Screening for autism in pre-school children in primary care: Systematic review of English Language tools

Elizabeth Mawle\textsuperscript{a}, Peter Griffiths\textsuperscript{b,*}

\textsuperscript{a}Health Visitor, Westminster Primary Care Trust, 158 Herne Hill Road, London SE24 0AH, UK
\textsuperscript{b}Primary and Intermediate Care, King's College London, Florence Nightingale School of Nursing and Midwifery, Waterloo Road, London SE1 8WA, UK

Received 6 June 2005; received in revised form 10 October 2005; accepted 15 November 2005

Abstract

Objectives: To review the accuracy of brief screening tools for autism in pre-school children.

Design: Systematic review of diagnostic accuracy studies.

Data sources: Medline, Embase, CinaHL and Psychlit plus references of identified papers and contact with authors.

Subjects: Children and infants aged 5 years or less without a prior diagnosis of autism or pervasive development delay.


Outcome measures: Sensitivity, specificity, positive and negative predictive value of screening tools and likelihood ratios relative to a diagnostic assessment made using either DSM-III/IV or ICD 10 diagnosis.

Results: Three studies considering two tools were identified. The CHecklist for Autism in Toddlers (CHAT) was tested on an appropriate population sample with moderate long-term follow-up but demonstrated poor sensitivity and positive predictive value. Weaker evidence suggested that the Modified CHecklist for Autism in Toddlers (M-CHAT) had high sensitivity but follow-up was of shorter term and less comprehensive.

Conclusions: The CHAT demonstrated a level of sensitivity unlikely to be useful for population screening purposes, however, its high specificity suggests it has utility in secondary screening. The M-CHAT is a parent only report and might be more sensitive, and therefore appropriate for population screening. However, full conclusions regarding its accuracy cannot be drawn until follow-up data has been collected.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Autistic disorder; Sensitivity and specificity; Infant; Child; Pre-school; Mass screening; Primary health care (all MESH)

What is already known about the topic?

• Public interest and concern about autism is currently high.
• Early intervention is believed to improve outcome for children.

What this paper adds

• Routine population screening of young children has been recommended by some in order to maximise the benefits of treatment.

• Routine screening in primary care requires accurate instruments that are easily administered. A number of candidate instruments exist.
This review examines the evidence base for such instruments and concludes that none is supported by sufficient evidence to recommend routine use in primary care.

1. Introduction

Autism is a developmental disorder characterised by impairments in socialisation, communication and demonstrating flexible behaviour (Baron-Cohen et al., 2000). Recent reviews suggest a prevalence in developed countries of 5.4–5.5/10,000 (Fombonne, 1999) for autism and between 18.7 (Fombonne, 1999) and 57 per 10,000 (Scott et al., 2002) for autism and pervasive developmental delay (PDD) combined. The prevalence of autism and autistic spectrum suggests that it is a disorder, that will present to professionals involved in the screening and surveillance of children in the primary care sector.

Increasingly there is agreement amongst clinicians that the treatment of autism is more effective if commenced prior to the age of 4 (Baron-Cohen et al., 2000; Tanguay, 2000) and this suggests that there could be prognostic benefit in early screening in pre-school children. Furthermore, siblings of children diagnosed as autistic have a 5% chance of developing the disease, information that may be extremely useful for parents planning a pregnancy (Baird et al., 2001). Study of early social development supports the increasingly popular hypothesis that autism could be detected at 18 months (usually pretend play and joint attention are evident at this time) or earlier (Filipek et al., 2000; Cox et al., 1999; Stone et al., 1999; Johnson et al., 1992). However, there is a lack of standardised practice in developmental screening in this country and others (Hall and Elliman, 2003; Filipek et al., 2000; Flanagan and Nuaillain, 2001).

In the USA, the Quality standards Subcommittee of the American Academy of Neurology and Child Neurology Society (Filipek et al., 2000) recommends that all children should be screened for autism using a validated screening tool. It recommends that any child failing routine surveillance should be screened using the CHildhood Autism Rating Tool (CHAT) (Baird et al., 2000) or The Autism Screening Questionnaire (Berument et al., 1999). At present in the UK screening for autism is not recommended (Hall and Elliman, 2003; NSC).

In the UK, the use of standardised checklists for developmental screening is not routine although details of screening tests are widely available on the world wide web and recommended to practitioners and parents by autism charities. In some areas children without previously identified problems receive little input from primary healthcare professionals after MMR vaccina-

tion at around 13 months. However, this appears to be the time at which the first symptoms of Autistic Spectrum might appear.

There is currently increased public awareness regarding autism in some countries, most notably the UK, due to a speculated link with the MMR vaccine. For these reasons primary healthcare professionals should be equipped to reassure and respond to parents who are concerned or whose children may be at risk. This is particularly important given evidence that half of parents whose children go on to a diagnosis of autism have recognised abnormality before the age of 2 (Howlin and Moore, 1997). Despite early parental concern, in this particular study, only 53% of children received a diagnosis by the age of 5 years.

