Testing two screening instruments for autism spectrum disorder in UK community child health services

TONY CHARMAN[1] | GILLIAN BAIRD[2] | EMILY SIMONOFF[1] | SUSIE CHANDLER[1,2] | ABI DAVISON-JENKINS[3] | AJAY SHARMA[4] | TONY O'SULLIVAN[5] | ANDREW PICKLES[1]

1 King's College London, Institute of Psychiatry, Psychology & Neuroscience, London; **2** Newcomen Centre, Guy's & St Thomas' NHS Foundation Trust, London; **3** UCL Institute of Child Health, London; **4** Sunshine House Community Services, Guy's and St. Thomas' NHS Trust, London; **5** Kaleidoscope, Lewisham Centre for Children & Young People, Lewisham Healthcare, London, UK.

Correspondence to Tony Charman at Department of Psychology, King's College London, Institute of Psychiatry, Psychology & Neuroscience, PO Box 077, De Crespigny Park, London SE5 8AF, UK. E-mail: tony.charman@kcl.ac.uk

This article is commented on by Williams on pages 324–325 of this issue.

PUBLICATION DATA
Accepted for publication 30th June 2015.
Published online 25th August 2015.

ABBREVIATIONS

ASD	Autism spectrum disorder
AUC	Area under the curve
M-CHAT	Modified Checklist for Autism in Toddlers
PPV	Positive predictive value
SCQ	Social Communication Questionnaire

AIM The aim of this study was to test the accuracy of two screening instruments in UK Community health services: Modified Checklist for Autism in Toddlers (M-CHAT) and Social Communication Questionnaire (SCQ) for autism spectrum disorder (ASD). A two-stage screening and in-depth assessment procedure, combined with sampling stratification and statistical weighting, allowed the accuracy of the screens to be estimated in the entire population of referred children.

METHOD The study included all referrals of children aged 18 to 48 months to community paediatric and speech and language therapy services in two London districts over a 12-month period between September 2004 and September 2005. Parents of 808 children were approached; screen data were obtained on 543 children (67.2%). A stratified subsample of 120 children received an in-depth assessment for ASD as defined by the International Statistical Classification of Diseases and Related Health Problems, 10th edition. Community clinician judgement of likely ASD was available for 98 out of the 120 children.

RESULTS The sensitivity and specificity were 64% (95% confidence intervals; range 51–80%) and 75% (63–85%) for the SCQ, and 82% (72–92%) and 50% (33–64%) for M-CHAT. There was no evidence that the area under the curve differed between the two screening instruments. There was also no evidence that clinician judgement of likely ASD differed from either of the screening tests. The screening tests did not perform well to confirm preliminary clinical judgement to refer (in series), nor as an alternative indicator for referral (in parallel).

INTERPRETATION While screening tests may provide useful information, their accuracy is moderate. Screening information in isolation should not be used to make referral decisions regarding specialized ASD assessment.

Autism spectrum disorders (ASDs) are neurodevelopmental conditions that significantly impact on the individual, their family, and society.[1,2] Recent prevalence estimates suggest that around 1% of children and young people have ASD.[3,4] There is an emerging, though still limited, evidence base for the effectiveness of early social communication and behavioural interventions.[5,6] Screening instruments have been developed to identify children at a young age.[7] Three have been tested in general population screening studies. The Checklist for Autism in Toddlers (CHAT)[8] was found to have low sensitivity (38%) in a population of 18-month-old children. The Early Screening of Autistic Traits[9] had low positive predictive value (PPV) for ASD (25%) at 14 months of age. In two recent studies in populations of 16- to 30-month-old children using a two-stage screen and re-screen design, the Modified Checklist for Autism in Toddlers (M-CHAT)[10] and the M-CHAT-Revised with Follow-up[11] were found to have moderate PPV (54% and 48% respectively). However, sensitivity remains to be established in these latter studies as only screen positives were followed-up.

In samples of young children referred with developmental concerns, where the prevalence of ASD is higher, screens have higher PPV. Snow and Lecavalier[12] reported that M-CHAT had a PPV of 79% in 18- to 48-month-old children referred for possible ASD. The Social Communication Questionnaire (SCQ)[13] has also been tested as a screen in referred samples of 2- to 6-year-old children, and has been found to have moderate sensitivity (70%[12] and 60%[14]) and specificity (52%[12] and 70%[14]). In the subsample of children on whom both screens were completed (n=39), both the M-CHAT and SCQ had PPVs of around 70%. Eaves et al.[15] reported similar PPVs (~65%) for the M-CHAT and SCQ in separate subsamples of children

(age 2–3y and 4–6y) referred to an ASD specialist clinic. Another study of children between 20 months and 40 months of age referred for a clinical assessment after a positive score on the Early Screening of Autistic Traits screen reported a PPV of 79%.[16] However, in all of these studies the specificity of both the M-CHAT (used in a one-stage manner) and the SCQ has been low (50% or lower), meaning that reliance on screening information to trigger a referral for specialist assessment could result in many screen 'false negatives'.

