Screening for Autism Spectrum Disorders in Children below the age of 5 years

A draft report for the UK National Screening Committee

Dr Martin Allaby
Dr Mohit Sharma

4 July 2011
This report has been compiled by
Dr Martin Allaby, Consultant in Public Health Medicine
Dr Mohit Sharma, Specialist Registrar in Public Health Medicine

This work was undertaken by Solutions for Public Health (SPH) who received funding from the Department of Health via the National Screening Committee. Views expressed in this publication are those of the authors and not necessarily of the Department of Health.

Solutions for Public Health
4150 Chancellor Court
Oxford Business Park South
Oxford
OX4 2GX

Tel: +44 (0)1865 334700
www.sph.nhs.uk

July 2011
Contents

Introduction ................................................................................................................. 2
The Condition ............................................................................................................... 4
The Test ....................................................................................................................... 9
The Treatment ........................................................................................................... 14
The Programme ......................................................................................................... 20
Conclusion ................................................................................................................. 21
Key research questions on screening for ASD ......................................................... 22
References ................................................................................................................ 23
Introduction

1. This paper reviews screening for autism spectrum disorders (ASD) in children below the age of five years against the UK National Screening Committee criteria for appraising the viability, effectiveness and appropriateness of a screening programme (UK National Screening Committee 2003). The appraisal stops short of most of the criteria for appraising the programme as a whole, because gaps in the evidence regarding the test and the treatment suggest that implementation of a screening programme would be premature. This paper is based on a literature search conducted by the National Screening Committee in November 2010. Full details of the search strategy are set out in Appendix A.

2. Autism spectrum disorders (ASD) are complex developmental disorders, behaviourally defined, that include a range of possible developmental impairments in reciprocal social interaction and communication, and also a stereotyped, repetitive or limited, behavioural repertoire. Classical autism was described by Kanner in 1944 (Matson et al 2007). In 1980 the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM III) introduced the concept of ASD, which includes people with some, but not all of the features of classical autism. ASD now includes autism, Asperger’s syndrome and pervasive developmental disorders – not otherwise specified (PDD-NOS). Studies of screening and early intervention for children with ASD below the age of five years rarely include children with Asperger’s syndrome, because this is not usually diagnosed till later in childhood. Table 1 (slightly modified from Levy et al 2009) summarises the main features of these three conditions.

Table 1: main features of autism, Asperger’s syndrome and PDD-NOS

<table>
<thead>
<tr>
<th>Feature</th>
<th>Autism</th>
<th>Asperger’s syndrome</th>
<th>PDD-NOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of recognition (diagnosis)</td>
<td>yrs (3-5 yrs)</td>
<td>&gt;3 yrs (6-8 yrs)</td>
<td>Variable</td>
</tr>
<tr>
<td>Regression?</td>
<td>About 25% (social or communication)</td>
<td>No</td>
<td>Variable</td>
</tr>
<tr>
<td>Sex ratio (M:F)</td>
<td>2:1</td>
<td>4:1</td>
<td>M&gt;F (variable)</td>
</tr>
<tr>
<td>Socialisation</td>
<td>Poor</td>
<td>Poor</td>
<td>Variable</td>
</tr>
<tr>
<td>Communication</td>
<td>Delayed, deviant; might be non-verbal</td>
<td>Variable (circumscribed interests)</td>
<td>Variable</td>
</tr>
<tr>
<td>Behaviour</td>
<td>More impaired than in Asperger’s syndrome or PDD-NOS (includes stereotypy)</td>
<td>Variable</td>
<td></td>
</tr>
<tr>
<td>Intellectual disability</td>
<td>&gt;60%</td>
<td>Mild to none</td>
<td>Mild to severe</td>
</tr>
<tr>
<td>Cause</td>
<td>More likely to establish genetic or other cause than in Asperger’s syndrome or PDD-NOS</td>
<td>Variable</td>
<td></td>
</tr>
<tr>
<td>Seizures</td>
<td>25% over lifespan</td>
<td>Roughly 10%</td>
<td>Roughly 10%</td>
</tr>
<tr>
<td>Outcome</td>
<td>Poor to fair</td>
<td>Fair to good</td>
<td>Fair to good</td>
</tr>
</tbody>
</table>
Current screening policy

3. A 2006 review of screening for ASD against the NSC criteria reported that there was no screening test suitable for use in a population setting that has been fully validated, and that there was insufficient evidence regarding the effectiveness of interventions (Williams and Brayne 2006). In July 2009 the Child Health Sub-Group of the NSC reviewed the evidence on screening for autism and decided that the introduction of screening could not be recommended to the UK NSC. The current UK National Screening Committee policy is that whole population screening for autism in children should not be offered (UK National Screening Committee 2011).