Arguably, detection of autism meets the criteria for screening. Autism has long-term negative effects, it is reasonably common and a cause of parental concern, screening is non-invasive, and early intervention can be effective in reducing negative symptoms (Baron-Cohen et al., 2000). The purpose of this review is to evaluate the available research evidence measuring the accuracy of screening for autism in children under 4 years old in the primary care population. Screening tools will be assessed for their utility in primary care populations, including resource implications.

2. Methods

The databases Medline (1966–2004), Cinahl, Embase (1990–2004) and Psychinfo (1872–2004) were systematically searched. Terms for autism were combined (and) with terms for children. A standard search for studies of diagnosis and screening (Greenhalgh, 2001) was used to identify relevant study types (see Table 1). The Health Technologies Assessment databases and National Screening Committee website were also searched (NSC). References lists of relevant studies were checked for further studies. The author of one uncompleted study was contacted to identify any subsequent and additional publications. No other steps were taken to locate unpublished material or grey literature.

3. Criteria

Studies considered were those making comparison between results of an English language screening/diagnostic tool specific for autism/autistic spectrum and a reference standard based on DSM-III/IV (American Academy of Paediatrics (APA), 1994; American Psychiatric Association (APA) 1994) or ICD 10 (World Health Organisation (WHO), 1993) criteria for diagnosing autism/autistic spectrum disorders. Participants must be under the age of four at screening and have
no prior diagnosis of developmental delay. The screening tool must be one that could be used in a general primary care setting (defined as the ability to score the tool in a brief consultation without extensive prior training). Tools that test only for Asperger’s Syndrome (a sub-category of the Autistic Spectrum with low prevalence) were not considered nor were studies where the assessment was for global developmental delay. One reviewer selected studies for review and extracted data. Decisions were checked and validated by a second reviewer. Where studies did not provide sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV), these were calculated from the data. Ninety-five percent confidence intervals for sensitivity and specificity were estimated where not reported in the paper using an exact (Clopper–Pearson) method for single proportions (Stats Direct 2.4.4). Data on reproducibility of test results was also sought.

4. Findings

The initial search (March 2004) yielded 666 citations (Medline 240; Embase 159; PsychInfo 244; Cinahl 23) of which 36 were broadly relevant. Scrutiny of reference lists of these papers identified further relevant publications and a total of 52 studies were subjected to detailed assessment. Three studies met the review criteria (Table 2). Two studies utilise the same screening tool (CHAT), which combines parent and professional’s reports and the third uses a tool based only on parent report (MCHAT). In all cases screening requires minimal training and takes approximately 15 min. Forty-nine studies were excluded. These studies were frequently concerned with discriminating between autism and other disorders and in these cases research populations consisted only of children diagnosed with autism or developmental delay of some form. Furthermore these studies tend to report only on the correlation between test results and diagnosis, without providing data from which to calculate the measures here required (for details and reasons for exclusion see Table 3).

In order to ensure currency the searches were run again in January 2005 prior to submission for publication. Thirteen additional unique citations were identified of which 5 were potentially relevant. None were included in the review. Two studied non-English language instruments (Bolte and Poustka, 2004; Wong et al., 2004) two, (de Bildt et al., 2004; Steinhausen and Metzke, 2004) studied groups already diagnosed with autism and/or developmental delay and one studied the STAT instrument (Stone et al., 2004), which was excluded as the instrument was too time consuming for primary care use. One instrument, which might be suitable, (the PDDST—Siegel, 1998) was not supported by any published evidence that we could find although it was highlighted in reference lists of some papers.

5. Study quality

All studies used a brief checklist, which is administered in approximately 15 min. Baron-Cohen et al. (1992) screened 91 subjects aged between 17 and 21 months using a brief parent/professional report checklist, the CHAT. Fifty infants were randomly selected from a London health centre attending for a routine check-up. The additional 41 subjects (group 2) were selected as a high-risk group as they have siblings diagnosed with autism. Group 2 was matched on age but no other variables are explicitly controlled for (e.g. IQ). This group was selected because in the absence of a matched high-risk group of subjects a much larger sample size would be required to ensure sufficient cases of autism. Screening was performed by GPs or Health Visitors. Ten subjects received only the parental report part of the CHAT due to GP’s refusals to screen. All subjects were contacted after 1 year. Follow-up is therefore very limited. Those children highlighted by this follow-up as suffering developmental delay were assessed according to DSM-III criteria.

