Psychometric properties of questionnaires and diagnostic measures for 
autism spectrum disorders in adults: a systematic review. 
Autism: International Journal of Research and Practice 2018. DOI: 
10.1177/1362361317748245

Copyright:
Copyright © 2018 The authors. Reprinted by permission of SAGE Publications.

DOI link to article:
https://doi.org/10.1177/1362361317748245

Date deposited:
15/02/2018
Psychometric properties of questionnaires and diagnostic measures for autism spectrum disorders in adults: a systematic review

Sarah Wigham, Jacqui Rodgers, Tom Berney,
Ann Le Couteur, Barry Ingham & Jeremy R Parr

Word count: 5959

Declaration of Interest
None.

Correspondence
Sarah Wigham, Institute of Neuroscience, Henry Wellcome Building, Faculty of Medical Sciences, Newcastle University, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK.
Email: sarah.wigham@ncl.ac.uk

Funding
We are grateful to the UK autism research charity Autistica (www.autistica.org.uk) for funding this study as part of the Autism Life Course and Ageing research programme.
Abstract
Accurately diagnosing autism spectrum disorders (ASD) in adulthood can be challenging. Structured questionnaires (SQs) and diagnostic measures (DMs) are frequently used to assist case recognition and diagnosis. This study reviewed research evidence on SQs and DMs published since the National Institute for Health and Care Excellence evidence update (NICE; 2014). The Cochrane library, Medline, Embase and PsycINFO were searched. Twenty studies met inclusion criteria. Sensitivity and specificity of SQs was best for individuals with previously confirmed ASD diagnoses, and reduced in participants referred for diagnostic assessments, with discrimination of ASD from mental health conditions especially limited. For adults with intellectual disability diagnostic accuracy increased when a combination of SQs were used. Evidence suggests some utility of DMs in identifying ASD amongst clinic referrals, though specificity for diagnosis was relatively low. In mental health settings the use of a single SQ is unlikely to accurately identify adults without ASD, or differentiate ASD from mental health conditions. This is important as adults seeking an ASD diagnostic assessment are likely to have co-existing mental health conditions. Robust ASD assessment tools specifically for use in adult diagnostic health services in the presence of co-occurring mental health and neurodevelopmental disorders is a research priority.
Abstract

Accurately diagnosing autism spectrum disorders (ASD) in adulthood can be challenging. Structured questionnaires (SQs) and diagnostic measures (DMs) are frequently used to assist case recognition and diagnosis. This study reviewed research evidence on SQs and DMs published since the National Institute for Health and Care Excellence evidence update (NICE; 2014). The Cochrane library, Medline, Embase and PsycINFO were searched. Twenty studies met inclusion criteria. Sensitivity and specificity of SQs was best for individuals with previously confirmed ASD diagnoses, and reduced in participants referred for diagnostic assessments, with discrimination of ASD from mental health conditions especially limited. For adults with intellectual disability diagnostic accuracy increased when a combination of SQs were used. Evidence suggests some utility of DMs in identifying ASD amongst clinic referrals, though specificity for diagnosis was relatively low. In mental health settings the use of a single SQ is unlikely to accurately identify adults without ASD, or differentiate ASD from mental health conditions. This is important as adults seeking an ASD diagnostic assessment are likely to have co-existing mental health conditions. Robust ASD assessment tools specifically for use in adult diagnostic health services in the presence of co-occurring mental health and neurodevelopmental disorders is a research priority.
Introduction

The importance of evaluating ASD diagnostic tools in adult populations

Research on the diagnosis of Autism Spectrum Disorder (ASD) has to date predominately focussed on childhood. Compared to the number of studies recruiting child populations, there are few studies (e.g. Lehnhardt et al., 2013; Brugha et al., 2015) specifically examining the psychometric properties of tools used in the diagnosis of suspected ASD in adulthood, for example the DISCO (Diagnostic Interview for Social and Communication Disorders; Wing et al., 2002). The tools used in the diagnosis of ASD are structured questionnaires (SQs) that are generally self-report or informant completed brief measures developed as ASD screening tools; and diagnostic measures (DMs), which are more in depth assessment tools that tend to involve semi-structured interviews and interactive tasks to inform an ASD diagnosis (National Institute for Health and Care Excellence (NICE) 2012; Scottish Intercollegiate Guidelines Network (SIGN) 2016). For adults with a suspected ASD diagnosis there are significant limitations in access to high quality diagnostic assessments. The reasons for this include both the availability of diagnostic services and psychometrically established measures (Lehnhardt et al., 2013; Powell & Acker, 2016; Brugha et al., 2015). SQs and DMs can assist clinicians making a diagnosis (NICE, 2012; SIGN, 2016) and can be used in conjunction with direct observation and other sources of information (ICD-10; APA, 2013). However childhood measures are not always suitable for adults where the presentation of ASD may be masked by maturation, learned compensatory skills and the presence of co-occurring mental health and neurodevelopmental disorders. Further, informants (particularly parents) and information about early history may be unavailable (Brugha et al., 2012; 2015).
The NICE (2014) guidelines on case recognition and diagnosis of ASD in adults

As part of the development of evidence-based clinical guidelines for the assessment and management of ASD in adults, the UK National Institute for Health and Care Excellence (NICE 2014) reviewed and published evidence on the use of SQs and DMs for case recognition and diagnosis in adults with suspected ASD. These guidelines recommended a battery of SQs and DMs that could contribute to an ASD diagnosis (see Table 1 for a summary).

In practice, clinicians and researchers use a range of SQs and DMs to aid diagnosis, not all of which were included in the NICE recommendations (Rogers et al., 2015; Rutherford et al., 2016). For example, when the NICE review (2014) was completed no diagnostic test accuracy studies were included regarding the use of the Diagnostic Interview for Social and Communication Disorders (DISCO) (Leekam et al., 2002; Wing et al., 2002) or the Developmental, Dimensional and Diagnostic Interview (3DI) (Skuse et al., 2004), in adult populations. In addition, some published SQs (usually in self-report format) have been developed to measure ASD traits that may not directly map onto diagnostic criteria, e.g. the AQ (Autism Spectrum Quotient; Baron-Cohen et al., 2001). Given the current availability of various different tools, the psychometric properties of the SQs and DMs used in clinical and research practice require further consideration (Rogers et al., 2015; Rutherford et al., 2016).

Aims of the study

The aims of this systematic review were therefore to:

(a) Identify studies published since the NICE (2014) guidelines update on evidence for the assessment of ASD in adults that describe the psychometric properties of the SQs and DMs available to clinicians and researchers for identifying ASD in adults.

(b) Examine the quality of the identified studies.
(c) Make recommendations about which of the measures, if any, might be most appropriate for the diagnostic assessment of ASD in adults, based on current evidence.

Table 1 around here

Method

NICE first published the adult autism clinical guideline (CG142) in 2012 (with an evidence update in 2014). This systematic literature review includes papers published between January 2013 and October 2017; and coincides with the publication of DSM-5 and the revised criteria for ASD that are used throughout this paper (APA, 2013). The review search terms and inclusion criteria are shown in Table 2.

Titles and abstracts were screened by SW with 20% screened independently by a second reviewer (TB). SW and TB reviewed full texts of selected articles independently; uncertainties about inclusion were resolved through discussion with the research team. The selection process to determine eligibility of articles for inclusion in the review is shown in Figure 1. Data were synthesised narratively and risk of bias was assessed using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) measure for examining the quality of diagnostic studies (Whiting et al., 2011; Campbell et al., 2015). The second reviewer independently rated 20%; reviewers 1 and 2 had 96% agreement on bias ratings.

