



An evaluation of the diagnostic validity of the structured questionnaires of the adult Asperger's Assessment

Anna Jones¹ · Katie Rogers^{2,3} · Keith Sullivan¹ · Nina Viljoen²

Accepted: 20 March 2022 / Published online: 19 April 2022
© Crown 2022

Abstract

The Adult Autism Quotient (AQ), the Empathy Quotient (EQ) and the Relative's Questionnaire (RQ) were used as part of the Adult Asperger's Assessment (AAA) by a diagnostic service for adults without an intellectual disability with suspected autism spectrum disorder (ASD). This service is part of the National Health Service (NHS) in England. Little is known about the utility of these structured questionnaires despite wide use in clinical practice. It was investigated whether the questionnaires could discriminate between individuals with and without a diagnosis of ASD. Receiver Operating Curve analysis showed good levels of sensitivity to detect a positive diagnosis, but the specificity to exclude those without a diagnosis was poor. A binary logistic regression showed that a combination of the questionnaires also showed limited diagnostic validity. These findings have clinical implications in reviewing the efficiency of the assessment process.

Keywords Diagnostic validity · Structured questionnaires · Autistic adults · Sensitivity · Specificity

Autism spectrum disorder (ASD) is a neurodevelopmental condition that is defined by persistent difficulties in social communication and interactional patterns of restrictive, repetitive, and stereotyped behaviours and interests (American Psychiatric Association (APA), 2013). Until recently, the diagnosis of ASD has focused predominantly on childhood. Yet there are a significant number of individuals for whom ASD is not diagnosed until later into adulthood. In particular, adults without intellectual disabilities tend to be overlooked (National Audit Office, 2009).

It is important to note that prior to the publication of the Diagnostic and Statistical Manual Version 5 (DSM-V; APA, 2013), the term Asperger Syndrome (AS) or Asperger's

Disorder was used when an individual reported no significant delay in language or cognitive development in early childhood. This term has since been subsumed by ASD in diagnostic manuals - the DSM-V (APA, 2013) and the 11th revision of the International Classification of Diseases (ICD-11; World Health Organisation, 2019). The debate on terminology used in a diagnostic process is beyond the scope of this article, but it is the view of the authors that the terminology used should reflect the diagnostic manual that was used in reaching a diagnosis. For example, if a prior version of a diagnostic manual such as the ICD-10 is used, the diagnosis should reflect the diagnostic categories and the terminology of the ICD-10. However, for the purposes of this article the term ASD will be used, both when referring to autism in general, and when referring to specific assessment tools and questionnaires that were originally developed for the assessment of AS.

Following an update to the government's 'Think Autism' strategy (Department of Health & Social Care, 2014) to support adults on the autism spectrum in England to get a diagnosis and access support, the service in this study was commissioned in 2014 to meet a gap in services. The service aims to provide diagnosis-only for individuals who are referred specifically for suspected ASD, but with no comorbid intellectual disability, and are not currently being seen by secondary or acute mental health services who might

✉ Anna Jones
aj18abj@herts.ac.uk

¹ School of Life and Medical Sciences, University of Hertfordshire, College Lane, AL10 9AB Hatfield, Hertfordshire, UK

² Mid and West Essex Autism Spectrum Disorder Diagnostic Service, Hertfordshire Partnership University NHS Foundation Trust, Tekhnicon House, Springwood Drive, CM7 2YN Braintree, Essex, UK

³ Emotional Wellbeing & Mental Health Service, Essex Eating Disorder Service, NELFT, 125 New London Road, CM2 0QT Chelmsford, Essex, UK

otherwise carry out the assessment. The service is part of the NHS in England.

In England and Wales, the National Institute for Health and Care Excellence (NICE), an executive non-departmental public body that is sponsored by the Department of Health and Social Care, provides evidence-based recommendations for the diagnosis and care management of adults and young people with ASD. Using the NICE guidance is not mandatory, but it is used as a guide in both the NHS and private practice in the UK. Hence diagnostic services in the NHS and private practice will differ or share similarities depending on the extent to which they adhere to NICE guidance. The service in this study was set up to adhere to this guidance.

To provide a thorough assessment of ASD in adults, the National Institute for Health and Care Excellence (NICE, 2012) guidelines recommend a combination of structured questionnaires that are typically self-report measures; and diagnostic measures that include more in-depth assessment through semi-structured clinical interview and structured tasks (NICE 2012). The Adult Asperger's Assessment (AAA), along with other diagnostic measures, was utilised by the service being one of the measures recommended by the NICE (2012) guidelines. The AAA was purposely designed to assess adults without an intellectual disability and is based on the diagnostic criteria of the DSM-IV for AS. The AAA is also more stringent than the DSM-IV, as a greater number of symptoms are required to be met for a diagnosis (Baron-Cohen, Wheelwright, Robinson & Woodbury-Smith, 2005).

The AAA comprises four sections that are completed during a clinical interview. In addition, the AAA includes three structured questionnaires that are routinely completed before the interview. The referred individuals complete the Adult Autism Quotient (AQ⁵⁰) and the Empathy Quotient (EQ), with the designated advocate (as recommended by NICE, 2012), and someone who knows the individual well, preferably a parent, completes the Relative's Questionnaire (RQ; Baron-Cohen et al., 2005).

Despite a 787% increase in the number of people in England diagnosed with ASD between 1998 and 2018 and the development of several specialist autism teams, there seems to be a waiting list of up to 2 years in many adult diagnostic services (Beresford et al., 2020; Westminster Commission on Autism, 2021). In the context of this increased demand in the assessment of adults, there remains a scarcity of studies that investigate the psychometric properties of diagnostic tools used in the adult population (Wigham et al., 2019). The review of existing measures specifically developed for adults, such as the AAA, as opposed to measures that were adapted from children's measures, seem important for future adult diagnostic services' consideration of diagnostic tools.