4. In the USA the Centers for Disease Control and Prevention takes a different view, recommending universal screening for both developmental delays and ASD. It recommends that all children should be screened for developmental delays and disabilities during regular well-child doctor visits at ages 9 months, 18 months, and 24 or 30 months. In addition, it recommends that all children should be screened specifically for ASD during regular well-child doctor visits at ages 18 months and 24 months (Centers for Disease Control and Prevention 2011a). This endorsement of universal screening for ASD is presumably based on confidence in the effectiveness of early intervention, since the CDC webpage on treatments for ASD claims that ‘research shows that early intervention treatment services can greatly improve a child’s development’ (Centers for Disease Control and Prevention 2011b). However, the two references cited in support of this statement (Handleman and Harris 2000, National Research Council 2001) are both a decade old and therefore predate almost all the randomised controlled trials (RCTs) of early intervention treatment services for ASD. The ‘treatment’ section of this review presents the findings, and limitations, of the 14 identified RCTs of early intervention for ASD.
The Condition

The condition should be an important health problem

5. Up to one per cent of children may have ASD. Most studies of the prevalence of autism and ASD include mainly school age children, with few studies measuring prevalence in children under five. There is wide variation in the prevalence estimates for autism and ASD across individual studies, and systematic reviews have produced somewhat varying estimates of prevalence. Williams et al (2006) estimated the prevalence of autism as 7.1 per 10,000 (95% CI 1.6-30.6) and the prevalence of ASD as 20.0 per 10,000 (4.9-82.1). Fombonne (2009) estimated the prevalence of autism as 20.6 per 10,000 (1.6-30.6) and the prevalence of PDD-NOS as around 30 per 10,000. In prevalence studies conducted in the UK, Chakrabarti and Fombonne (2005) estimated the prevalence of autism among 4-6 year olds in part of the Midlands as 18.9 per 10,000 (14.1–25.0), and the prevalence of ASD as 59.8 per 10,000 (50.8-69.9). Baird et al (2006) produced somewhat higher estimates for the South Thames region, with the prevalence of autism among 9-10 year olds as 38.9 per 10,000 (29.9-47.8) and the prevalence of ASD as 116.1 per 10,000 (90.4-141.8).

6. There has been a rise in the recognised prevalence of autism and ASD over time, but it remains uncertain whether this reflects an increase in the true prevalence, or other factors. Fombonne (2009) concluded that the rise is at least partly explained by broadening of the diagnostic concept and criteria for diagnosing autism and ASD. King and Bearman (2009) concluded that diagnostic substitution (from categories such as ‘mental retardation’ to ‘autism’) accounted for a quarter of the increase in the prevalence in California from 1992 to 2005. Nassar et al (2009) concluded that the rise in the incidence of ASD in Western Australia was related to changes in diagnostic practices and service provision. Age at diagnosis has also been reducing and this may also contribute to a rising prevalence in children (Parner et al 2008, Hertz-Picciotto and Delwiche 2009, Leonard et al 2010). Increased awareness amongst parents and clinicians has also been suggested as a cause of increased assessment and diagnosis of autism (Leonard et al 2010).

7. Notwithstanding these uncertainties regarding the prevalence of autism and ASD, the cost of ASD to individuals, families and society is substantial. Knapp et al (2009) estimated the cost of supporting children with ASD in the UK as £2.7 billion per year, and the cost of supporting adults with ASD as £25 billion per year. The largest costs for children with ASD are for education; the largest costs for adults are the opportunity cost of lost employment for individuals with ASD, and the cost of accommodation for those with intellectual disability.

The epidemiology and natural history of the condition, including development from latent to declared disease, should be adequately understood and there should be a detectable risk factor, disease marker, latent period or early symptomatic stage

8. Studies of the natural history of ASD throw light on the feasibility of giving an accurate diagnosis and prognosis in early childhood. To provide the best possible information about the natural history of ‘ASD’ in children who are given this diagnosis in early childhood through a population-based screening programme, studies should ideally include all the children with ASD in a defined population, including those who have been detected through a population-based screening programme, and not just those who have been referred to a clinic (Centre for Reviews and Dissemination 2009:113). This is important because cohorts of toddlers who have been referred to a clinic may be more severely affected, and hence easier to diagnose reliably, than toddlers who have not been referred to clinic but might be detected through a population-based screening programme.
9. In these studies the assessments of the children’s condition at follow-up should also ideally be made by people who are blind to the original diagnosis. The currently accepted ‘gold standard’ for diagnosing ASD in children under the age of five years is clinical judgement (Kleinman et al 2008a), so unless the clinicians making this judgement at follow-up are blind to the original diagnosis it is impossible to rule out an unconscious bias towards endorsing the original diagnosis.

