ORIGINAL ARTICLE

Practical applications of brief screening questionnaires for autism spectrum disorder in a psychiatry outpatient setting

Yukiko Kuru[1] | Takeshi Nishiyama[2,3] | Satoshi Sumi[3] | Futoshi Suzuki[3,4] | Tomoko Shiino[4] | Takuma Kimura[5] | Kou Hirai[3] | Miho Kuroda[6] | Yoko Kamio[7,8] | Shogo Kikuchi[1]

[1]Department of Public Health, Aichi Medical University School of Medicine, Nagakute, Japan

[2]Department of Public Health, Nagoya City University Graduate School of Medicine, Nagoya, Japan

[3]Department of Child and Adolescent Psychiatry, Kamibayashi Memorial Hospital, Ichinomiya, Japan

[4]Research Center for Child Mental Development, University of Fukui, Fukui, Japan

[5]Department of Early Childhood Education and Care, Nagoya Management Junior College, Owariasahi, Japan

[6]Department of Human Care, Nagoya University of Arts and Sciences, Nagoya, Japan

[7]Institute of Education and Human Development, Ochanomizu University, Tokyo, Japan

[8]National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan

Correspondence
Takeshi Nishiyama, Department of Public Health, Nagoya City University Graduate School of Medicine, 1 Kawasumi, Mizuho-ku, Nagoya 467-8602 Japan.
Email: p-gen@umin.ac.jp

Abstract

Objectives: This study was designed to examine the diagnostic performance of the social and communication disorders checklist (SCDC) and strength and difficulties questionnaire (SDQ) to detect autism spectrum conditions (ASC), along with the social responsiveness scale-second edition (SRS-2) as reference, in a psychiatry outpatient setting.

Methods: We translated the SCDC into Japanese since its Japanese version was unavailable. We examined its test-retest reliability as well as the internal consistency reliability and diagnostic performance of the three questionnaires among 41 Japanese psychiatric outpatients, using the best-estimate diagnosis of ASC based on the diagnostic interview for social and communication disorders, as a gold standard.

Results: The test-retest reliability was high for the SCDC. Although the internal consistency reliability was high for the SCDC and SRS-2, that was low for the prosocial and peer problem subscales of the SDQ. The performance of the SCDC, SDQ, and SRS-2 to detect ASC was moderate: the area under the ROC curve of 0.78, 0.78, and 0.84, respectively.

Conclusions: Although questionnaires to detect ASC, including the three examined, generally have only moderate performance in this setting, these can be successfully applied to high-risk populations such as psychiatry outpatients, when multi-level rather than dichotomous likelihood ratios are used.

KEYWORDS
autism spectrum conditions, multi-level likelihood ratio, questionnaire, reliability, validity

1 | INTRODUCTION

Routine screening for autism spectrum condition (ASC), also known as autism spectrum disorder (ASD), in the general population is recommended at younger age, for example, at 18 and 24 months (Guthrie et al., 2019; Johnson et al., 2007), although it is still controversial (Crowe & Salt, 2015; Siu et al., 2016). However, diagnosis of ASC is often made at a later age (Baio et al., 2018; Shattuck et al., 2009; Williams, Thomas, Sidebotham, & Emond, 2008), and besides, ASC often remains undetected and thus untreated (Saemundsen, Magnússon, Georgsdóttir, Egilsson, & Rafnsson, 2013). This highlights the need to screen for older children and adolescents with ASC in a psychiatric outpatient setting, where they are referred for diagnostic assessments through less specialized instances such as primary care physician's recommendations and parental (or their own) concerns. For screening in such a clinical practice setting, parent-or teacher-administered questionnaires are typically chosen that are inexpensive and simple to administer, score, and interpret.

Currently, there is limited evidence on psychometric properties of screening questionnaires of ASC in such settings. For example, in a systematic review of screening measures for children aged 3 or above at risk for ASC, only five measures had enough published evidence to be included, though the evidence was not well established (Norris & Lecavalier, 2010): the social communication questionnaire (SCQ; Berument, Rutter, Lord, Pickles, & Bailey, 1999), Gilliam autism rating scale/-second edition (Gilliam, 2006), social responsiveness scale/-second edition (SRS/SRS-2; Constantino & Gruber, 2012), autism spectrum screening questionnaire (Ehlers, Gillberg, & Wing, 1999), and Asperger syndrome diagnostic scale (Myles, Bock, & Simpson, 2001). These measures have 40, 42, 65, 27, and 50 items, respectively, and are all relatively lengthy in clinical practice; thus, there is a need for briefer and psychometrically sound measures that warrant further diagnostic assessments. As such, we chose the social and communication disorders checklist (SCDC; Skuse et al., 2005) and strength and difficulties questionnaire (SDQ) (Goodman, 1997) in the present study.

The SCDC is a caregiver or teacher-administered questionnaire of 12 items. It was originally developed to measure the degree of the social and communication deficits in Turner's syndrome (Skuse et al., 1997) and has later been used to measure that deficits in autism traits (Skuse et al., 2005). However, only one study examined the performance of the SCDC to detect ASC in a psychiatry outpatient setting (Bölte, Westerwald, Holtmann, Freitag, & Poustka, 2011). The SDQ is a screening measure of 25 items that form five subscales (composed of five items each): emotional symptoms, hyperactivity/inattention, prosocial behavior, peer problems, and conduct problems. The SDQ discriminates well between children with and without psychiatric disorders in the general population (Goodman, 1997, 1999), but only limited studies examined its screening performance to detect specific psychiatric disorders (Goodman, Renfrew, & Mullick, 2000; Russell, Rodgers, & Ford, 2013; Salayev & Sanne, 2017). As the SDQ has been widely used in clinical practice, it is clinically significant and useful to know how well its subscales can predict ASC

(Russell et al., 2013). In the present study, therefore, we aimed to examine and compare the performance of the SCDC and SDQ, along with the SRS-2 as reference, for the diagnosis of ASC in a psychiatry outpatient setting.