Although the extent of the current use of the AAA in adult diagnostic services is unclear, and the term AS is no longer formally used as a diagnosis, the structured questionnaires developed alongside the use of the AAA may well continue to be used in the NHS as well as private practice as the AAA is recommended by NICE. The Westminster Commission on Autism (2021) makes reference to the low levels of satisfaction regarding the diagnostic process and identifies the high volume of paperwork that usually form part of the assessment as one of the challenges of the diagnostic process. Hence, the value that the lengthy EQ, AQ and RQ add to the diagnostic process is important to investigate. Furthermore, the EQ, AQ and RQ are freely available online, and lay people and researchers are likely to continue to use the questionnaires. Although Wigham et al. (2019) stress that structured questionnaires alone should not be used to exclude further autism assessment, the use of structured questionnaires that do not require a trained assessor can be useful in the community, and in primary and secondary healthcare settings to gather information about a possible neurodevelopmental disability. Therefore, further research has relevance for any future or current diagnostic service using the AAA, as well as for healthcare settings that wish to gather information, and for lay people and researchers who wish to use the AQ, EQ and RQ.

While a number of studies have considered the utility of the AQ (e.g. Ashwood et al., 2016), few studies have investigated the clinical usefulness of the combined use all three of the structured questionnaires. The original study that proposed the AAA as a diagnostic method (Baron-Cohen et al., 2005) used a clinical population, however the sample size was small ($N=42$), with a particularly small sample of those who did not receive a diagnosis ($N=8$). In addition, the sensitivity (i.e., ability to detect those with ASD) and specificity (i.e., ability to exclude those without ASD) of the structured questionnaires was not considered (i.e., discriminant validity).

Further studies have been carried out to investigate the discriminant validity of the AQ⁵⁰, however the findings are inconclusive (Ashwood et al., 2016; Booth et al., 2013; Woodbury-Smith, Robinson, Wheelwright and Baron-Cohen, 2005). In an early study by Woodbury-Smith et al. (2005), the AQ⁵⁰ was found to be able to discriminate between those who were diagnosed with ASD and those who were not with moderate accuracy in their diagnostic service for adults without an intellectual disability. The AQ⁵⁰ was moderately good in terms of sensitivity (0.77) and specificity (0.72) for the cut off of > 32 . In a later study comparing participants with a preexisting diagnosis of ASD with a general population control group, the sensitivity and specificity of the AQ⁵⁰ was good ($\geq 80\%$; Booth et al., 2013); however, the ecological validity is limited compared

to studies involving all participants presenting to diagnostic services for a first-time assessment, meaning that diagnostic accuracy may be inflated (Wigham et al., 2019). Ashwood et al. (2016) more recently reported findings in a more ecologically valid setting with individuals referred to an ASD assessment clinic that used the Autism Diagnostic Interview – Revised (ADI-R; Lord et al., 1997) and the Autism Diagnostic Observation Schedule-Generic (ADOS-G; Lord et al., 2000) to diagnose ASD in all participants. For both the diagnostic cut off (> 32) of the AQ⁵⁰ and the screening cut off (> 26), the sensitivity was above 71%, however specificity was less than 38%. In addition, the self-rated measure using the AQ⁵⁰ did not correlate with the clinician rated ASD behaviours (i.e. ADOS-G), and the relationship of the AQ⁵⁰ with the early-life symptoms (i.e. ADI-R) was weakly correlated.

The discriminant validity of the other structured questionnaires that form part of the AAA – the EQ and the RQ – has not previously been assessed in adults without an intellectual disability. This is important to consider given that the AAA continues to be used in clinical practice to aid diagnosis of ASD. The current study also provides a context in which to interpret the use of the questionnaires in previous studies and assessments. The service in this study is meeting a previously unmet need for individuals without an intellectual disability who reach adulthood with undiagnosed ASD. However, due to the specific nature of the service and therefore sample, it might be expected that the measures would yield low levels of specificity in this study's population, similar to the diagnostic service reported for the AQ⁵⁰ by Ashwood et al. (2016), due to most individuals being referred having at least ASD traits, although not all enough to meet diagnostic criteria. A further important consideration to highlight is that the questionnaires were not originally designed to fully capture ASD. For example, the EQ (Baron-Cohen & Wheelwright, 2004) specifically considers empathy, and does not cover the full ASD phenotype. Hence, it needs to be considered whether a combination of the questionnaires may better predict ASD diagnosis.

The purpose of this study was to examine the structured questionnaires of the AAA as part of a service evaluation process with the hope of reviewing their utility and how the measures are currently being used. Therefore, the key aim was to address the validity of the structured questionnaires used as part of the Adult Asperger's Assessment, through answering the following questions:

- To assess discriminant validity, can each of the three structured questionnaires, the Adult Autism Quotient, the Relative's Questionnaire and Empathy Quotient, individually discriminate between those with a diagnosis of ASD and those without?

- Do the clinical cut-off points on each of the structured questionnaires have good sensitivity (i.e., how good is the test at picking up people with ASD?) and specificity (i.e., how good is the test at excluding people without ASD)?
- To assess combined predictive power, does a combination or all three measures enable the ability to discriminate between those with and without an ASD diagnosis?

Methodology

Study Population

Our study population consisted of 259 individuals who underwent a full diagnostic assessment with the Mid and West Essex Adult ASD Diagnostic Service between September 2014 and March 2019. Of this sample, 38 had not completed the RQ measure, and were therefore excluded from the study. A further 6 individuals were recorded as receiving a “probable diagnosis”. As a significant number of assessed individuals were excluded, a comparison of diagnosis rate was made to determine whether the sample excluded was representative. There was no significant difference between number of individuals diagnosed with ASD from the full population assessed during this period and the sample included in this study. Individuals were adults referred to the service via their GP without concomitant intellectual disability. Referral criteria included that they must be 18 years or over, and not open to either a secondary or acute Mental Health Team or a Community Learning Disabilities Team.

The participants' demographic data (age, gender and ethnicity) is summarized in Table 1. The overall sample comprised 215 individuals (118 male) with a mean age of 31 years at the time of assessment. In terms of ethnicity, the majority were White British ($N=148$; 69%). Of those assessed, 179 (83%) received a positive diagnosis of ASD, and 36 were assessed as not meeting requirements for a diagnosis. The ASD participant group (96 male) had a mean age of 29 years 11 months, and the non-ASD participant group (22 male) had a mean age of 36 years 1 month. There was no significant difference in proportion of gender between groups ($\chi^2=0.68$, $df=1$, $p=.41$), however the mean age of the non-ASD group was significantly older than the ASD group ($t(213) = -3.08$, $p=.002$).