Table 2 around here

Results

3887 articles were identified after removal of duplicates (Figure 1). 83 articles were read in full and twenty articles were selected for inclusion in the review. A brief summary of the characteristics of the measures from the selected articles is provided in Table 3.

Figure 1 around here
Bias ratings are shown online in the data supplement Table (DS1) and summarized with the current review’s findings for each measure. Most studies were conducted in Europe, two in Japan (Nishiyama et al., 2014; Takei et al., 2014) and four in the USA (Hus and Lord, 2014; Grodberg et al., 2014; Pugliese et al., 2015; Maddox et al., 2017). Two studies were multi-site with participants recruited from the USA, UK and Germany (Derks et al., 2017; Sappok et al., 2017). Six studies focussed on adults with intellectual disabilities (Sappok et al., 2014, 2015a, 2017; Mutsaerts et al., 2016; Derks et al 2017; Heinrich et al., 2017). Study designs were case-control, cross sectional or retrospective.

Table 3 around here

Structured questionnaires

SQs are usually self-report (but can be informant completed), do not require training for administration and have been used in epidemiological studies, and to gather information prior to a full diagnostic assessment (NICE 2014; Mutsaerts et al., 2016). Evidence for SQs available at the time of the NICE (2014) recommendations is shown in Table 1; new evidence identified in the current review from studies evaluating SQs since the NICE (2014) guideline is shown in Table 4.

Autism Spectrum Quotient (AQ) (Baron-Cohen et al., 2001)

The evidence for the psychometric properties of the AQ published since the NICE (2014) guidelines and covered in the current review is not consistent, varying according to study design and recruitment source. The sensitivity and specificity of the AQ-50 and the AQ-10 was good (≥80%) when comparing archival clinical data from adults with ASD, against a general population group (Booth et al., 2013). However this is not equivalent to the
population of subjects presenting to ASD diagnostic assessment settings. In such settings case recognition and the need to consider both differential diagnosis and the identification of additional mental health problems can be challenging.

In a Japanese study (Nishiyama et al., 2014) recruiting individuals with a known diagnosis of ASD, and a comparison general population sample, the internal consistency and construct validity of short versions of the AQ were inconsistent. A strength of the study was that a battery of screening measures were administered to all participants, including the general population comparator group, to identify any co-occurring mental health conditions.

A further study recruiting individuals with schizophrenia (Lugnegard et al., 2015) from a psychiatric clinic, adults with ASD and individuals from the general population, found that ASD and schizophrenia were both significantly associated with a higher AQ-50 score. In addition, only the AQ version with a binary response option discriminated reliably between ASD and schizophrenia and only on one subscale (attention switching) (Lugnegard et al., 2015). There was no significant difference in AQ-50 scores between individuals with ASD and those with schizophrenia when using the 4-point response option. Participants were diagnosed prior to the study, though diagnoses of ASD and schizophrenia were confirmed in this study using the DISCO-11 (Wing et al 2002) and the Structured Clinical Interview for DSM-IV diagnosis (SCID-I; First & Gibbon 2004) respectively. However, such purposive recruitment of individuals with a defined specific diagnosis is not comparable to the use of the AQ in ASD diagnostic assessment clinical settings.

In the fourth paper (Sizoo et al., 2015), individuals attending an ASD diagnostic clinic completed abbreviated versions of the AQ prior to a clinical assessment. The comparison group again consisted of individuals from the general population. Reported specificity (over 66%) and sensitivity (57-62%) was lower than in the Booth et al (2013) study. A strength of
the study was the recruitment of participants from ASD clinics and administration of the AQ prior to their full assessment, which is similar to procedures in clinical settings.

Finally, Ashwood et al., (2016) reported findings using the subject and informant versions of the AQ-50 and the AQ-10. Participants were consecutively referred to an ASD assessment clinic, and had high rates of comorbid mental health conditions. Across both AQ versions sensitivity was above 71% but specificity was less than 38%, and the relationship between the AQ scores and the ADI-R and the ADOS-G was weak. Thus the AQ did not reliably identify those individuals who did not have ASD. The presence of mental health conditions (such as a generalised anxiety disorder) also increased the risk of false positive ASD diagnoses using the AQ10 (Ashwood et al., 2016). This study highlights the limitations of the AQ in a more ecologically valid setting compared to studies using non-clinical general population comparison groups.

Thus overall, the findings from these studies suggest that due to low levels of specificity the AQ is not a reliable indicator of which people should progress to a full ASD assessment.

Ritvo Autism Asperger Diagnostic Scale-Revised (RAADS-R) (Ritvo et al., 2011)

When the RAADS-R was used in an ASD assessment service (Sizoo et al., 2015) and compared to a general population group, sensitivity was good (73%) but specificity was low (58%). This indicated some problems accurately identifying those without ASD (Sizoo et al., 2015). Limitations in this study are similar to those reporting on the AQ, including the use of a case-control design, and purposive recruitment of a general population comparison group that lacked mental health or developmental assessment. A study strength was that participants were recruited from ASD assessment services and administered the RAADS-R prior to full
ASD assessment. The low levels of specificity reported suggest the RAADS-R was not a reliable indicator of those who should progress to a full ASD assessment.

The psychometric properties of a 14-item version (the RAADS-14) were tested by Eriksson et al. (2013) in an ASD group, a clinical control group (individuals with ADHD, mood, psychotic and borderline personality disorders), and a general population group. The ASD group had significantly higher scores compared with the clinical control group. However, although both sensitivity (97%) and specificity (95%) was good with general population controls, the specificity for comparison with the clinical control group was reduced (to 46% with ADHD) (Eriksson et al., 2013). Comparison to a psychiatric control group was clinically relevant, but study limitations again included case-control design, and the lack of a consistent use of ASD and/or other mental health assessments in the control group; in addition some participants were recruited online and for these individuals diagnoses were self-reported and not confirmed.

Social Responsiveness Scale 2 (SRS-2) (Constantino & Gruber, 2005)

Full and shortened versions of the adult self-report SRS-2 were translated into Japanese (Nishiyama et al., 2014) and completed by individuals recruited from the general population and from clinical services. These versions had good internal consistency but were correlated with symptoms of psychoticism and distress as measured by the Brief Symptom Inventory (BSI) (Derogatis & Melisaratos 1983) and the K10 (Kessler et al., 2002). A strength of the study was that the general population group were administered a battery of mental health screening measures; a limitation was diagnosis of participants prior to the study.

A Japanese translation of the SRS informant version (Takei et al., 2014) demonstrated good sensitivity in both genders, and good specificity in men with ASD, compared to individuals in a control group with mental health difficulties. However, specificity in women
was reduced (61%) indicating limitations in identifying women who did not have ASD (Takei et al., 2014). A strength of this study was confirmation of diagnosis by the research team; a limitation was that only individuals in the ASD group completed a reference standard (the ADOS), the control group had no clinical assessment to exclude the presence of ASD.

Social Communication Questionnaire (SCQ) (Berument et al., 1999)

The SCQ (German version) was completed by parents or professional carers (in the majority of cases), of adults with intellectual disability and suspected ASD as part of routine clinical assessments of consecutive admissions to an intellectual disability mental health service (Sappok et al., 2015a). Interestingly, the SCQ current version (assesses the last 3 months) was found to be more effective than the lifetime version; the best combination of sensitivity (89%) and specificity (66%) was achieved by raising the SCQ cut-score for ASD from 15 to 18. When the focus was on lifetime development, specificity decreased to 48%. This finding may have been because the SCQ was completed by a professional carer who may not have known about a subject’s early development, rather than a family member. 87% of included participants had a moderate/severe intellectual disability, so the findings may not generalise to adults with milder intellectual disability (Sappok et al., 2015a). A strength of the study was that no exclusions were made on grounds of neurodevelopmental or mental health comorbidities, increasing the clinical validity of the sample.

