Predictive validity of self-report questionnaires in the assessment of autism spectrum disorders in adults

Sizoo, B.B.; Horwitz, E.H.; Teunisse, J.P.; Kan, C.C.; Vissers, C.T.W.M.; Forceville, E.J.M.; van Voorst, A.J.P.; Geurts, H.M.

Link to publication

Predictive validity of self-report questionnaires in the assessment of autism spectrum disorders in adults

Bram B Sizoo[1], EH Horwitz[2], JP Teunisse[3,4], CC Kan[5], CTWM Vissers[6,7,8], EJM Forceville[9], AJP Van Voorst[10] and HM Geurts[1,3,11]

Abstract

While various screening instruments for autism spectrum disorders are widely used in diagnostic assessments, their psychometric properties have not been simultaneously evaluated in the outpatient setting where these instruments are used most. In this study, we tested the Ritvo Autism Asperger Diagnostic Scale–Revised and two short versions of the Autism-Spectrum Quotient, the AQ-28 and AQ-10, in 210 patients referred for autism spectrum disorder assessment and in 63 controls. Of the 210 patients, 139 received an autism spectrum disorder diagnosis and 71 received another psychiatric diagnosis. The positive predictive values indicate that these tests correctly identified autism spectrum disorder patients in almost 80% of the referred cases. However, the negative predictive values suggest that only half of the referred patients without autism spectrum disorder were correctly identified. The sensitivity and specificity of each of these instruments were much lower than the values reported in the literature. In this study, the sensitivity of the Ritvo Autism Asperger Diagnostic Scale–Revised was the highest (73%), and the Autism-Spectrum Quotient short forms had the highest specificity (70% and 72%). Based on the similar area under the curve values, there is no clear preference for any of the three instruments. None of these instruments have sufficient validity to reliably predict a diagnosis of autism spectrum disorder in outpatient settings.

Keywords

autism, Autism-Spectrum Quotient, casefinding, RAADS-R-NL, self-report questionnaires

Introduction

Over the past years, we have seen an increase in referrals of adults for assessment of autism spectrum disorders (ASDs; Ritvo et al., 2010). Interviews with patients and informants as the basis for assessment can be supported by diagnostic tools. Bishop and Seltzer (2012) noted that self-report instruments may not be suitable for everyone with ASD and called for further research into the validity of self-report tools (Bishop and Seltzer, 2012). As a result of the increase in referrals, specialized services for adults with ASD are stretched to their limits. The pressure on these services would be alleviated if valid tools were available for making a reliable first evaluation on which to base the decision of whether or not to proceed with time-consuming and expensive assessment procedures. After all, without valid tools, every referral has to undergo the full assessment. Although several screening tools are available for ASD in adults, very few have been evaluated for use in outpatient settings where the instruments are needed most.

[1]Dimence, The Netherlands
[2]University Medical Center Groningen, The Netherlands
[3]Dr Leo Kannerhuis, The Netherlands
[4]Hogeschool van Arnhem en Nijmegen, The Netherlands
[5]Radboud University Medical Center, The Netherlands
[6]Kentalis Academy, The Netherlands
[7]Vincent van Gogh Institute for Psychiatry, The Netherlands
[8]Radboud University Nijmegen, The Netherlands
[9]GGZ Noord-Holland-Noord, The Netherlands
[10]GGz Centraal, The Netherlands
[11]University of Amsterdam, The Netherlands

Corresponding author:
Bram B Sizoo, Center for Developmental Disorders, Dimence, PO box 5003, 7400 GC Deventer, The Netherlands.
Email: b.sizoo@dimence.nl