Ethical considerations

At the start of the ASD assessment process, services users were informed that anonymized data could be used for service evaluation: a leaflet provided to explain the use of data and consent was obtained. Ethical approval was not needed

for this service evaluation, due to the use of retrospective anonymized data, and a Privacy Impact Analysis (PIA) was carried out with Information Governance in line with the local trust policy. This PIA detailed how the data was to be stored, kept confidential and anonymised.

Measures

The measures focused on in this study were routinely carried out by the ASD Diagnostic Service. The *Adult Asperger's Assessment* (Baron-Cohen et al., 2005) comprises a semi-structured interview and three structured questionnaires: the Autism Quotient, the Empathy Quotient and the Relative's Questionnaire.

The *Autism Quotient* (AQ⁵⁰; Baron-Cohen, Wheelright, Skinner, Martin & Clubley, 2001) contains 50 statements assessing five key areas: social skills, attention switching, attention to detail, communication and imagination. Individuals score in the range 0–50, with a clinical cut-off of 32 or more. The questionnaire is a self-report measure that is completed by the referred individual. The respondent must answer 'definitely agree', 'slightly agree', 'slightly disagree' or 'definitely disagree' to each question. The scores are binarized into 'ASD responses' (1) and non-ASD responses (0), scoring one point if the respondent records the ASD related behaviour as either 'mildly' or 'strongly' (Woodbury-Smith et al., 2005). Moderate to high internal consistency for the subscales has been reported (Cronbach's alpha = .63–.77), and test-retest reliability was good ($r = .70$; Baron-Cohen et al., 2001).

The *Empathy Quotient* (EQ; Baron-Cohen & Wheelright, 2004) comprises 60 statements, 40 of which assess empathy and 20 are control items. It has a range of 0–80, with a cut-off score of 30 or less. The questionnaire is similarly completed by the assessed individual, and the statements are endorsed ranging from 'strongly agree' to 'strongly disagree'. Each of the empathy items are given a score of 1 point if the empathic behaviour is recorded as 'slightly' or 2 points if it is responded to with 'strongly'; half of the items are worded to be 'agree' responses and half are 'disagree' (Baron-Cohen & Wheelright, 2004). This measure has been shown to have good internal consistency (Cronbach's alpha ranging from 0.88 – 0.99; Allison, Baron-Cohen, Wheelright, Stone & Muncer, 2011; Wakabayashi et al., 2006).

The *Relatives' Questionnaire* (RQ; adapted from CAST; Scott, Baron-Cohen, Bolton & Brayne, 2002) contains a range of questions that cover social impairments, communication impairments and repetitive or stereotyped behaviour that a relative (preferably parent) completes retrospectively thinking of the referred individual as a child. A "yes" or "no" response is given to each question. A score of 0–31 is given, with a clinical cut-off of 15 or more. The CAST has

been shown to have good internal consistency (Cronbach's alpha ranging from 0.71 to 0.81 dependent on age and rater; Holmboe et al., 2014).

Assessment procedure

The questionnaires were completed by the assessed individual and their family with support from an independent advocate, and returned in advance of the face-to-face assessment. Independent advocates are assigned by the service, who offer pre- and post-diagnostic support. The questionnaires are then scored by entering the responses onto a Microsoft Excel document and then running a macro.

After receiving the questionnaires back, the person referred for the assessment is questioned using the semi-structured interview of the AAA, as well as a parent or family member who knows them well and who can provide additional data about their developmental history and day-to-day functioning. The responses on the questionnaires shape some of the questions asked during the clinical interview, but did not directly determine the diagnostic decision making. This interview is conducted by specialist professionals who have formal training and several years of experience in the assessment of ASD and is usually completed as part of a full day assessment. In addition, a sensory screening tool, selected structured tasks from Module 4 of the Autism Diagnostic Observation Schedule (e.g. story telling task) (ADOS; Lord, Rutter, DiLavore & Risi, 2008), and informal observations made while having lunch with the referred individuals, are used to inform the decision around a diagnosis of ASD. Diagnosis is made with reference to the ICD-10 diagnostic criteria, and in consultation with other clinicians in the diagnostic service. For the purpose of this study, the assessment outcome is reported as positive diagnosis (1), or no diagnosis (0).

Data analysis

All analyses were carried out by using SPSS version 24.0. To determine the discriminant validity of each of the structured questionnaires (AQ⁵⁰, EQ, and RQ), the mean scores of ASD and sub-threshold group (i.e., non-ASD) were first compared with independent *t*-tests. Secondly, the area under the curve (AUC) scores were examined from nonparametric receiver operating curve (ROC) analyses (see Figs. 1, 2 and 3). This plots true positives (i.e., the likelihood of an individual with an ASD diagnosis meeting the clinical cut-off point on the measure) against false positives (i.e., the likelihood of an individual without a diagnosis of ASD meeting the clinical cut-off point on the measure; Habibzadeh et al., 2016). Streiner & Cairney (2007) suggested the following benchmark for interpreting AUC scores: 0.50–0.70 (low

accuracy), 0.70–0.90 (moderate accuracy), and > 0.9 (high accuracy). The diagnostic accuracy of the AQ⁵⁰, the EQ and RQ at previously reported cut-offs was then reported. Chi-squared tests were used to report whether the various cut-off points significantly predicted ASD diagnosis better than chance, with meeting the threshold vs. receiving a clinical diagnosis both as binary variables (i.e., with ‘1’ representing a score above the cut-off point/positive ASD diagnosis, and ‘0’ being a score on the measure being below each cut-off point/ no ASD diagnosis). The positive predictive value (PPV) and negative predictive value (NPV) were also calculated for each cut-off score.