Cross-cultural validity of the SCQ current version was examined in a study recruiting adults with intellectual disability from mental health services in Germany, the UK and the USA (Sappok et al., 2017). In contrast to the previous study (Sappok et al., 2015a) a lower cut-score of 13 was found optimal and made for sensitivity of 87% and specificity of 58%. More severe levels of intellectual disability were associated with higher scores. In some cases administration of the SCQ was prior to an ASD assessment so informants completing the
measure were blind to diagnosis; however variations in administration of the SCQ in different countries may have introduced bias. SCQ scores were influenced by gender and country with males and those recruited in Germany having higher scores.

Structured questionnaires developed for use with adults with intellectual disability

*Social Communication Questionnaire for Adults with Intellectual Disabilities (SCQ-AID)*

*(Derks et al., 2017)*

International data (from Germany, the USA and UK) were collected from individuals recruited via intellectual disability mental health settings with the SCQ-AID completed by carers or researchers. The new algorithm with a cut-point of 9 demonstrated sensitivity of 81% suggesting the shortened version has clinical utility though as found by Sappok et al., (2015a, who used the original version) specificity was low (62%). A strength of the study was the international recruitment of participants; however the authors noted some differences in diagnostic process across sites. Further independent evaluation of this new measure is needed.

*The Diagnostic Behavioural Assessment for ASD – Revised (DiBAS-R) (Sappok et al., 2015b)*

When the DiBAS-R was completed by a relative or staff member supporting an adult accessing an intellectual disability service, rates of 81% sensitivity and specificity were found when comparing those with and without ASD (Sappok et al., 2014). These findings using this new measure are encouraging, and a strength of the study is that diagnoses were made contemporaneously therefore replicating circumstances from clinical settings. Further evaluation in other populations of adults with intellectual disability is required.

In a study of consecutive referrals to a specialist intellectual disability psychiatric clinic in Germany, DiBAS-R sensitivity was 82% and specificity 67%. Diagnostic accuracy
was better in adults with mild/moderate intellectual disability than those with severe intellectual disability (for whom specificity dropped to 34%) (Heinrich et al., 2017). A strength of the study was that the DiBAS-R was completed by carers, prior to a full diagnostic assessment, and was therefore representative of the clinical process.

One study used a combination of the DiBAS-R and the Autism Checklist (ACL) and is described below (Mutsaerts et al., 2016).

**Autism Checklist (ACL) (Sipes and Matson, 2014)**

In a psychiatric clinic for adults with intellectual disability, Mutsaerts et al., (2016) reported that combined use of the ACL and the DiBAS-R led to improved sensitivity of 95% (when there was a positive screen on at least one measure) compared to when each measure was used alone (91%/75% respectively). Combined use also led to an increase in specificity to 88% (when there were positive screens on both measures) compared to using each measure separately (75%) (Mutsaerts et al., 2016). The inclusion of data collected from people referred to clinic was a study strength.

Table 4 around here

**Diagnostic measures and observational assessments**

Findings from studies of DMs and observational assessments, are shown in Table 5.

**Diagnostic measures: interviews**

**Autism Diagnostic Interview-Revised (ADI-R) (Lord et al 1994)**

The ADI-R was evaluated with adults without intellectual disability who were consecutively referred (mostly by self or a relative) to an ASD diagnosis clinic in Italy (Fusar-Poli et al., 2017). Overall the ADI-R had low sensitivity (43%) but good specificity (95%). In contrast
across the subscale domains sensitivity was acceptable (over 60%) and specificity was good (over 79%), with the exception of the restricted, repetitive and stereotyped behaviours domain (just 37%).

When the ADI-R and ADOS-G were combined sensitivity was still low (42%). The authors noted that the ADI-R may work less well with higher-functioning adults because a significant number of items rely on developmental history, and so more subtle presentations later in life may be missed (Fusar-Poli et al., 2017). It was a strength of the study that those administering the ADI-R were blind to DSM-5 based clinical consensus and that a third of participants already had a psychiatric diagnosis.

Diagnostic measures: observational

Autism Diagnostic Observation Schedule (second edition) (ADOS-2) Module 4 revised algorithm (Lord et al., 1999; Lord et al., 2012)

Hus & Lord (2014) reported sensitivity and specificity of the new revised Module 4 algorithm to be above 80% in a study including previous research participants and individuals presenting to a developmental disabilities clinic (the control group comprised individuals with a variety of DSM-IV-TR disorders). The severity scores were mostly independent of verbal IQ and race. A limitation of the study was the retrospective design, and exclusion of some individuals with a relative with ASD, so these new algorithms do require replication in other groups.

In a Dutch study of individuals with ASD and three different control groups de Bildt et al. (2016) reported a sensitivity of 61% for the ADOS Module 4. However, specificity varied according to the comparator group: 95-100% for individuals recruited from non-clinical or forensic settings, compared to 22-50% for individuals with schizophrenia (de Bildt et al., 2016). Items that were endorsed by the ASD group, but not by individuals with
schizophrenia included ‘quality of rapport’, ‘conversation’, ‘quality of social responses’ and
‘highly specific topics’. Limitations in the study design included the small sample size, and
the use of archival video data with individuals being diagnosed prior to the study.

In the third study Pugliese et al., (2015) investigated individuals referred for an ASD
assessment and previous research participants (many of whom met criteria for a range of
diagnoses) using the ADOS-2 Module 4, and a cut-point of 8. The study found that while
sensitivity was 85% (80-89% across research sites), specificity was somewhat lower at 72%
(62-91%). The results were best for women and those with above or below average verbal IQ.
Limitations in the study design included variability in the reference standards used to confirm
a diagnosis of ASD, and omission of any description of blinding during administration of the
tests.

Langmann et al., (2017) investigated the performance of the original and revised
ADOS-2 Module 4 algorithms using retrospective data. Participants were adults and
adolescents with suspected ASD who had been referred to specialist ASD diagnostic clinics
in psychiatry services. The revised algorithm demonstrated slightly better performance
overall (sensitivity 86% and specificity 80% at cut point 8) compared to the original
algorithm. However specificity for adults with a diagnosis of personality disorder compared
to ASD was poor (individuals with personality disorder comprised 58% of false positive
cases). In addition diagnostic accuracy was less for women and older individuals. However a
possible limitation to the generalisability of these findings was that less than a quarter of
those recruited were female and only 25% were over age 24 years.

Fusar-Poli et al., (2017) evaluated the ADOS-2 Module 4 with adults whose IQ was
over 70 and who were consecutively referred to an ASD diagnosis clinic in Italy. The
majority of participants were self or relative-referrals. The original ADOS-2 Module 4
algorithm (Lord et al 2012) had sensitivity of over 86% across all domains. Sensitivity of the
revised algorithm (Hus & Lord 2014) was 87%. Specificity was 83% for the original algorithm but reduced on individual domains (63%) and was 74% on the revised algorithm. Strengths of the study included the avoidance of a case-control design, and that those administering the ADOS-2 were blind to DSM-5 based clinical consensus. In addition, and as would be expected in a clinical setting, a proportion of those recruited had existing psychiatric diagnoses (37%).

Maddox et al., (2017) found the original and revised algorithms for the ADOS-2 Module 4 (Lord et al 2012; Hus & Lord 2014) had a sensitivity of 100% while specificity was 74% and 70% respectively. A strength of the study was recruitment of participants from psychiatric settings; however study findings indicated a high number of participants with psychosis falsely met criteria for ASD and that the social-communication domain was particularly limited in discriminating between the two.