Stability of ASD diagnoses in screen-detected children

10. This review did not identify any studies that meet the criteria described in paragraphs eight and nine i.e. that include all the children with ASD in a defined population, and use assessors at follow-up who are blind to the original diagnosis. Four studies (Cox et al 1999, Sutera et al 2007, Kleinman et al 2008a, Van Daalen et al 2009) assessed the stability of diagnosis in screen-detected ASD, but in none of them were the diagnoses at follow-up made by people who were blind to the initial diagnosis. In these four studies the stability of diagnoses made around the age of two years, up to re-assessment around the age of four years, was higher for ‘autism’ (range of reported stability = 63% to 70%) than for ‘PDD-NOS’ (33% to 67%); the more inclusive category ‘ASD’ is naturally more stable (75% to 100%) than either ‘autism’ or ‘PDD-NOS’ (Table 2).
<table>
<thead>
<tr>
<th>Screen-derived cohort?</th>
<th>T2 assessors blind to diagnosis at T1?</th>
<th>Number of children</th>
<th>Median age at T1 assessment (months)</th>
<th>Median age at T2 assessment (months)</th>
<th>'Autism': proportion of T1 diagnoses of that were stable till T2</th>
<th>'PDD-NOS': proportion of T1 diagnoses of that were stable till T2</th>
<th>'ASD' (includes both autism and PDD-NOS): proportion of T1 diagnoses of that were stable till T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
<td>12</td>
<td>20</td>
<td>42</td>
<td>6/9 (67%)</td>
<td>2/3 (67%)</td>
<td>12/12 (100%)</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>73</td>
<td>27</td>
<td>53</td>
<td>-</td>
<td>-</td>
<td>60/73 (83%)</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>77</td>
<td>24</td>
<td>48</td>
<td>32/46 (70%)</td>
<td>5/15 (33%)</td>
<td>46/61 (75%)</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>53</td>
<td>23</td>
<td>42</td>
<td>25/40 (63%)</td>
<td>7/13 (54%)</td>
<td>46/53 (91%)</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>130</td>
<td>24</td>
<td>108</td>
<td>71/84 (85%)</td>
<td>14/46 (30%)</td>
<td>124/130 (95%)</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>26</td>
<td>24</td>
<td>84</td>
<td>22/26 (85%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>25</td>
<td>31</td>
<td>109</td>
<td>16/18 (89%)</td>
<td>2/7 (29%)</td>
<td>22/25 (88%)</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>48</td>
<td>29</td>
<td>53</td>
<td>20/38 (53%)</td>
<td>3/10 (30%)</td>
<td>30/48 (63%)</td>
</tr>
<tr>
<td>No</td>
<td>no</td>
<td>61</td>
<td>22</td>
<td>47</td>
<td>32/43 (74%)</td>
<td>15/18 (83%)</td>
<td>61/61 (100%)</td>
</tr>
</tbody>
</table>

Table 2: Diagnostic stability of diagnoses of autism, PDD-NOS, and ASD made at around two years of age.

Note: Median age at T1 assessment (months): 20, 27, 24, 23, 24, 24, 24, 29, 31, 22, 22, 22.
11. Some studies of the stability of diagnoses of ASD in clinic-referred children have much longer periods of follow-up than those in screen-detected children, the longest being the cohort followed to nine years of age by Lord et al (2006). This study also has the merits of a large sample size and use of follow-up assessors who were blind to the original diagnoses. With the exception of a rather low stability for diagnoses of PDD-NOS (30%), this study reported figures for stability of diagnoses given at age two years that are slightly more favourable than those found in the four studies of screen-detected ASD described above: 85% for autistic disorder, and 95% for ASD as a whole. Diagnostic change was primarily accounted for by movement from PDD-NOS to autism.

12. Across the group of studies conducted with clinic-referred children who received their initial diagnosis around the age of two years, the ranges of estimates of the stability of diagnoses overlaps with those reported for screen-detected children: 53% to 89% for ‘autism’, 29% to 83% for ‘PDD-NOS’; and 63% to 100% for ‘ASD’.

13. If one relies only on the data obtained from screen-detected cohorts, and overlooking the problem that none of them used blind assessment at follow-up, it is probably safe to conclude that about a third of children who are given a diagnosis of ‘autism’ at 20-23 months of age as a result of a screening programme, and up to a quarter of those identified as being within the broader category of ‘ASD’, are likely to lose these diagnoses by the age of four years.

14. These figures could reflect either the impact of early intervention, assuming it is effective, or over-diagnosis at age two. Whether early intervention can account for the movement of a third of two-year-olds out of the category of ‘autism’ depends on evidence from RCTs of the effectiveness of such intervention (see the section on ‘treatment’ below).

15. Risk of missed diagnoses during screening

16. Variability of prognosis within diagnostic category

17. Within diagnostic category, prognosis is very variable. Anderson et al (2009) assessed the development of adaptive social skills in 192 children who were diagnosed at age 2 years with autism, PDD-NOS or non-ASD developmental disabilities. They found that children with autism had the weakest social skills, but in all diagnostic categories improvement in social skills ranged from minimal to very dramatic. Strong expressive language skills were associated with better outcome in the autism group, and strong receptive language skills were associated with better outcome in the PDD-NOS group. The authors claimed that ‘children
with autism most at risk for problems with social adaptive abilities later in life can be identified with considerable accuracy at a very young age’, but this conclusion is premature until the performance of these predictors of outcome has been validated in an independent sample of children with ASD.

All the cost-effective primary prevention interventions should have been implemented as far as practicable

17. Opportunities for primary prevention of ASDs are constrained by limited knowledge of their causes. About 10-15% of cases of ASD are associated with known genetic causes, such as fragile X syndrome and tuberous sclerosis (Levy et al 2009), but this knowledge does not lend itself to primary prevention strategies.

If the carriers of a mutation are identified as a result of screening the natural history of people with this status should be understood, including the psychological implications.

The Test

There should be a simple, safe, precise and validated screening test and the distribution of test values in the target population should be known and a suitable cut-off level defined and agreed.

19. Table 3 summarises the findings of studies that have assessed a variety of approaches to general population screening for ASD in early childhood. Most studies have assessed a specific screening tool, but two (Tebrugge et al 2004, Barbaro et al 2010) have evaluated routine child surveillance by health professionals. Tebrugge et al (2004) used community medical files to conduct a retrospective study of children aged 9-10 years in one district. This design allowed accurate assessment of the sensitivity of surveillance for detecting ASD (64% at the 2-year check, 94% at the 3.5-year check), but the authors did not describe any data from which positive predictive value might be estimated. Barbaro et al (2010) conducted a prospective study with limited follow-up, from which the positive predictive value of screening can be estimated (81%), but not sensitivity.