Despite being intended as a screening tool, the Autism-Spectrum Quotient (AQ) is frequently used in the diagnosis and assessment of ASD in adults, and its use is mentioned in various research papers to include and exclude participants. Recent British and Dutch guidelines on the diagnosis and treatment of adults with ASD suggest using the AQ for case identification purposes and for diagnosis and assessment (Kan et al., 2013). This instrument was, however, not developed as a diagnostic tool, but to quantify ASD symptoms in the population because these symptoms were assumed to lie on a continuum. The AQ has five subscales: social skills, attention switching, attention to detail, communication, and imagination. With the original 50-item version, 80% of adults with Asperger's syndrome (AS) of high-functioning autism (HFA) scored above a critical minimum of 32 using a dichotomous scoring method, whereas only 2% of control adults did (Baron-Cohen et al., 2001). These results were replicated in Japan (Wakabayashi et al., 2006). For both studies, a sensitivity of 88% and a specificity of 97% could be calculated using a cut-off score of 32. Including pervasive developmental disorder not otherwise specified (PDD-NOS) patients, thereby focusing on a broader diagnostic group than the two former studies, both the sensitivity and specificity were somewhat lower (76% and 71%, respectively; Kurita et al., 2005). The AQ was also shown to discriminate between clinical samples of ASD and attention deficit hyperactivity disorder (ADHD) with or without comorbid substance use disorder (SUD), with 73% correct classifications using a cut-off score of 26 (Sizoo et al., 2009). Woodbury-Smith et al. (2005) were the first to investigate the use of AQ in a clinical setting among 100 patients referred for ASD assessment. In this study, where AS was used as the outcome measure, a cut-off of 26 yielded the best predictive value (83%), with a sensitivity of 95% and a specificity of 52%. Other studies showed no significant difference in AQ scores obtained during assessment, between a group of referred patients later diagnosed with ASD and a group later diagnosed with another disorder (Ketelaars et al., 2008). Hence, the only studies that reported good sensitivity and specificity figures for the AQ were conducted in a non-clinical setting, comparing patients with HFA or AS to non-psychiatric controls. The majority of studies, therefore, compared non-blinded patients with a narrow-spectrum ASD diagnosis to non-clinical individuals, while there is need for knowledge on how these instruments differentiate broad-spectrum ASD patients from non-ASD patients within a clinical setting.

In the meantime, short versions of the AQ have appeared, like the AQ-28 (Hoekstra et al., 2011) and the AQ-10 (Allison et al., 2012) intended for screening purposes. To our knowledge, these short forms of the AQ have not yet been examined in a naturalistic outpatient setting, which means that their practical use as screener for ASD in the clinical setting is not known.

Recently, a revised version of the 80-item self-report Ritvo Autism Asperger Diagnostic Scale (RAADS-R) was developed and shown to be "a valid and reliable instrument to assist the diagnosis of adults with Autism Spectrum Disorders (ASD)" (Ritvo et al., 2010). The potential advantage of the RAADS-R over other instruments such as the AQ could be that it includes items referring to hypo- and hypersensitivity, in line with the new diagnostic criteria of *Diagnostic and Statistical Manual of Mental Disorder* (5th ed.; DSM 5). In a Swedish study on the RAADS-R, comparing 75 adults with ASD to 197 comparison cases, a sensitivity of 91% and a specificity of 93% were found with an optimum cut-off score of 72. The RAADS-R was compared to the AQ in 39 ASD patients and 49 comparisons, but only with respect to correlations between the (sub-)scales of both instruments. The psychometric properties of the AQ were not reported (Andersen et al., 2011). Recently, an abbreviated 14-item version of the RAADS was evaluated in Sweden by Eriksson et al. (2013), by comparing patients with ASD to other psychiatric patients and a non-psychiatric control group. In the various phases of the study, a cut-off score >13 reached a sensitivity of 97% and a specificity of 46%–64% (Eriksson et al., 2013). Like the short versions of the AQ, the RAADS-R has not been assessed in a clinical setting.

The objective of this study was to explore the predictive validity of three self-report screening instruments for ASD in adults (RAADS-R, AQ-28, and AQ-10). This is relevant since there is the potential for real clinical utility should any of the instruments show good predictive validity, which would call for integration into the referral protocols for diagnostic clinics.

