Comparison of three screening tests for autism in preterm children with birth weights less than 1,500 grams

Background: Preterm children seem to be at increased risk for autism spectrum disorders (ASD).

Methods: Parents of 157 children with birth weights less than 1,500 g (age 2 years, corrected for prematurity; 88 boys, 69 girls) completed screening questionnaires. The screening battery included the Modified Checklist for Autism in Toddlers (M-CHAT), Communication and Symbolic Behavior Scales Developmental Profile Infant-Toddler Checklist (CSBS-DP-ITC), and the Infant/Toddler Sensory Profile (ITSP). Children with disabilities were excluded. All children who screened positive on any of the screening tools were subsequently assessed by clinical examination including the Autism Diagnostic Observation Schedule.

Results: Fifty-six children (35.7%) screened positive on at least one of the parental screening questionnaires. Of the 56 children who tested positive, 33 participated in the detailed clinical follow-up assessment. A diagnosis of ASD was confirmed in 13 of the 33 children. The ASD prevalence was 9.7% of the sample. Analysis of children with and without an ASD diagnosis found significant differences relative to gestational age (26.9 weeks vs 28.3 weeks, P=0.033) and length of the stay in hospital (89.5 days vs 75.4 days, P=0.042). The screening tool with the most positive results was CSBS-DP-ITC (42 positive screens [PS]), followed by M-CHAT (28 PS), and ITSP (22 PS). Differences in the frequency of PS among the tests were significant (P=0.008). CSBS-DP-ITC had the highest sensitivity (0.846), followed by M-CHAT (0.692) and ITSP (0.462).

Conclusion: Our results indicate a higher prevalence of autism in children with birth weights <1,500 g at 2 years of age compared to the general population prevalence. The ASD diagnosis was associated with shorter gestation times and longer hospital stays. Our findings support the simultaneous use of more than one screening tests in order to increase screening sensitivity.

Keywords: autism spectrum disorders, preterm children, screening, Modified Checklist for Autism in Toddlers, Communication and Symbolic Behavior Scales Developmental Profile Infant-Toddler Checklist, Infant/Toddler Sensory Profile

Introduction

There is a growing body of evidence that preterm children with very low birth weight (VLBW; 1,000–1,500 g) or extremely low birth weight (ELBW; under 1,000 g) are at increased risk for autism spectrum disorders (ASD). Recent studies on this topic, in which screening results were validated with clinical examinations and/or diagnostic instruments, have described the prevalence of ASD among preterm births to be in the range 3.65%–12.9%,1–4 whereas the ASD prevalence in the general pediatric population has been found to be in the range 1%–1.5%.5,6
The autistic phenotype seen in preterm children is thought to represent a milder form of the disorder than that seen in full-term children. Moreover, extremely preterm children have greater symptoms on the dimension of impaired social interaction and communication than on the dimension of repetitive or stereotyped behavior, the latter of which is a core symptom domain in diagnostic classifications. Studies performed on school-age children who were born preterm have revealed a markedly lower prevalence of ASD than would have been expected from the positive screen rate during infancy. The change in the expected prevalence may suggest that this population has the potential to “recover from autism”.

ASD have generally been regarded as life-long conditions. However, in recent years a significant minority of children with well-documented ASD have achieved recovery from the disorder (terms “best outcome” and “optimal outcome” have also been used). In a comprehensive review, Helt et al found that 3%–25% of children reportedly “lost” their ASD diagnosis and attained a normal range of cognitive, adaptive, and social skills. However, we do not yet have any specific data on recovery from autism in preterm children.