The American Academy of Pediatrics has recommended routine use of screens (including the CHAT and M-CHAT) at well-child checks.[17] In contrast, the UK National Institute of Health and Care Excellence (NICE) guidelines[18] and the UK National Screening Committee[19] concluded that there was insufficient evidence to recommend any ASD-specific screening instrument. The threshold set by NICE was for a sensitivity and specificity of at least 80%, with a lower confidence interval of 70% or greater. In line with this, Al-Qabandi et al.[20] have criticized the American Academy of Pediatrics recommendation for universal screening on the basis of the limited research evidence.

There is evidence from both the USA[21] and the UK[22] that in many communities the age of diagnosis is highly variable and in some cases unacceptably late. If there was evidence that screening instruments might help identify those at greatest risk of ASD when concerns are first raised, this might aid prompt referral and earlier diagnosis. In some areas the wait for specialist autism assessments is considerable and reducing inaccurate referral via screening could help remove this 'bottleneck' to the benefit of children, parents, and local services.[23] However, if the sensitivity of screening instruments is low, there is a cost to children and families in potential delayed assessment and diagnosis, as well as the distress of unnecessary referrals.

The present study aimed to test the accuracy of the most widely studied screening instruments – the M-CHAT and the SCQ – in community child health services in an urban setting in the UK. To test whether ASD screens might improve the accuracy of referrals to specialist ASD diagnostic teams we independently collected screening information on preschool children who had been referred to community child health and speech and language therapy services. On a stratified subsample we conducted independent diagnostic assessments. In addition, we collected routine clinician judgement at the initial appointment (blinded to screening information) about likely ASD, so that we could compare whether combining clinical judgement with screening information improved the accuracy of referral.

METHOD

The Guy's Hospital Research Ethics Committee approved the study (03/08/07). Informed consent was obtained from all parents/carers. A STARD checklist is included as Appendix S1 (online supporting information).

What this paper adds

- The Social Communication Questionnaire and Modified Checklist for Autism in Toddlers screens performed only moderately well in accurately identifying cases of autism spectrum disorder (ASD) in children referred to community child health services.
- Screening information in isolation should not be used to decide whether or not to make a referral for specialized ASD assessment.

The SCQ[13] and the M-CHAT[11] are questionnaires that ask parents to endorse (Yes/No) symptom descriptions that cover the full range of behaviours described in the classification systems (SCQ 40 items, M-CHAT 23 items). The established cut-off on the SCQ for ASD is a score of ≥15.[13] The M-CHAT has two high/low risk thresholds: scoring on 2 or more of 6 'critical' items and scoring ≥3 overall.[11]

Screening stage

In the UK, when parents or primary care practitioners are concerned about a child's development, the child is referred to community child health services, including paediatrician-led child development teams and speech and language therapy teams. If ASD is suspected then it is common for a referral to be made to a more specialist, often multidisciplinary, paediatric-led ASD diagnostic team. Across a 12-month period (September 2004–September 2005), all parents of children between the ages of 18 months and 48 months (mean 32.9mo [SD 8.0]) referred to community services for any developmental problem in two inner London NHS Trusts were approached by the research team, independently of the clinical services.

After the initial appointment, the clinicians (paediatricians, speech and language therapists) were asked to complete a checklist of concerns about a possible ASD and/or whether a referral had been initiated for an ASD assessment (both districts had specialist multidisciplinary ASD diagnostic teams). When clinicians had not completed the checklist the research team accessed the file and recorded whether an ASD concern and/or referral had been noted. The research team independently collected the screening information, which was not available to the community clinicians. The research diagnostic process (described below) was conducted independently of the local community assessments and was blind to the screening information and the community clinical judgement of likely ASD.