Methods

Participants

The study sample consisted of 285 adults aged between 18 and 55 years, recruited between April 2012 and December 2013 in six outpatient departments for ASD in the Netherlands. These outpatient departments are organized within universities (2) or institutes for mental health (4): University of Groningen Medical Center (UGMC), Radboud University of Nijmegen Medical Center (RADBUMC), Center for Developmental Disorders at Dimence Institute of Mental Health (COSDIM), Vincent van Gogh Institute (VVGI), GGZ Centraal (GGZCEN), and GGZ Noord-Holland-Noord (GGZNHN). The participating centers are part of a nationwide collaboration of clinicians involved in the treatment of adults with ASD. The protocol for the assessment of ASD in adults is similar in the participating centers and follows the national guideline for ASD in adults (Kan et al., 2013). The assessment is based on a thorough developmental history and diagnostic interviews by experienced clinicians. In each center, a

sequential cohort was constructed consisting of adults aged between 18 and 55 years, referred for ASD diagnosis by general practitioners or other psychiatric units. Exclusion criteria were the following: no proficiency in the Dutch language, illiteracy, major auditory and visual impairments, SUD, intellectual impairment, and current psychotic or manic episode. Eligible patients were informed about the study by their own clinicians and asked to give informed consent. This resulted in 218 participants referred for ASD assessment over the six outpatient departments. The 63 non-psychiatric control subjects were recruited from the general population through social media and matched to the clinical group for age and education level.

Instruments

The Dutch version of the RAADS-R (RAADS-R-NL) is based on a procedure of translating and back-translating with permission from the original authors. The RAADS-R is a revised version of the Ritvo Autism and Asperger Diagnostic Scale, which is designed as a self-report measure to support the diagnosis of ASD in adults of normal intelligence. The original version was revised in 2008 following a pilot study, resulting in the current 80-item RAADS-R. The RAADS-R has four subscales: 39 questions on social relatedness, 14 questions on circumscribed interests, 7 questions concerning language, and 20 questions on sensory–motor symptoms. For each question, one of the four answers can be chosen, ranging from mild to severe. In total, 64 questions describing specific symptoms of ASD are scored in order of severity as follows: "true now and when I was young" = 3, "Only now true" = 2, "True only when I was young" = 1, and "never true" = 0. The 16 questions describing non-symptomatic (normative) behaviors are scored as follows: "true now and when I was young" = 0 to "never true" = 3. Therefore, the range of scores is 0–240. Based on a psychometric study among 201 adults with ASD and 578 control subjects (with no diagnosis or with another-than-ASD axis-I DSM diagnosis), the conclusion was drawn that the psychometric properties were excellent (sensitivity = 97%, specificity = 100% at a cut-off > 65, test–retest reliability = 0.987, accuracy = 98.5%) (Ritvo et al., 2010).

The AQ-28 and the AQ-10 are derived from the 50-item AQ, originally designed by Baron-Cohen et al. (2001). The AQ is a self-report tool with 25 ASD-congruent and 25 ASD-incongruent questions. The respondent can choose from four answers: "1 = definitely agree," 2 = "slightly agree," 3 = "slightly disagree," and 4 = "definitely disagree." After dichotomization, this results in a scoring range of 0–50. For the ASD-congruent questions, the scores are reversed.

The 28-item version of the AQ was designed by reducing the number of items, while maintaining high validity, using the item retention procedure (Hoekstra et al., 2011).

Using exploratory and subsequently confirmatory factor analysis, two higher order factors were obtained with 28 items in total. The factor social behavior consists of 23 of the original AQ-50 questions, and the second factor numbers and patterns consisted of 5 of the original questions. The scoring range of the AQ-28 with the 4-point Likert scale is 28–112. A comparison between Dutch and English control groups, and an English AS group, indicated that a score > 65 was associated with a sensitivity of 97% and a specificity of 82% (Hoekstra et al., 2011).

The 10-item version of the AQ was constructed by selecting the two items in each AQ-50 subscale with the highest discrimination index (DI = the proportion of participants scoring on an ASD-congruent question in the control group minus the proportion of participants scoring on an ASD-congruent question in the ASD group) (Allison et al., 2012). The AQ-10 uses dichotomized scores where the two "agree" answers obtain 1 point and the "disagree" answers no point, leading to a scoring range of 0–10. The AQ-10 consists of the following questions from the AQ-50 (ASD-incongruent questions underlined): 5, 28, 32, 37, 27, 31, 20, 41, 36, 45. This 10-item version was examined using 173 adult patients with AS or HFA from three (partly anonymous) samples and 134 controls (Booth et al., 2013). The authors reported a sensitivity of 80% and a specificity of 87% using 6 as cut-off (dichotomous scoring method).