| Table 1  |
| Search strategy on OVID databases |

<table>
<thead>
<tr>
<th>Search terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disorder</td>
</tr>
<tr>
<td>Autis$^a$</td>
</tr>
<tr>
<td>Autistic spectrum</td>
</tr>
<tr>
<td>Exp Autistic Disorder$^b$</td>
</tr>
<tr>
<td>1 or 2 or 3</td>
</tr>
<tr>
<td>Population</td>
</tr>
<tr>
<td>Child$</td>
</tr>
<tr>
<td>Infant$</td>
</tr>
<tr>
<td>Toddler$</td>
</tr>
<tr>
<td>Preschool$</td>
</tr>
<tr>
<td>exp CHILD/</td>
</tr>
<tr>
<td>exp INFANT/</td>
</tr>
<tr>
<td>exp CHILD,PRESCHOOL/</td>
</tr>
<tr>
<td>5 or 6 or 7 or 8 or 9 or 10 or 11</td>
</tr>
<tr>
<td>4 and 12</td>
</tr>
<tr>
<td>Methodological filter</td>
</tr>
<tr>
<td>Sensitivity and specificity</td>
</tr>
<tr>
<td>Value$^c$</td>
</tr>
<tr>
<td>Exp “Sensitivity and Specificity”/</td>
</tr>
<tr>
<td>14 or 15 or 16 or 17</td>
</tr>
<tr>
<td>17 and 13</td>
</tr>
</tbody>
</table>

*S Ovid truncation.

*bExp “term” — indicates an exploded index term.