Short observational assessments

Autism Mental State Examination (AMSE) (Grodberg et al., 2012)

The AMSE was reported to have good sensitivity and specificity for discriminating ASD from other disorders using the ADOS-G or the ADI-R (Grodberg et al., 2014). Strengths of the study design included blinded administration of the measure, and the consecutive recruitment of individuals who self-referred for assessment of ASD.

Discussion

Key Findings

The two key findings of the review are that, overall, there is very limited evidence to support the use of structured questionnaires (SQs: self-report or informant completed brief measures
developed to screen for ASD) in the assessment and diagnosis of ASD in adults. Further the evidence regarding the use of diagnostic measures (DMs: more in depth assessment tools involving semi-structured interviews and interactive tasks) suggests some utility in identifying ASD amongst clinic referrals, though specificity for eventual diagnosis of ASD was still relatively low. For both types of assessment tools, more evidence is required regarding their use in diagnostic assessments. We consider that tools are useful, but only as an aid to diagnostic decision making within a broader multidisciplinary team ASD assessment. These conclusions are in keeping with the NICE (2014) guidelines that suggest using a battery of tools to support the diagnostic process.

Structured Questionnaires. The new information reported compliments but does not change the overall NICE (2014) recommendations regarding SQs. Most SQs show limited utility in identifying those who did not have ASD (Ashwood et al., 2016). This finding was evident in studies recruiting participants in clinically realistic settings (e.g. Ashwood et al., 2016) compared to studies using a case-control design with a general population comparator (e.g. Booth et al., 2013). There were particular problems differentiating ASD from schizophrenia, which may be in part a consequence of the potential overlap in symptoms (e.g. flattened affect and aspects of social-communication) and/or underlying genetic phenotype (Cross-Disorder Group of the Psychiatric Genomics, 2013). Given the low specificity, the AQ and the RAADS-R should not in our opinion be used on their own for screening and case recognition in clinical settings. Based on published data this would apply to all versions of the AQ including the abbreviated 10 item version recommended by NICE (2014). Regarding the SRS, at the time of the NICE (2014) recommendations there were no diagnostic accuracy studies that met guideline inclusion criteria. New evidence suggests the accuracy of the SRS (Japanese translation) for adults may be limited in women, and in the presence of symptoms
of psychoticism or distress as measured using screening tools (Nishiyama et al., 2014; Takei et al., 2014). Evidence for the diagnostic accuracy of the SCQ with adults was not available previously. In the current review the SCQ was found to be more effective when using a higher cut-off score in a German cohort (Sappok et al., 2015a). However when an international cohort was recruited (Germany, USA and the UK) a lower cut-score improved diagnostic accuracy (Sappok et al., 2017). Where professional carers completed the current SCQ (without reference to developmental history) for adults with intellectual disability in a clinical setting, this showed better sensitivity and specificity than the lifetime version (Sappok et al., 2015a). This highlights the potential difficulty of gathering a developmental history to support the diagnosis of ASD in adults, which has implications for the choice of diagnostic tools used. A new and shortened version (the SCQ-AID) used in an international study had good sensitivity but low specificity, however the authors note some variation in diagnostic rates and processes across countries (Derks et al., 2017). This review found no new evidence regarding the ASDI or the AAA.

**Diagnostic Measures.** Studies assessing the psychometric properties of DMs (such as the ADOS-2 module 4) reported some success identifying ASD among clinic referrals though limitations remained. The ADOS-G and ADI were recommended by NICE for use with adults with and without intellectual disability. New evidence shows the new ADOS-G algorithms had good sensitivity and specificity in a study with a case-control design (Hus & Lord 2014). This was to a certain extent replicated (specificity was acceptable) in individuals consecutively referred to a psychiatric clinic with a variety of diagnoses (Fusar-Poli et al., 2017). However new studies also indicated that in clinical groups, specificity was reduced for schizophrenia, personality disorder and psychosis (de Bildt et al., 2016; Maddox et al., 2017; Langmann et al., 2017). Scores were additionally found to interact with gender and IQ.
although further studies are required to replicate these findings before definite conclusions can be made (Pugliese et al., 2015; Langmann et al., 2017). The ADI-R demonstrated mixed results in adult ASD clinic referrals, having good specificity but low sensitivity (Fusar-Poli et al., 2017). The AMSE showed promising evidence as a new measure; further studies in other settings and by teams independent of the authors are required (Grodberg et al., 2014).

Combinations of tools. NICE (2014) recommended using a battery of tools to support the diagnostic process. This was based on evidence from studies using SQs and DMs in isolation, and at the time the guidelines were published there were no diagnostic accuracy studies regarding use of tools in combination. In the current review only two studies evaluated the use of a combination of tools. In adults with ASD and intellectual disability, using the ACL and the DiBAS-R improved the sensitivity and specificity of these tools (Mutsaerts et al., 2016). For adults without an intellectual disability combining the ADI-R and the ADOS-G improved specificity but reduced sensitivity (Fusar-Poli et al., 2017). Sensitivity of the combined measures was lower than to using the ADOS-G alone; the authors suggest the ADI-R may be less sensitive to the more subtle presentation of ASDs diagnosed in adulthood (Fusar-Poli et al., 2017). This new evidence whereby combined use of these measures improved diagnostic effectiveness is in accordance with NICE (2014) recommendations regarding a battery of measures (Mutsaerts et al., 2016).

Strengths and limitations of the review

A strength of this review is that it includes new evidence about measures for which at the time of the NICE (2014) review publication there was no or limited evidence available (the SCQ, DiBAS-R, ACL, SRS, and the AMSE). A further strength was that only studies reporting sensitivity and specificity, or ROC analysis were included. A limitation was that the
number of studies and the sample sizes reported were relatively small, and heterogeneity in
the design of studies was considerable. Additionally case-control design was highlighted as a
potential source of bias for the existing evidence on tools when the NICE (2014) guidelines
were published; this design limitation was also present in a number of the studies included in
this review. However, the studies reported here that used more ecologically valid designs (for
example, consecutive clinic referrals) also reported poor psychometric properties of most
measures, particularly around specificity.

**Implications**

The diagnosis of ASD during adulthood is likely to pose challenges to clinicians and
researchers as people’s presentation may be more subtle (Brugha et al., 2015; Rogers et al.,
2015). The presence of other mental health conditions, which is highly likely in individuals
presenting to a specialist mental health service, challenges the accuracy of case recognition of
ASD, and complicates the ability of clinicians and researchers to identify the core symptoms
of ASD (Underwood et al., 2015). The risk of diagnostic overshadowing (i.e. misattributing
behaviours to an existing diagnosis leading to a failure in identifying other diagnoses) as a
consequence of the presence of associated psychiatric disorders may be increased (Cross-
Disorder Group of the Psychiatric Genomics Consortium, 2013; Ford & Crewther, 2014). For
many seeking a diagnostic assessment for suspected ASD in adulthood, the likely scenario
will be an assessment in the context of co-occurring mental health needs, additional
neurodevelopmental difficulties, social/relationship difficulties and/or problems with
employment or education (NICE, 2012; Russell., et al 2016). This review highlights the
limitations of the psychometric properties of existing diagnostic tools (SQs and DMs) to aid
the accuracy of diagnostic practice, in adults particularly in the presence of co-occurring
mental health diagnoses. The use of SQs that do not require a trained assessor may be useful
particularly in community primary and secondary health care settings (e.g. GPs and mental health services), to gather information about a possible neurodevelopmental disability (Booth et al., 2013; Allison et al., 2012). However, as already described, SQs alone should not be used to exclude further ASD assessment other than if the scores are extremely low (Brugha et al., 2012).