The differences between preterm ASD and genuine ASD in the autism phenotype and/or prognosis may arise from a different causal pathway, ie, one that is nongenetic and stems from brain injuries and altered neurodevelopment associated with a very premature birth. Some authors have found a direct relationship between shorter gestation times and increased risk of ASD. Losh et al in a same-sex twin study estimated that every 100 g increase in birth weight provided a 13% reduction in the risk of ASD. Kuzniewicz et al found that ASD was approximately three times more prevalent in infants born before 27 weeks of gestation compared with term infants and that each 1-week reduction in gestation was associated with further increases in the risk of ASD. Leavey et al also observed a gradual increase in ASD risk linked to shorter gestation times, which was especially apparent for cutoffs between 29 weeks and 37 weeks. The results were not affected by sex or measures of fetal growth. Buchmayer et al observed a different relationship. They found that the increased risk of ASD related to preterm birth was mediated not by the gestation weeks but primarily by prenatal and neonatal complications, which occur more commonly among preterm infants. However, the largest study on the topic, performed by Mackay et al, did not identify any significant association between ASD risk and obstetric factors.

Other authors have emphasized the role of fetal growth on the risk of ASD. Moore et al reported that autism risk was increased in preterm small-for-gestational age (SGA; <5th percentile) infants 23–33 weeks and term large-for-gestational age (LGA; >95th percentile) 39–41 weeks, but decreased in preterm LGA infants 23–31 weeks. In a study by Abel et al ASD risk was seen to increase with fetal growth 1.50 standard deviations below and ≥2.00 standard deviations above the mean gestational age.

There were three aims of our study: 1) to estimate ASD prevalence in preterm children aged 2 years (corrected for prematurity); 2) to identify potential association of ASD diagnosis in preterm children with specific demographic and medical factors, such as parental age at delivery and education, child’s birth weight and length, gestational age at delivery, length of the stay in hospital, and the use of corticoids; and 3) to compare the efficacy of three different screening tests in this population. The preliminary results have already been published, and now we present final data from our study.

**Methods**

**Sample**

Children with birth weights less than 1,500 g were consecutively recruited from March 2012 to June 2014. Three centers for “Newborns and Infants at Risk” participated in the study: the Department of Pediatrics, University Hospital Motol, Prague; the Department of Pediatrics and Adolescent Medicine, General University Hospital, Prague; and the Department of Pediatrics, University Hospital, Hradec Kralove. Families were informed about the research project during routinely scheduled checkups for 2-year-old children (age corrected for prematurity). The 2-year (age-corrected) checkup is usually the last examination and assessment of preterm children in specialized centers for “Newborns and Infants at Risk” and further care is then decentralized to pediatricians along with the switch from corrected age to chronological age. Children with substantial disabilities, such as cerebral palsy or major vision and/or hearing impairments, were excluded.

Parents of 247 children with birth weights less than 1,500 g agreed to participate in the study and signed informed consents. Of these, families of 157 children (63.6%) completed the screening questionnaires and returned them to the Department of Child Psychiatry. The sample consisted of 88 boys and 69 girls, aged 2 years (corrected for prematurity).

**ASD screening tools**

The screening battery included the Modified Checklist for Autism in Toddlers (M-CHAT), the Communication and
Symbolic Behavior Scales Developmental Profile Infant-Toddler Checklist (CSBS-DP-ITC), and the Infant/Toddler Sensory Profile (ITSP). General descriptions and detailed information of the psychometric properties of these tests are given in our previous report.

In those with positive M-CHAT screens, follow-up interviews, consisting of additional corresponding questions, are recommended. In our study, we used a follow-up clinical evaluation instead of the recommended telephone interview; however, we think this substitution improved the accuracy of our results.

ITSP was not originally considered a screening tool, and this study is probably the first to evaluate its use as a screening tool. Although our initial results did not seem to overly promising, we decided to retain ITSP in the battery in order to make a final analysis.

Procedure
The study was approved by the Ethics Committees of all three participating hospitals. Parents of VLBW and ELBW children (2 years of age, corrected for prematurity) who agreed to participate in the study signed informed consents and received test materials with written instructions on how to fill out the questionnaires. No recommendations regarding the order of test completion were given. Parents completed the screening battery of questionnaires at home and returned them by regular mail to the Department of Child Psychiatry, Motol University Hospital.