The in-depth measures consisted of the Autism Diagnostic Interview-Revised[24] and the Autism Diagnostic Observation Schedule-Generic[25] as the core assessments of autism. In addition, IQ was assessed using the Mullen Scales of Early Learning Scales,[26] language using the Preschool Language Scale,[27] and adaptive behaviour using the Vineland Adaptive Behaviour Scales.[28]

The Autism Diagnostic Interview-Revised is an investigator-led parent interview that has an established cut-off for childhood autism[24] and a recommended cut-off for broader ASD.[29] The Autism Diagnostic Observation Schedule-Generic is a structured, play-based experimenter–child

observation that generates an algorithm score with established cut-offs for ASD.[25] In the current study 58 children completed Module 1 (for non-verbal children and children with single words) and 62 completed Module 2 (for children with phrase speech).

Diagnostic stage

The research team scored the assessments and made an initial clinical diagnosis. The principal clinical investigators (GB, ES, TC) reviewed comprehensive clinical material for every case, including the Autism Diagnostic Interview and Autism Diagnostic Observation Schedule summaries, clinical vignette, and psychometric results. A consensus clinical diagnosis of childhood autism or other ASD (other pervasive developmental disorder, pervasive developmental disorder-unspecified, atypical autism) as defined by the International Statistical Classification of Diseases and Related Health Problems 10th edition (ICD-10)[30] was made on the basis of all sources of information. Because of the relatively young age of children at assessment and the modest sample size of the childhood autism and 'other ASDs' subgroups, the subgroups were merged into one ASD diagnosis group for analysis.

Analysis

All reported frequencies and means are unweighted. All other statistics, such as proportions and percentages, are target population estimates calculated using inverse probability weighting to take account of the differences in sampling proportions and the selective participation in in-depth assessment across the sample stratification (see 'Sample stratification and selection'). Confidence intervals and standard errors were calculated using the linearization version of the robust parameter covariance matrix as implemented by the *svy* procedures of Stata Statistical Software release 9.0 (StataCorp LP, College Station, TX, USA).[31] A receiver–operator–characteristic (ROC) area under the curve (AUC) analysis was performed to assess and compare the discriminant power of the screens/clinician judgement in distinguishing ASD cases from non-ASD cases.[32,33] Confidence intervals for weighted AUC estimates and tests were obtained by bootstrap resampling receiver–operator–characteristic procedures of Stata release 9.0.[31]

RESULTS

A total of 347 referrals to child health services and 461 referrals to speech and language therapy services were received. Of these 808 referrals, 722 gave primary reasons for referral: 398 (55%) language difficulties; 138 (19%) social communication difficulties; 82 (11%) behaviour; 48 (7%) general development; and 56 (8%) medical or other problems. No reason for referral was recorded for 86 children.

Parents of all 808 referrals (556 males, 237 females, 15 not recorded) were approached (initially by post; followed-up by a telephone call) to complete the screening pack containing the SCQ and M-CHAT. The presentation of the two screens was counterbalanced in the questionnaire pack across the sample. A total of 543 screens were completed (67.2%), 276 by post and 267 by telephone. Characteristics of participants for whom screening data were achieved are shown in Table I.

Clinician checklist information was available for 369 out of 543 (68%) children for whom screening information was available (233 completed by clinicians, 132 by the research team, four cases had missing information). In 101 cases (19%), concern about possible ASD and/or referral to the specialist team was noted at the initial appointment. Initial clinical appointments took place sometimes before (n=132) but more often after (n=298) the screens were completed (mean [SD] lag=1.2mo [3.4], range −13.4mo to 17.2mo); the appointment date was missing on 131 cases.

Sample stratification and selection

For the in-depth assessment we sampled children from across all cells of consistency and inconsistency of measures. The sample was thus stratified not only by service (equal proportions: child development team versus speech and language therapy teams) and referral to ASD diagnostic team (two levels: Yes vs No), but also by SCQ score (four levels: low score <8; moderately low score 8–14; moderately high score 15–21; high score ≥22) and M-CHAT (binary: high vs low risk). The sampling fraction for each stratum was then used as an inverse probability weight such that the weighted selected sample was then representative of the whole screened sample. Of 181 families selected, 61 were not seen (21 opted out, 40 were not contactable) and 120 children (100 males, 20 females) received in-depth assessment (66.3%) (see Fig. 1). There was no evidence that children selected for in-depth assessment who did not participate further were more likely to be screen positive versus screen negative ($\chi^2(1)=0.46$, p=0.50). The mean age at the time of assessment was 51.6 months (SD 8.8, range 32–73) as the direct

Table I: Characteristics of participants screened

Sample size (n)	543
Males (%)	72
Mean age (mo) (SD; range)	35.2 (8.3; 18–56)
Parent-reported child ethnicity (%)[a]	
Black	49
White	38
Asian	3
Other	8
Mixed ethnic group	6
Language spoken at home (%)[b]	
English only	59
English plus and additional language	33
Additional language only	8
Maternal highest qualification (%)[c]	
No formal qualifications	17
Vocational qualifications	19
GCSE	21
A levels	14
Degree or above	29

[a]n=10 missing. [b]n=15 missing. [c]n=27 missing.