2 | METHODS

2.1 | Participants

This study was conducted in companion with a validation study of the Kiddie Schedule for Affective Disorders and Schizophrenia Present and Lifetime Version for Diagnostic and Statistical Manual of Mental Disorders (DSM-5) in Japan (Nishiyama et al., 2020). We consecutively recruited all outpatients seen by the second author (Takeshi Nishiyama) in the department of child and adolescent psychiatry at the Kamibayashi Memorial Hospital from March 3 to April 21, 2018. The only inclusion criterion was age (between 5 and 18 years in both samples).

Informed consent was obtained from the parents, and assent was obtained from the participating children. Of the 41 patients we approached, all consented to participate in the present study. The procedures of the present study were approved by the institutional ethics committee of the Kamibayashi Memorial Hospital (No.2017015) and Aichi Medical University (No.13-159).

3 | MEASURES

3.1 | Gold-standard diagnosis

The second author, blind to questionnaire responses, made the diagnoses of all the patients based on DSM-5 criteria using all available data sources (including previous records, a significant other, psychological assessments, laboratory test results, and information provided from the multidisciplinary clinical stuff, a caregiver, and a teacher of the patient when possible). He followed each patient longitudinally for at least 3 months to assess the stability of the original diagnoses assigned. We used a "best estimate" diagnosis (Leckman, Sholomskas, Thompson, Belanger, & Weissman, 1982) thus obtained as a gold standard in the present study. This procedure included the Diagnostic Interview for Social and Communication Disorders (DISCO) for the DSM-5 diagnosis of ASD (Kent et al., 2013). The DISCO interview was conducted by the second author, who had been certified to use the instrument for research purposes, before, during or after the sampling period.

3.2 | Questionnaires

In the present study, we used three questionnaires: the SCDC, SDQ, and SRS-2. Whereas the Japanese version of the SDQ and SRS-2 was already developed, that of the SCDC was not (Kamio et al., 2013;

Matsuishi et al., 2008). We translated and adapted the SCDC into Japanese according to the procedure described in the section of translation and cultural adaptation. Then we used the Japanese versions of the SDQ and SRS-2 and a Japanese translation of the SCDC. Although the SDQ and SRS-2 are available in self-reported and caregiver-rated (or teacher-rated) forms, we used the caregiver-rated form in the present study. To examine the test-retest reliability of the SCDC, we administered this measure twice to the subsample of 27 participants, with a mean interval 5.0 weeks.

The SCDC is an informant-administered questionnaire of 12 items that measures autistic traits. Each item of the SCDC is rated on a 3-point Likert scale (0–2): "not true," "quite or sometimes true," or "very or often true" with a maximum score of 24. The SRS-2 is a 65-item rating scale that measures autistic traits. Each item of the SRS-2 is rated on a 4-point Likert scale (0–3) that ranges from "not true" to "almost always true" with a maximum score of 195. The SRS-2 provides the T-score, a standardized score with a mean of 50, and a standard deviation of 10 in the Japanese population by age and sex (Kamio et al., 2013). By standardizing the SRS-2 raw scores, it is possible to represent the severity of autistic symptomology across different sex vand age groups. T-scores indicate: nonclinical (T-score < 60), mild (T-score of 60–65), moderate (T-score of 66–75), and severe (T-score ≥76; Constantino & Gruber, 2012).

Unlike the former two questionnaires focusing on autistic traits, the SDQ is a measure of 25 items that covers five broad areas: emotional symptoms, hyperactivity/inattention, prosocial behavior, peer problems, and conduct problems. Each item of the SDQ is rated on a 3-point Likert scale (0–2): "not true," "somewhat true," or "certainly true." A score for prosocial behavior is in the opposite direction of scores for the other four areas in the SDQ. To use the prosocial score with other scores in a unified manner, we reversed the items for prosocial behavior by (a maximum prosocial score of 10)–(an original unreversed prosocial score), hereinafter referred to as "an unsocial score." The extended version of the SDQ includes the impact supplement that enquires about overall distress, social impairment, burden, and chronicity (Goodman, 1999), which we also used in the present study. The impact supplement consists of eight items: perceived difficulties, chronicity, distress, social impairment in four domains (home life, friendships, classroom learning, and leisure activities), and burden to others. All the items are rated on a 4-point Likert scale: "not at all (0)," "only a little (0)," "quite a lot (1)," or "a great deal (2)." Since the item on chronicity did not predict clinical status in a previous study (Goodman, 1999), we did not include the item in the present study. An impact score is computed by adding five scores of a distress and four social impairment items, ranging between 0 and 10 in the caregiver-rated form.