c*tw search for keyword in title or abstract only.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Title</th>
<th>Participants</th>
<th>Screening test</th>
<th>Diagnostic criteria</th>
<th>Blinded assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baron-Cohen et al. (1992)</td>
<td>Can autism be detected at 18 months?</td>
<td>N 91, age 18 months (17–21 months) 50 randomly selected at check-up, 41 younger siblings of autistic children</td>
<td>Checklist by GP or HV (A) 9 parent report items, (B) 5 professional observations</td>
<td>DSM III criteria. Two independent psychiatrists</td>
<td>Not explicit. Probably blinded screening, Unblinded diagnosis</td>
</tr>
<tr>
<td>Baird et al. (2000)</td>
<td>A screening instrument for Autism at 18 months of age a 6 year</td>
<td>N 16,235, age 18 months ± 2 months birth cohort Children with profound developmental delay excluded</td>
<td>Checklist by GP or HV (A) 9 parent report items, (B) 5 professional observations</td>
<td>All children re-screened 3.5 years 12770 checklists returned.</td>
<td>Blinded screening followed by unblinded screening followed by probable unblinded diagnosis</td>
</tr>
<tr>
<td>Robins et al. (2001)</td>
<td>The Modified Checklist For Autism in Toddlers</td>
<td>N 1293, 1122, 18 or 24-month-old screened in well baby clinics 171 referred from early intervention sites</td>
<td>30 reduced to 23 yes/no items for parental report</td>
<td>DSM IV, Vineland Adaptive Scales, Bayley Scales, CARS, Communication and Symbolic Behaviour Scale</td>
<td>Screening blind</td>
</tr>
<tr>
<td>Authors</td>
<td>Study</td>
<td>Subjects</td>
<td>Screening Tool</td>
<td>Reference standard</td>
<td>Reason for exclusion</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>Adrien et al. (1992)</td>
<td>Validity and Reliability of the Infant Behavioural Summarised Evaluation (IBSE): A Rating scale of young children with Autism and developmental disorders</td>
<td>N 89 children referred due to concern over developmental delay</td>
<td>33 behavioural items and classified under 6 headings, administered by experienced raters used to autistic children, using a video of 20 min</td>
<td>DSM III-R</td>
<td>French Language Scale</td>
</tr>
<tr>
<td>Barthelemy et al. (1997)</td>
<td>Validation of the Revised Behaviour Summarised Evaluation Scale</td>
<td>N 136 developmentally disabled 20–139 months</td>
<td>5 days direct observation</td>
<td>Blinded assessment using Expert Severity Score, 2 psychiatrists</td>
<td>Population characteristics</td>
</tr>
<tr>
<td>Brereton et al. (2002)</td>
<td>Screening young people for Autism with the Developmental Behaviour Checklist</td>
<td>N 360 Consecutively presenting individuals referred to autism assessment centres by primary care practitioners 4–18 years, 180 with autism matched with 180 with intellectual disability</td>
<td>Development behaviour checklist</td>
<td>DSM IV criteria</td>
<td>Population sample age</td>
</tr>
<tr>
<td>Berument et al. (1999)</td>
<td>Autism Screening Questionnaire: diagnostic validity</td>
<td>N 200</td>
<td>40 question screen postal questionnaire</td>
<td>Autistic diagnostic interview and clinical observation</td>
<td>Age</td>
</tr>
<tr>
<td>DiLavore et al. (1995)</td>
<td>The Pre-Linguistic Autism diagnostic observation schedule</td>
<td>N 63, 14-61 months</td>
<td>Semi-structured observation schedule 30 min to administer</td>
<td>ICD 10 and DSM IV criteria</td>
<td>Population characteristics</td>
</tr>
<tr>
<td>Authors</td>
<td>Study</td>
<td>Subjects</td>
<td>Screening Tool</td>
<td>Reference standard</td>
<td>Reason for exclusion</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Vance et al. (2005)</td>
<td>N 70 years, Age 11.2</td>
<td>Checklist can be administered by parents, lay persons and nurses</td>
<td>2 psychiatrists and/or psychologists made a rating of psychopathology</td>
<td>Insufficient subject information</td>
</tr>
<tr>
<td></td>
<td>Freeman et al. (1978)</td>
<td>N 89 children aged 23-65 months</td>
<td>‘disturbed and non-disturbed’</td>
<td>Not a measure of autism specifically</td>
<td>Insufficient information for outcome data.</td>
</tr>
<tr>
<td></td>
<td>The Behaviour Observation Scale for Autism</td>
<td>36 autistic, 23 normal, 30 mental retardiation</td>
<td>67 behaviours observed and rated from 0-3</td>
<td>Consensus between 2 child psychiatrists</td>
<td>Insufficient information for outcome data.</td>
</tr>
<tr>
<td>Freeman (1986)</td>
<td>A scale for rating symptoms of patients with the syndrome of autism in real life settings</td>
<td>N 229 patients with Autism</td>
<td>Ritco-Freeman real life rating scale</td>
<td>DSM III</td>
<td>Population characteristics</td>
</tr>
<tr>
<td>Gillberg et al. (1993)</td>
<td>Screwing methods, epidemiology and evaluation of intervention in DAMP in Pre-school Children</td>
<td>N unknown, Aged 6-7 years</td>
<td>Questionnaire and motor examination screening</td>
<td>Not known</td>
<td>Age of population</td>
</tr>
<tr>
<td>Krug et al. (1980)</td>
<td>Behaviour Checklist for identifying severely handicapped individuals</td>
<td>N 1049 aged between 18 months and 35 years some without a diagnosis (most with) and 62 3-23 years already diagnosed as autistic</td>
<td>Survey checklist details of administration not clear how it is performed</td>
<td>Not specified</td>
<td>Insufficient subject scoring data to establish sensitivity and specificity</td>
</tr>
<tr>
<td>Lord et al. (2000)</td>
<td>The Autism Diagnostic Observation Schedule-Generic: Standard measure of social and communication deficits associated with the spectrum of Autism</td>
<td>N 74, module 1 N 55, module 2</td>
<td>30 min semi-structured observational schedule</td>
<td>Clinical diagnosis with no clear reference standard of DSM IV or ICD 10</td>
<td>Need for data (IQ), and not readily available in primary care setting</td>
</tr>
<tr>
<td>Makita and Umezu (1973)</td>
<td>An objective evaluation technique for autistic children: an introduction of CLAC scheme</td>
<td>0</td>
<td>Theoretical set of screening tools</td>
<td>None</td>
<td>Not research study</td>
</tr>
<tr>
<td>Study</td>
<td>Authors</td>
<td>Year</td>
<td>Sample</td>
<td>Age</td>
<td>Tool</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>------</td>
<td>--------</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>Nordin and Lee (1998)</td>
<td>The Swedish Version of the Childhood Autism Rating Scale in a clinical setting</td>
<td>N 25 4–12 years</td>
<td>Childhood autism rating scale</td>
<td>Swedish Language tool</td>
<td>No VARCHAR</td>
</tr>
<tr>
<td>Rimland (1971)</td>
<td>The Modified Checklist for Autism in Toddlers (M-CHAT)</td>
<td>N 1293 children</td>
<td>Brief parent checklist</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Robins and Lee (2003)</td>
<td>The Modified Checklist for Autism in Toddlers (M-CHAT)</td>
<td>N 1293 children</td>
<td>Brief parent checklist</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Schopler et al. (1980)</td>
<td>Toward Objective Classification of childhood Autism: Childhood Autism Rating Scale</td>
<td>N 537 0–10 years</td>
<td>15 item rating scale</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Sevin et al. (1991)</td>
<td>A comparison of three commonly used autism scales</td>
<td>N 24 aged 2–22 with autism or autism symptomatology</td>
<td>19 item rating scale</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>South et al. (2002)</td>
<td>Utility of the Gilliam Autism rating scale</td>
<td>N 199 children 3–10.5 years</td>
<td>42 item rating scale</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Stone and Hogan (1993)</td>
<td>A structured parent interview for identifying young children with autism</td>
<td>N 165 under 6 years</td>
<td>104 item interview</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Stone (2000)</td>
<td>A Validity Analysis of selected instruments used to assess children with autism</td>
<td>N 36 2–15 years</td>
<td>32 item interview</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>South et al. (2002)</td>
<td>Utility of the Gilliam Autism rating scale</td>
<td>N 199 children 3–10.5 years</td>
<td>42 item rating scale</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>South et al. (2002)</td>
<td>Utility of the Gilliam Autism rating scale</td>
<td>N 199 children 3–10.5 years</td>
<td>42 item rating scale</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Stone (2000)</td>
<td>A Validity Analysis of selected instruments used to assess children with autism</td>
<td>N 36 2–15 years</td>
<td>32 item interview</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>South et al. (2002)</td>
<td>Utility of the Gilliam Autism rating scale</td>
<td>N 199 children 3–10.5 years</td>
<td>42 item rating scale</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Stone (2000)</td>
<td>A Validity Analysis of selected instruments used to assess children with autism</td>
<td>N 36 2–15 years</td>
<td>32 item interview</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Wadden et al. (1991)</td>
<td>A closer look at the autism behaviour checklist</td>
<td>N 123, 67 autistic and 56 mentally retarded aged 6-15 years</td>
<td>22 item checklist</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Volkmar (1998)</td>
<td>The Autism Diagnostic Behaviour Checklist</td>
<td>N 157 Autistic and non-autistic children aged 6–15 years</td>
<td>42 item interview</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>
Baird et al. (2000) screened a cohort of children from 10 health districts in the South East of the UK. In all, 16, 235 (40%) of the 40,818 eligible children were screened using the CHAT instrument at routine checks (mean age 18.7 months, SD 1.1 months). Otherwise eligible subjects with profound developmental delay were excluded on the basis of practitioners’ discretion. Screening was performed by primary healthcare workers during a routine 18-month check-up; however, in 15.7% of the cases questionnaires were mailed directly to parents to complete. No analysis is given of differences between professional and parental CHAT scores and parent only CHAT scores. Children who screened positive were re-screened by the research team using the CHAT (CHAT 2) 1 month later (unblinded). Due to resource constraints only 50% of medium-risk screens were re-screened and outcome data were calculated only using those who were. Children developing autism over a 7-year follow-up period were identified in several stages. At age 3.5 (78.7% response rate) and 5.5 years (47.8% of original sample responded) parents completed further checklists. If children had been referred for assessment or failed a checklist they were assessed by the research team. Outside referrals to a clinical centre were also tracked. Children identified in these ways were directly assessed by the research team and diagnosed according to ICD 10 Criteria. Records of Autistic children were identified through special needs registers and special schools and professionals were consulted. Children identified in this way were assigned an ICD 10 diagnosis at 7–8 years, agreed from notes by researchers. Subjects have not all received the ICD 10 assessment but there has been follow-up, case detection, and consistent diagnosis of cases found.