Procedure

Prior to the routine assessment in the outpatient departments, patients completed the abridged versions of the AQ and the RAADS-R-NL. The AQ-28 and the AQ-10 overlap with six questions. To avoid answering the same six questions twice, the duplicate questions were left out. The scores for both instruments were thus inferred from 32 (= 28 + 10 − 6) original AQ-questions, using the 4-point Likert scale for the AQ-28 and the dichotomized score for the AQ-10. The results of the AQ-28, the AQ-10, and the RAADS-R were not disclosed to the clinicians conducting the routine assessment and were in no way part of the assessment procedure. The three questionnaires were also administered to 63 control subjects with no known psychiatric disorder from the general population. The administration was done in the outpatient departments using a secure Internet link. The electronic data were kept on the server of the University of Amsterdam (Psychology department) in compliance with university regulations. Patients entered their participant codes in the data file but not their names. The name-code keys were kept by their own clinicians and were not available to the researchers. After completion of the questionnaires, patients and control subjects received a €10 voucher as a token of appreciation for participating. For the analysis, the clinical diagnosis was designated as the outcome measure and matched to the anonymized questionnaires in the database.

Statistics

The groups were compared with t-tests or chi-squared tests where appropriate. Analysis of the scores on the three questionnaires was done with analysis of variance (ANOVA) computations, with the score as the dependent variable and location (six outpatient departments) and diagnosis (ASD or non-ASD) as between-subject factors. For each questionnaire, Cronbach's alpha was calculated as a measure of internal consistency. The predictive value of each questionnaire was calculated using receiver operating characteristics (ROC) curves (Figure 1). The area under the curve (AUC) is a measure of predictive validity, with an AUC = 0.50 indicating a random prediction and an AUC > 0.90 indicating excellent validity. The AUC can also be interpreted as an average value of sensitivity for all the possible values of specificity and is therefore a useful measure for comparing diagnostic tests. The cut-off score was determined with a method that is referred to as the Youden index. This method uses the maximum of vertical distance of the ROC curve from the point (x, y) on the diagonal line (chance line). The Youden index maximizes the difference between true positive fraction (TPF; sensitivity) and false positive fraction (FPF; 1 − specificity). Therefore, the Youden index = TPF − FPF = Sensitivity + Specificity − 1. Thus, by maximizing the sum of the sensitivity and specificity across various cut-off points, the optimal cut-off point is calculated. The positive predictive value (PPV) and negative predictive value (NPV) were also calculated for this cut-off score. Computations were made using IBM© SPSS© version 20 software, using two-tailed analyses with alpha set at 0.05.

Results

The sample consisted of 281 participants: 218 clinical referrals and 63 control subjects. Data were missing in the electronic dataset for 8 referred patients, leaving 210 patients for assessment. All referred patients had completed the RAADS-R-NL, but 18 participants had not completed the AQ questionnaires. Hence, the analyses concerning the AQ were run with a total sample of 192 participants.

There was a main group effect for gender; the clinical group consisted of more males than the control group. Age, education level, living status, and number of children were similar between the groups (Table 1). There was no main effect for location of assessment and no interaction effect for location × group.

The scores on the three questionnaires were normally distributed, both in the combined group and in the clinical and control group separately. The internal consistencies in the clinical/control groups were good to excellent for the RAADS-R-NL (0.94/0.92), the AQ-28 (0.90/0.80), and the AQ-10 (0.72/0.45).

After assessment in the six outpatient departments, 139 were diagnosed with ASD and 71 with another diagnosis: ADHD (n = 9), anxiety disorder (n = 8), depression (n = 10), personality disorder (n = 11), and other diagnoses (n = 33).

The scores on all three instruments were higher in the clinical group upon referral (before assessment) compared to the control group, with large effect sizes (Pearson's r > 0.5; Table 2).

There was also a main effect for diagnosis on scores on all three instruments. Scores of the ASD patients were higher than in the non-ASD patients in all locations. There was no main effect for location of assessment and no interaction effect for location × diagnosis.

A post hoc t-test showed that the difference between the non-ASD group and the controls was also significant (RAADS-R-NL: $F_{(118.9)} = 8.58$, $p = 0.000$; AQ-28: $F_{(112.2)} = 10.06$, $p = 0.000$; AQ-10: $F_{(114.67)} = 8.60$, $p = 0.000$).