3.3 | Translation and cultural adaptation

To preserve conceptual equivalence with the original version, we carefully followed *the principles of good practice for the translation and cultural adaptation process for patient-reported outcomes measures*

(Wild et al., 2005). After permission from the original developer (DS), the SCDC was translated into Japanese independently by two Japanese investigators fluent in English (the eighth and ninth authors). Together with the second author, the translators compared the two translations and produced a single, reconciled version. Then, it was back-translated into English by a professional native English translator fluent in Japanese (the first author), who was blind to the original sources of the measures before or during back-translation. The resultant back-translations were sent to the original developer for any translation discrepancies that arose between English and Japanese versions. We repeated these procedures several times until the Japanese versions were approved by the original developer.

3.4 | Statistical analysis

To describe the clinical and demographic variables in the samples, we reported numbers and percentages for categorical variables, and means and standard deviations for continuous variables. To examine the internal consistency reliability of the three questionnaires, we estimated Cronbach's alpha. To examine the test-retest reliability of the SCDC, we estimated intraclass correlation (ICC) based on a one-way random-effects model, single measurement, and absolute agreement (Boateng, Neilands, Frongillo, Melgar-Quiñonez, & Young, 2018). To assess the criterion validity of each questionnaire, we conducted receiver operating characteristics (ROC) analysis to evaluate the association between the gold-standard diagnosis and each scale (Mokkink et al., 2019). We calculated the area under the ROC curve (AUC) by the nonparametric method (DeLong, DeLong, & Clarke-Pearson, 1988). For criterion validity of each questionnaire, we also estimated multiple-level likelihood ratios (LRs; Peirce & Cornell, 1993), which indicate how much more likely or less likely a specific test result is for individuals with a disease than for individuals without this disease. In clinical practice, LR > 10 indicates strong evidence for diagnosis and LR < 0.1 exclusion of diseases. To achieve the optimum number of levels, we followed the rules proposed by a previous study as follows: (1) provide sufficient disordered and nondisordered subjects in each level to allow the LRs to be monotonically related and (2) collapse levels where the LRs are close to one another and their 95% CIs easily overlap (Peirce & Cornell, 1993). In order to examine whether sex, age, and comorbidity of intellectual disability (ID) could have influence on the SCDC and SDQ scores, multivariate regression analysis was performed. Omega squared (ω^2) was used as a measure of effect size to describe the amount of total variation in the scores that could be explained by each predictor. The effect sizes ω^2 of 0.01, 0.06, and 0.14 are considered cut-off points for small, medium and large effect sizes, respectively (Kirk, 2016).

Prior to the study, we estimated the required sample size for test-retest reliability, based on the former observed ICC of 0.81 (Skuse et al., 2005) and found that the sample size of 25 is required to attain the prespecified width of 95% CI of 0.3 (Supporting information 1;

Giraudeau & Mary, 2001). We also estimated the required sample size for a ROC analysis, based on the former observed AUC for the SDQ: 0.714 (standard error: 0.039; Salayev & Sanne, 2017) and that for the SCDC: 0.64 (Bölte et al., 2011), and found that the sample size of 51 is required to attain the prespecified width of 95% CI of 0.3 (Supporting information 1; Hajian-Tilaki, 2014).

All analyses were performed using R version 3.4.0 for Windows. The package *ltm* was used for computing Cronbach's alpha; the package *psych* was used for computing ICC; the package *pROC* was used for computing AUC (Robin et al., 2011); and the package *effectsize* for estimating omega squared.

4 | RESULTS

The sample profile indicated a predominance of males (boys: 78.0%) and school-aged children (mean age $\pm$ SD: 11.1 $\pm$ 3.1 years). The sample revealed low IQ (82.4 $\pm$ 24.6) partly because of eight participants (19.5%) with ID included. A diagnosis of ASC was assigned to 14 participants (34.1%), a diagnosis of Attention deficit hyperactivity disorder to 24 participants (58.5%), and other diagnoses to 17 participants (41.4%). 13 participants (31.7%) were assigned to comorbid diagnoses, and thus, the total does not sum up to 100% (Figure 1). Based on the SRS-2 T-scores, 78.6% ($n = 11$) of ASC subjects were in the severe clinical range (T-score $\geq$ 76), 14.3% ($n = 2$) were in the moderate clinical range (T-score of 66–75), none were in the mild clinical range (T-score of 60–65), and 7.1% ($n = 1$) were in the nonclinical range (T-score < 60).

The SCDC and SRS-2 demonstrated high internal consistency, as shown by Cronbach's alpha coefficients of 0.90 (95% CI: 0.84, 0.93) and 0.95 (95% CI: 0.92, 0.97), respectively (Nunnally & Bernstein, 1994). The SCDC also demonstrated good test-retest reliability, as shown by one-way random ICC of 0.69 (95% CI: 0.43, 0.84) (Fleiss, 2011). On the contrary, the unsocial (= reversed prosocial) and peer problem subscales of the SDQ indicated low internal consistency, as shown by Cronbach's alpha coefficients of 0.69 (95% CI: 0.50, 0.80) and 0.25 (95% CI: 0.00, 0.53), respectively. Especially, the latter internal consistency was unacceptably low. Cronbach's alpha coefficients of other SDQ subscales were 0.62 (95% CI: 0.38, 0.75) for emotional symptoms, 0.46 (95% CI: 0.09, 0.66) for conduct problems and 0.80 (95% CI: 0.64, 0.88) for hyperactivity/inattention.