All children who had screened positive on any of the screening tools were subsequently invited for a detailed follow-up assessment. The assessment involved testing using the Autism Diagnostic Observation Schedule (ADOS) and a clinical examination by two experienced child psychiatrists with expertise in autism. The concept of best estimate clinical diagnosis (BECD), by consensus of two experienced specialists, was used as the gold standard. In cases of disagreement between the ADOS diagnosis and BECD, the latter was preferred. The International Classification of Diseases, Tenth Edition (ICD-10), was used for clinical diagnoses.

Data analysis
Statistical analysis was performed using the Statistical Package for the Social Sciences (IBM SPSS, version 22.0) and statistical software R with the RVAideMemoire and pROC libraries. Descriptive statistics for the samples were used. The Mann–Whitney U-test and \( \chi^2 \)-test were used for analyzing the differences between subgroups with and without an ASD diagnosis. Cochran’s \( Q \)-test was used for analyzing differences in positivity among tests. More detailed pair comparisons between tests were performed using the paired Wilcoxon sign tests, with the false discovery rate adjusted for multiple testing. Receiver operating characteristic (ROC) analysis was used in the evaluation of the three screening tests.

Results
ASD prevalence in the sample
Fifty-six children (35.7%) screened positive on at least one of the screening questionnaires. Parents of 33 of the 56 children agreed to participate in the follow-up assessment. A diagnosis of ASD was indicated in 15 children based on the ADOS, and ASD was ultimately confirmed in 13 of the 33 children. Diagnoses, based on the ICD-10, were childhood autism (n=7) and atypical autism (n=6). In two cases (2 of 15), there was a disagreement between the ADOS diagnosis and the BECD. In the first case, ADOS indicated ASD but the BECD was that of mild mental retardation. In second case, ADOS indicated ASD; however, no psychiatric diagnosis (BECD) was established. ASD prevalence calculated from those examined (33 children) and those with negative screening results (101 children) was estimated to be as high as 9.7% of the sample.

Demographic and medical factors
Table 1 shows the family and child characteristics for the total sample and for the subgroups. Based on the final diagnosis, two subgroups were formed. The subgroup with an ASD diagnosis (N=13) consisted of children who screened positive and when subsequently examined (ADOS and BECD) fell into the category of ASD. The other subgroup was without an ASD diagnosis (N=121) and consisted of 101 children who initially screened negative on all three screening tests as well as 20 children who initially screened positive; however, after subsequent examination, BECD failed to confirm a diagnosis of ASD. Children who were positive on at least one screening test but whose parents did not agree with the clinical examination (N=23) were not included in the subgroup analysis.

A comparison of children with and without an ASD diagnosis showed significant differences relative to gestational age (\( P=0.033 \)) and length of the stay in hospital (\( P=0.042 \)) between the two subgroups (Table 1). We found no significant differences relative to maternal and paternal age at delivery, level of maternal and paternal education, mean birth weight and length, or the use of corticoids.
Table 1 Family and child characteristics of the preterm population (age 2 years)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean (SD) or frequency (%)</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (N=157)</td>
<td>With ASD diagnosis (N=13)</td>
</tr>
<tr>
<td>Maternal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at delivery (years)</td>
<td>32.6 (4.4)</td>
<td>32.3 (3.2)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary</td>
<td>2 (1.3%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>High school</td>
<td>89 (56.7%)</td>
<td>6 (46.2%)</td>
</tr>
<tr>
<td>University</td>
<td>65 (41.4%)</td>
<td>7 (53.8%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (0.6%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Paternal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at delivery (years)</td>
<td>34.8 (4.9)</td>
<td>33.5 (2.9)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary</td>
<td>1 (0.6%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>High school</td>
<td>90 (57.3%)</td>
<td>9 (69.2%)</td>
</tr>
<tr>
<td>University</td>
<td>61 (38.9%)</td>
<td>4 (30.8%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>5 (3.2%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Child</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth weight (g)</td>
<td>1.034 (283.4)</td>
<td>964.5 (300.4)</td>
</tr>
<tr>
<td>Birth length (cm)</td>
<td>35.6 (3.6)</td>
<td>35.6 (4.1)</td>
</tr>
<tr>
<td>Gestational age (weeks)</td>
<td>28.2 (2.6)</td>
<td>26.9 (2.0)</td>
</tr>
<tr>
<td>Stay in hospital (days)</td>
<td>75.8 (29.7)</td>
<td>89.5 (23.9)</td>
</tr>
<tr>
<td>Use of corticoids</td>
<td>39 (24.8%)</td>
<td>5 (38.5%)</td>
</tr>
</tbody>
</table>

Abbreviations: SD, standard deviation; ASD, autism spectrum disorders; N, number of children.