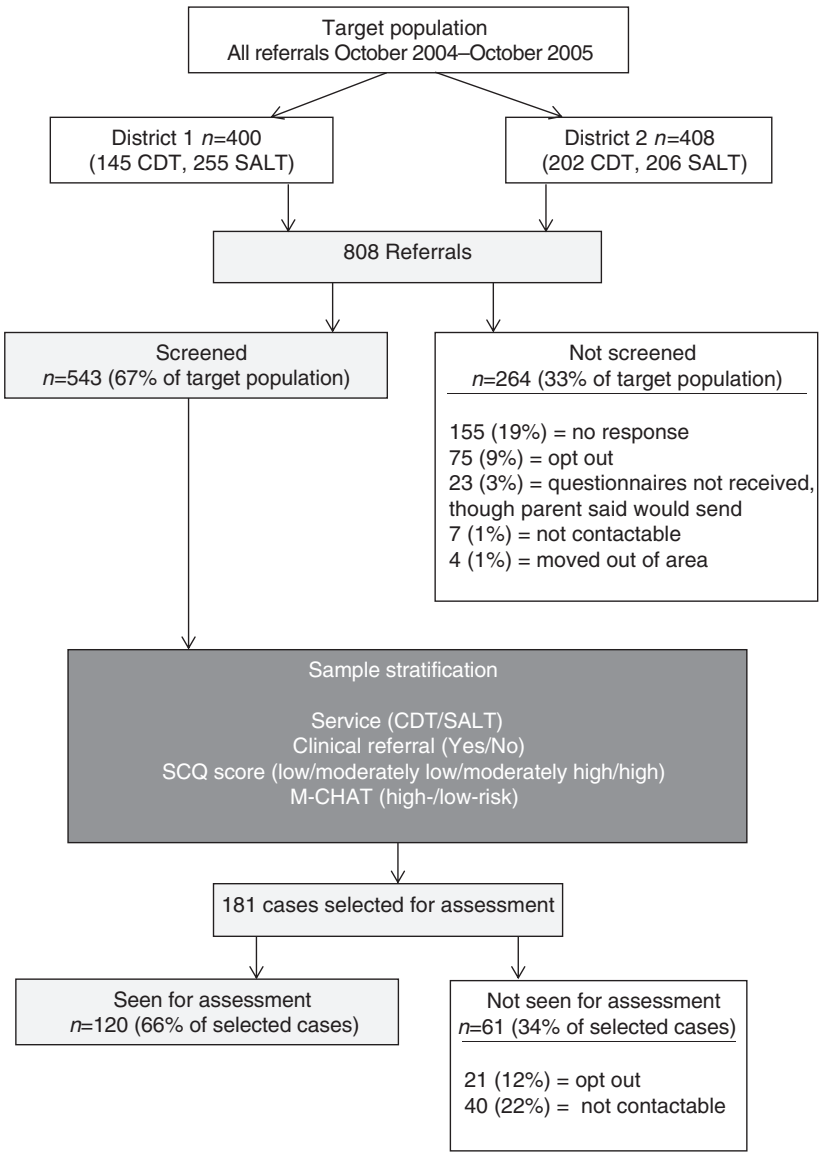

Figure 1: Stratum selection and participation. CDT, Child Development Team (paediatrician-led); SALT, Speech and Language Therapy service; SCQ, Social Communication Questionnaire; M-CHAT, Modified Checklist for Autism in Toddlers.

assessments only begun once the screening stage of the study was completed.

Within the sample of 120 children seen for in-depth assessment, 55 met criteria for a consensus ICD-10 clinical diagnosis of ASD. Of the 65 children who did not meet the criteria for an ASD diagnosis, 16 had an intellectual disability (IQ<70), 11 a language delay, four a hyperkinetic disorder, four an oppositional defiant disorder, one cerebral palsy, one Holt-Oram syndrome, and 28 did not meet criteria for any of the clinical conditions that were assessed as part of the ASD-focused assessment. Because the assessments did not comprehensively test for all possible psychiatric or medical disorders we consider possible diagnoses (or lack of them) in this group uncertain. However, in many of these children there had been an earlier developmental concern (e.g. speech delay or unclear speech) that prompted the initial referral that had resolved or was below clinical threshold for a developmental disorder by the time of the research assessment. Characteristics of the children seen for assessment are shown in Table II. Some children were not able to complete the direct assessments and others fell below the test basal required to calculate standardized scores. The sample numbers for each assessment are shown in the table.