Robins et al. (2001) screened 1293 children in the USA using a parent report version of the CHAT (M-CHAT). In all, 1122 children between 18 and 24 months were recruited at well baby clinics and 171 children aged between 18 and 30 months were recruited from those who had been referred for early intervention without any diagnosis. Methods of recruitment are not explicit. Children with severe motor deficits and lack of expressive language were excluded (1 child excluded). No analysis is given of differences between professional and parental CHAT scores and parent only CHAT scores. Children who failed the screen were re-screened (unblinded) over the telephone and those who passed at this stage were analysed as part of the screen negative group. Those who failed the repeat screen were invited for developmental assessment (none refused but one was lost to follow-up) and children were assessed and diagnosed (unblinded) according to the DSM-IV criteria, and for the others, follow-up is still in progress but interim results reported in abstract only form have been obtained from the authors (Robins and Lee, 2003). Thus far, 537 participants have been re-screened aged between 3.5 and 4 years old, 42% of the original participants. These participants represent 484 from the unselected population and 51 from the selected population. Children were re-screened using the M-CHAT and telephone interview. Developmental evaluation was offered to all children failing the two stages of the re-screen and those who had failed the two-stage screen originally. A total of 26 children received developmental evaluation, all of whom had been evaluated at the age of 2 years (45% of the original screen positive group). Children were evaluated against DSM-IV criteria, among others.

Thus no study tested all research participants against the reference standard. Either long-term follow-up to identify cases that develop or short-term assessment of test positives only was used. In the Baron-Cohen et al. (1992) and Baird et al. (2000) studies, no details are given of whether researchers applying the reference standard are aware of (blinded) to the screening results. In Robin et al.’s (2001) study, researchers were aware of screen results. In all studies where further assessment is selective based on adverse result at re-screening, those applying the reference standard are (by definition) aware of the result of the screening. Thus all the studies are potentially vulnerable to ‘verification bias’ (Jaeschke et al., 1994) whereby clinical judgement is influenced by knowledge of a previous test result.