The following abbreviations are used for screening tools in Table 3:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CESDD</td>
<td>Checklist for Early Signs of Developmental Disorders</td>
</tr>
<tr>
<td>CHAT</td>
<td>CHecklist for Autism in Toddlers</td>
</tr>
<tr>
<td>ESAT</td>
<td>Early Screening of Autistic Traits questionnaire</td>
</tr>
<tr>
<td>M-CHAT</td>
<td>Modified CHecklist for Autism in Toddlers</td>
</tr>
<tr>
<td>YACHT-18</td>
<td>Young Autism and other developmental disorders CHeckup Tool</td>
</tr>
</tbody>
</table>
### Table 3: Studies of screening tools or child surveillance for ASD

<table>
<thead>
<tr>
<th>Screening tool</th>
<th>Users of the tool</th>
<th>Setting</th>
<th>Number screened</th>
<th>Mean months of age (range) at screen</th>
<th>Age in months up to which false negatives could be identified</th>
<th>Number of screen-positive cases of ASD</th>
<th>Sens</th>
<th>PPV</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Routine Child Surveillance</td>
<td>Health professionals</td>
<td>UK</td>
<td>2,536</td>
<td>24 and 42</td>
<td>96-108</td>
<td>20</td>
<td>63%</td>
<td>94%</td>
<td>• Sensitivity data are for ‘concerns noted at surveillance check’. Sensitivity lower at 2yr check than at 3.5yr check.</td>
</tr>
<tr>
<td>ESAT, then psychological assessment</td>
<td>Physicians, then psychologists</td>
<td>Netherlands</td>
<td>31,724</td>
<td>15 (13-23)</td>
<td>42 (further follow-up due at age 6yrs)</td>
<td>18</td>
<td>?</td>
<td>25%</td>
<td>• Only 1.2% of children failed the initial screen with ESAT. • 31% refused psychological assessment, and another 27% refused clinical evaluation. • ASD diagnosis was dropped for 2/16 children who were available for assessment at 42 months.</td>
</tr>
<tr>
<td>M-CHAT, then phone interview for ‘failures’</td>
<td>Caregivers, then phone interview by investigator</td>
<td>USA</td>
<td>3,309</td>
<td>21 (16-30)</td>
<td>59 (48-88)</td>
<td>20</td>
<td>?</td>
<td>65%</td>
<td>• USA lacks a good surveillance system to detect missed cases of ASD</td>
</tr>
<tr>
<td>M-CHAT, then phone interview for ‘failures’</td>
<td>Caregivers, then phone interview by investigator</td>
<td>USA</td>
<td>4,265</td>
<td>19</td>
<td>59?</td>
<td>10</td>
<td>?</td>
<td>28%</td>
<td>• Update on Kleinman 2008. • 11% refused phone interview, another 37% refused clinical evaluation after interview</td>
</tr>
<tr>
<td>M-CHAT, then phone interview for ‘failures’</td>
<td>Caregivers, then phone interview by investigator</td>
<td>USA</td>
<td>1,785</td>
<td>25</td>
<td>59?</td>
<td>19</td>
<td>?</td>
<td>61%</td>
<td>• Update on Kleinman 2008. • 15% refused phone interview, another 21% refused clinical evaluation after interview.</td>
</tr>
<tr>
<td>Screening tool</td>
<td>Users of the tool</td>
<td>Setting</td>
<td>Number screened</td>
<td>Mean months of age (range) at screen</td>
<td>Age in months up to which false negatives could be identified</td>
<td>Number of screen-positive cases of ASD</td>
<td>Sens</td>
<td>PPV</td>
<td>Comments</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------</td>
<td>---------</td>
<td>----------------</td>
<td>-------------------------------------</td>
<td>-----------------------------------------------------</td>
<td>------------------------------------------</td>
<td>------</td>
<td>-----</td>
<td>----------</td>
</tr>
<tr>
<td>M-CHAT, then phone interview for ‘failures’</td>
<td>Parents, then phone interview by investigator</td>
<td>USA</td>
<td>4,797</td>
<td>24 (17-34)</td>
<td>?</td>
<td>21</td>
<td>?</td>
<td>57%</td>
<td>• 22% refused phone interview, another 39% refused clinical evaluation after interview.</td>
</tr>
<tr>
<td>Infant-Toddler Checklist</td>
<td>Families</td>
<td>USA</td>
<td>5,385</td>
<td>6-24 (repeated)</td>
<td>48?</td>
<td>56</td>
<td>≤93%</td>
<td>6%</td>
<td>• Infant-Toddler Checklist does not screen specifically for ASD, so 18% of children were screen-positive.</td>
</tr>
<tr>
<td>YACHT-18, then phone call, home visit, psychological consultation and weekly group for ‘failures’</td>
<td>Public health nurses</td>
<td>Japan</td>
<td>2,814</td>
<td>18</td>
<td>?</td>
<td>11</td>
<td>79%</td>
<td>3% for YACHT-18</td>
<td>• ‘Screening’ involves a lot of follow-up assessment for children who turn out not to have a problem.</td>
</tr>
<tr>
<td>Developmental surveillance</td>
<td>Maternal and Child Health nurses</td>
<td>Australia</td>
<td>20,770</td>
<td>8-24 (repeated)</td>
<td>24</td>
<td>89</td>
<td>?</td>
<td>81%</td>
<td>• Only 1.0% of children were considered ‘at risk’ by the developmental surveillance programme and referred.</td>
</tr>
<tr>
<td>CESDD</td>
<td>Childcare workers</td>
<td>Flanders</td>
<td>6,808</td>
<td>17 (3-39)</td>
<td>?</td>
<td>27</td>
<td>?</td>
<td>?</td>
<td>• Authors’ claims for test performance are invalid because true number of false negatives is unknown.</td>
</tr>
</tbody>
</table>