The AUC, sensitivity, specificity, PPV and negative predictive value of the two screens in predicting consensus diagnosis ASD versus non-ASD status are shown in Table III. There was no evidence that the AUC differed between the SCQ (70%) and M-CHAT screens (66%) (p=0.47). The sensitivity and specificity for the SCQ were

Table II: Characteristics of the participants directly assessed

	Non-ASD		ASD	
	n	Mean (SD)	*n*	Mean (SD)
Mullen ELC (SS)	60	84.4 (2.6)	22	68.3 (3.3)
PLS TLS (SS)	61	82.4 (2.5)	24	72.2 (3.9)
VABS ABC (SS)	64	78.3 (1.7)	48	61.5 (1.7)
ADOS-Social	65	2.60 (0.27)	55	8.56 (0.45)
ADOS-Comm	65	1.91 (0.17)	55	4.35 (0.31)
ADI-Social	65	4.42 (0.45)	54	15.2 (0.81)
ADI-Verbal	58	4.31 (0.42)	26	12.1 (0.83)
ADI-Non-verbal	7	3.71 (1.32)	28	9.79 (0.57)
ADI-Repetitive	65	1.74 (0.24)	54	4.89 (0.31)

ASD, autism spectrum disorder; Mullen ELC, Mullen Early Learning Composite; SS, Standard Score; PLS TLS, Preschool Language Scale Total Language Scale; VABS ABC, Vineland Adaptive Behavior Scale Adaptive Behavior Composite; ADOS, Autism Diagnostic Observation Schedule; ADI, Autism Diagnostic Interview.

64% and 75% and for the M-CHAT 82% and 50% respectively. For clinical judgement alone the AUC was 77%, which did not differ from the SCQ (p=0.27) or the M-CHAT (p=0.11), sensitivity was 85% and specificity 69%. When the screened sample was split by median age (≤33mo vs >33mo) the M-CHAT had the highest sensitivity for younger children (93% with accompanying low specificity of 42%) and clinical judgement had the highest sensitivity for older children (97% with a specificity of 72%).

Clinician judgement of a likely ASD was then combined with screen positive results on the two screens in two ways (see Table III). First, they were considered in series and the sensitivity and specificity calculated for children who were both screen positive and for whom clinicians had indicated a possible ASD. Sensitivity was reduced (clinician *and* SCQ 51%; clinician *and* M-CHAT 69%) but specificity increased (clinician *and* SCQ 87%; clinician *and* M-CHAT 82%). Next, they were considered in parallel and the sensitivity and specificity calculated for children who were either screen positive or for whom clinicians had indicated a possible ASD. Sensitivity was greatly increased (clinician *or* SCQ 98%; clinician *or* M-CHAT 98%) but specificity was low (clinician *or* SCQ 47%; clinician *or* M-CHAT 27%).

DISCUSSION

Both the SCQ and the M-CHAT performed only moderately well at identifying cases who went on to meet the diagnostic criteria for ASD. The M-CHAT had higher sensitivity than the SCQ but lower specificity. Based on their performance in this study, reliance on either of the screening instruments in isolation would lead to considerable under identification of children likely to go on to receive an ASD diagnosis (i.e. an unacceptable rate of screen 'false negatives', especially the SCQ) and also considerable over-identification of children who would be unlikely to go on to receive an ASD diagnosis (i.e. an unacceptable rate of screen 'false positives', especially the M-CHAT). There was no evidence that clinician judgement of likely ASD differed from either screening instrument. Also, the screens did not perform well either to confirm preliminary clinical judgement to refer (in series) or as an alternative indicator for referral (in parallel). As would be expected, if *both* clinical judgement *and* a positive screen result was required to trigger a referral, sensitivity was unacceptably low (leading to under identification of cases), whereas if referral was based on *either* a clinical judgement of likely ASD *or* a screen positive result specificity was unacceptably low (leading to many non-ASD cases being referred on). Finally, in younger children (≤33mo) the M-CHAT had the highest sensitivity and in older children (>33mo) clinical judgement had the highest sensitivity; but the current study lacks the power to robustly test whether the screens worked differently in children of different ages.