6. Results

Results of the studies are summarised in Table 4. The Baron-Cohen study (1992) gives high values of sensitivity and specificity for the CHAT but this can be given little weight due to the limitations of study size, and lack of use of the reference standard (Baird et al., 2000). The apparently good performance in terms of sensitivity may simply be an artefact due to short follow-up as more false negatives (generally children with less frank symptoms who would be harder to detect) might subsequently be diagnosed. In any case a wide confidence interval indicates a large degree of imprecision. Perfect specificity is likely to simply be an artefact of sample size.

Baird et al. (2000) report results in two stages using the data from screen 1 (first screen of all subjects) and screen 2 (re-screen using CHAT of high and medium risk scorers on screen 1) separately. Outcome measures for Childhood Autism (CA) and for an autistic spectrum approach (including PDD) are also reported separately. For all outcomes on both high and medium risk cut-offs the sensitivity is extremely low even for the highest estimates and allowing for the relatively wide confidence intervals. This implies a high level of false negatives in the practice setting. If the sensitivity of this instrument is transposed to a national population sample, the number with autism that would not be identified would be large. Furthermore, the reference standard was not applied to...
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Diagnosis</th>
<th>Study prevalence (%)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baron-Cohen et al. (1992)</td>
<td>N 91</td>
<td>Autistic spectrum&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4.4</td>
<td>100% (40–100)</td>
<td>100% (96–100)</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Baird et al. (2000) Screen 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>N 16,235</td>
<td>Childhood autism</td>
<td>0.3</td>
<td>H 20% (10–34)</td>
<td>99.8% (99.8–99.9)</td>
<td>26.3%</td>
<td>99.7%</td>
</tr>
<tr>
<td>Baird et al. (2000) Screen 2</td>
<td>N 16, 235</td>
<td>Childhood autism</td>
<td>0.3</td>
<td>M 38% (25–53)</td>
<td>M 97.6% (97–98)</td>
<td>M 4.7%</td>
<td>M 99.8%</td>
</tr>
<tr>
<td>Baird et al. (2000) screen 1</td>
<td>N 16,235</td>
<td>Autistic spectrum</td>
<td>0.579</td>
<td>H 18% (9–31)</td>
<td>100% (100–100)</td>
<td>75%</td>
<td>99.7%</td>
</tr>
<tr>
<td>Baird et al. (2000) screen 2</td>
<td>N 16, 235</td>
<td>Autistic spectrum</td>
<td>0.579</td>
<td>M 20% (10–34)</td>
<td>99.9% (99.8–99.9)</td>
<td>29.4%</td>
<td>99.7%</td>
</tr>
<tr>
<td>Baird et al. (2000) screen 1</td>
<td>N 16,235</td>
<td>Autistic spectrum</td>
<td>0.579</td>
<td>H 12% (6–20)</td>
<td>99.8% (99.8–99.9)</td>
<td>28.9%</td>
<td>99%</td>
</tr>
<tr>
<td>Baird et al. (2000) screen 2</td>
<td>N 16, 235</td>
<td>Autistic spectrum</td>
<td>0.579</td>
<td>M 35% (26–46)</td>
<td>97.7% (97–98)</td>
<td>8.1%</td>
<td>99%</td>
</tr>
<tr>
<td>Robin’s et al. (2001)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>N 1293</td>
<td>Autistic spectrum</td>
<td>3</td>
<td>NA</td>
<td>98% (98–99)</td>
<td>67%</td>
<td>NA</td>
</tr>
<tr>
<td>Robins and Lee (2003)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>N 537</td>
<td>Autistic spectrum&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.8</td>
<td>100% (100–78)</td>
<td>98% (97–99)</td>
<td>62%</td>
<td>100%</td>
</tr>
</tbody>
</table>

<sup>a</sup>Childhood Autism and PDD screening are considered as Autistic Spectrum.

<sup>b</sup>Baird et al. record data for both CHAT screen 1 and CHAT screen 2. The second screen is given to all those who have failed the first screen, 1 month later.

<sup>c</sup>H indicates ‘high’ where a high-risk cut-off score was used, M indicates ‘medium’ where a medium risk cut-off score was used.

<sup>d</sup>Early follow-up calculated from data in the paper.

all subjects participating in this study so it is possible
that the false negative rate is greater than here indicated,
reducing sensitivity further. Specificity is high indicating
a low false positive rate meaning concern would rarely
be raised inappropriately and this is evidenced by the
high NPV on all outcomes. PPV is high only for the
screen 2 high/ medium risk groups for PDD and screen 2
high risk for CA. This does suggest that a screen positive
at CHAT 2 strongly indicates a future diagnosis of
PDD. Confidence intervals demonstrate a narrow range
indicating little uncertainty about the precision of the
result.