The AUCs obtained from the ROC curves indicate the average value of sensitivity for all the possible values of specificity for each test. The AUCs were lower than 90% (Table 4). The optimum cut-off points were defined as the score for which the sum of the associated sensitivity and specificity value is the largest, using the output of the three ROC analyses. This resulted in the following cut-offs: RAADS-R-NL = 98, AQ-28 = 80, and AQ-10 = 7 (Table 4).

Discussion

The aim was to investigate the psychometric properties of the RAADS-R-NL, the AQ-28, and the AQ-10 in predicting a correct diagnosis in adults referred for ASD assessment in outpatient departments. The mean prevalence of ASD among the referred patients in the six participating centers was 66% (Table 3). Based on this prevalence, the PPV and the NPV could be calculated for the cut-off points of each of the three instruments. The PPVs indicate that these tests correctly identified ASD patients in almost 80% of the referred cases. However, the NPVs suggest that only half of the referred patients without ASD were correctly identified. These inferences must be treated with much caution, however, because it cannot be readily assumed that this prevalence rate also applies to other outpatient settings. Indeed, even within our sample, the prevalence rate varied between 31% and 83%, although the differences were not statistically significant as main effect. An overall comparison of the three instruments using the AUCs shows little difference (Table 4). Based on these figures, there is no preference for any of the three instruments. For the clinical practices where these referred patients were seen, the psychometric figures indicated that it was not possible to make a clear decision early in the diagnostic process on who is eligible for confirmatory full assessment and who does not require further assessment. The sensitivity and specificity results in this study contrast
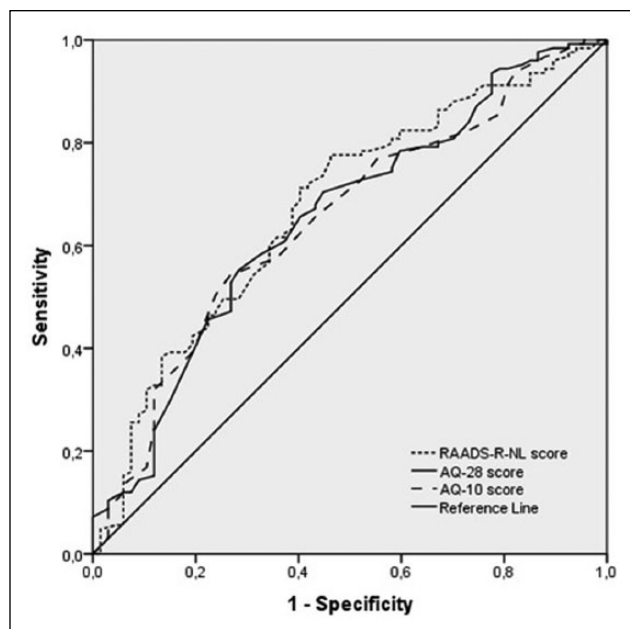
Figure 1. ROC curves for three self-report questionnaires.

with the high values for sensitivity and specificity as reported in studies mentioned in the introduction. These studies were carried out in the general population or by comparing adults without psychopathology to adults who were already aware of their ASD diagnosis. However, in line with these studies, post hoc ROC analysis comparing controls with those referred patients later diagnosed with ASD also resulted in excellent psychometric properties, represented by very high AUCs (RAADS-R-NL=0.93, AQ-28=0.95, and AQ-10=0.93). Unfortunately, these excellent psychometric properties have little clinical value. They are after all derived from a group comparison that does not represent a clinically significant situation.

Whereas Bishop and Seltzer (2012) advocated a focus on ASD samples instead of on non-ASD general population in order to obtain meaningful psychometric data for the AQ (Bishop and Seltzer, 2012), this study takes the argument a step further by proposing that for clinical practice, meaningful psychometric data can only be obtained with samples consisting of adults not yet diagnosed but referred for assessment of ASD. Although the AQ questionnaires and the RAADS-R are not designed to make an accurate prediction in the clinical situation, these instruments are nevertheless often used for that purpose in clinical practice. The questionable predictability obtained for the three instruments in the naturalistic clinical setting of this study does not mean, however, that instruments like the RAADS-R-NL, the AQ, or its short forms should not be used anymore in the diagnostic procedure. Ritvo et al. (2010) state that the RAADS is useful as an adjunct clinical diagnostic tool that must be completed in the presence of the clinician. The additional value of the procedure

suggested by Ritvo is that valuable spontaneous remarks made by patients provide valuable material for further diagnostic exploration upon completion of the questionnaires.