While screens may provide useful information to aid the decision to refer for a specialized ASD assessment, their accuracy is moderate and does not meet the criterion set of 80% sensitivity and specificity used in the recently published UK NICE guidance[18] and recommended in general for screens for developmental disabilities.[34] In children about whom a developmental concern has been raised, ASD screening in isolation should not be used to make a referral for specialized ASD assessment; neither should they be combined with clinical opinion about likely ASD either to confirm this opinion (which leads to a rise in missed cases) or as an alternative to clinical opinion (which leads to over-inclusion). The current findings do not support the American

Table III: The AUC, sensitivity, specificity, PPV, and NPV of the SCQ, M-CHAT, and clinical judgement in isolation and combined (weighted values)

	AUC Weighted value % (95% CI)	Sensitivity Weighted value % (95% CI)	Specificity Weighted value % (95% CI)	PPV Weighted value % (95% CI)	NPV Weighted value % (95% CI)
SCQ	70 (59–80)	64 (51–78)	75 (63–85)	60 (48–73)	78 (65–88)
M-CHAT	66 (55–76)	82 (72–92)	50 (33–64)	49 (33–62)	82 (67–92)
Clinical judgement	77 (68–87)	85 (76–94)	69 (52–81)	73 (61–82)	83 (68–93)
SCQ plus clinical judgement	69 (60–78)	51 (37–63)	87 (75–94)	79 (64–91)	65 (51–78)
M-CHAT plus clinical judgement	76 (66–85)	69 (54–80)	82 (68–90)	79 (64–89)	73 (58–84)
SCQ or clinical judgement	73 (51–95)	98 (95–100)	47 (27–65)	65 (54–75)	97 (85–100)
M-CHAT or clinical judgement	63 (54–72)	98 (94–100)	27 (11–49)	57 (44–68)	94 (67–100)

AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; SCQ, Social Communication Questionnaire; M-CHAT, Modified Checklist for Autism in Toddlers.

Academy of Pediatrics recommendation that ASD screening instruments should be used at universal well-child checks,[17] and are in line with more cautionary opinion.[20] It might be that the two-stage screen and re-screen procedure recently reported for the M-CHAT/M-CHAT-Revised[10,11] provides a model for a more accurate screening procedure. However, in these studies the rate of initial (first stage) screen positives were 9%[10] and 7%,[11] which would place considerable demand on services required to conduct the follow-up screening, as well as careful explanation to parents as to what the initial test result meant. Even if such service provision was available this would only be justified if the sensitivity was high. However, a recent very large population study (n=52 026) using the M-CHAT in a single-stage manner that included population follow-up to identify all known cases of ASD, found that only one-third of ASD cases were screen positive on the initial M-CHAT administration at 18 months of age.[35]

Competing motivations lie behind the increasing interest in early screening for ASD. On the one hand, there is a concern to identify ASD as soon as possible in young children in order to provide support to the family and access to appropriate interventions and services. On the other hand, waiting lists for multidisciplinary assessment for ASD and other complex developmental disorders are under pressure in many countries so there is a desire to reduce the referral of children who may not require this specialist service.[23] This needs to be balanced with the relative costs and benefits of false positives and false negatives. Broadly these costs fall on different parties: false positives may lead to unneeded specialized assessment and parental concern, whereas false negatives may lead to under identification and late diagnosis and intervention.

In many ways the moderate performance of the screens to accurately identify ASD is perhaps not surprising. ASD is a developmental disorder whose presentation, perhaps particularly in young children, can vary over time. It also commonly co-occurs with other developmental disorders such as language delay/disorder[36] and intellectual disability[37] and at a later age with child psychiatric disorders,[38] making differential diagnosis challenging. Population studies indicate that a significant number of children may display characteristics of ASD at a preschool age but go undiagnosed, and although their trajectory of symptoms into mid-childhood is less severe than children who receive a diagnosis, they have elevated rates of educational and behavioural problems.[39]

Another important consideration is whether developing accurate screens *specifically* for ASD is either a desirable or a realistic goal. At an early age emerging symptoms of ASD overlap with those seen in other groups of children who would also benefit from a more comprehensive assessment and intervention, such as children with general developmental delay, language impairment, and those with early emerging emotional and behavioural difficulties. Some of the interventions that such children require, and the support and advice to be provided to parents, are based on similar principles. Few studies have directly compared general developmental screens versus ASD-specific screens, and such studies should be undertaken to further inform child healthcare policy and practice (for rare exceptions see Glascoe et al.[40] and Pinto-Martin et al.[41]). More generally, the services and advice that parents of children with a range of neurodevelopmental disabilities require (that affect general development, language and communication, and commonly motor and other aspects of adaptive functioning) are best served by professional teams that have expertise in both ASD and in a range of neurodevelopmental conditions.