Robins et al.’s (2001) report outcome measures on the
basis of discriminant function analysis which suggests a
high sensitivity, specificity and PPV for the M-CHAT.
These results are, however, only theoretical, pertain to a
selected sample for a low prevalence disorder and
require validation. Sensitivity cannot be meaningfully
ascertained from the data due to limited follow-up of
screen negatives in the preliminary report. Specificity
and positive predictive value based on the assessed
screen positives are good, with a narrow confidence
interval. However, the sample is skewed by selection of a
high-risk group on the basis of referral. The spectrum of
disorder in this group may be different from that found
in cases in the general population as frank and severe
symptoms lead to referral for early intervention. Thus
the performance of the test may differ in the general
population.

Partial follow-up data from this cohort at age 3.5
(Robins and Lee, 2003 unpublished data) suggests a high
specificity, sensitivity, and NPV. These results suggest
that the M-CHAT is a highly accurate tool for
population screening as no cases of autism were missed.
However, there is incomplete follow-up and unlike the
Baird et al. study the ascertainment of cases is limited to
evaluation of those who fail the MCHAT at some point
during the study. Furthermore, two apparently very
different groups (selected and unselected) have been
analysed together.

No study reports on reproducibility of test results
between observers.

7. Discussion

None of the included studies meet rigorous standards
for researching the accuracy of tools used for screening.
These studies do, however, represent the best available
evidence regarding the focussed clinical question raised.
Evidence for two potential tools, the CHAT and M-
CHAT, was identified. Despite being recommended by the
American Academy of Pediatrics (2001a,b) we
found no published evidence for a further tool, the
PDDST. Although this may point to limitations of our
review methods we do not believe that this means that
we have missed any published evidence since these
authoratative guidelines do not cite any.

Neither the CHAT nor the M-CHAT screens require
significant resources or training to implement, although
consideration needs to be given to the availability of
services to evaluate those identified by the screen. The
checklists present no risk to children in themselves,
although parental anxiety is often an unwelcome side
effect of screening programmes (Hall and Elliman,
2003).

Both sensitivity and specificity are important outcome
measures in a screening tool, however a high sensitivity
is most desirable (Greenhalgh, 2001). A screening tool
needs to be sensitive in order to rule out as accurately as
possible the risk of the disorder (Sackett et al., 2000).
Specificity is potentially less important in terms of
accuracy as a further assessment is triggered and
diagnosis can be ruled out later. However, a false
positive may cause distress and further assessment on
large numbers without the diagnosis is potentially
expensive.

The M-CHAT provides some promising results in
terms of the balance between sensitivity and specificity
but again these are difficult to translate to clinical
practice due to sample size and selection, and limited
follow-up. Evidence for the CHAT from the Baird et al.
study is based on a more appropriate sample for
application to population screening and demonstrates
intensive follow-up and case ascertainment. However,
flaws in reference standard comparison, blinding and
follow-up remain and results from this study need to be
interpreted cautiously.

Both the CHAT and the M-CHAT seem to offer high
specificity. However, if these were applied to a low
prevalence population (as opposed to the skewed
samples used for the study) the absolute numbers of
false positives might still be high relative to the number
of true positives if large numbers of children were
routinely screened. Furthermore, the CHAT does not
demonstrate a desirable level of sensitivity with a
consequence that a high proportion of cases on the
Autistic spectrum would be undetected within a screened
population. Due to the follow-up methods utilised in the
study, sensitivity might be even less than here reported
as some cases may remain undetected. However, the
authors argue that a high false negative rate could be
acceptable, as autism is not a life-threatening disorder
(Baron-Cohen et al., 2000). CHAT 2 also has high
specificity, PPVs, NPVs, suggesting that although the
screen will not accurately rule out autism, a positive
result strongly suggests a diagnosis of pervasive
developmental disorder. These qualities of the CHAT screen
suggest that it may have greater utility for secondary
level screening.

Thus the best evidence suggests that the CHAT
performs poorly as a screening instrument. Although
the high NPV appears intuitively useful and a possible
means of providing reassurance the test provides little
additional discrimination than a rule based purely on
population prevalence and thus little additional reassur-
ance. Evidence for the M-CHAT is currently limited but
its characteristics seem more promising.

The M-CHAT differs from the CHAT as it is a parent
only checklist with responses clarified by phone for
those ‘failing’ the screen. The potential accuracy of
this approach reflects other work regarding parent
report screens (Glascoe, 1999; Galscoe, 2000; Daniels
et al., 2003). Parent report screening tools appear to
be a useful area for further research, and the
M-CHAT appears particularly worthy of further in-
vestigation.