Table 1 shows the performance of the SCDC, SDQ, and SRS-2 to detect ASC. In the ROC analysis, the unsocial and conduct problem subscales of the SDQ, the SCDC, and SRS-2 revealed AUC significantly larger than 0.5, and the peer problem subscale of the SDQ revealed a marginally significant AUC. Therefore, we did not address the other two subscales of the SDQ (emotional symptoms and hyperactivity/inattention) in subsequent analyses. To examine how well a total score of two from the chosen three subscales of the SDQ predict ASC status, we conducted ROC analyses and found that summing the unsocial and peer problem subscales gave the largest AUC of 0.75 (95% CI: 0.58, 0.91) and the only AUC significantly
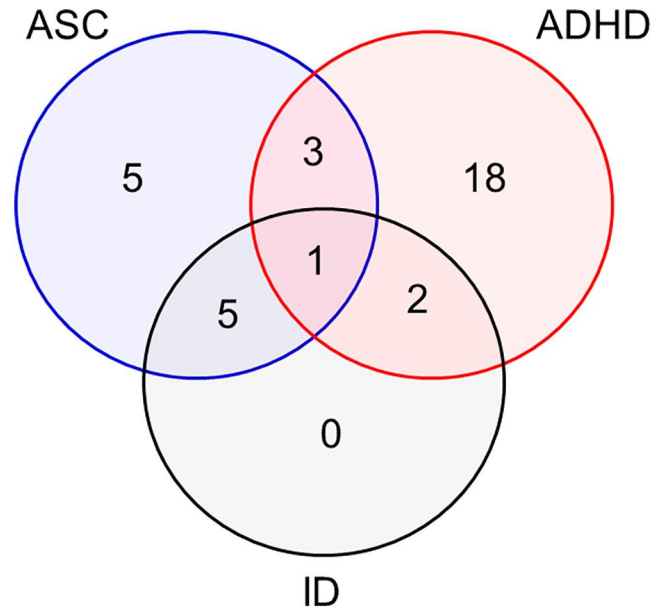

FIGURE 1 Venn diagrams showing the diagnostic overlaps between autism spectrum disorder (ASD), Attention deficit hyperactivity disorder (ADHD), and intellectual disability (ID). The numbers of participants diagnosed with the three most prevalent diseases, ASD, ADHD, and ID were 14, 24, and 8, respectively. The numbers of participants with more than one disorder are represented in the overlapping portions of the circles, while those with one disorder are represented in the non-overlapping portions of the circles

larger than 0.5. Further adding the impact score to the sum of the unsocial and peer problem subscales resulted in a decreased AUC of 0.73 (95% CI: 0.58, 0.89). Then, to examine how well each item in the impact supplement along with the total of the unsocial and peer problem scores predicts ASC status, we conducted a logistic regression analysis and found that only two items, perceived difficulties and social impairment in leisure activities, had a statistically significant and substantial influence on ASC status (Table 2). Therefore, we conducted the ROC analysis to examine how well the total of the unsocial and peer problem scores plus perceived difficulties and social impairment in leisure activities in the impact supplement predicts ASC status. This resulted in an increased AUC of 0.78 (95% CI: 0.62, 0.94), which is essentially identical to that of 0.78 (95% CI: 0.63, 0.94) for the SCDC, making this the overall best scoring method of the SDQ. The SRS-2 showed a slightly higher AUC of 0.84 (95% CI: 0.71, 0.97) than the SCDC and SDQ (Figure 2).

Table 3 shows the multi-level LRs of the SDQ based on the best scoring method and the SCDC for ASC status. Here, the number of levels was set to three according to the rule of thumb described in the statistical analysis section. Both questionnaires had informative levels of LRs, that is, around 0.1 or 10 at both ends of the score range.

Table 4 shows the influence of sex, age, and ID on the SCDC and SDQ scores, conditional on ASC status in the multiple linear regression analysis. Whereas the influence of ASC status was significant and large (0.14 < ω^2), the influence of sex, age, and ID were all found to be non-significant and negligible ($\omega^2 \leq 0.01$).

TABLE 1 The performance of the SCDC, SDQ, and SRS-2 to detect ASC[a]

Scale		AUC (95% CI)
SDQ	Emotional symptoms	0.48 (0.29, 0.67)
	Conduct problems	**0.68 (0.51, 0.86)**
	Hyperactivity/Inattention	0.58 (0.40, 0.76)
	Peer problems	0.65 (0.48, 0.82)
	Unsocial behavior[b]	**0.69 (0.50, 0.87)**
	Impact score	**0.68 (0.52, 0.85)**
	Unsocial behavior + peer problems	**0.75 (0.58, 0.91)**
	Unsocial behavior + conduct problems	0.56 (0.37, 0.74)
	Peer problems + conduct problems	0.45 (0.27, 0.63)
	Unsocial behavior + peer problems + impact score	**0.73 (0.58, 0.89)**
	Unsocial behavior + peer problems + 2 impact items[c]	**0.78 (0.62, 0.93)**
SCDC		**0.78 (0.63, 0.94)**
SRS-2		**0.84 (0.71, 0.97)**

Abbreviations: ASC, autism spectrum conditions; AUC, area under the curve; SCDC, social and communication disorders checklist; SDQ, strength and difficulties questionnaire; SRS-2, social responsiveness scale-second edition.

[a]The area under the ROC curve (AUC) was used to show the performance of the SCDC, SDQ and the SRS-2 to detect ASC. Bold value represents AUC significantly larger than 0.5.

[b]Unsocial behavior: the reversed prosocial behavior scale.