In our study, the scores for the ASD and non-ASD participants were comparable in all six diagnostic centers, as there was no main effect for location or an interaction effect of location × diagnosis. This indicates a certain degree of uniformity in diagnostic decision making, compensating for the lack of inter-rater reliability measures between the six centers. In addition, the comparable means for the ASD and non-ASD groups indicated that differences in the percentage of referred patients diagnosed with ASD can probably be attributed to factors like (a) familiarity among referring professionals with ASD in adults as a possible explanation for psychological distress, (b) regional availability of ASD assessment services, or (c) differences in perceived likelihood that women may have ASD. On average, adults referred for ASD diagnosis (i.e. before assessment) appear to have higher scores on the three questionnaires compared to controls, although in this study only 66% of the referred patients were later diagnosed with ASD. The fact that the scores in the non-ASD group differed significantly from the scores in the control group on all three instruments could indicate that referred adults with non-ASD psychiatric problems have scores higher in the ASD spectrum, but not as high as patients with ASD.

This is to our knowledge the first study to explore psychometric data of multiple instruments used in the assessment of broad-spectrum ASD in adults in a realistic and natural setting. The results are very relevant for clinical practice as they emphasize that the scores of these instruments must be treated with great caution in the context of classifying ASD. The study is limited by relying on the clinical diagnosis as the outcome variable in six diagnostic centers in the Netherlands. In addition, group sizes were not equal. The control group was smaller and showed a larger variance compared to the experimental group, which may have led to a too conservative interpretation of the results. Therefore, we expect that in the case of matched sample sizes, the differences between the control group and the clinical group would be even more pronounced than they are now. Another possible limitation was that the control group did not match the referred patients sample for gender because male controls appeared less willing to participate. This gender difference was accepted, however, because there was no significant difference in the scores on the three instruments between males and females in the clinical and control groups, which suggests that the imbalance in gender does not affect the overall results. We therefore expect that if the samples had been matched for gender, this would not have led to different results.

In summary, this study indicated that the predictive validity of the RAADS-R-NL, the AQ-28, and the AQ-10 is not high enough to accurately predict the outcome in adults

Table 1. Characteristics of 63 control subjects and 210 adult patient groups referred to six specialized outpatient departments for assessment of autism spectrum disorder (ASD).

	Control	Clinical[a]	GGZCEN	GGZNHN	RADBUMC	UGMC	VVGI	COSDIM	Statistics
N	63	210	26	46	51	16	22	49	(diagnosis)
Age	39.3 (13.80)	39.4 (12.50)	41.2 (10.68)	35.8 (11.74)	39.0 (14.00)	41.0 (12.55)	40.0 (14.49)	41.2 (11.31)	ns
Male	58.7%	75.7%	84.6%	71.7%	76.5%	93.8%	86.4%	63.3%	$\chi^2_1 = 6.90$, p = 0.009
Living alone	65.1%	62.4%	53.8%	69.6%	70.6%	56.2%	59.1%	55.1%	ns[b]
Education[c]	4.6 (0.79)	4.5 (0.94)	4.9 (0.73)	4.2 (0.86)	4.6 (0.98)	4.9 (0.62)	4.3 (0.95)	4.4 (1.07)	ns[b]
Children	1.2 (1.34)	1.1 (1.34)	1.3 (1.34)	.7 (1.07)	1.0 (1.23)	1.4 (2.22)	1.2 (1.30)	1.2 (1.32)	ns

[a]Clinical group was assessed in GGz Centraal (GGZCEN), GGz Noord Holland-Noord (GGZNHN), Radboud University Medical Center (RADBUMC), University of Groningen Medical Center (UGMC), Vincent van Gogh Institute of Mental Health (VVGI), Center for developmental disorders at Dimence Institute of Mental Health (COSDIM).
[b]Non-parametric chi-squared test.
[c]Education levels on a 5-point scale from "no education" = 1 to "university degree" = 5.

Table 2. Total scores on three self-report questionnaires for control subjects and referred adults before assessment of possible autism spectrum disorder.