Screening battery

The screening tool with the most positive results was CSBS-DP-ITC (42 positive screens). M-CHAT had 28 positive screens and ITSP had 22 positive screens. For details, see Table 2. The difference in positive results among the screening tests was significant (P=0.008). In pair comparisons, using the paired Wilcoxon sign tests with the false discovery rate adjusted for multiple testing, CSBS-DP-ITC was found to be significantly more positive than both ITSP (P=0.022) and M-CHAT (P=0.022). The difference between M-CHAT and ITSP was not significant.

ROC analysis was calculated from a sample of 134 children which consisted of the subgroup with an ASD diagnosis (N=13) and the subgroup without an ASD diagnosis (N=121); for a more detailed description of both subgroups see the “Demographic and medical factors” section. Table 3 summarizes the results of the ROC analyses of individual tests in relation to the ASD diagnosis. CSBS-DP-ITC turned out to be the most sensitive tool; similarly, M-CHAT and ITSP also showed high levels of specificity and accuracy. Additionally, the negative predictive value was high for all three screening questionnaires; however, the positive predictive value (PPV) tended to be relatively low for all three tests. The second half of Table 3 displays the psychometric values derived from the combined use of the three screening tests. The first line in Table 3 shows results for the combined use, which is how it was used in our study (ie, if one of the tests was positive, then the screen was considered positive). The next two lines demonstrate how psychometric values would change relative to changes in the “screen positive” criteria (ie, if 2 of 3, or 3 of 3 tests must be positive for the screen to be positive).

Discussion

Our study estimated the prevalence of ASD to be as high as 9.7%, a number that was lower than in our preliminary data (12.9%). The ASD prevalence might have been slightly higher if all the children with positive screens had undergone the clinical examination. Unfortunately, parents of a large
proportion of children (41.1%) who screened positive did not bring their children in for the follow-up examination. Our estimated ASD prevalence in preterms was several times higher than in general population, which was in agreement with previous findings from studies that validated the screening results with clinical examinations.1–3 Moreover, our sample of VLBW and ELBW children allowed us to establish an increased frequency of ASD at the age of 2 years, whereas the studies cited above examined either school-age children1,2 or adolescents and those in early adulthood.3

Some may argue that certain developmental trajectories in preterm children are at variance with those in term children. Children with a “preterm behavioral phenotype” may present with symptoms that resemble, overlap, or even meet the criteria for ASD diagnosis. Thus, more prospective studies with comparison groups of term children are needed in order to identify precisely the optimal ASD screening age for preterm children.13 We think that early identification of autism in preterms (eg, at 2 years of corrected age, as it was in our study) and prospective follow-up of this group will provide valuable information on the stability of the diagnosis with time and on possible rates of “recovery from autism” in this specific group. Such a study is one of our long-term research aims.

Many previous studies have examined the association between gestational age and ASD, and some of them observed an increased association with autism in preterm children as well as post-term children. Studies by Kuzniewicz et al16 and Leavey et al17 found a higher prevalence of ASD associated with shorter gestational periods. In accordance with these results, we also observed an association between shorter gestational age and an increased occurrence of ASD. An association between birth weight and ASD frequency has also been reported in a large population-based study.15 Our study also found a difference in birth weight between children with and without ASD; however, it was not statistically significant.