[c]Perceived difficulties and social impairment in leisure activities in the impact supplement.

TABLE 2 The result of the logistic regression analysis to examine the influence of each item in the impact supplement of the SDQ on ASC status[a]

Item	OR (95% CI)
Unsocial behavior + peer problems[b]	1.35 (0.94, 2.15)
Item 1: perceived difficulties	**17.4 (1.18, 604.0)**
Item 2: distress	0.42 (0.02, 4.76)
Item 3: social impairment in home life	0.45 (0.03, 6.18)
Item 4: social impairment in friendships	0.12 (0.01, 0.91)
Item 5: social impairment in classroom learning	0.45 (0.08, 2.12)
Item 6: social impairment in leisure activities	**13.8 (1.87, 216.3)**
Item 7: burden to others	0.55 (0.05, 3.77)

Abbreviations: ASC, autism spectrum conditions; SDQ, strength and difficulties questionnaire.

[a]Bold value represents OR significantly larger than 1.0.

[b]The total score of the unsocial (= reversed prosocial) and peer problems subscales of the SDQ.

5 | DISCUSSION

The present study sought to assess and compare diagnostic performance of very short questionnaires, the SCDC and SDQ to detect ASC along with the SRS-2 as reference. For this purpose, we first developed the Japanese adaptation of the SCDC (Supporting Information 2). This yielded slightly different but essentially the same findings as reported in former studies using the original English and German versions: Cronbach's alpha coefficient of 0.90, compared to 0.93 (Skuse et al., 2005); the ICC of 0.69, compared to 0.81 (Skuse et al., 2005) for test-retest reliability; and the AUC of 0.78, compared to 0.64 (Bölte et al., 2011) to detect ASC in clinical samples. This indicates the success in producing a functionally equivalent measure from English into Japanese via the standard back-translation procedure.

Our study finding that the SCDC and SDQ have worse discriminatory power than the SRS-2 in detecting ASC is no wonder, as the SCDC and SDQ subscales for screening for ASC include much fewer items than the SRS-2: whereas the SRS-2 has 65 items, both the SCDC and the SDQ subscales for screening for ASC have 12 items, where the SDQ subscales for screening for ASC in this study consists of the unsocial subscale, the peer problem subscale and the items of perceived difficulties and social impairment in leisure activities in the impact supplement. We chose two subscales of the SDQ, the unsocial and peer problem subscales as the best predictive subscales of ASC status through exploring possible combinations of subscales. This finding is consistent with the only study that examined the performance of the SDQ to detect ASC in a clinical setting (Salayev & Sanne, 2017), which showed that the two subscales best predicted ASC status with the AUC of 0.71, close to the value of 0.75 in our study. To enhance the diagnostic performance of ASC, we examined the influence of adding the impact supplement of the SDQ using the impact score. Contrary to our expectation, this slightly weakened the diagnostic performance to the AUC of 0.73. On the contrary, adding the most predictive items in the impact supplement alone, the items of perceived difficulties and social impairment in leisure activities,
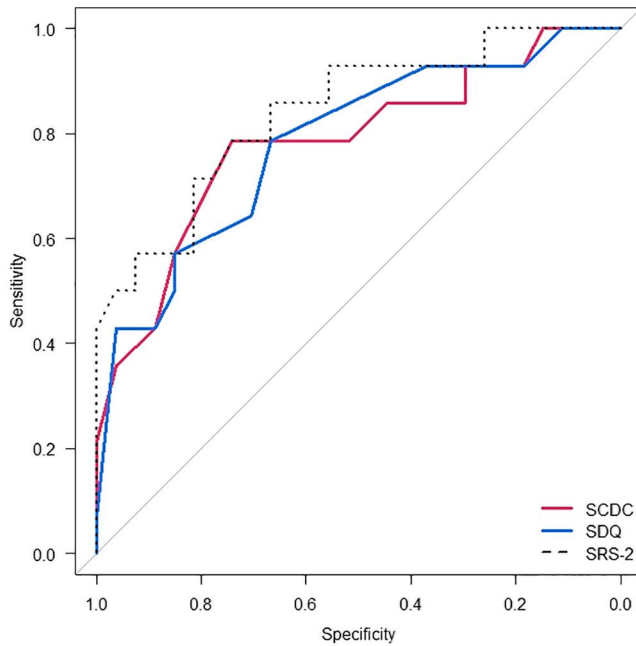
FIGURE 2 Receiver operating characteristics (ROC) curves for the social and communication disorders checklist (SCDC), strength and difficulties questionnaire (SDQ), and social responsiveness scale-second edition (SRS-2) to discriminate between patients with and without autism spectrum conditions. The ROC analysis of the SDQ was conducted based on the total of the unsocial (= reversed prosocial) and peer problem scores plus perceived difficulties and social impairment in leisure activities in the impact supplement. The total score was used for the ROC analysis of the SCDC and SRS-2

TABLE 3 LRs of the SDQ based on the best scoring method and the SCDC for the diagnosis of ASC

Questionnaire	Score	LR (95% CI)
SCDC	19–23	9.64 (1.78, 52.2)
	7–18	0.85 (0.52, 1.42)
	0–6	0.24 (0.05, 1.21)
SDQ1[a]	17–24	11.6 (2.21, 60.7)
	11–16	0.84 (0.47, 1.51)
	0–10	0.19 (0.04, 0.94)
SDQ 2[b]	15–18	9.64 (1.78, 52.2)
	10–16	1.10 (0.63, 1.93)
	0–9	0.16 (0.03, 0.77)