**Comments**
- 22% refused phone interview, another 39% refused clinical evaluation after interview.
- Authors are still in the process of rescreening this cohort, to determine sensitivity.
- Infant-Toddler Checklist does not screen specifically for ASD, so 18% of children were screen-positive.
- ‘Screening’ involves a lot of follow-up assessment for children who turn out not to have a problem.
- PPV of 100% was only achieved after 17 months of further assessment.
- Only 1.0% of children were considered ‘at risk’ by the developmental surveillance programme and referred.
- 49% refused assessment after being referred.
- Authors’ claims for test performance are invalid because true number of false negatives is unknown.
- Second step of screening (parent questionnaire) failed for two-thirds of screen-positive children because parents did not complete the questionnaire.
Among studies that have assessed a specific screening tool, the approach that has yielded the highest positive predictive values for ASD (around 60%) involves parents or caregivers using the Modified Checklist for Autism in Toddlers (M-CHAT), followed by a phone interview for those who fail this initial screen (Kleinman et al 2008, Pandey et al 2008, Robins et al 2008). Pandey et al (2008) found that the positive predictive value was much better when the M-CHAT was used at 25 months rather than 19 months of age (61% vs 28%). Dietz (2006) attempted screening at an even younger age (15 months) using the Early Screening of Autistic Traits (ESAT) questionnaire and found a similarly low positive predictive value at this age (25%). These positive predictive values are for confirmation of diagnosis shortly after screening; up to a quarter of children who are counted as true positives shortly after screening will lose their diagnosis of ‘ASD’ by the age of four years (see sections 10-14 above).

Few studies of specific screening tools for ASD have attempted to estimate sensitivity, because detection of missed cases requires excellent surveillance systems and several years' follow-up of the screened cohort. Such surveillance systems are not widely available in the USA, where most of the population-based screening studies have been performed. None of the papers on M-CHAT have data from which sensitivity in the general population can be estimated. The sensitivity of ESAT is also unknown, though by comparing the number of cases detected in their study with recent prevalence figures in the literature Dietz (2006) concluded that it is probably 'low'.

Approaches to screening for which authors have claimed high levels of sensitivity have used the Young Autism and other developmental disorders CHeckup Tool (YACHT-18) (Honda et al 2009) and the Infant-Toddler Checklist (Wetherby et al 2008). However, these levels of sensitivity (79-93%) were only obtained by using approaches to screening that had very low initial positive predictive values (3-6%), and children who failed the initial screen required multiple follow-up assessments over about 18 months before the outcome of screening was decided and sensitivity estimates were made.

Boyd (2010) reports that two other screening tools are undergoing testing: a revision of the original CHAT tool, called the Quantitative Checklist for Autism in Toddlers (QCHAT) designed for use in toddlers aged 18-24 months; and the First Year Inventory, which focuses on screening infants at 12 months of age.

In summary, it is possible that routine surveillance of child development by health professionals may offer the best trade-off between sensitivity and positive predictive value, though no study has reported both these measures. Among screening tools that can be used by parents or caregivers, M-CHAT seems to be the most promising, in that it offers reasonable positive predictive values (provided the screened children are aged at least two years). However, the sensitivity of M-CHAT in a general population sample has not yet been reported.

The test should be acceptable to the population

This review did not find any studies that directly assessed acceptability. However, studies of screening for ASD in the general population typically report that parents of between one third and one half of all children who fail the initial screening test drop out of the screening process before it has completed (Dietz et al 2006, Van Den Heuvel et al 2007, Kleinman et al 2008, Pandey et al 2008, Robins et al 2008, Barbaro et al 2010). Approaches to screening for ASD used in recent studies are clearly not accepted by a substantial proportion of parents.
There should be an agreed policy on the further diagnostic investigation of individuals with a positive test result and on the choices available to those individuals


If the test is for mutations the criteria used to select the subset of mutations to be covered by screening, if all possible mutations are not being tested, should be clearly set out.

27. Not relevant to screening for autism.
The Treatment

There should be an effective treatment or intervention for patients identified through early detection, with evidence of early treatment leading to better outcomes than late treatment

28. This section deals exclusively with RCTs, for the following reasons. Screening differs from routine clinical care because the process is initiated by the state or professionals, not by patients or parents. In the context of routine clinical care it is appropriate for professionals to use the best evidence available, even if it is of questionable validity, to guide their response. In the context of screening, it is not appropriate for professionals or the state to initiate contact with the public unless there is very strong evidence that available treatments are effective. RCTs are the gold standard for assessing effectiveness; the only context in which non-randomised designs can produce very strong evidence of effectiveness is when the effect of treatment is large in relation to the effects of all the possible biases, and that is not the case with treatments for ASD.