Baird et al. (2000) suggests that their work with
professionals has raised awareness and skills around
recognition of the symptoms of autistic spectrum
disorders and that this awareness might increase the
likelihood of early referral and treatment. It seems
reasonable to recommend, therefore, that familiarisation
with the CHAT or M-CHAT might provide an
opportunity for primary healthcare providers working
with young children to better recognise and understand
the manifestations of autism, and that use of either tool
for secondary screening in cases of early concern might
benefit a subgroup of the population through allowing
more rapid and appropriate diagnosis. The evidence for
the CHAT is stronger in this regard. However, use of the
tools in routine childhood surveillance is not yet
warranted based on the current evidence. The M-
CHAT, as a parent report screen, appears from early
investigation to be the more promising instrument for
future research.

References

Adrien, J., Barthelemy, C., Perrot, A., Roux, S., Lenoir, P.,
Hameury, L., Sauvage, D., 1992. Validity and reliability of
the Infant Behavioural Summarised Evaluation (IBSE): a
rating scale for the assessment of young children wit autism
and developmental disorders. Journal of Autism and
Developmental Disorders 22 (1), 375–389.

American Academy of Paediatrics, 1994. Screening infants and
young children for developmental disabilities. Pediatrics 93
(5), 863–865.

American Academy of Pediatrics, 2001a. The pediatrician’s role
in the diagnosis and management of autistic spectrum

pediatrician’s role in the diagnosis and management of
autistic spectrum disorder in children. Pediatrics 107 (5),
E85.

American Psychiatric Association, 1994. Diagnostic and
Statistical Manual of Mental Disorders, fourth ed. (DSM-

Baird, G., Charman, T., Baron-Cohen, S., Cox, A., Swetten-
instrument for autism at 18 months of age: a 6 year follow-
up study. Journal of the American Academy of Child and
Adolescent Psychiatry 39 (6), 694–702.

Baird, G., Charman, T., Cox, A., Baron-Cohen, S., Swetten-
ham, J., Wheelwright, S., Drew, A., 2001. Screening and
surveillance for autism and pervasive developmental dis-
orders. Archives of Diseases in Childhood 84, 468–475.

detected at 18 months? The needle, the haystack and the
CHAT. British Journal of Psychiatry 161, 158–163.

Baron-Cohen, S., Wheelwright, S., Cox, A., Baird, G., Char-
Early identification of autism by the Checklist for Autism in
Toddlers (CHAT). Journal of The Royal Society of Medicine 93,
521–525.

Barthelemy, C., Adrien, J.L., Tanguay, P., Garreau, B.,
the behavioural summarised evaluation: validity and reliabil-
ity of a scale for the assessment of autistic behaviours.
Journal of Autism and Developmental Disorders 22 (1),
189–204.

Barthelemy, C., Adien, J., Roux, S., Garreau, B., Perrot, A.,
Lelord, G., 1992. Sensitivity and specificity of the Beha-
vioural Summarised Evaluation Scale (BSE) for the assess-
ment of autistic behaviours. Journal of Autism and develop-
mental Disorders 22 (1), 23–31.

Barthelemy, C., Roux, S., Adrien, L., Hameury, L., Guerin, P.,
Garreau, B., Fermanian, J., Lelord, G., 1997. Validation of
the Revised Behaviour Summarised Evaluation Scale.
Journal of Autism and Developmental Disorders 27 (2),
139–163.

Berument, S., Rutter, M., Lord, C., Pickles, A., Bailey, A.,

Diagnostic Observation Schedule (ADOS): first results on
reliability and validity. Zeitschrift fur Kinder-und Jugend-
psychiatrie und Psychotherapie 32 (1), 45–50.

Screening young people for autism with the Developmental
Behaviour Checklist. American Academy of Child and
Adolescent Psychiatry 41 (11), 1369–1375.

Cox, A., Klein, K., Charman, T., Baird, G., Baron-Cohen, S.,
spectrum disorders at 20 and 42 months of age: stability of
clinical and ADI-R diagnosis. Journal of Child Psychology
and Psychiatry 40, 719–732.

autism be identified in young children through parental
report? Dissertation Abstracts International 63(5-A),
1706

De Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Mulder, E.,
Volkmar, F., Minderaa, R., 2004. Interrelationship between
Autism Diagnostic Observation Schedule-Generic (ADOS-
G), Autism Diagnostic Interview-Revised (ADI-R), and the
Diagnostic and Statistical Manual of Mental Disorders
(DSM-IV-TR) classification in children and adolescents
with mental retardation. Journal of Autism & Develop-
mental Disorders 34 (2), 129–137.


Gudaitis, Lindsay, J., 2003. The early Childhood Inventory-4 and Child Symptom Inventory-4 as screening measures for pervasive developmental disorders: a validity study. Dissertation Abstracts International: Section B: the Sciences and Engineering 63 (9-B), 4411


South, M., Wiliams, B., McMahon, W., Owley, T., Filipek, P., Shernoff, E., Corsello, C., Lainhart, J., Landa, R., Ozonoff,


Further reading