To determine whether using the three standard questionnaires in combination enables the ability to correctly predict ASD diagnosis (i.e. combined predictive power), a binary logistic regression was performed, using scores on the AQ⁵⁰, EQ and RQ as predictor variables (where ‘1’ equals a score above cut-off and ‘0’ equals a score below cut off). These variables were entered into logistic regression in a single step with age and gender also being entered as potential covariate predictors of an ASD diagnosis. Age was included as a covariate since the ASD and non-ASD groups were not matched on this criterion and the range was wide as is expected in an adult diagnostic service where services for adults born pre 1980 have only been recently available (Brugha et al., 2011). Gender was also included since prevalence rates are higher in males than females (Werling & Geschwind, 2013). The goodness of fit of the model was evaluated with chi-squared with a significant threshold of 0.05, and the sensitivity and specificity of the model was evaluated by reviewing how many participants were correctly assigned to the diagnosis, or no diagnosis group.

Results

Discriminant validity of AQ⁵⁰, EQ and RQ

Table 2 shows skew and kurtosis statistics and standard errors for each of the structured questionnaires. These scores and visual inspection of the distributions suggested that they did not depart significantly from normality in this sample. Table 2 also displays the mean and standard deviation scores for the ASD and non-ASD group on each of the structured questionnaires. Independent *t*-tests demonstrated that the ASD group had significantly higher mean scores on the AQ⁵⁰, and the RQ and a significantly lower mean score on the EQ than the non-ASD group. Cohen’s *d* effect sizes (all ≥ 0.61) suggested moderate practical significance of these findings. However, it is important to note that the AQ⁵⁰ and EQ mean scores for *both* the ASD and non-ASD groups reached the clinical cut-off.

Table 1 Demographic characteristics for whole sample, and ASD and non-ASD subgroups

	Whole group (<i>N</i> =215)	ASD sample (<i>N</i> =179)	Non-ASD sample (<i>N</i> =36)
Percentage of sample		83%	17%
Mean Age (years; months)	31;0	29;11	36;1
Age range (years; months)	18;3– 68.3	18;3– 61;6	18;4–68;3
Gender	118 male (55%) 97 female	96 male (54%) 83 female	22 male (61%) 14 female
Ethnicity			
White British	148 (69%)		
Other White background	3 (1%)		
Black, Asian and Minority Ethnic groups	5 (2%)		
Mixed White and Asian	3 (1%)		
Unknown	55 (26%)		

Table 2 Mean and standard deviation of AQ⁵⁰, EQ and RQ for ASD and non-ASD group; Skewness and Kurtosis for distribution of total dataset for each questionnaire

	Range scores	Mean Score (SD)		<i>t</i> -test (<i>df</i> =213)	Sig. level	Effect size Cohen’s <i>d</i>	Skew- ness (SE)	Kur- tosis (SE)
		ASD (<i>N</i> =179)	Non- ASD (<i>N</i> =36)					
Autism Quotient (AQ ⁵⁰)	0–50	36.84 (6.71)	32.03 (7.18)	3.88	<i>p</i> < .001	0.69*	–0.54 (0.17)	–0.08 (0.33)
Empathy Quotient (EQ)	0–80	19.21 (10.37)	26.11 (12.3)	–3.53	<i>p</i> = .001	0.61*	1.0 (0.17)	1.19 (0.33)
Relative’s Questionnaire (RQ)	0–31	18.95 (6.27)	13.86 (8.42)	4.17	<i>p</i> < .001	0.69*	–0.70 (0.17)	–0.13 (0.33)

NB. SD, standard deviation, and SE, standard error are in parenthesis; Df, degrees of freedom. Significant differences in *t*-tests are marked in bold; moderate effect sizes are marked with an *

The AUCs obtained from the ROC curves for each structured questionnaire (Fig. 1) present the average value of sensitivity for each possible value of specificity (Table 3).

For all three questionnaires, the AUC values reached significance, meaning the structured questionnaires predicted ASD diagnosis better than chance; however, the AQ⁵⁰ just