29. Hundreds of studies have attempted to assess the effectiveness of various treatments for ASD, but this review identified only 14 RCTs of interventions for children under the age of 5 years with ASD. Three were for Early Intensive Behavioural Intervention / Applied Behaviour Analysis (EIBI / ABA) (Table 4) and 11 were for focused behavioural interventions (Table 5). No RCTs were found for pharmacological interventions in children under 5 years with ASD. Most of these RCTs have reported some changes in response to early intervention. Whether such changes lead to significant improvements in adulthood, in terms of greater independence and vocational and social functioning, is unknown (Rogers and Vismara 2008).

Early Intensive Behavioural Intervention / Applied Behaviour Analysis

30. Interventions under this heading seek to address multiple core deficits in ASD, including linguistic, social, and cognitive problems (Vismara and Rogers 2010). The three RCTs of EIBI / ABA all involved intensive treatment (at least 25 hours per week) over a long period (at least two years) and periods of follow-up of at least two years. A total of 100 children with ASD have been studied in these three RCTs. The first RCT (Smith 2000) concluded that EIBI was effective, the second (Sallows and Graupner 2005) concluded that it made no difference. A systematic review that incorporated these two RCTs and nine non-randomised studies concluded that ‘overall, the quality and consistency of this body of evidence are weak. Consequently, no conclusions can be drawn from this literature about how well EIBI works’ (Blue Cross and Blue Shield Association 2009). The authors recommended that RCTs with larger sample sizes and longer follow-up should be done. A review by Spreckley and Boyd (2009) reached similar conclusions.
Table 4: RCTs of Early Intensive Behavioural Intervention / Applied Behaviour Analysis

<table>
<thead>
<tr>
<th>Reference</th>
<th>Number of children</th>
<th>Median age (range) at baseline in months</th>
<th>Case mix</th>
<th>Test intervention</th>
<th>Control intervention</th>
<th>Duration of intervention (months)</th>
<th>Length of follow up (months)</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith 2000</td>
<td>28</td>
<td>36</td>
<td>Autism, PDD-NOS</td>
<td>Intensive behavioural treatment at home for 25 hrs/wk</td>
<td>Parent training: 5 hrs/wk for 3-9 months, plus 10-15 hrs/wk in public school special education classes</td>
<td>24-36</td>
<td>57</td>
<td>Improvement in IQ, visual-spatial skills, language &amp; academics, but not adaptive functioning or behaviour problems. Those with autism improved less than those with PDD</td>
</tr>
<tr>
<td>Sallows 2005</td>
<td>24</td>
<td>34</td>
<td>Autism</td>
<td>Intensive behavioural treatment: 38 hrs/wk of direct treatment, 6-10 hrs/wk of in-home therapist supervision</td>
<td>32 hrs/wk of direct treatment, 1.5-2.5 hrs/wk of in-home therapist supervision</td>
<td>48</td>
<td>48</td>
<td>No effect on any measure (cognitive, language, adaptive, social, or academic). Control intervention may have been too similar to test intervention.</td>
</tr>
<tr>
<td>Dawson 2010</td>
<td>48</td>
<td>24</td>
<td>Autism, PDD-NOS</td>
<td>Intensive behavioural treatment (Early Start Denver Model): 15 hrs/wk of treatment by trained therapist, 16 hrs/wk of treatment by parents</td>
<td>Routine community care: 9 hrs/wk of individual therapy, 9 hrs/wk of group therapy</td>
<td>24</td>
<td>≥24</td>
<td>Improvement in IQ and adaptive behaviour. After including parental ratings, proportion with diagnosis of 'autism' improved, but ADOS severity scores and repetitive behaviour scores did not improve. Parents had invested a great deal of themselves in delivering the treatment for two years, so they might have been very keen to see a change in diagnostic severity and unconsciously biased their outcome assessments in favour of the intervention being effective.</td>
</tr>
</tbody>
</table>
31. Other reviewers have been more generous in their interpretation of the evidence base. For example, although Howlin et al. (2009) concluded that ‘there is strong evidence that EIBI is effective for some, but not all, children with ASD’, though they also acknowledged that ‘there remains a dearth of RCTs, which are needed in order to provide unbiased evidence of efficacy’.

32. It is important to point out that the claim in a 2009 Lancet review article that EIBI / ABA is ‘highly effective for up to half of children enrolled in about ten randomised clinical trials done in the past 20 years’ (Levy 2009) is incorrect. The authors of this statement (Mandell et al. 2010) claimed that a previous publication (Rogers 1998) had reviewed five randomised trials. In fact, no RCTs of these interventions had been published by 1998, and none are cited in the article by Rogers (1998). Dawson and Gernsbacher (2010) were therefore correct in writing that ‘the claims made by Levy and colleagues, with respect to intensive Applied Behaviour Analysis programmes for autistic children, have no basis’.

33. Dawson and Gernsbacher (2010) were mistaken, however, in stating that the intended comparison between randomised groups in the RCT reported by Sallows and Graupner (2005) was not done. The comparison was done, and the authors found no benefit from Applied Behavioural Analysis. Although the authors took the unusual step of combining data from the two arms of the trial for many of their analyses, data from the two arms were also analysed separately and the abstract of their paper states that ‘outcome after 4 years of treatment, including cognitive, language, adaptive, social, and academic measures, was similar for both groups’. Although the study by Sallows and Graupner (2005) is an RCT, some authors of systematic reviews of EIBI (for example Eldevik et al. 2009) have excluded it from their analyses on the grounds that children in both arms of the trial received a form of EIBI. They argue that the negative result does not therefore imply that EIBI/ABA is ineffective.

