 (interview* or questionnaire* or report*))
#35 #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34
#36 #9 and #35
#37(infant* or child* or toddler* or preschool* or pre-school*)
#38 #36 and #37 in Other Reviews

12. National Autistic Society: Library Catalogue (previously AutismData)

National Autistic Society – Library Catalogue

Searched: 21 July 2016 (15 records after duplication with previous records)

Title: GILLIAM OR GARS OR ADOS OR CARS OR 3di OR DISCO OR ADI-R OR DIAGNOSTIC

AutismData

Searched: 11 February 2011 (137 records); 2 April 2012 (23 records after duplication with previous records); 14 May 2013 (21 records after duplication with previous records)

KEYWORDS ="Gilliam Asperger s Disorder Scale" / ="Gilliam Autism Rating Scale" / ="Gilliam Autism Rating Scale 2" / ="DISCO" / ="Autism diagnostic interview" / ="Autism Diagnostic Interview Revised" / ="Autism Diagnostic Observation Schedule" / ="Autism Diagnostic Observation Schedule Revised" / ="Developmental diagnostic and dimensional interview 3Di" / ="Developmental Dimensional and Diagnostic Interview 3di" / ="Childhood Autism Rating Scale" / ="Childhood Autism Rating Scale" / ="Diagnostic instruments" / ="Diagnostic markers"

Appendix 2. Glossary

3di: Developmental, Dimensional, and Diagnostic Interview.
AAN: American Academy of Neurology.
AAP: American Academy of Pediatrics.
ABC: Autism Behaviour Checklist.
AD: autistic disorder.
ADI-R: Autism Diagnostic Interview - Revised.
ADOS-G: Autism Diagnostic Observation Schedule - Generic.
ASD: autism spectrum disorder.
ASQ: Autism Screening Questionnaire.
CARS: Childhood Autism Rating Scale.
CHAT: Checklist for Autism in Toddlers.
DISCO: Diagnostic Interview for Social and Communication Disorders.
DSM-IV: *Diagnostic and Statistical Manual of Mental Disorders.*
GABA: gamma-aminobutyric acid.
GARS: Gilliam Autism Rating Scale.
ICD-10: *International Statistical Classification of Diseases - Tenth Revision.*
M-CHAT: Modified Checklist for Autism in Toddlers.
NLGN 3/4: neuroligin-3/4.
PDD-NOS: pervasive developmental disorder not otherwise specified.
PEDS: Parent Evaluation of Developmental Status.
PTEN: phosphatase and tensin homolog.
QUADAS: Quality Assessment of Studies of Diagnostic Accuracy.
SHANK3: SH3 and multiple ankyrin repeat domain 3.
SIGN: Scottish Intercollegiate Guidelines Network.
STAT: Screening Tool for Autism in Two-Year-Olds.

Appendix 3. Conflict of interest information

For Risi 2006 Study 1 ADOS Cohort A and Risi 2006 Study 1 ADOS Cohort B, and Kim 2012b ADOS Cohort A and Kim 2012b ADOS Cohort B, study authors declared that "profits related to this study were donated to charity".

For Le Couteur 2008 ADOS and Lord 2000, no conflict of interest statement was found.

For Gray 2008 ADI-R and Gray 2008 ADOS, study authors conduct training for the Autism Diagnostic Observation Schedule (ADOS) - Generic, ADOS - Second Edition (ADOS-2), and the Autism Diagnosis Interview - Revised, which raises the potential for conflict of interest, but no information by which to assess this was provided.

CONTRIBUTIONS OF AUTHORS

Aarti Samtani undertook the initial review of articles in 2011. She assessed titles and abstracts for inclusion, assessed the quality of included articles, and extracted data from the first search undertaken in February 2011. Melinda Randall and Nuala Livingstone undertook the same for the top-up searches run in 2012 and 2013. Melinda Randall, Kristine Egberts, and Nuala Livingstone undertook screening of the 2016 search. Nuala Livingstone transferred data from the original QUADAS format into QUADAS-2. Rob Scholten and Lotty Hooft assisted with quality assessment, data extraction, and statistical analyses. Susan Woolfenden and Katrina Williams acted as arbitrators if differences of opinion occurred with regard to inclusions, quality assessment, and data extraction, and provided advice regarding clinical relevance. Kristine Egberts reviewed and altered the text of the final draft, revised the QUADAS-2 figure and flow diagram, reviewed inclusion of analyses and referencing, and co-ordinated the efforts of the other review authors to progress the Review to completion.

Katy Sterling-Levis died on 29 January 2015. She had assisted with protocol development and the first draft of the Review, as well as exclusion of references and QUADAS allocation for results of the first search.

DECLARATIONS OF INTEREST

The review author team was established with support from the William Collie Trust, and their work was administered by the University of Melbourne.

Melinda Randall - none known.
Kristine J Egberts - Editor with the Cochrane Developmental, Psychosocial and Learning Problems Group (CDPLPG).
Aarti Samtani - none known.
Rob JPM Scholten and Lotty Hooft - work for Cochrane Netherlands (Dutch Cochrane Centre; DCC). The DCC carried out a systematic review in which Rob and Lotty participated for a Dutch guideline regarding the diagnosis of ASD (www.youthpolicy.nl). Some of the results thereof were used for this Cochrane Review. Rob and Lotty confirm that they were not involved in any primary studies included in the systematic review. The DCC regularly prepares commissioned systematic reviews for the Dutch Health Insurance Council, the Dutch Health Council, the Belgian Health Care Knowledge Centre, and various other parties. Rob and Lotty declare there is no relationship with the current work.
Nuala Livingstone - Editor with the Cochrane Developmental, Psychosocial and Learning Problems Group (CDPLPG) and Associate Editor with the Cochrane Editorial Unit.
Katy Sterling-Levis - author deceased; declarations of interest published in the protocol as "none known".
Susan Woolfenden - none known.
Katrina Williams - Editor with the Cochrane Developmental, Psychosocial and Learning Problems Group (CDPLPG).

SOURCES OF SUPPORT

Internal sources

- None, Other.

External sources

- None, Other.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

1. We modified our primary objective by collapsing the first three aims into one because we found only one interview test. Prior wording was as follows:
   a. "Which of the parent or carer interview tools (ADI-R, GARS, DISCO, or 3di) has the best diagnostic test accuracy?
   b. How does the diagnostic test accuracy of the best performing interview tool compare to the diagnostic test accuracy of the CARS?
   c. How does the diagnostic test accuracy of the ADOS-G compare to the CARS?
   d. Is the diagnostic test accuracy of any one test sufficient for it to be suitable as a sole assessment tool for preschool children?"
2. We removed the second secondary objective, "Does any diagnostic test have greater diagnostic test accuracy for the different diagnostic subgroups, that is, in differentiating Autistic Disorder/Childhood Autism from other ASD?", as this is no longer relevant with the publication of DSM-5 and the global trend to not distinguish autism diagnostic subgroups.
3. Web searching has been removed from the search methods section, as this was not performed for this Review.
4. We did not need to conduct analyses for different test cutoff points, as consistent cutoffs were used in all studies except one.
5. Investigations of heterogeneity could only be conducted for age of study participants for two tests, ADOS and CARS, as this was the only source of heterogeneity available that was sufficiently different between studies to be explored.
6. We could not pool study results using the bivariate normal method for tests with small numbers of studies. Instead, we performed separate meta-analyses for sensitivity and specificity (via logit transformations) according to the methods described by Takwoingi 2017.
7. Based on the suggestions of one of the peer reviewers, we also performed sensitivity analyses by including only studies at low risk of bias for the reference standard.