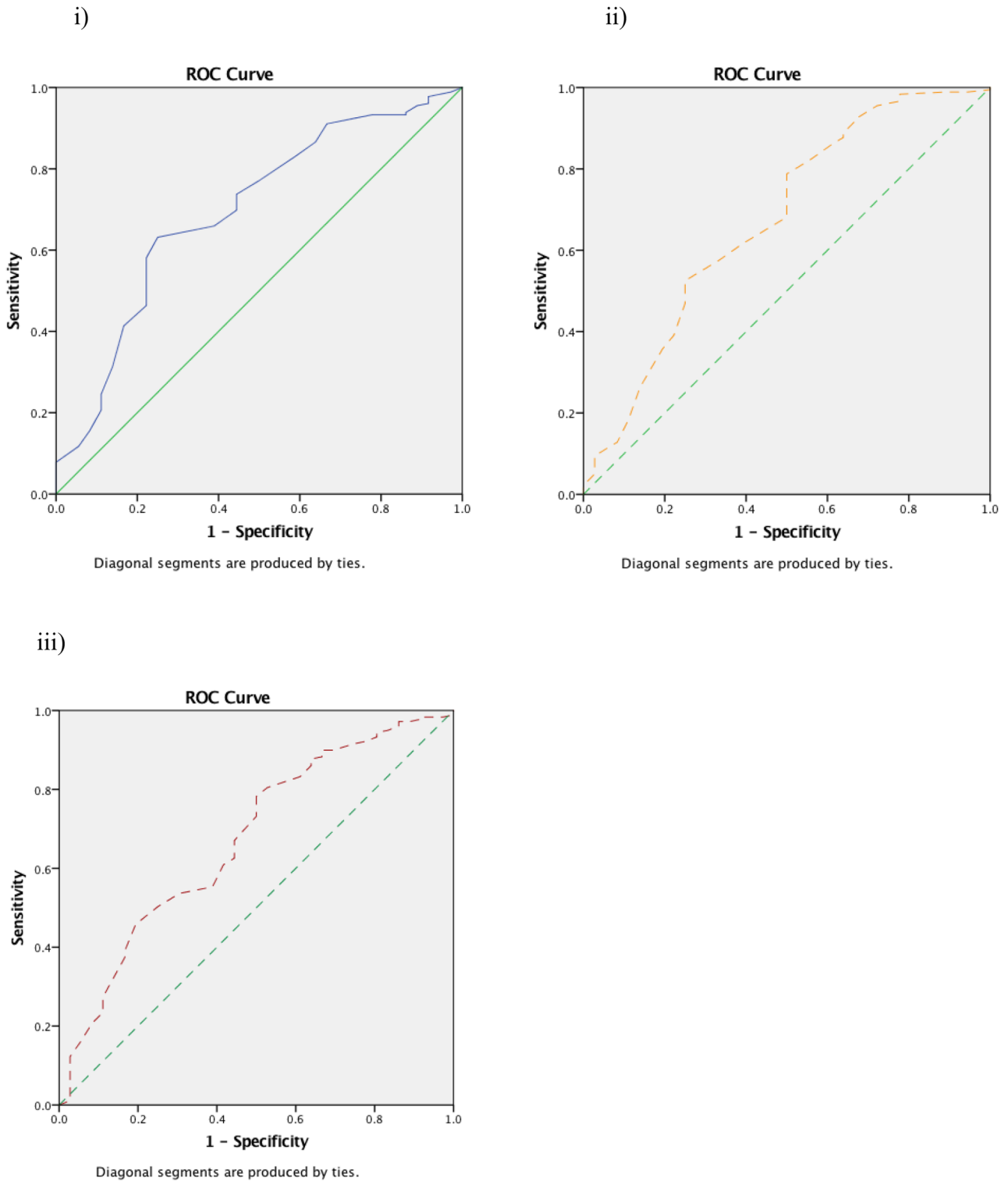


Fig. 1 Receiver Operator Characteristic Curves for the AQ⁵⁰(i) EQ (ii), and RQ (iii)

Table 3 Area Under the ROC Curve for structured questionnaires (AQ⁵⁰, EQ and RQ) predicting ASD diagnosis

	Area Under the ROC Curve (SE)	95% CI	Significance level, <i>p</i>
AQ ⁵⁰	0.70 (0.05)	0.60-0.79	< .001
EQ	0.67 (0.05)	0.57-0.78	= .001
RQ	0.67 (0.05)	0.57-0.78	= .001

NB. ROC, Receiver-Operating Characteristic; SE, Standard Error is in parenthesis; Significant differences in t-tests are marked in bold

met the threshold for a moderate level of accuracy (0.70), and the EQ and RQ were in the poor range in terms of level of accuracy.

Sensitivity, Specificity, and Positive and Negative Predictive Values for cut-off points

Table 4 summarizes the diagnostic accuracy of the AQ⁵⁰, the EQ and RQ at various cut-off points. The AQ⁵⁰, with the original diagnostic cut off score of ≥ 32 , predicted ASD diagnosis better than chance ($\chi^2(1, N=259)=11.05, p<.001$). It showed a fair level of sensitivity in predicting those receiving a diagnosis when scoring above cut-off, although its specificity was less accurate (Table 4). The positive predictive value was high, meaning that more than four-fifths of those scoring ≥ 32 on the AQ⁵⁰ did receive a diagnosis of ASD. However, the low negative predictive value suggests that more than two-thirds of individuals who scored below 32, who were predicted not to receive a diagnosis, did go on to get a diagnosis of ASD. The “screening” cut-off score of ≥ 26 (Baron-Cohen et al. 2001; Woodbury-Smith et al., 2005) was not able to predict ASD diagnosis better than chance ($\chi^2(1, N=259)=2.12, p=.14$). While it yielded a higher level of sensitivity, its specificity was very low. The pattern was similar in showing a high positive predictive value, and low negative predictive value.

The EQ, with the previously established clinical cut-off score of ≤ 30 , and the RQ, with a clinical cut-off score of ≥ 15 , both predicted ASD diagnosis better than chance (EQ: $\chi^2(1, N=259)=54.65, p<.001$; RQ: $\chi^2(1, N=259)=12.88,$

$p<.001$). The sensitivity in predicting diagnosis on the EQ was good and the RQ was fair, but as per the AQ⁵⁰, the specificity of the measures was poor. While the positive predictive values were high, the negative predictive values were poor.

Combined predictive power of the AQ⁵⁰, EQ and RQ

To test how well a combination of the structured questionnaires to determine how the measures worked in combination to identify an ASD diagnosis, these variables were entered as a single step into binary logistic regression with diagnosis (ASD or no ASD) as the dependent variable. The goodness of fit of the model was statistically significant ($\chi^2(5) 35.88=p<.001$), but the effect size was small, with the model being able to explain 25.8% of variance in the data (Nagelkerke $R^2=0.26$). The model correctly classified 85% of the cases; however, although there was a very high level of sensitivity (97%) the specificity was very low (25%). In terms of the covariables, increasing age (Wald=5.1, $p=.02$; Exp (β)=1.0, 95% CI=0.94, 1.06) was a significant predictor of a positive ASD diagnosis, but gender was not (Wald=1.77, $p=.18$; Exp (β)=0.57, 95% CI=0.24, 1.3). The AQ⁵⁰ (Wald=5.95, $p=.02$; Exp (β)=0.91, 95% CI=0.85, 0.98) and the RQ (Wald=7.32, $p=.01$; Exp (β)=4.03, (5% CI=0.87, 0.98) were both significant predictors, however the EQ did not significantly predict ASD diagnosis (Wald=0.55, $p=.46$; Exp (β)=1.2, 95% CI=1.0, 1.01).