34. The third RCT of EIBI / ABA (Dawson et al. 2010), published after all the systematic reviews cited above, found that intervention produced significant improvements in IQ and adaptive behaviour. Diagnostic severity improved when parental assessments were taken into account, but not when assessment was based solely on objective rating scales. Given that the parents had invested a great deal of themselves in delivering the treatment for two years, they might have been very keen to see a change in diagnostic severity and unconsciously biased their outcome assessments in favour of the intervention being effective.

35. Focused behavioural interventions are specific teaching procedures that practitioners or parents use to promote children’s learning and development in specific areas, or to decrease challenging behaviours. Service providers select specific focused interventions to address individual objectives for children and their families (Boyd et al. 2010). These interventions are less intensive than EIBI / ABA.

36. This review identified 11 RCTs of various types of focused behavioural interventions for young children with autism or ASD. They all reported some beneficial effects, though in the trial by Yoder (2006) these had disappeared by 12 months. However, only one of these studies involved more than 60 children, and most of them followed up the children for only a year or less (Table 5). The one larger study, with 152 children, found no effect of treatment on autism symptoms (Green et al. 2010). The authors noted that larger trial sizes generally produce smaller effects (see McMahon et al. 2008), and suggested that the optimistic results from other studies should be reassessed. In the one trial with longer follow up (two years), a third of the included children had diagnoses of global developmental delay or language delay, not ASD, and the published data do not permit an assessment of whether there were significant benefits for children with ASD (Rickards et al. 2009).
<table>
<thead>
<tr>
<th>Number of children</th>
<th>Median age (range) at baseline in months</th>
<th>Case mix</th>
<th>Test intervention</th>
<th>Control intervention</th>
<th>Duration of intervention (months)</th>
<th>Length of follow up (months)</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>54</td>
<td>Autism</td>
<td>Touch therapy from a volunteer student for 15 minutes per day, 2 days per week</td>
<td>A volunteer student sat with the child on her lap and engaged the child in a game</td>
<td>1</td>
<td>1</td>
<td>Reduction in stereotypic behaviours, increase in initiative behaviours</td>
</tr>
<tr>
<td>35</td>
<td>43 (24-72)</td>
<td>Autism, PDD</td>
<td>Day care plus parent-focused intervention: 5 weekly 3-hr classes; on-site consultation 3 hrs/wk for 10 weeks; 3 case conferences</td>
<td>Day care only</td>
<td>3</td>
<td>3</td>
<td>Improvement in language age score (5.3 vs 1.1 months)</td>
</tr>
<tr>
<td>24</td>
<td>23</td>
<td>Autism</td>
<td>Train parents to develop joint attention skills and action routines; speech and language therapist visited parents at home for 3 hrs ever 6/52</td>
<td>Standard local services (but 3 children started intensive behavioural intervention)</td>
<td>12</td>
<td>12</td>
<td>Marginal improvement in words understood</td>
</tr>
<tr>
<td>28</td>
<td>48 (29-60)</td>
<td>Autism</td>
<td>Social communication intervention targeting parental communication; regular monthly therapist contact for 5 months with a further 6 months of 2-monthly consolidation sessions</td>
<td>Routine care</td>
<td>12</td>
<td>12</td>
<td>Improvement in ADOS score</td>
</tr>
<tr>
<td>36</td>
<td>33 (21-54)</td>
<td>Autism, PDD-NOS</td>
<td>Communication intervention: Picture Exchange Communication System (PECS)</td>
<td>Communication intervention: Responsive Education &amp; Prelinguistic Milieu Teaching</td>
<td>6</td>
<td>12</td>
<td>No difference between treatments at 12 months</td>
</tr>
<tr>
<td>58</td>
<td>43</td>
<td>Autism</td>
<td>Joint attention or symbolic play for 30 min/day from educational psychologists experienced in autism, plus the (substantial) control intervention</td>
<td>30 hrs/wk at an applied-behavioural-analysis-based day hospital</td>
<td>1.5</td>
<td>12</td>
<td>Improvement in expressive language (mod to large effect size). Controls were receiving other intensive therapies, so uncertain whether children would benefit from this test while receiving standard community care.</td>
</tr>
<tr>
<td>Number of children</td>
<td>Median age (range) at baseline in months</td>
<td>Case mix</td>
<td>Test intervention</td>
<td>Control intervention</td>
<td>Duration of intervention (months)</td>
<td>Length of follow up (months)</td>
<td>Results</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------------------</td>
<td>----------</td>
<td>------------------</td>
<td>----------------------</td>
<td>----------------------------------</td>
<td>-----------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>54</td>
<td>45</td>
<td>Autism, PDD-NOS, global dev. Delay, language delay</td>
<td>Home visits from specialist pre-school teacher for one hr/wk during school terms, plus control intervention</td>
<td>5 hrs/wk at multidisciplinary centre</td>
<td>12</td>
<td>24</td>
<td>Less deterioration in mean IQ, but no effect on behaviour.</td>
</tr>
<tr>
<td>152</td>
<td>45 (24-60)</td>
<td>Autism</td>
<td>Parent-mediated communication-focused intervention (PACT): 18 x 2 hr therapist-parent sessions, 0.5 hr/day practice at home</td>
<td>Routine care</td>
<td>12</td>
<td>13</td>
<td>No improvement in primary outcome (ADOS-G score). Improvement in parent-child social communication.</td>
</tr>
<tr>
<td>21</td>
<td>39 (24-47)</td>
<td>Autism</td>
<td>Reciprocal Imitation Training (RIT) for 3 hrs/wk, to teach children with autism to imitate during play</td>
<td>Routine care</td>
<td>2.5</td>
<td>2.5</td>
<td>Improvement in elicited and spontaneous imitation</td>
</tr>
<tr>
<td>38</td>
<td>31 (21-36)</td>
<td>Autism</td>
<td>Caregiver-mediated joint attention: 24 coaching sessions delivered by graduate students</td>
<td>Delayed test intervention</td>
<td>2</td>
<td>12</td>
<td>Improvement in joint attention and diversity of functional play</td>
</tr>
<tr>
<td>17</td>
<td>26 (17-36)</td>
<td>Autism</td>
<td>Teach parents to train children in eye-contact, gestures and words: 0.5 hr session x 10</td>
<td>Routine care</td>
<td>0.5</td>
<td>2</td>
<td>Improvement in language &amp; communication, reciprocal social interaction and symbolic play</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
There should be agreed evidence based policies covering which individuals should be offered treatment and the appropriate treatment to be offered