The extent to which these results are applicable to other populations is unclear. The two health districts in which the study was conducted are inner London areas in the UK with high levels of social and economic deprivation, individuals from ethnic minorities, and families where English is an additional language. From personal knowledge, both districts had also worked for many years to train their work force about the early signs of ASD. However, the sensitivity and specificity levels on the SCQ and M-CHAT in this study are comparable to those found in referred preschool samples in very different communities in the USA, Australia, and the Netherlands.[12,14–16] No previous study has achieved screening data on such a large community clinical sample, none has directly compared two different screening instruments, and none has systematically also captured referring clinician judgement about possible ASD.

CONCLUSION

We found that both when used in isolation, and also when combined with clinician judgement, neither the SCQ nor the M-CHAT performed sufficiently well to be recommended for universal adoption within UK community paediatric services. As with much clinical practice, clinical judgement in combination with information gained from administering specific tests and assessments (including information from ASD screening instruments), as well as the presence of parental concerns[42] and the impact the developmental problems are having on the child and family, should inform the decision to refer for more specialized assessment rather than reliance on the result of a screening instrument in isolation.

ACKNOWLEDGEMENTS

We are grateful to the families who participated in all phases of the study. We are also grateful to the clinician teams who co-operated so willingly in the research. We are grateful to Kristelle Hudry, Sally Clifford, Fiona May, Victoria Bird, and Elisabeth Ireland for help with data collection. The authors have stated that they had no interests that might be perceived as posing a conflict or bias. This work was funded by a grant from the Health Foundation (formerly PPP Health Foundation). Tony Charman received support from the ESF COST Action BM1004 (ESSEA). Emily Simonoff and Andrew Pickles receive support from the National Institute for Health Research (NIHR) Mental Health

Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

SUPPORTING INFORMATION

The following additional material may be found online:

Appendix S1: STARD checklist.