The studies included highlight that performance of SQs and DMs depend on the circumstances (countries, cultures and clinical setting) in which they are used and evaluated. For example, using ASD measures to identify individuals with ASD within the general population is a very different task to using the same measures in mental health and intellectual disability services, where ASD prevalence would be expected to be higher (Brugha et al., 2016). The interpretation of data derived from SQs and DMs including the values for sensitivity and specificity will be affected by both the prevalence of ASD in particular populations, and by the presence of coexisting conditions including intellectual disability and mental health conditions (Leeflang et al., 2009; Campbell et al, 2015). In addition, deciding how best to interpret the significance of particular SQ and DM scores is not without problems given the possibility that there may be cultural and/or country differences in scores (Sappok et al 2015a; 2017) in addition to the existing reported different clinical and research cut-off scores for many measures (Pugliese et al., 2015). Sappok et al., (2017) note that variations regarding the SCQ scores across countries may relate to differences in measure administration practices, ASD reference standards and referral processes. This has implications for international multi-site studies when researchers should agree cut-offs in advance (Sappok et al., 2017).

An inclusion criterion for the review was that study participants had completed a SQ or DM as part of a comprehensive ASD diagnostic assessment. However, in some studies, participants were diagnosed prior to the study and in others participants were consecutive
new referrals to a diagnostic clinic. Where participants were diagnosed prior to the study it was not always clear whether the ASD diagnosis was made in childhood or adulthood. This is important for several reasons: first, the presentation of adults diagnosed in childhood may be different to that of people referred for a first diagnosis in adulthood. Second, selecting adults with an existing diagnosis for studies of SQs and DM rather than adults referred for a suspected diagnosis may influence the sensitivity and specificity values (Leeflang et al., 2009; Campbell et al., 2015). Therefore the findings from studies in which individuals were first time referrals for an ASD assessment might arguably have more validity for real life clinical settings.

Future research

Further research is needed to investigate the accuracy, efficiency and comparative utility of all the tools identified in this review. Tools may require modification for use in adult ASD diagnosis services with individuals with or without intellectual disability whose neurodevelopmental conditions were not recognised in childhood, and for whom the risk of additional mental health comorbidities is high. Research should compare the sensitivity and specificity between tools when appropriately used in the diagnostic pathway with ecologically valid populations, such as individuals presenting to adult mental health settings. Only then can the clinical utility of individual SQs and DMs be directly compared. The strongest tools can then be developed further as needed – and others redeveloped, or discarded. New tools that have robust psychometric properties and that have been constructed specifically for use in adult diagnostic health services may be more useful than those adapted from childhood measures. In addition, clinicians and researchers should examine the utility of the combined use of measures to aid the assessment, differential diagnosis of ASD, and recognition of co-occurring neurodevelopmental and mental health conditions in the context
of an overall clinical formulation. An important next step will be investigating consensus views of adults with ASD and clinicians in different clinical contexts on the most effective, efficient and acceptable combinations of diagnostic measures. Whether SQs and DMs are sensitive to clinical change in response to interventions and/or progress over time is an area for future research (Brugha et al., 2015; Bolte & Diehl, 2013), as is the impact of gender and also culture on the psychometric properties of SQs and DMs – particularly as international collaborations are increasingly used to gather data from different settings (Pinto et al., 2014).

Regarding specific measures, no studies published since 2014 were found to have evaluated the ADOS modules 1-3 for adults with intellectual disability. No studies investigating the psychometric properties of the DISCO (Leekam et al., 2002; Wing et al., 2002) or peer-reviewed publications evaluating the 3DI (Skuse et al., 2004; Santosh et al., 2009) were identified that met the inclusion criteria for this review. Research on the clinical utility of these diagnostic measures that are used in a number of UK and European clinical settings is required.

This review did not include any existing interview measures of the more subtle presentation of the broader autism phenotype (BAP) (Parr et al., 2015). Given the possibility that presentations of ASD in adulthood may be more subtle, the potential value of interviews originally designed to identify the BAP for use as ASD diagnostic tools needs to be explored.

In conclusion, the findings of the current review compliment the NICE (2014) guidelines recommendations on the use of SQs and DMs in combination rather than in isolation. The current review suggests that SQs have limitations when used alone as screening measures and are unlikely to be reliable indicators of who should go on to have a full ASD diagnostic assessment. Future studies should examine the combined use of SQs alongside DMs for case recognition within the multidisciplinary diagnostic assessment in order to identify the most efficient and effective (clinical and cost) ways to ‘streamline’ diagnostic
practice and intervention planning in different settings for adults with suspected ASD. Research should also focus on which SQs and DMs both separately and in combination might be most appropriate for particular client groups – for example those with co-existing mental health conditions and those with intellectual disability.

References


Constantino JN and Gruber CP. (2005) *Social responsiveness scale (SRS)*: Western Psychological Services Los Angeles, CA.


Table 1 ASD screening and diagnostic measures for adults and evidence recommended by NICE (NICE 2012; 2014)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Structured questionnaires for case recognition</strong></td>
<td></td>
</tr>
<tr>
<td>NICE (2014) recommended the use of the following SQs (with adults without intellectual disability**) as part of a multidisciplinary diagnostic assessment, but not as diagnostic measures in isolation. The measures are easily available and training is not required</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conciseness and utility for case identification prior to a full diagnostic assessment in individuals when ASD was already suspected, were reported as strengths; case-control design (general population comparators) was a limitation*</td>
</tr>
<tr>
<td>Ritvo Autism Asperger Diagnostic Scale – Revised (RAADS-R) (Ritvo et al., 2011)</td>
<td>Sensitivity/ specificity (cut-off): 97%/ 100% (65).</td>
</tr>
<tr>
<td></td>
<td>Evidence of reliability (test-retest, internal) and criterion validity.</td>
</tr>
<tr>
<td></td>
<td>Case-control design (individuals with/without DSM-1V-TR conditions) was a limitation*</td>
</tr>
<tr>
<td>Adult Asperger Assessment (AAA) comprises the Autism-Spectrum Quotient (AQ) and Empathy Quotient (EQ) (Baron-Cohen 2005).</td>
<td>Sensitivity/ specificity (cut-off) 92% (10)</td>
</tr>
<tr>
<td></td>
<td>Concerns about administration blinding to the reference standard (ASD assessment clinic)</td>
</tr>
<tr>
<td>Asperger Syndrome Diagnostic Interview (ASDI) (Gillberg et al., 2001)</td>
<td>Sensitivity/specificity (cut-off): 100%/ 91% (5/6 algorithm criteria)</td>
</tr>
<tr>
<td></td>
<td>Some limitations in reliability and validity evidence; general population comparators</td>
</tr>
<tr>
<td><strong>Diagnostic measures and observational assessments</strong></td>
<td></td>
</tr>
<tr>
<td>NICE (2014) recommended the use of the following DMs for adults with &amp; without intellectual disability; though noted that the measures must be purchased and require training (Lord et al 1997; 2000; Rutter et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>Autism Diagnostic Interview – Revised (ADI-R) (Lord et al., 1997)</td>
<td>Sensitivity/specificity (cut-off): 91%/96% (communication 8; social reciprocity 10; restricted and repetitive behaviour 4)</td>
</tr>
</tbody>
</table>
| Autism Diagnostic Observation Schedule – Generic (ADOS-G) (superseded by ADOS-2) (Lord et al., 2000) | Sensitivity/specificity (cut off): 80%/87% (7); 70%/94% (10); 90%/93% (13)  
Some evidence of reliability (inter-rater, test-retest, internal) and criterion validity |