37. The Scottish Intercollegiate Guidelines Network guideline 98 (SIGN 2007) includes recommendations on clinical interventions for children and young people with ASD. Treatment is not covered in the forthcoming NICE Clinical Guideline on ASD in children and young people.

38. Not surprisingly, given the limited evidence available from RCTs, all but one of the recommendations in favour of specific treatments in the 2007 SIGN guideline are based on non-analytic studies, expert opinion or clinical experience, rather than scientific studies of effectiveness. The exception is a grade B recommendation that 'behavioural interventions should be considered to address a wide range of specific behaviours in children and young people with ASD, both to reduce symptom frequency and to increase the development of adaptive skills'.

Clinical management of the condition and patient outcomes should be optimised in all health care providers prior to participation in a screening programme

39. This review did not identify any literature that informs appraisal against this criterion.
The Programme

There should be evidence from high quality Randomised Controlled Trials that the screening programme is effective in reducing mortality or morbidity

40. This review did not identify any RCTs of screening for ASD in the general population. A Dutch general population screening study with a geographic control area (Oosterling et al 2010) found that a screening programme based on the ESAT tool reduced the mean age at diagnosis of ASD from 84 to 64 months, but the study did not assess whether there was any impact on morbidity or mortality.

41. An RCT of screening and early intervention among siblings of children with ASD is being conducted at the University of Washington and is due to complete in July 2012 (King 2009). The children will be screened from age 6 months and followed up to age 24 months, with assessment of the impact of early intervention on autism symptoms, language, communication and symbolic behaviour. Judgement will be required to assess the extent to which the results of this RCT can be generalised to whole population screening, because there may be genetic differences between ASD in single-incidence families compared with multiple-incidence families (Zwaigenbaum 2010).
Conclusion

This review has identified the following reasons for caution regarding a national screening programme for autism and autism spectrum disorder (ASD) in children aged less than five years:

1. Studies of the natural history of these conditions indicate that about a third of children who are given a diagnosis of ‘autism’ at 20-23 months of age as a result of a screening programme, and up to a quarter of those identified as being within the broader category of ‘ASD’, are likely to lose these diagnostic labels by the age of four years. It is not clear whether these figures reflect the impact of early intervention (assuming it is effective) or over-diagnosis at 20-23 months of age.

2. No approach to screening for ASD has demonstrated acceptable performance, in terms of both sensitivity and positive predictive value, in a general population screening study.

3. Approaches to screening for ASD used in recent studies are not accepted by a substantial proportion of parents. Parents of between one third and one half of all children who failed the initial screening test dropped out of the screening process before it had completed.

4. This review identified only three RCTs of Early Intensive Behavioural Intervention / Applied Behaviour Analysis, in which a total of 100 children have been studied. The claim made in a 2009 Lancet review article that EIBI/ABA is ‘highly effective for up to half of children enrolled in about ten randomised clinical trials done in the past 20 years’ (Levy 2009) is incorrect. The authors’ conclusion that ‘screening strategies for early identification could enable early treatment and improved outcomes’ therefore lacks an adequate foundation.

5. The effect of EIBI/ABA on outcomes varied across the three identified RCTs. The most consistent effect (in two RCTs) was an improvement in IQ. The duration of follow-up in the largest trial (Dawson et al 2010) was limited to two years.

6. The review identified 11 RCTs of various focused behavioural interventions, most of which reported some benefit from intervention. However, only one of these studies involved more than 60 children, and in most of them the children were followed up for only one year or less.

7. Whether the short-term effects reported in these RCTs lead to significant improvements later in childhood, or greater independence and improved vocational and social functioning in adulthood, is unknown.
Key research questions on screening for ASD

1. Can any approach to screening for ASD demonstrate acceptable performance, in terms of both sensitivity and positive predictive value, in a general population based study?

2. Why do so many parents of children who fail initial screening tests for ASD drop out of the screening process before it has completed, and can the process be refined so that the drop-out rate is reduced?

3. Does early intervention lead to significant improvements later in childhood, or greater independence and improved vocational and social functioning in adulthood?
References


Mandell DS, Levy S, Schultz RT. Authors reply Lancet 2010;375:723