In a study by Buchmayer et al18 the increased risk of autistic disorders related to preterm birth was mediated primarily by prenatal and neonatal complications, which occurred more commonly among preterm infants. Kuzniewicz et al16 retrospectively evaluated changes in ASD risks relative to changes in gestational age. Additionally, the study sought to correlate specific gestational age–related neonatal risk factors (ie, conditions and interventions) with changes in the prevalence of ASD. The study found that high-frequency ventilation and intracranial hemorrhage were associated with ASD in infants with gestation ages <34 weeks. Our study found an association between ASD and length of hospital stay, which indicated an association between ASD and neonatal risk factors. In a population-based cohort study by Moore et al20 the authors found many factors associated with ASD, eg, male sex, advanced maternal and paternal age, and twin gestation. In our study, we found no significant differences relative to maternal and paternal age at delivery or level of maternal and paternal education; however, this could easily be a reflection of a small sample size.

Our study may be the first to compare the efficacy of three well-established tests for use with preterms (M-CHAT and CSBS-DP-ITC have been well established in screening; the use of ITSP for screening was experimental but the test itself has been reliably studied). M-CHAT is the most frequently used parental screening test for autism. Authors of the test have demonstrated its high sensitivity (0.87) and specificity (0.99).22 Kleinman et al31 found a PPV of 0.36 for the initial screening which improved to 0.74 when combined with a follow-up interview. We used the screening questionnaire without a follow-up interview and found the sensitivity lower (0.69) than in Robins et al’s study,22 although the specificity was almost as high (0.93). The PPV, which tends to be low when the screening questionnaire is used without a follow-up

### Table 3 Psychometric values of three screening test for autism based on ROC analysis

<table>
<thead>
<tr>
<th>Single test use</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>PPV</th>
<th>NPV</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-CHAT (N=134)</td>
<td>0.926</td>
<td>0.692</td>
<td>0.500</td>
<td>0.966</td>
<td>0.903</td>
</tr>
<tr>
<td>CSBS-DP-ITC (N=132)</td>
<td>0.849</td>
<td>0.846</td>
<td>0.379</td>
<td>0.981</td>
<td>0.848</td>
</tr>
<tr>
<td>ITSP (N=133)</td>
<td>0.942</td>
<td>0.462</td>
<td>0.462</td>
<td>0.942</td>
<td>0.902</td>
</tr>
</tbody>
</table>

Combined use of the tests with requested positivity in at least

- One of the tests (N=131) 0.797 1.000 0.351 1.000 0.817
- Two of the tests (N=131) 0.924 0.769 0.526 0.973 0.908
- All three tests (N=131) 0.992 0.231 0.750 0.921 0.916

Notes: Due to missing data, ROC calculations for some tests did not include the full number of 134 probands.

Abbreviations: ROC, receiver operating characteristic; M-CHAT, Modified Checklist for Autism in Toddlers; CSBS-DP-ITC, Communication and Symbolic Behavior Scales Developmental Profile Infant-Toddler Checklist; ITSP, Infant/Toddler Sensory Profile; PPV, positive predictive value; NPV, negative predictive value; N, number of probands whose data entered the ROC analysis.
were comparable to those of M-CHAT (Table 3). As with the high (0.90), and so was its specificity (0.94); both parameters too strict. Test accuracy, with regard to ASD diagnosis, was positive responses in comparison to the other two questionnaires, it seems that the established criteria might have been positive screening. The criterion was that participants were considered to have screened positive if the results were deficient in ASD diagnosis was low in our study (0.38) compared to the PPV presented in Pierce et al’s report which used CSBS-DP-ITC for the detection of a wide range of disorders connected to global developmental impairment (PPV =0.75).