	Control group			Clinical group			Statistic
	N	Mean (SD)	Median (range)	N	Mean (SD)	Median (range)	
RAADS-R-NL score	63	39.9 (27.30)	38 (6–189)	210	110.0 (44.55)	114 (12–206)	$F_{(1)} = 139.8$, p = 0.000, r = 0.58
AQ-28 score	63	49.2 (9.60)	46 (37–79)	192	76.8 (14.98)	78 (34–110)	$F_{(1)} = 187.6$, p = 0.000, r = 0.65
AQ-10 score	63	2.0 (1.62)	2 (0–6)	192	6.2 (2.59)	7 (0–10)	$F_{(1)} = 150.6$, p = 0.000, r = 0.61

SD: standard deviation; RAADS-R-NL: Dutch version of Ritvo Autism Asperger Diagnostic Scale–Revised; AQ: Autism-Spectrum Quotient.

Table 3. Mean scores for diagnosed ASD and non-ASD adults after assessment in six outpatient centers.

Total patients assessed		GGZCEN	GGZNHN	RADBUMC	UGMC	VVGI	COSDIM	Total	Diagnosis (test, sign., partial η^2)
ASD, N (%)[a]		N=26	N=46	N=51	N=16	N=22	N=49	N=210	
		16 (61)	38 (83)	32 (63)	5 (31)	11 (50)	37 (76)	139 (66)	$\chi_5^2 = 119.3$, p = 0.002
RAADS-R-NL (n=210)	ASD	135.2 (33.0)	119.0 (41.2)	121.3 (49.75)	130.8 (23.2)	109.6 (44.5)	102.6 (37.5)	118.3 (42.5)	$F_{(1)}=12.6$, p=0.000, $\eta^2=0.07$
	Non-ASD	94.1 (53.1)	102.5 (39.9)	80.8 (39.1)	94.9 (34.7)	106.1 (61.5)	85.1 (36.9)	92.4 (44.2)	
AQ-28 (n=192)	ASD	84.5 (11.9)	78.7 (13.1)	80.5 (18.0)	90.4 (10.6)	77.4 (11.9)	74.9 (11.2)	79.5 (14.1)	$F_{(1)}=13.9$, p=0.000, $\eta^2=0.07$
	Non-ASD	72.0 (13.6)	74.8 (10.3)	68.6 (16.8)	75.0 (14.3)	72.6 (18.8)	68.6 (16.5)	71.5 (15.3)	
AQ-10 (n=192)	ASD	7.6 (1.93)	6.6 (2.41)	6.6 (2.61)	8.2 (1.92)	6.6 (2.40)	6.2 (3.00)	6.7 (2.50)	$F_{(1)}=13.7$, p=0.000, $\eta^2=0.08$
	Non-ASD	5.7 (2.36)	5.9 (2.70)	5.1 (2.30)	5.9 (2.60)	4.6 (3.41)	5.0 (1.93)	5.3 (2.52)	

GGZCEN: GGz Centraal; GGZNHN: GGz Noord Holland-Noord; RADBUMC: Radboud University Medical Center; UGMC: University of Groningen Medical Centre; UGMC: University of Groningen Medical Centre; VVGI: Vincent van Gogh Institute of Mental Health; COSDIM: Center for developmental disorders at Dimence Institute of Mental Health; ASD: autism spectrum disorder; RAADS-R-NL: Dutch version of Ritvo Autism Asperger Diagnostic Scale–Revised; AQ: Autism-Spectrum Quotient.

[a]Number and percentage of ASD diagnoses per center after assessment.

Table 4. Correct classification ability of three self-report questionnaires for ASD in adults.

	Cut-off[a]	Sensitivity (%)[b]	Specificity (%)[b]	PPV (%)[c]	NPV (%)[c]	AUC at optimum cut-off point[d]
RAADS-R-NL (n=210)	85	81	45	74	45	
	98	**73**	**58**	**77**	**53**	**0.674**
	115	59	64	76	44	
AQ-28 (n=192)	70	77	41	72	47	
	80	**57**	**70**	**79**	**45**	**0.653**
	90	23	88	78	37	
AQ-10 (n=192)	5	79	36	71	47	
	7	**62**	**66**	**78**	**47**	**0.650**
	9	27	91	86	39	

AUC: area under the curve; RAADS-R-NL: Dutch version of Ritvo Autism Asperger Diagnostic Scale–Revised; AQ: Autism-Spectrum Quotient; PPV: positive predictive value; NPV: negative predictive value.