Discussion

The findings of this study suggested that the AQ⁵⁰, RQ and EQ did not discriminate well between those with and without a diagnosis of ASD. While those diagnosed with ASD scored significantly higher on the AQ⁵⁰ and the RQ, and significantly lower on the EQ, the mean scores on the AQ⁵⁰ and EQ for both those with and without an ASD diagnosis were above previously determined clinical cut-off points. In addition, both groups’ mean scores were higher

Table 4 Diagnostic accuracy of the AQ⁵⁰, the EQ and RQ

Predictor	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy, %
AQ ⁵⁰	≥ 32	0.77 (0.71-0.83)	0.50 (0.34-0.66)	0.88 (0.83-0.93)	0.31 (0.18-0.42)	73%
AQ ⁵⁰	≥ 26	0.93 (0.89-0.97)	0.14 (0.03-0.35)	0.84 (0.79-0.89)	0.29 (0.08-0.51)	80%
EQ	≤ 30	0.88 (0.84-0.93)	0.33 (0.18-0.49)	0.87 (0.82-0.92)	0.36 (0.20-0.53)	79%
RQ	≥ 15	0.77 (0.70-0.83)	0.50 (0.33-0.66)	0.88 (0.83-0.93)	0.30 (0.18-0.42)	72%

NB. CI, Confidence Interval; PPV, Positive Predictive Value; NPV, Negative Predictive Value; 95% CIs are shown in parenthesis for sensitivity, specificity, PPV and NPV

compared to mean scores of typical populations (without ASD) in previous studies on the AQ⁵⁰, and lower on the EQ (e.g., in Wheelwright et al., 2006, mean AQ⁵⁰ = 16.3; mean EQ = 44.3). This study tested whether these predetermined cut-off points on the AQ⁵⁰ (≥ 32), RQ (≥ 15) and EQ (≤ 30) (Baron-Cohen et al., 2005) had good sensitivity in picking up those with ASD, and specificity in excluding those without ASD. While this has been investigated in previous studies for the AQ⁵⁰ (e.g. Ashwood et al., 2016), the sensitivity and specificity for the cut-off scores has not been considered for the EQ and RQ in adults without an intellectual disability. This is important to consider as these cut-off points are recommended for clinical use. The three structured questionnaires showed a similar pattern of results in demonstrating the ability to predict diagnosis better than chance, with fair to good levels of sensitivity, meaning those with a diagnosis were likely to score above clinical cut-off. However, specificity was poor, suggesting that the ability of the questionnaires to correctly identify people without ASD was reduced, thus the likelihood that those without an ASD diagnosis scored below the cut-off was poor. High positive predictive values, and low negative predictive values, meant that while the structured questionnaires were able to accurately predict diagnosis in those scoring above clinical cut-off (> four-fifths), around two-thirds of those scoring below clinical-cut off scores were ‘false negatives’, who did have a diagnosis of ASD. Furthermore, the area under the ROC curve tests found that the AQ⁵⁰ just met the threshold for a moderate level of accuracy, and the RQ and EQ’s levels of accuracy were in the low range, suggesting limited evidence of the diagnostic validity of the three structured questionnaires was found within this sample.

As each questionnaire describes a different aspect of the autism phenotype, this study also assessed how well a combination of all three measures could discriminate between those with and without an ASD diagnosis. The analysis found that a combination of the AQ and RQ were able to accurately identify those with a diagnosis of ASD. However, there were high number false positives (i.e., poor specificity), where 75% of those who did not receive a diagnosis also scored above clinical cut-off on the questionnaires. In addition, within this model, the EQ did not significantly discriminate between those with and without ASD. Therefore, these findings suggest that using the questionnaires collectively also showed limited diagnostic validity.

To our knowledge, this is the first study to consider the discriminant validity of the EQ, and the RQ in an adult population without an intellectual disability. The findings for the AQ⁵⁰ are similar to other studies conducted in similar clinical settings (Ashwood et al., 2016), but contrasted to findings of a larger study that found high levels of sensitivity and specificity, and a greater ability to discriminate between

ASD adults and healthy controls (Booth et al., 2013). One explanation for this variability is the nature of the non-ASD group. The AQ⁵⁰ is better at discriminating individuals with ASD when compared to a control group comprising of individuals from the general population. However, case-control studies have been shown to inflate the accuracy of diagnostic tests, known as ‘spectrum bias’ (Ashwood et al., 2016; Lijmer et al., 1999).

Indeed, psychometric research involving carefully screened comparison groups are not necessarily representative for a clinical setting (Bezemer, Blijd-Hoogewys, & Meek-Heekelaar, 2021). In such studies, the ASD group are already aware of their diagnosis before completing the measures (e.g. Booth et al., 2013). Testing whether the measures have predictive power within an ecologically valid ASD assessment service where they are completed before a clinical diagnosis is then made, is important to determine. Data from the present study and other clinical samples suggests that group differences (ASD vs. non-ASD) in the constructs measured in these questionnaires are weaker in clinical practice (Bezemer et al., 2021), despite being originally developed through carefully designed studies comparing to matched control groups.

Interestingly, a further recent study conducted with a clinical sample (Bezemer et al., 2021) showed higher levels of specificity in the AQ⁵⁰ (0.90 for cut off ≥ 32 and 0.72 for cut off of ≥ 26) than we found in the present study, yet still lower overall predictive value compared with studies comparing ASD groups to the general population. The key difference between the sample in the present study and Bezemer et al.’s (2021) sample was a significantly lower prevalence rate of ASD (63% vs. 83%) and a lower rate of missed patients. Therefore, even within clinical settings there is a wide variability in predictive power of the AQ⁵⁰, making it challenging to draw comparisons.

Within a clinical setting, why then do the AQ⁵⁰, EQ and RQ have limited ability to discriminate between those with and without a diagnosis of ASD? In accounting for the ‘false positives’ (i.e., poor specificity), it could be due to the nature of the sample, as some individuals without a positive diagnosis had a number of ‘traits’, but not a sufficient number to warrant a diagnosis. It may also be that for those in the non-ASD group but who have a number of ‘traits’, a greater level of public awareness of ASD over the last 5–10 years (Dillenburger et al., 2013) may have led to people being more vigilant in ‘noticing’ ASD related difficulties. This may lead to a ‘confirmation bias’ when completing the questionnaire measures, and potentially explain why both the ASD and the non-ASD group’s mean scores met the cut-off points, which were higher than in previous studies carried out before adult autism assessment services were more widely available (e.g. Woodbury et al., 2005). Even

if the diagnosis is yet to be made, all participants are aware that they are being referred for an ASD assessment when completing the measures. There is also some evidence that the AQ is sensitive to anxiety symptoms (Ashwood et al., 2016), whereas clinicians are able to distinguish between ASD and ‘ASD-like’ symptoms. While the participants in this study were not open to a community mental health team, it does not exclude the possibility that some had mental health difficulties.