Abbreviations: ASC, autism spectrum conditions; LRs, likelihood ratios; SCDC, social and communication disorders checklist; SDQ, strength and difficulties questionnaire.
[a]SDQ 1: the total of the unsocial and peer problem scores plus perceived difficulties and social impairment in leisure activities in the impact supplement.
[b]SDQ 2: the total of the unsocial and peer problem scores.

increased the diagnostic performance to the AUC of 0.78, identical to that of the SCDC. From this finding, although the impact supplement includes eight items to cover diverse aspects of social impairment

TABLE 4 Results of multiple regression analysis for the influence of sex, age, and ID on the SCDC and SDQ scores

	SCDC			SDQ[a]		
	β[b]	ω²[c]	p-value	β[b]	ω²[c]	p-value
ASC	4.32	0.14	0.033	2.96	0.19	0.017
Sex: male	−1.03	<0.01	0.614	0.17	<0.01	0.894
Age	−0.30	<0.01	0.315	−0.08	<0.01	0.654
ID	2.73	<0.01	0.279	1.73	0.01	0.261

Abbreviations: ASC, autism spectrum conditions; ID, intellectual disability; SCDC, social and communication disorders checklist; SDQ, strength and difficulties questionnaire.
[a]The SDQ score: the total of the unsocial (= reversed prosocial) and peer problem scores plus perceived difficulties and social impairment in leisure activities in the impact supplement.
[b]β: an estimated regression coefficient for each variable.
[c]ω²: Omega squared represents how much variance in the score of each questionnaire is accounted for by each explanatory variable.

and distress, the items could be redundant, and the two items were only found enough to enhance the diagnostic performance. Overall, our results showed that the diagnostic performance of both the SDQ and SCDC was moderate for screening for ASC in a psychiatry outpatient setting. Nevertheless, both questionnaires can be reasonably used in the setting as below.

In general, a trade-off exists between high sensitivity (high false positives) and high specificity (high false negatives). If the multi-level LRs are instead used for test performance, it is not necessary to tolerate the cost of high false positives or negatives. The approach retains as much information as possible that is originally contained in the test by deriving multiple level indices instead of reducing the test into a dichotomous value below or above the cutoff. At the same time, it is important to notice that patients are referred to a psychiatry outpatient setting because of a perceived high risk for psychiatric disorders (ASC here in the study), while general population screening is conducted in a lower-risk population. A more accurate screening test is of greater importance in a low-risk population than in a high-risk population such as psychiatry outpatients: based on the result of the SCDC in the current study (Table 3), a patient with a pretest probability of 33.3% for ASC (≒ the ASC prevalence in our sample) who shows the SCDC score ≥18 has a posttest probability of 83% for ASC, whereas a patient with a pretest probability of 1% who shows the SCDC score ≥18 has a post-test probability of 1.1%. Note that this calculation was conducted using the formula: pretest odds × LR = posttest odds, where odds can be converted to probabilities and vice versa using the following formulae of odds = probability/(1 − probability) and probability = odds/(1 + odds). Thus, such a questionnaire with moderate accuracy can be successfully applied only to high-risk populations such as psychiatry outpatients when multi-level rather than dichotomous LRs are used. As shown in the diagnostic performance of the SRS-2, whose AUC was 0.84 in this study, even questionnaires lengthier than the SCDC and SDQ have essentially moderate diagnostic performance for ASC in clinical

samples: for the SCQ (current form), AUC of 0.77 (Corsello et al., 2007), 0.67 (Snow & Lecavalier, 2008), and 0.56 (Hollocks et al., 2019); and for the SRS, AUC of 0.81 (Bölte et al., 2011) and 0.92 (Duvekot, van der Ende, Verhulst, & Greaves-Lord, 2015). Thus, these lengthier questionnaires for screening for ASC can be applied only to a high-risk population in a clinically meaningful way, just like the SCDC and SDQ.

There have been concerns that several factors such as sex, age, and ID could potentially affect the presentation of autistic symptomatology and thus the accuracy of ASC diagnostic or screening instruments. However, studies on this topic were very limited. In the present study, we found that all of age, sex, and comorbid ID had nonsignificant and negligible influence on the SCDC and SDQ scores, conditional on ASC status. This finding is in agreement with the previous work investigating sex bias in the 10-item autism spectrum quotient (AQ-10), where although individual items showed some sex bias, these biases were found to cancel out to give an overall unbiased test score (Murray et al., 2017). However, our finding is in disagreement with the former studies showing lower accuracy of the SDQ in a severe ID group (Sappok, Diefenbacher, Gaul, & Bölte, 2015) and in a lower age group (Barnard-Brak, Brewer, Chesnut, Richman, & Schaeffer, 2016). The disagreement regarding the influence of age might be partly due to use of an adult sample in the former study.

