



Published in final edited form as:

Autism. 2019 January ; 23(1): 29–38. doi:10.1177/1362361317718482.

Autism spectrum disorder screening with the CBCL/1½–5: Findings for young children at high risk for autism spectrum disorder

Leslie A Rescorla¹, Breanna M Winder-Patel², Sarah J Paterson³, Juhi Pandey³, Jason J Wolff⁴, Robert T Schultz³, Joseph Piven⁵

¹Bryn Mawr College, USA

²University of California, Davis, USA

³University of Pennsylvania, USA

⁴University of Minnesota, USA

⁵The University of North Carolina at Chapel Hill, USA

Abstract

The screening power of the CBCL/1½–5's Withdrawn and *Diagnostic and Statistical Manual of Mental Disorders*-Pervasive Developmental Problems (DSM-PDP) scales to identify children diagnosed with autism spectrum disorder at 24 months was tested in a longitudinal, familial high-risk study. Participants were 56 children at high risk for autism spectrum disorder due to an affected older sibling (high-risk group) and 26 low-risk children with a typically developing older sibling (low-risk group). At 24 months, 13 of the 56 high-risk children were diagnosed with autism spectrum disorder, whereas the other 43 were not. The high-risk children diagnosed with autism spectrum disorder group had significantly higher scores on the CBCL/1½–5's *Diagnostic and Statistical Manual of Mental Disorders*-Pervasive Developmental Problems and Withdrawn scales than children in the low-risk and high-risk children not diagnosed with autism spectrum disorder groups ($\eta_p^2 > 0.50$). Receiver operating characteristic analyses yielded very high area under the curve values (0.91 and 0.89), and a cut point of $T \geq 60$ yielded sensitivity of 77% and specificity of 97% to 99% between the high-risk children diagnosed with autism spectrum disorder and the combination of low-risk and high-risk children not diagnosed with autism spectrum disorder. Consistent with several previous studies, the CBCL/1½–5's *Diagnostic and Statistical Manual of Mental Disorders*-Pervasive Developmental Problems scale and the Withdrawn syndrome differentiated well between children diagnosed with autism spectrum disorder and those not diagnosed.

Keywords

autism spectrum disorder screening; baby sibling paradigm; CBCL/1½–5; *Diagnostic and Statistical Manual of Mental Disorders*-Pervasive Developmental Problems scale; familial high-risk

The diagnostic assessment of autism spectrum disorder (ASD) can be a lengthy, involved, and costly process because it often involves administration of the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) (Lord and Rutter, 2012) and the Autism Diagnostic Interview-Revised (ADI-R) (Rutter et al., 2003), integrated with expert clinical opinion. For this reason, in-depth assessment for ASD is most practical and cost-effective when previous screening with a less expensive assessment tool suggests a high likelihood of ASD. Such screening tools typically take the form of parent-report checklists, which can be completed and scored quickly at minimal expense.

As noted by Barton et al. (2012) and Norris and Lecavalier (2010), Level 1 screening is designed to identify potential cases of ASD in unselected populations, whereas Level 2 screening aims to differentiate children with ASD from those with other disorders and/or developmental disabilities. The priority in Level 1 screening is to use a quick, low-cost instrument that will not miss children who may have ASD (i.e. few false negatives) even if some children are false positives (i.e. they turn out to have some other developmental disability or behavioral/emotional problem rather than ASD). In contrast, cost and administration time may be less important considerations in Level 2 screening, given that the next step for flagged cases may be costly evaluation with the ADOS-2.

Most of the instruments used for either Level 1 or Level 2 screening of ASD are ASD specific, meaning that most or all of their items describe ASD symptoms rather than symptoms of psychopathology more broadly. Norris and Lecavalier (2010) reported screening accuracy results for five such instruments, including the Social Communication Questionnaire (SCQ) (Berument et al., 1999) and the Gilliam Autism Rating Scale, Second Edition (GARS-2) (Gilliam (2006). The authors concluded that the SCQ performed the best of the five scales, although most SCQ studies showed a commonly found trade-off between sensitivity (SENS) and specificity (SPEC), namely, that they incur many false positives to obtain few false negatives, or vice versa.

Barton et al. (2012) summarized results of several studies using the Checklist for Autism in Toddlers (CHAT) (Baron-Cohen et al., 1992) and the Modified Checklist for Autism in Toddlers (M-CHAT) (Robins et al., 2001) to screen for ASD. They noted that the SENS of the initial CHAT was only 0.38 (Baird et al., 2000). Barton et al. also reported that the high false-positive rate of the initial 23-item M-CHAT prompted the authors to add a 15-min follow-up interview with the parents of children flagged by the screening tool (Robins et al., 2009). However, even with the interview, the M-CHAT had a positive predictive value (PPV) of only 57%.