The rate of ‘false negatives’ could be explained by some individuals with ASD potentially having poor insight into their condition, meaning that it is challenging for them to self-report their own behaviours and symptoms that could lead to an underreporting of difficulties (Bishop & Seltzer, 2012). This may partly explain the findings for the self-reporting AQ⁵⁰ and EQ, yet similar results were found for the parent-rated RQ. A difficulty with the RQ is that it requires a parent to consider childhood behaviours retrospectively, which may also impact the accuracy of reporting. In addition, as there is a high level of genetic heritability in ASD (Hallmayer, Cleveland & Torres, 2011), the parent may struggle with similar issues. This means that they may not experience their child’s difficulties as unusual. While the questionnaires can act as a prompt to ask clarifying questions when completing the AAA (Baron Cohen et al. 2005), the rate of false negatives suggests that these measures may be less useful as a screening measure to determine who would need an assessment. The use of these questionnaires as screening measures could result in a number of service users not being assessed who may have gone on to receive a diagnosis.

Although not an explicit aim, a strength of the study within its clinical population is the proportion of females (approximately half) who have participated where other studies typically include significantly more males. An additional interesting finding in terms of gender is that within the sample in this study, women were at least as likely to receive a diagnosis (85%) as males (81%). Staff members in the diagnostic service informally reported that they are mindful of the gender differences in ASD presentation when they conduct the AAA’s clinical interview, which appears to be reflected in the gender balance in diagnosis. This contrasts with research that has shown a gender bias with females scoring significantly lower on items of early childhood restricted and repetitive behaviours on the Autism Diagnostic Interview-Revised (Tillmann et al., 2018). A more fine-grained analysis of sex differences on items within the AAA’s semi-structured interview would need to be carried out to examine this finding in more depth.

Limitations

While a strength of this study is its ecologically valid population of individuals presenting to a naturalistic setting of an ASD diagnostic service, this comes with some caveats. The high rate of diagnosis (83%) means that it cannot be assumed that the same prevalence rate applies to other outpatient settings. For example, prevalence was lower at 73% in Ashwood et al. (2016), bringing into question the generalizability of the results. This also meant that the group sizes were not equal: the non-ASD group sample was small ($N=36$) with a larger variance compared to the ASD group. If the sample sizes were matched, the differences between groups would probably have been more pronounced. It was not possible to perform *a priori* power calculations, as retrospective data were used to evaluate the use of questionnaires within a particular service, as opposed to a prospectively designed study aiming to address a particular question.

A second limitation is the non-blind design, as the clinicians who made the decision in terms of ASD diagnosis were privy to the results of the structured questionnaires, meaning the results may have influenced the diagnostic decision (i.e. incorporating bias; Hajian-Tilaki, 2013). The measures are considered initially in the referral process, but on the day of the assessment the diagnosis is not made based on the score on the questionnaire, particularly as clinicians are aware of the limitations of the measures. However, the measures give qualitatively useful information as they contain certain characteristics pertaining to diagnostic criteria that will be further explored during the interview, which certainly limits the extent to which there is independence between the variables. This was unavoidable due to the questionnaires being part of routine clinical practice and is an inherent limitation of retrospective data collected within clinical services. Ideally, clinicians would be blind to the responses, or at least as to whether the scores meet the cut-off points. Although this would be challenging to maintain in clinical practice with limited numbers of staff working in the service. Importantly, however, each clinical decision was made on the basis of a detailed clinical interview, ICD-10 criteria and team discussion.

Clinical implications

The results of this study suggest limited discriminant validity of the structured questionnaires of the AAA which has relevance for clinical practice, as it highlights that they should be used with caution in the context of diagnosing or screening of ASD. The results also confirm previous findings that suggest that pre-assessment structured questionnaires are not currently a suitable substitute in replacement

of a clinician's judgment in a diagnostic interview (Ashwood et al., 2016).

The findings also bring into question whether the information gathered from the structured questionnaires generates enough useful information to warrant the time taken for individuals to complete the measures pre-assessment. Service users have commented that they are overwhelming and time consuming to complete. Clinicians working in the service mentioned that answers to specific questions sometimes lead them to probe further during assessment, and that individuals sometimes add written responses that provides useful information. However, the findings in this study raise doubt as to how the questions are interpreted by the person completing the questionnaire, hence the 'construct validity' of the answers, and therefore how useful the structured questionnaires can really be as part of an assessment battery. It might be that an alternative ASD trait questionnaire might prove to have better quality information and hence diagnostic validity. Potentially, a novel scale could be developed through exploring how clinicians discriminate ASD from 'ASD-like' characteristics so as to develop a pre-assessment measure with better diagnostic utility (Ashwood et al., 2016).

As aforementioned, the questionnaires were not originally designed to fully capture ASD (e.g., the EQ measures empathy). However, a combination of all three questionnaires did not improve the ability to discriminate between those with and without a diagnosis. The NICE guidelines (2012) propose that a combination of structured questionnaires and diagnostic tools should be used to aid diagnosis of ASD, but these findings raise potential limitations in the use of the AAA's structured questionnaires as part of a diagnostic battery in clinical practice and research. Further research is warranted to determine the most effective, efficient and acceptable combinations of questionnaires and diagnostic methods.

Conclusions

In conclusion, the present study found that the AQ⁵⁰, EQ and RQ questionnaires, both on their own and in combination, were limited in their ability to discriminate between those with and without a diagnosis in an ASD diagnostic service assessing adults without a learning disability. The structured questionnaires showed good levels of sensitivity in picking up those with ASD, but poor levels of specificity at excluding those without ASD. Therefore, the question of whether they are limited in their use as part of the diagnostic assessment process is raised. However, these findings should be interpreted with caution in light of the small sample size of the non-ASD group. These findings have

implications for the service in reviewing the efficiency of the assessment process.