*In studies using a case-control design participants are recruited from different populations to form the ‘cases’ and ‘controls’, which does not replicate ‘real life’ clinical settings and may inflate reported diagnostic accuracy (Leeflang et al, 2009; Campbell et al, 2015).  
** No case identification SQs were identified for adults with intellectual disability by NICE (2014)
### Table 2 Search terms, study sources including databases, and study inclusion criteria

<table>
<thead>
<tr>
<th>Searches</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Search terms</td>
<td>(asd or asperg$ or autis$) and (screen$ or diagnos$) and (adult$ or adolescen$)</td>
</tr>
<tr>
<td>Electronic database scoping searches</td>
<td>Health and Social Care Information Centre (HSIC)</td>
</tr>
<tr>
<td></td>
<td>Health Management Information Centre (HMIC)</td>
</tr>
<tr>
<td></td>
<td>Cinahl</td>
</tr>
<tr>
<td>Electronic database formal searches</td>
<td>Cochrane Library</td>
</tr>
<tr>
<td></td>
<td>Medline</td>
</tr>
<tr>
<td></td>
<td>Embase</td>
</tr>
<tr>
<td></td>
<td>PsycINFO</td>
</tr>
<tr>
<td>Other study sources</td>
<td>Reference lists of included articles</td>
</tr>
<tr>
<td></td>
<td>Google search undertaken for the names of the measures listed in NICE (2014)</td>
</tr>
<tr>
<td></td>
<td>Lead researchers in the field contacted about recently completed studies/ forthcoming publications</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Articles published in English</td>
<td></td>
</tr>
<tr>
<td>Adult participants (where &gt;50% of participants were over 18 years old)</td>
<td></td>
</tr>
<tr>
<td>Participants had completed a structured questionnaire (SQ) or diagnostic measure (DM) as part of a comprehensive ASD diagnostic assessment</td>
<td></td>
</tr>
<tr>
<td>A comprehensive ASD diagnostic assessment was considered to have taken place if the assessment was carried out by trained clinicians and incorporated (wherever possible) a developmental history e.g. from a family</td>
<td></td>
</tr>
</tbody>
</table>
To maximise the clinical/research utility of this review, evidence of sensitivity (true positives), specificity (true negatives) or receiver operating characteristics (ROC) analyses using a diagnostic threshold was required (Campbell et al., 2015)

**Exclusion criteria**

Studies that only evaluated measures of one of the two core ASD symptom domains, for example, the Adult Repetitive Behaviours Questionnaire-2 (Barrett et al 2015), as a self-report measure of restricted and repetitive behaviours; and the Empathy Quotient (EQ: Baron-Cohen & Wheelwright 2004) as a measure of social cognition were excluded.
Table 3 Summary of characteristics of SQs and DMs evaluated in the studies included in this review

<table>
<thead>
<tr>
<th>Structured questionnaires</th>
<th>Diagnostic measures and observational assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Autism-Spectrum Quotient (AQ-10)</strong> (Allison et al., 2012)</td>
<td><strong>Autism Checklist (ACL)</strong> (Sipes and Matson, 2014)</td>
</tr>
<tr>
<td>Abridged from the 50-item version (Baron-Cohen et al., 2001). 10 items; 5 subscales: imagination, social skills, communication, attention to detail, and attention switching; cut-off = 6</td>
<td>The ACL is a 10 minute observational measure that assesses the three ICD-10 core ASD domains of social interaction, social communication, and stereotyped and restrictive behaviours. There are four items per domain scored using ordinal response options of present, partly present or not present. Training is not required; however, it is intended for administration by clinicians with ‘ASD expertise’.</td>
</tr>
<tr>
<td><strong>Ritvo Autism Asperger Diagnostic Scale – Revised (RAADS-R)</strong> (Ritvo et al., 2011)</td>
<td><strong>Diagnostic Behavioural Assessment for ASD – Revised (DiBAS-R)</strong> (Sappok et al., 2015b)</td>
</tr>
<tr>
<td>80-items; 4 subscales: social relatedness, circumscribed interests, sensory–motor and language</td>
<td>The DiBAS-R was developed based on DSM-5 and ICD-10 ASD criteria. It is a 20-item observational screening assessment for adults with intellectual disability that can be administered by carers without training and is scored on a 4-point Likert scale.</td>
</tr>
<tr>
<td><strong>Social Responsiveness Scale 2 (SRS-2)</strong> (Constantino &amp; Gruber, 2005)</td>
<td><strong>Social Communication Questionnaire for Adults with Intellectual Disabilities (SCQ-AID)</strong> (Derks et al., 2017)</td>
</tr>
<tr>
<td>Standardized 65-item four-response option measure of autism traits (with five subscales) that has both self-report and informant versions for adults.</td>
<td>The SCQ-AID was developed from the SCQ (current version) by reducing the number of items to 24 and creating a new algorithm.</td>
</tr>
<tr>
<td><strong>Social Communication Questionnaire (SCQ)</strong> (Berument et al., 1999)</td>
<td><strong>Social Communication Questionnaire for Adults with Intellectual Disabilities (SCQ-AID)</strong> (Derks et al., 2017)</td>
</tr>
<tr>
<td>The SCQ is a 40-item screening measure of current and lifetime symptoms of ASD developed from the ADI-R (Lord et al., 1994; Rutter et al., 2003). It is an informant questionnaire usually completed by parents/carers.</td>
<td><strong>Diagnostic Behavioural Assessment for ASD – Revised (DiBAS-R)</strong> (Sappok et al., 2015b)</td>
</tr>
<tr>
<td><strong>Social Communication Questionnaire for Adults with Intellectual Disabilities (SCQ-AID)</strong> (Derks et al., 2017)</td>
<td>The DiBAS-R was developed based on DSM-5 and ICD-10 ASD criteria. It is a 20-item observational screening assessment for adults with intellectual disability that can be administered by carers without training and is scored on a 4-point Likert scale.</td>
</tr>
</tbody>
</table>

5
**Autism Diagnostic Interview – Revised (ADI-R)**  
The ADI-R is a semi-structured interview conducted with parents or carers which focuses on current presentation and lifelong developmental history. The ADI-R focuses on social communication and interaction, plus restricted, repetitive and stereotyped behaviours (Lord et al. 1994).

**Autism Diagnostic Observation Schedule – Generic (ADOS-G) (superseded by ADOS-2)**  
(Lord et al., 1999; Lord et al., 2012)  
The ADOS-2 is a standardized semi-structured diagnostic assessment. It is conducted through one-to-one interaction and direct observation of an individual with suspected ASD, using a range of activities and is delivered by a trained examiner. Ratings include non-verbal behaviours, mannerisms, restricted and repetitive behaviours (including repetitive speech), and aspects of social interaction (including conversational style, gestures and eye contact, imagination and creativity). Four modules are available for use with adults across a range of ability from little or no expressive language (Module 1) through to verbally fluent (Module 4). A revised set of algorithms has been published which comprise domains for social affect and restricted and repetitive behaviours. For the revised Module 4 algorithm there is a single cut-off score to differentiate ASD from non-ASD classifications in line with DSM-5, together with a measure of ASD symptom severity (Hus & Lord, 2014).

**Short observational assessments**

**Autism Mental State Examination (AMSE)**  
(Grodberg et al., 2012)  
The AMSE is a brief observational tool for clinicians about the presence of ASD. There is an online training package for the AMSE, which comprises eight items (including pragmatics, preoccupations, interactions, repetitive behaviours and sensitivities).