REFERENCES

1. Lai MC, Lombardo MV, Baron-Cohen S. Autism. *Lancet* 2014; **383**: 896–910.
2. Buescher AVS, Cidav Z, Knapp M, Mandell DS. Costs of autism spectrum disorders in the United Kingdom and the United States. *JAMA Pediatr* 2014; **168**: 721–28.
3. Baird G, Simonoff E, Pickles A, et al. Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet* 2006; **368**: 210–15.
4. Centers for Disease Control and Prevention (CDC). Prevalence of autism spectrum disorder among children aged 8 years – autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR Surveill Summ* 2014; **63**: 1–21.
5. National Institute for Health and Care Excellence (NICE CG170). The Management and Support of Children and Young People on the Autism Spectrum. London: NICE, 2013.
6. Warren Z, McPheeters ML, Sathe N, et al. A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics* 2011; **127**: E1303–11.
7. Charman T, Gotham K. Measurement issues: screening and diagnostic instruments for autism spectrum disorders – lessons from research and practise. *Child Adol Mental Health* 2013; **18**: 52–63.
8. Baird G, Charman T, Baron-Cohen S, et al. A screening instrument for autism at 18 months of age: a 6-year follow-up study. *J Am Acad Child Adol Psychiatry* 2000; **39**: 694–702.
9. Dietz C, Swinkels S, van Daalen E, et al. Screening for autistic spectrum disorder in children aged 14–15 months. II: population screening with the early screening of autistic traits questionnaire (ESAT). Design and general findings. *J Autism Dev Disorders* 2006; **36**: 713–22.
10. Chlebowski C, Robins DL, Barton ML, et al. Large-scale use of the Modified Checklist for Autism in low-risk toddlers. *Pediatrics* 2013; **131**: E1121–7.
11. Robins DL, Casagrande K, Barton M, et al. Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics* 2014; **133**: 37–45.
12. Snow AV, Lecavalier L. 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* 2008; **12**: 627–44.
13. Rutter M, Bailey A, Lord C. Social Communication Questionnaire (SCQ). Los Angeles, CA: Western Psychological Services, 2003.
14. Allen CW, Silove N, Williams K, Hutchins P. Validity of the social communication questionnaire in assessing risk of autism in preschool children with developmental problems. *J Autism Dev Disord* 2007; **37**: 1272–78.
15. Eaves LC, Wingert HD, Ho HH, et al. Screening for autism spectrum disorders with the Social Communication Questionnaire. *J Dev Behav Pediat* 2006; **27**: S95–103.
16. Oosterling I, Rommelse N, de Jonge M, et al. How useful is the Social Communication Questionnaire in toddlers at risk of autism spectrum disorder? *J Child Psychol Psychiatry* 2010; **51**: 1260–68.
17. Johnson CP, Myers SM. Identification and evaluation of children with autism spectrum disorders. *Pediatrics* 2007; **120**: 1183–215.
18. National Institute for Health and Care Excellence (NICE CG170). Recognition, Referral and Diagnosis of Children and Young People on the Autism Spectrum. London: NICE, 2011.
19. The UK NSC recommendation on Autism screening in children. 2012. Available from: http://www.screening.nhs.uk/autism (accessed 10 January 2015).
20. Al-Qabandi M, Gorter JW, Rosenbaum P. Early autism detection: are we ready for routine screening? *Pediatrics* 2011; **128**: E211–17.
21. Shattuck PT, Durkin M, Maenner M, et al. Timing of identification among children with an autism spectrum disorder: findings from a population-based surveillance study. *J Am Acad Child Adol Psychiatry* 2009; **48**: 474–83.
22. Williams E, Thomas K, Sidebotham H, et al. Prevalence and characteristics of autistic spectrum disorders in the ALSPAC cohort. *Dev Med Child Neurol* 2008; **50**: 672–77.
23. Palmer E, Ketteridge C, Parr JR, et al. Autism spectrum disorder diagnostic assessments: improvements since publication of the National Autism Plan for Children. *Arch Disease Child* 2011; **96**: 473–75.
24. Rutter M, LeCouteur A, Lord C. The Autism Diagnostic Interview-Revised (ADI-R). Los Angeles, CA: Western Psychological Services, 2003.
25. Lord C, Risi S, Lambrecht L, et al. The Autism Diagnostic Observation Schedule-Generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disorders* 2000; **30**: 205–23.
26. Mullen B. Mullen Scales of Early Learning. Circle Pines, MN: American Guidance Service, 1995.
27. Zimmerman IL, Steiner V, Pond RE. Preschool Language Scales. 3rd edn. London: The Psychological Corporation, 1997.
28. Sparrow SS, Cicchetti DV, Balla DA. Vineland Adaptive Behavior Scales, second edition: Survey interview form. Circle Pines, MN: American Guidance Service, 2005.
29. Risi S, Lord C, Gotham K, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adol Psychiatry* 2006; **45**: 1094–103.
30. World Health Organisation. Mental Disorders: A Glossary And Guide To Their Classification In Accordance With The 10th Revision Of The International Classification Of Disease – Research Diagnostic Criteria: ICD-10. Geneva: WHO, 1993.
31. StataCorp. Stata 9 Base Reference Manual. College Station, TX: Stata Press, 2005.
32. Dunn G. Statistics in Psychiatry. London: Arnold, 2000.
33. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic curve. *Radiology* 1983; **214**: 29–36.
34. Glascoe FP. Toward a model for an evidenced-based approach to developmental/behavioral surveillance, promotion and patient education. *Amb Child Health* 1999; **5**: 197–208.
35. Stenberg N, Bresnahan M, Gunnes N, et al. Identifying children with autism spectrum disorder at 18 months in a general population sample. *Paed Peri Epidem* 2014; **28**: 255–62.
36. Michelotti J, Charman T, Slonims V, Baird G. Follow-up of children with language delay and features of autism from the pre-school years into middle childhood. *Dev Med Child Neurol* 2002; **44**: 812–19.
37. Charman T, Pickles A, Simonoff E, Chandler S, Loucas T, Baird G. IQ in children with autism spectrum disorders: population data from the SNAP Project. *Psychol Medicine* 2011; **41**: 619–27.
38. Simonoff E, Pickles A, Charman T, Loucas T, Chandler S, Baird G. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity and associated factors in a population-derived sample. *J Am Acad Child Adol Psychiatry* 2008; **47**: 921–29.
39. Russell G, Golding J, Norwich B, Emond A, Ford T, Steer C. Social and behavioural outcomes in children diagnosed with autism spectrum disorders: a longitudinal cohort study. *J Child Psychol Psychiatry* 2012; **53**: 735–44.
40. Glascoe FP, Macias MM, Wegner LM, Robertshaw NS. Can a broadband developmental-behavioral screening test identify children likely to have autism spectrum disorder? *Clin Pediatrics* 2007; **46**: 801–05.
41. Pinto-Martin JA, Young LM, Mandell DS, Poghosyan L, Giarelli E, Levy SE. Screening strategies for autism spectrum disorders in pediatric primary care. *J Dev Behav Pediatrics* 2008; **29**: 345–50.
42. Glascoe FP. Parents' concerns about children's development: prescreening technique or screening test? *Pediatrics* 1997; **99**: 522–28.