The strengths of the current study are the use of a sample with a broad spectrum of patients typically seen in clinical practice and of the best-estimate diagnosis of ASC using a semi-structured interview, blinded to the screening tests. The possibility of the so-called "spectrum bias" and "work up bias" is thus minimized (Ransohoff & Feinstein, 1978). Nevertheless, our findings should be viewed with some degree of caution. First, the sample size was relatively small, which led to rather large confidence intervals for AUC and LRs. Especially, since we could not achieve the required sample size to attain the prespecified width of 95% CI, the ROC analyses in the present study were low-powered. The second issue concerns the generalizability of our study findings, especially for the SDQ. In the present study, we examined the best discriminatory scoring method of the SDQ to detect ASC, but the finding should be examined in other samples to confirm the performance. Third, the "best estimate" diagnosis used in the present study relied on all available data sources including the semi-structured interview, DISCO, but not a semi-structured observation instrument such as the Autism Diagnostic Observation Schedule second edition (Lord et al., 2012). In general, diagnostic classification of ASC should rely on the integration of different sources of information, including a parental interview as well as a child observation from different contexts (Constantino & Charman, 2016). Thus, no reliance on an observation instrument might compromise the validity of the "best estimate" diagnosis. Finally, we found that all of age, sex, and comorbid ID had nonsignificant and negligible influence on the SCDC and SDQ scores, conditional on ASC status in the multiple linear regression analysis. However, our sample size was prohibitively small to conduct a subgroup analysis to confirm these factors could influence accuracy of the questionnaires to detect ASC.

In conclusion, the SCDC and SDQ demonstrated moderate screening performance to detect ASC among Japanese psychiatry outpatients. Although questionnaires to detect ASC, including the three examined in this study, generally have only the moderate performance, these can be successfully applied to high-risk populations such as psychiatry outpatients, when multi-level rather than dichotomous LRs are used.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Takeshi Nishiyama https://orcid.org/0000-0002-0032-6508

REFERENCES

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., …, Durkin, M. S. (2018). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *MMWR. Surveillance Summaries, 67*(6), 1–23.

Barnard-Brak, L., Brewer, A., Chesnut, S., Richman, D., & Schaeffer, A. M. (2016). The sensitivity and specificity of the social communication questionnaire for autism spectrum with respect to age. *Autism Research: Official Journal of the International Society for Autism Research, 9*(8), 838–845.

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *The British Journal of Psychiatry: The Journal of Mental Science, 175*, 444–451.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, 149.

Bölte, S., Westerwald, E., Holtmann, M., Freitag, C., & Poustka, F. (2011). Autistic traits and autism spectrum disorders: The clinical validity of two measures presuming a continuum of social communication skills. *Journal of Autism and Developmental Disorders, 41*(1), 66–72.

Constantino, J. N., & Charman, T. (2016). Diagnosis of autism spectrum disorder: Reconciling the syndrome, its diverse origins, and variation in expression. *The Lancet. Neurology, 15*(3), 279–291.

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale* (2nd ed.). Torrance, CA: Western Psychological Services.

Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H., Leventhal, B. L., & Lord, C. (2007). Between a ROC and a hard place: Decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 48*(9), 932–940.

Crowe, B. H. A., & Salt, A. T. (2015). Autism: The management and support of children and young people on the autism spectrum (NICE Clinical Guideline 170). *Archives of Disease in Childhood. Education and Practice Edition, 100*(1), 20–23.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44*(3), 837–845.

Duvekot, J., van der Ende, J., Verhulst, F. C., & Greaves-Lord, K. (2015). The screening accuracy of the parent and teacher-reported social responsiveness scale (SRS): Comparison with the 3Di and ADOS. *Journal of Autism and Developmental Disorders, 45*(6), 1658–1672.

Ehlers, S., Gillberg, C., & Wing, L. (1999). A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children. *Journal of Autism and Developmental Disorders, 29*(2), 129–141.

Fleiss, J. L. (2011). *The design and analysis of clinical experiments*, New York, NY: John Wiley & Sons, Ltd.

Gilliam, J. E. (2006). *GARS-2: Gilliam autism rating scale* (2nd ed.). Austin, TX: PRO-ED.

Giraudeau, B., & Mary, J. Y. (2001). Planning a reproducibility study: How many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Statistics in Medicine, 20*(21), 3205–3214.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 38*(5), 581–586.

Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 40*(5), 791–799.

Goodman, R., Renfrew, D., & Mullick, M. (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child & Adolescent Psychiatry, 9*(2), 129–134.

Guthrie, W., Wallis, K., Bennett, A., Brooks, E., Dudley, J., Gerdes, M., ... Miller, J. S. (2019). Accuracy of autism screening in a large pediatric network. *Pediatrics, 144*(4), e20183963.

Hajian-Tilaki, K. (2014). Sample size estimation in diagnostic test studies of biomedical informatics. *Journal of Biomedical Informatics, 48*, 193–204.

Hollocks, M. J., Casson, R., White, C., Dobson, J., Beazley, P., & Humphrey, A. (2019). Brief report: An evaluation of the social communication questionnaire as a screening tool for autism spectrum disorder in young people referred to child & adolescent mental health services. *Journal of Autism and Developmental Disorders, 49*(6), 2618–2623.

Johnson, C. P., Myers, S. M., & American Academy of Pediatrics Council on Children with Disabilities. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics, 120*(5), 1183–1215.

Kamio, Y., Inada, N., Moriwaki, A., Kuroda, M., Koyama, T., Tsujii, H., ... Constantino, J. N. (2013). Quantitative autistic traits ascertained in a national survey of 22 529 Japanese schoolchildren. *Acta Psychiatrica Scandinavica, 128*(1), 45–53.

Kent, R. G., Carrington, S. J., Le Couteur, A., Gould, J., Wing, L., Maljaars, J., ... Leekam, S. R. (2013). Diagnosing autism spectrum disorder: Who will get a DSM-5 diagnosis? *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 54*(11), 1242–1250.