Acknowledgements This service evaluation was carried out as part of the requirements of Anna Jones' doctorate in Clinical Psychology at the University of Hertfordshire (funded by Cambridge and Peterborough NHS Partnership Trust). No specific funding was given for this project.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author
- Allison, C., Baron-Cohen, S., Wheelwright, S. J., Stone, M. H., & Muncer, S. J. (2011). Psychometric analysis of the Empathy Quotient (EQ). *Personality and Individual Differences*, 51(7), 829–835
- Ashwood, K. L., Gillan, N., Horder, J., Hayward, H., Woodhouse, E., McEwen, F. S. ... Cadman, T. (2016). Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire. *Psychological Medicine*, 46(12), 2595–2604
- Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34(2), 163–175
- Baron-Cohen, S., Wheelwright, S., Robinson, J., & Woodbury-Smith, M. (2005). The Adult Asperger Assessment (AAA): a Diagnostic Method. *Journal of Autism and Developmental Disorders*, 35(6), 807–819
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5–17
- Beresford, B., Mukherjee, S., Mayhew, E., Heavy, E., Park, A., Allgar, S. L., V., & Knapp, M. (2020). Evaluating specialist autism teams' provision of care and support for autistic adults without learning disabilities: the SHAPE mixed-methods study. *Health Services and Delivery Research*, 8 (48).
- Bishop, S. L., & Seltzer, M. M. (2012). Self-reported autism symptoms in adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 42, 2354–2363
- Booth, T., Murray, A. L., McKenzie, K., Kuenssberg, R., O'Donnell, M., & Burnett, H. (2013). Brief report: An evaluation of the AQ-10 as a brief screening instrument for ASD in adults. *Journal of Autism and Developmental Disorders*, 43(12), 2997–3000
- Brugha, T. S., McManus, S., Bankart, J., Scott, F., Purdon, S., Smith, J. ... Meltzer, H. (2011). Epidemiology of autism spectrum disorders in adults in the community in England. *Archives of General Psychiatry*, 68, 459–465. <https://doi.org/10.1017/S0033291716001082>
- Bezemer, M. L., Blijd-Hoogewys, E. M. A., & Meek-Heekelaar, M. (2021). The Predictive Value of the AQ and the SRS-A in the Diagnosis of ASD in Adults in Clinical Practice. *Journal of Autism and Developmental Disorders*, 51(7), 2402–2415
- Department of Health & Social Care. (2014). *Think Autism. Fulfilling and Rewarding Lives, the Strategy for Adults with Autism in England: An Update*. London: DHSC
- Dillenburger, K., Jordan, J. A., McKerr, L., Devine, P., & Keenan, M. (2013). Awareness and knowledge of autism and autism interventions: A general population survey. *Research in autism spectrum disorders*, 7(12), 1558–1567

- Habibzadeh, F., Habibzadeh, P., & Yadollahie, M. (2016). On determining the most appropriate test cut-off value: the case of tests with continuous results. *Biochemia Medica*, 26(3), 297–307
- Hallmayer, J., Cleveland, S., & Torres, A. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry*, 68, 1095–1102
- Holmboe, K., Rijdsdijk, F. V., Hallett, V., Happé, F., Plomin, R., & Ronald, A. (2014). Strong genetic influences on the stability of autistic traits in childhood. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(2), 221–230
- Lijmer, J. G., Mol, B. W., Heisterkamp, S., Bossel, G. J., Prins, M. H., Van Der Meulen, J. H., & Bossuyt, P. M. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*, 282(11), 1061–1066
- Lord, C., Pickles, A., McLennan, J., Rutter, M., Bregman, J., Folstein, S., ... Minshew, N. (1997). Diagnosing autism: analyses of data from the Autism Diagnostic Interview. *Journal of Autism and Developmental Disorders*, 27(5), 501–517
- Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (2008). *Autism Diagnostic Observation Schedule: ADOS Manual*. Western Psychological Services
- National Audit Office (2009). Supporting people with autism through adulthood. Report by the Comptroller and Auditor General. HC 556 Session 2008–2009. London
- National Institute for health and Care Excellence. (2012). *Autism: Recognition, Referral, Diagnosis and Management of Adults on the Autism Spectrum (Clinical Guideline 142)*. Leicester: British Psychological Society
- Scott, F. J., Baron-Cohen, S., Bolton, P., & Brayne, C. (2002). The CAST (Childhood Asperger Syndrome Test): preliminary development of a UK screen for mainstream primary-school-age children. *Autism*, 6(1), 9–31
- Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry*, 52(2), 121–128
- Tillmann, J., Ashwood, K., Absoud, M., Bölte, S., Bonnet-Brilhault, F., Buitelaar, J. K. ... Bildt, D., A (2018). Evaluating sex and age differences in ADI-R and ADOS scores in a large European multi-site sample of individuals with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 48(7), 2490–2505
- Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Goldenfeld, N., Delaney, J., Fine, D. ... Weil, L. (2006). Development of short forms of the Empathy Quotient (EQ-Short) and the Systemizing Quotient (SQ-Short). *Personality and individual differences*, 41(5), 929–940
- Werling, D. M., & Geschwind, D. H. (2013). Sex differences in autism spectrum disorders. *Current opinion in neurology*, 26(2), 146
- Westminster Commission on Autism (2021). Support Surrounding Diagnosis. An inquiry into Pre- and Post-Support for the Autism Diagnosis Pathway. <https://barrysheerman.files.wordpress.com/2021/11/westminster-commission-on-autism-report-support-surrounding-diagnosis.pdf> (accessed 18 January 2022)
- Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R. ... Wakabayashi, A. (2006). Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain Research*, 1079, 47–56
- Wigham, S., Rodgers, J., Berney, T., Le Couteur, A., Ingham, B., & Parr, J. R. (2019). Psychometric properties of questionnaires and diagnostic measures for autism spectrum disorders in adults: A systematic review. *Autism*, 23(2), 287–305
- Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders*, 35(3), 331–335
- World Health Organisation (2018). International classification of diseases and related health problems (11th edition) <https://icd.who.int/en>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.