[a]Optimum cut-off (in bold) is defined as the score for which the sum of the sensitivity and the specificity are maximum.

[b]Sensitivity and specificity values and the cut-off points are derived from ROC calculations (ASD = 139, non-ASD = 71).

[c]PPV and NPV calculated for mean prevalence of ASD in the sample (66%).

[d]The AUC indicates the average value of sensitivity for all the possible values of specificity.

referred to specialized outpatient departments for ASD assessment. The PPV and NPV suggest that, with these instruments, one in five referrals in outpatient settings score above the cut-off and yet do not have ASD; conversely, almost half of the referrals with a score below cut-off do in fact have ASD. The decision whether or not to assign patients for further full assessment should, therefore, be based not only on these instrument scores but also on a careful clinical scrutiny of all additional information available at these early stages of referral, especially in the screen negatives.

Funding

References

Allison C, Auyeung B and Baron-Cohen S (2012) Toward brief "Red Flags" for autism screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist for Autism in toddlers in 1,000 cases and 3,000 controls [corrected]. *J Am Acad Child Adolesc Psychiatry* 51(2): 202–212.

Andersen LM, Naswall K, Manouilenko I, et al. (2011) The Swedish version of the Ritvo autism and Asperger diagnostic scale: revised (RAADS-R). A validation study of a rating scale for adults. *Journal of Autism Developmental Disorder* 41(12): 1635–1645.

Baron-Cohen S, Wheelwright S, Skinner R, et al. (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: 5–17.

Bishop SL and Seltzer MM (2012) Self-reported autism symptoms in adults with autism spectrum disorders. *Journal of Autism and Developmental Disorder* 42: 2354–2363.

Booth T, Murray A, McKenzie K, et al. (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: 2997–3000.

Eriksson JM, Andersen LM and Bejerot S (2013) RAADS-14 Screen: validity of a screening tool for autism spectrum disorder in an adult psychiatric population. *Molecular Autism* 4: 49.

Hoekstra RA, Vinkhuyzen AA, Wheelwright S, et al. (2011) The construction and validation of an abridged version of the autism-spectrum quotient (AQ-Short). *Journal of Autism and Developmental Disorder* 41: 589–596.

Kan CC, Geurts H, Bosch KVD, et al. (2013) *Multidisciplinaire richtlijn diagnostiek en behandeling van autismespectrumstoornissen bij volwassenen*. Utrecht: De Tijdstroom.

Ketelaars C, Horwitz E, Sytema S, et al. (2008) Brief report: adults with mild autism spectrum disorders (ASD): scores on the autism spectrum quotient (AQ) and comorbid psychopathology. *Journal of Autism and Developmental Disorder* 38: 176–180.

Kurita H, Koyama T and Osada H (2005) Autism-Spectrum Quotient-Japanese version and its short forms for screening normally intelligent persons with pervasive developmental disorders. *Psychiatry and Clinical Neurosciences* 59: 490–496.

Ritvo RA, Ritvo ER, Guthrie D, et al. (2010) The Ritvo Autism Asperger Diagnostic Scale–Revised (RAADS-R): a scale to assist the diagnosis of autism spectrum disorder in adults: an international validation study. *Journal of Autism and Developmental Disorder* 41(8): 1076–1089.

Sizoo BB, Van den Brink W, Gorissen-van Eenige M, et al. (2009) Using the autism-spectrum quotient to discriminate autism spectrum disorder from ADHD in adult patients with and without comorbid substance use disorder. *Journal of Autism and Developmental Disorder* 39(9): 1291–1297.

Wakabayashi A, Baron-Cohen S, Wheelwright S, et al. (2006) The autism-spectrum quotient (AQ) in Japan: a cross-cultural comparison. *Journal of Autism and Developmental Disorder* 36: 263–270.

Woodbury-Smith MR, Robinson J, Wheelwright S, et al. (2005) Screening adults for Asperger syndrome using the AQ: a preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorder* 35: 331–335.