*Recommended in the NICE (2014) guidelines on case recognition and diagnosis of ASD in adults*

**At the time of publication of the NICE (2014) guideline no evidence on the use of the measure with adults was available that met inclusion criteria**
Table 4 Findings from the studies evaluating Structured Questionnaires

<table>
<thead>
<tr>
<th>Study (country)</th>
<th>Structured Questionnaire</th>
<th>Participants (n): Recruitment source</th>
<th>ASD group Mage (SD) range</th>
<th>ASD group % Males</th>
<th>Findings</th>
<th>Version (cut-point)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC</th>
<th>Reliability and validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Booth et al (2013) (UK)</td>
<td>AQ</td>
<td>ASD (149): clinical &amp; university disability service archival data. General population (134): university &amp; social media.</td>
<td>15-75 NR (mostly males)</td>
<td>AQ 10 (6)</td>
<td>80%</td>
<td>87%</td>
<td>AQ 10 (6)</td>
<td>80%</td>
<td>87%</td>
<td>&gt;90%</td>
</tr>
<tr>
<td>Nishiyama et al (2014) (Japan)</td>
<td>AQ</td>
<td>ASD (60): psychiatry &amp; paediatric clinics. General population (3147): students &amp; workers.</td>
<td>25.5 (11); 16-65 67%</td>
<td>AQ50</td>
<td>AQ28</td>
<td>AQ10</td>
<td>AQ-J21</td>
<td>AQ-J10</td>
<td>AQ-m</td>
<td>All versions &gt;0.84; AQ-J10, AQ-J21, AQ20 &amp; AQ10 = 0.64-.75. Though item level discriminant validity only good for AQ28 &amp; AQ10. Low item total correlations on all versions of AQ.</td>
</tr>
<tr>
<td>Study</td>
<td>AQ</td>
<td>Sample Description</td>
<td>N (Age)</td>
<td>Percentage</td>
<td>AQ Score</td>
<td>Effect Size</td>
<td>Other Findings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>----</td>
<td>------------------------------------------------------------------------------------</td>
<td>---------</td>
<td>------------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lugnegard</td>
<td>AQ</td>
<td>AS (51): neuro-developmental clinic. Schizophrenia (36): psychiatry outpatients. General population (49): students</td>
<td>27 (4)</td>
<td>47%</td>
<td>64%</td>
<td></td>
<td>ASD &amp; SP: 89% ASD &amp; GP: 50% scores significantly higher in schizophrenia and AS than general population; &amp; higher in AS group (27) than schizophrenia (23) using binary scoring*; difference not significant on four point scale. ASD subscale scores significantly higher than GP but not significantly higher than schizophrenia except for ‘attention switching’**; No gender effects except for ‘imagination’ (women lower).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sizoo et al</td>
<td>AQ</td>
<td>ASD (210): ASD diagnostic clinic referrals. General population (63): via social media.</td>
<td>39 (12.5) 18-55</td>
<td>76%</td>
<td>65%</td>
<td></td>
<td>ASD group alpha = AQ 28: 0.90; AQ 10: 0.72. Significantly higher scores for ASD than controls***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ashwood et al (2016) (UK)</td>
<td>AQ</td>
<td>Suspected ASD diagnostic clinic referrals</td>
<td>32 (11)</td>
<td>75%</td>
<td>Self-report</td>
<td>AQ 50 (≥26)</td>
<td>88%</td>
<td>20%</td>
<td>AQ 50 (≥32)</td>
<td>71%</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----</td>
<td>----------------------------------------</td>
<td>--------</td>
<td>-----</td>
<td>-------------</td>
<td>-------------</td>
<td>-----</td>
<td>-----</td>
<td>-------------</td>
<td>-----</td>
</tr>
<tr>
<td>Sizoo et al (2015) (Netherlands)</td>
<td>RAADS-R</td>
<td>ASD diagnostic clinic referrals. General population (63): via social media.</td>
<td>39 (12.5)</td>
<td>76%</td>
<td>(98)</td>
<td>73%</td>
<td>58%</td>
<td>67%</td>
<td>ASD group alpha = 0.94; significantly higher scores for ASD than controls***</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Measure</td>
<td>Population</td>
<td>Score $\pm$</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Alpha</td>
<td>Additional Information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>------------</td>
<td>-------</td>
<td>------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eriksson et al (2013) (Sweden)</td>
<td>RAADS-14</td>
<td>ASD (77); ADHD and other psychiatric disorders (370); psychiatric outpatients or online mental health forums; general population (590); professionals via mental health lectures.</td>
<td>35 (11); 16-58</td>
<td>42% ($\geq 14$)</td>
<td>97%</td>
<td>95% (GP) 64% (PC) 46% (ADHD)</td>
<td>Full scale alpha = 0.9 Significantly higher total/item scores in ASD V controls*** Overlap with social anxiety (2 items). Females: significantly higher sensory reactivity scores across all groups.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nishiyama et al (2014) (Japan)</td>
<td>SRS 2</td>
<td>ASD (60); psychiatry &amp; paediatric clinics. General population (3147); students &amp; workers.</td>
<td>25.5 (11); 16-65</td>
<td>67%</td>
<td>SRS2</td>
<td>$&gt;87%$</td>
<td>SRS-2 (65, 30 &amp; 11 item versions): all versions alpha $&gt;0.80$. Correlations with psychological distress and psychoticism.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Questionnaire</td>
<td>Sample Description</td>
<td>N</td>
<td>Mean Age (SD)</td>
<td>ASD Only</td>
<td>ASD &amp; Psychiatric</td>
<td>ID Only</td>
<td>ID &amp; Psychiatric</td>
<td>Alpha</td>
<td>Factor Structure</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>---</td>
<td>---------------</td>
<td>-----------</td>
<td>-------------------</td>
<td>---------</td>
<td>------------------</td>
<td>-------</td>
<td>------------------</td>
</tr>
<tr>
<td>Takei et al (2014) (Japan)</td>
<td>SRS</td>
<td>ASD (65); psychiatric disorders (78): all psychiatry research volunteers and developmental clinic. General population (458): students &amp; workers</td>
<td>27(8)</td>
<td>68%</td>
<td>65 (men)</td>
<td>84%</td>
<td>81% (PC)</td>
<td>90%</td>
<td>Single or two factor structure. Significantly higher scores in ASD group than psychiatry group***</td>
<td>No correlation with IQ. Convergent validity with ADOS ($r = 0.34$).</td>
</tr>
<tr>
<td>Sappok et al (2015a) (Germany)</td>
<td>SCQ</td>
<td>ID (68); ASD &amp; ID (83: 13% mild; 87% moderate/severe ID): psychiatry clinic for ID.</td>
<td>35 (11)</td>
<td>75%</td>
<td>Current (15)</td>
<td>98%</td>
<td>47%</td>
<td>85%</td>
<td>Current and lifetime scores higher in ASD + ID than ID**; discriminated across ID severity (except lifetime: only in mild ID).</td>
<td></td>
</tr>
<tr>
<td>Sappok et al (2017) (Germany, UK &amp; USA)</td>
<td>SCQ</td>
<td>ASD &amp; ID (220); ID (231)</td>
<td>Across 29-41 (M)</td>
<td>&gt;70%</td>
<td>Current (13)</td>
<td>87%</td>
<td>58%</td>
<td>80%</td>
<td>Higher scores in Germany* Higher likelihood of diagnosis in UK (OR=3)</td>
<td></td>
</tr>
<tr>
<td>Derks et al (2017) (Germany, UK &amp; USA)</td>
<td>SCQ-AID</td>
<td>N=225 ASD &amp; severe ID: 41% Germany; Across sites 27-37 (5-12)</td>
<td>Across sites 70-80%</td>
<td>Reduced algorithm (9)</td>
<td>81%</td>
<td>62%</td>
<td>81%</td>
<td>SCQ-AID higher in ASD/ID group than ID group***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In/out patient mental health services. 93% diagnostic agreement on full and reduced algorithm.