ITSP is a norm-referenced questionnaire, in which the caregiver rates the frequency at which the individual being engaged in the described response to sensory input. In extensive meta-analysis of sensory modulation symptoms in persons with ASD, different versions of the sensory profile, including ITSP, were found to be the most frequently used methods for measuring sensory processing. Unlike M-CHAT and CSBS-DP-ITC, ITSP is not generally used for screening; however, early on we concluded that ITSP might be a useful screening tool because autistic children often display unusual behavioral responses to sensory input, such as hypersensitivity to auditory, visual, tactile, and olfactory stimuli. Our study may be the first to evaluate its use as a screening tool. We also established a new criterion for a positive screening. The criterion was that participants were considered to have screened positive if the results were definitely abnormal (ie, results outside two standard deviations of population norms) on at least two scores involving section and/or quadrant scores. ITSP had the lowest number of positive responses (14.3%). With regard to the lowest number of positive responses in comparison to the other two questionnaires, it seems that the established criteria might have been too strict. Test accuracy, with regard to ASD diagnosis, was high (0.90), and so was its specificity (0.94); both parameters were comparable to those of M-CHAT (Table 3). As with the other two tests, the negative predictive value was high (0.94) and the PPV was low (0.46). However, of the three tests, ITSP turned out to be the least sensitive (0.46) and this handicaps its use for general screening.

To conclude this part of the discussion, M-CHAT and CSBS-DP-ITC showed good psychometric properties for screening a preterm population. CSBS-DP-ITC had the highest sensitivity of the three tests we used, which could be outweighed by faster and easier administration of M-CHAT. Therefore, the selection of an appropriate screening test for clinical purposes should be based not only on psychometric properties of the tests but also on specific aims and broader circumstances of each screening.

To our knowledge, our results can only be compared to the 2012 study by Stephens et al who also used a screening battery instead of a single tool. They found a markedly smaller percentage of positive screens, with only 20% of infants (excluding those with disabilities) having at least one positive screen, whereas in our study it was 35.7% of children. Stephens et al reported that only 1% of the sample screened positive on all three screens, whereas our study found that 9% of participants were positive on all three screens. It must be noted that Stephens et al used less common screening instruments such as the Pervasive Developmental Disorder Screening Test, Second Edition Stage 2, and two parts from the ADOS (ie, response to name and response to joint attention). Despite using different screening tools, they reached the same conclusion as we did: that is, simultaneous use of more than one screening test results in a higher number of positive screens, which means better screening sensitivity.

The most noteworthy limitation of our study was the large number of children with positive screens who did not undergo clinical evaluation because of their parent’s lack of interest (23 from 56 or 41.1%). Additionally, we did not clinically examine the children who had negative screening results; therefore we were not able to make a valid differentiation between true and false negatives in our ROC analyses. However, we can model this differentiation using an analysis of the combined use of the three tests, because there were children who were negative on one screening test and positive on one or both of the other screening test(s) (Table 3). The assessment of the combined tests are done using three steps, each using a different criterion for a “positive screen” (positive on at least one test, on at least two tests, on all three tests), shows us, despite the first step where the sensitivity and NPV are unrealistically high (this is direct logical
consequence of the lack of false negatives in our sample), that there is a potential for substantial improvement of ROC values (sensitivity, specificity, PPV, and NPV) when using a combination of screening tests. However, taking into account the second limitation, the psychometric values given in Table 3 are much stronger for mutual comparisons among the listed screening tests than for use as independent values.

Another issue, which was stated in the publication of our preliminary results, was that we relied solely on the clinical judgment of pediatricians and child neurologists regarding disabilities. This reliance could have affected the number of children excluded for major disabilities. Finally, there was no structured examination of motor and cognitive functions in our sample of children, which can be analyzed and presented.

Conclusion
Our results indicate a higher prevalence of autism in children with birth weights <1,500 g at 2 years of age compared to the prevalence in the general population. We found that the ASD diagnosis was associated with shorter gestational age and longer hospital stays. Our findings support the simultaneous use of more than one screening test in order to increase screening sensitivity.

Acknowledgments
This work was supported by the Ministry of Education, Youth and Sports, Czech Republic (research grant COST LD11028), by the Ministry of Health, Czech Republic (research grant IGA NT/14200 and conceptual development of research organization, University Hospital Motol, Prague, Czech Republic 00064203), and by the ESF (COST Action ESSEA BM1004). The authors would like to thank Thomas Secret for his assistance with the English version of the manuscript.

Disclosure
The authors report no conflicts of interest in this work.

References