Kirk, R. E. (2016). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*(5), 746–759.

Leckman, J. F., Sholomskas, D., Thompson, W. D., Belanger, A., & Weissman, M. M. (1982). Best estimate of lifetime psychiatric diagnosis: A methodological study. *Archives of General Psychiatry, 39*(8), 879–883.

Lord, C., Rutter, M., Dilavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism diagnostic observation schedule, second edition (ADOS-2) manual (Part 1) modules 1-4*, Torrance, CA: Western Psychological Services.

Matsuishi, T., Nagano, M., Araki, Y., Tanaka, Y., Iwasaki, M., Yamashita, Y., ... Kakuma, T. (2008). Scale properties of the Japanese version of the strengths and difficulties questionnaire (SDQ): A study of infant and school children in community samples. *Brain & Development, 30*(6), 410–415.

Mokkink, L. B., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., De Vet, H. C., & Terwee, C. B. (2019). *COSMIN study design checklist for Patient-reported outcome measurement instruments*, Amsterdam, The Netherlands: Department of Epidemiology and Biostatistics.

Murray, A. L., Allison, C., Smith, P. L., Baron-Cohen, S., Booth, T., & Auyeung, B. (2017). Investigating diagnostic bias in autism spectrum conditions: An item response theory analysis of sex bias in the AQ-10. *Autism Research: Official Journal of the International Society for Autism Research, 10*(5), 790–800.

Myles, B. S., Bock, S. J., & Simpson, R. L. (2001). *Asperger syndrome diagnostic scale (ASDS)*, Austin, TX: PRO-ED.

Nishiyama, T., Sumi, S., Watanabe, H., Suzuki, F., Kuru, Y., Shiino, T., ... Hirai, K. (2020). The kiddie schedule for affective disorders and schizophrenia present and lifetime version (K-SADS-PL) for DSM-5: A validation for neurodevelopmental disorders in Japanese outpatients. *Comprehensive Psychiatry, 96*, 152148.

Norris, M., & Lecavalier, L. (2010). Screening accuracy of level 2 autism spectrum disorder rating scales. A review of selected instruments. *Autism: The International Journal of Research and Practice, 14*(4), 263–284.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Peirce, J. C., & Cornell, R. G. (1993). Integrating stratum-specific likelihood ratios with the analysis of ROC curves. *Medical Decision Making: An International Journal of the Society for Medical Decision Making, 13*(2), 141–151.

Ransohoff, D. F., & Feinstein, A. R. (1978). Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *The New England Journal of Medicine, 299*(17), 926–930.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 12*, 77.

Russell, G., Rodgers, L. R., & Ford, T. (2013). The strengths and difficulties questionnaire as a predictor of parent-reported diagnosis of autism spectrum disorder and attention deficit hyperactivity disorder. *PLoS One, 8*(12), e80247.

Saemundsen, E., Magnússon, P., Georgsdóttir, I., Egilsson, E., & Rafnsson, V. (2013). Prevalence of autism spectrum disorders in an Icelandic birth cohort. *BMJ Open, 3*(6), e002748.

Salayev, K. A., & Sanne, B. (2017). The strengths and difficulties questionnaire (SDQ) in autism spectrum disorders. *International Journal on Disability and Human Development, 16*(3), 275–280.

Sappok, T., Diefenbacher, A., Gaul, I., & Bölte, S. (2015). Validity of the social communication questionnaire in adults with intellectual disabilities and suspected autism spectrum disorder. *American Journal on Intellectual and Developmental Disabilities, 120*(3), 203–214.

Shattuck, P. T., Durkin, M., Maenner, M., Newschaffer, C., Mandell, D. S., Wiggins, L., ... Cuniff, C. (2009). Timing of identification among children with an autism spectrum disorder: Findings from a population-based surveillance study. *Journal of the American Academy of Child and Adolescent Psychiatry, 48*(5), 474–483.

Siu, A. L., US Preventive Services Task Force (USPSTF), Bibbins-Domingo, K., Grossman, D. C., Baumann, L. C., Davidson, K. W., Ebell, M., ... Pignone, M. P. (2016). Screening for autism spectrum disorder in young children: US preventive services task force recommendation statement. *JAMA, 315*(7), 691–696.

Skuse, D. H., James, R. S., Bishop, D. V., Coppin, B., Dalton, P., Aamodt-Leeper, G., ... Jacobs, P. A. (1997). Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature, 387*(6634), 705–708.

Skuse, D. H., Mandy, W. P. L., & Scourfield, J. (2005). Measuring autistic traits: Heritability, reliability and validity of the social and communication disorders checklist. *The British Journal of Psychiatry: The Journal of Mental Science, 187*, 568–572.

Snow, A. V., & Lecavalier, L. (2008). Sensitivity and specificity of the modified checklist for autism in toddlers and the social

communication questionnaire in preschoolers suspected of having pervasive developmental disorders. *Autism: The International Journal of Research and Practice, 12*(6), 627–644.

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P., & ISPOR Task Force for Translation and Cultural Adaptation. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research, 8*(2), 94–104.

Williams, E., Thomas, K., Sidebotham, H., & Emond, A. (2008). Prevalence and characteristics of autistic spectrum disorders in the ALSPAC cohort. *Developmental Medicine and Child Neurology, 50*(9), 672–677.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.