### Diagnostic Behavioural Assessment for ASD – Revised (DiBAS-R)

<table>
<thead>
<tr>
<th>Study</th>
<th>DiBAS-R</th>
<th>ID (142); ASD &amp; ID (77; 18% mild; 82% moderate-severe); ID mental health clinic</th>
<th>35 (12)</th>
<th>69%</th>
<th>Total (29); SCI (21); SRS (5)</th>
<th>88%</th>
<th>72%</th>
<th>89%</th>
<th>Two factors. Alpha: SCI = 0.91; SRS = 0.84; total = 0.91. Significantly higher total and subscale scores ASD group*** Correlations*** with SCQ, PDD-MRS &amp; ACL.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sappok et al</td>
<td>DiBAS-R</td>
<td>ID (289); ASD &amp; ID (92; 42% mild/moderate; 58% severe); ID psychiatric clinic.</td>
<td>40(13)</td>
<td>68%</td>
<td>82%</td>
<td>67%</td>
<td>81%</td>
<td>Sensitivity/specificity &amp; ID level: mild/moderate (79%/84%); severe (83%/34%).</td>
<td></td>
</tr>
<tr>
<td>(2014) (Germany)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heinrich et al</td>
<td>DiBAS-R</td>
<td>Combined ACL and DiBAS-R 84 ASD/ID: 18% mild; 82% moderate/severe . 64 ID: ID psychiatric clinic.</td>
<td>38 (12)</td>
<td>67%</td>
<td>ACL DiBAS-R 91%</td>
<td>75%</td>
<td>75%</td>
<td>85%</td>
<td>Significantly higher scores on DiBAS-R and ACL in ASD than ID***</td>
</tr>
<tr>
<td>(2017) (Germany)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DiBAS-R 75%</td>
<td>75%</td>
<td>81%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutsaerts et al</td>
<td>Combined</td>
<td>84 ASD/ID: 18% mild; 82% moderate/severe . 64 ID: ID psychiatric clinic.</td>
<td>38 (12)</td>
<td>67%</td>
<td>ACL+ DiBAS-R 70%</td>
<td>88%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2016) (Germany)</td>
<td>ACL and</td>
<td>DiBAS-R</td>
<td></td>
<td></td>
<td>DiBAS-R 75%</td>
<td>75%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DiBAS-R</td>
<td></td>
<td></td>
<td></td>
<td>At least one measure 95%</td>
<td>63%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
NR: not reported; AUC: area under curve; AS: Asperger Syndrome; GAD: generalised anxiety disorder; GP: general population; PC: psychiatric controls; SP: schizophrenia. ID: intellectual disability; SCI: Social Communication and Interaction; ADHD: attention deficit hyperactivity disorder; SRS: Stereotypy, Rigidity, and Sensory Abnormalities; AQ-J: Japanese version; AQ-m: modified response option of 1-4; PDD-MRS: Pervasive Developmental Disorder in Mental Retardation Scale; *P<0.05; **P<0.01; ***P<0.001.
Table 5 Findings from studies evaluating diagnostic measures and observational assessments

<table>
<thead>
<tr>
<th>Study (country)</th>
<th>Measure</th>
<th>Participants (n)</th>
<th>ASD Mage: years (SD) range</th>
<th>ASD Male</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fusar-Poli et al (2017) (Italy)</td>
<td>ADI-R</td>
<td>ASD (78); other psychiatric diagnoses (35). ASD diagnosis clinic consecutive referrals.</td>
<td>26(9) 18-55</td>
<td>73%</td>
<td>ADI-R CL (8) 69% SI (10) 71% RRB (3) 86% ADI-R &amp; ADOS-G 42% 100%</td>
</tr>
<tr>
<td>Hus &amp; Lord (2014) (USA)</td>
<td>ADOS 2 Module 4</td>
<td>ASD (347); psychiatric (90): all research participants &amp; ASD clinic referrals</td>
<td>21 (7) 9-55</td>
<td>80% (8)</td>
<td>&gt;80% &gt;80%</td>
</tr>
<tr>
<td>Study</td>
<td>ADOS Module 4</td>
<td>Participants</td>
<td>Mean total and subscale scores significantly higher for ASD than psychopathy &amp; controls***; not compared to schizophrenia.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Bildt et al (2016)</td>
<td>ASD (38); schizophrenia (18); psychopathy (16); non-clinical controls (21); all previous research participants</td>
<td>32 (11) 18-66</td>
<td>100% (8) 61% SP 50% SP 22% PP 100% 66% ASD &amp; SP &amp; PP. 86% ASD &amp; PP 88% ASD &amp; GP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Netherlands)</td>
<td></td>
<td></td>
<td>Mean total and subscale scores significantly higher for ASD than psychopathy &amp; controls***; not compared to schizophrenia.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pugliese et al (2015) (USA)</td>
<td>ASD (253); non-ASD (68: 43 having psychiatric diagnosis); all research participants &amp; ASD clinic referrals</td>
<td>19 (7) 11-61</td>
<td>77% 85% 72%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Significantly higher scores in ASD on all except 7 items. Majority of SA &amp; RRB inter-item correlations significant. Sensitivity/ specificity slightly higher &gt;16 years, VIQ &lt;85 or &gt;115. Cross gender sensitivity &gt;80%; specificity lower for males (65 %) than females (86 %). CSS correlated with SRS*, negatively with SCQ*. No significant correlation with ADI or adaptive behaviour.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Langmann et al., (2017)</td>
<td>ASD (165); Non ASD/other diagnosis (191)</td>
<td>21(7) 13-54</td>
<td>82% Original 82% 83% Revised 57% 92% Revised 86% 80% Revised 75% 86%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Best discrimination: &lt;18 years; lower and average IQ. Sensitivity females cut-off 10: 35% (original); 59% (revised).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fusar-Poli et al (2017) (Italy)  | ADOS 2 Module 4  | ASD (78); other psychiatric diagnoses (35). ASD diagnosis clinic consecutive referrals. | 26(9) | 18-55 | 73% | Original  | 86% | 82% | 84% | Discriminant validity***  
Maddox et al (2017) (USA)  | ADOS 2 Module 4  | ASD (6); Psychosis (57); Mood disorder/ other (12) Outpatient psychiatry service. | 31(14) | 67% | Original Revised  | 100% | 74%  
|  |  | ADOS 2 & ADI-R  | 42% | 100% |  

**Observational tools to aid clinical judgement**

**Autism Mental State Examination (AMSE)**

Grodberg et al (2014) (USA)  | AMSE  | ASD (23); psychiatric diagnosis (27): all ASD assessment clinic consecutive self-referrals | 18-45 | NR | DSM-5 reference ADOS reference (≥ 5)  | 91% | 93% | 97% | Optimal cut score for those with DSM-5 or ADOS diagnosis=5. Floor effect on language and pointing items.  

RRB: Restricted and Repetitive Behaviour; CL: communication & language; SI: social interaction; Comm: communication; RSI: reciprocal social interaction; SA: social affect; SP: schizophrenia; PP: psychopathy; GP: general population; VIQ: verbal IQ; *P<0.05, ** P<0.01, *** P<0.001