Despite mixed results for the M-CHAT, the American Academy of Pediatrics recommends it for early ASD screening and hence it has been widely studied in large pediatric samples. For

example, Chlebowski et al. (2013) reported findings for 18,989 pediatric patients between 16 and 30 months of age scored on the M-CHAT and the follow-up interview. PPV was only 54%, but 98% of the screen positive cases manifested developmental issues warranting intervention. With another large pediatric sample ($N = 16,071$) screened at 18 or 24 months, Robins et al. (2014) reported that a slightly revised M-CHAT plus follow-up interview (M-CHAT-R/F) with cutoffs ≥ 3 on the M-CHAT-R and ≥ 2 on the follow-up interview yielded SENS of 85% and PPV of 48% for ASD. However, 95% of the screened positive cases manifested significant developmental delay (DD) or concern, parallel to Chlebowski et al. (2013). SPEC was very high (99%), probably because only about 7% of the sample screened positive on the M-CHAT-R, with >15,000 children screening negative.

Although ASD screening for young children (i.e. up to the age of 6 years) has mostly been studied using ASD-specific instruments such as the SCQ or M-CHAT, some studies have tested the ability of a broadband assessment instrument to screen for ASD in this age group. The two most commonly used instruments of this type are the Child Behavior Checklist for Ages 1½–5 (CBCL/1½–5) (Achenbach and Rescorla, 2000) and the Behavior Assessment System for Children, Second Edition (BASC-2) (Reynolds and Kamphaus, 2004). Both the CBCL/1½–5 and the BASC-2 contain items reflecting many kinds of behavioral and emotional problems. Ratings on these items yield standardized scores on several narrowband scales (e.g. Attention Problems), as well as on a few broadband scales (e.g. Total Problems). On the CBCL/1½–5, two scales have been most commonly used to screen for ASD. One is the empirically based Withdrawn syndrome, derived via factor analysis, and the other is the *Diagnostic and Statistical Manual of Mental Disorders*-Pervasive Developmental Problems (DSM-PDP) scale, constructed based on expert judgment regarding correspondence with *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; DSM-IV; American Psychiatric Association, 1994) criteria. Several BASC-2 scales have been identified as possible ASD screeners in young children, including Social Skills, Functional Communication, Atypicality, and most recently, the Developmental Social Disorders (DSD) scale, constructed by drawing items considered to reflect ASD from numerous other BASC-2 scales (Bradstreet et al., 2017).

Myers et al. (2014) studied a sample of 156 clinically referred young children, 70 (45%) of whom were diagnosed with ASD. The CBCL/1½–5 was completed for the full sample, with 82 parents also completing the BASC-2 and 74 also completing the Clinical Assessment of Behavior (CAB) (Bracken and Keith, 2004). ASD was diagnosed based on a clinical interview conducted by one psychologist, with a follow-up ADI-R administered if ASD was suspected. The non-ASD group of 86 children had communicative or cognitive delays and/or behavioral concerns, but all received a diagnosis of “Developmentally Delayed.” On the CBCL/1½–5, children with ASD had significantly higher scores on the Withdrawn and DSM-PDP scales ($d = 0.99$ for Withdrawn and 0.81 for DSM-PDP). On the BASC-2, the ASD group scored significantly lower than the non-ASD group on the Social Skills and Functional Communication scales (d s of 0.88 and 0.64 , respectively), but their scores were not significantly higher on the Atypicality scale, considered to tap ASD-like behaviors. No significant group differences emerged on the CAB, including on the Autism Spectrum Behaviors scale. On both the CBCL/1½–5 and the BASC-2, SENS and negative predictive value (NPV) were high (SENS = 89%–93%, NPV = 81%–85%), whereas SPEC and PPV

were much lower (SPEC = 28%–52%, PPV = 50%–60%). This pattern of decision statistics is common when the non-ASD group has other developmental disabilities and the base rate of ASD in the sample is very high, as was the case in the Myers et al.'s (2014) study.

Bradstreet et al. (2017) used the BASC-2 as part of a diagnostic evaluation with 224 young children who had all screened positive on the M-CHAT or the M-CHAT-R. Of this sample, 117 received a diagnosis of ASD, 55 received a non-ASD diagnosis, and 52 had typical development (TD) ($n = 14$) or mild DD ($n = 38$) (No Diagnosis group). When the ASD and No Diagnosis groups were compared, the DSD scale yielded SENS = 76%, SPEC = 73%, PPV = 86%, and NPV = 57%. When the ASD and non-ASD Diagnosis groups were compared, all four decision statistics were lower (SENS = 62%, SPEC = 63%, PPV = 78%, and NPV = 44%). PPV values were high in both comparisons because the ASD group was more than twice the size of the contrast group. It is also worth noting that the PPV for the M-CHAT/M-CHAT-R appears to have been only 52% (117/224) in this study.

The CBCL/1½–5, which takes about 15 min for a parent to complete and <5 min to score, is not a diagnostic instrument but has been used in several studies for ASD screening of young children. Because the CBCL contains a wide variety of behavioral/emotional problems, a parent's pre-existing disposition to endorse or deny features of ASD may be less likely to influence ratings than might be the case on an ASD-specific instrument. Additionally, the CBCL can identify a wide range of other behavioral and emotional problems in young children that may require clinical attention. Furthermore, the age range of the CBCL/1½–5 spans the full period in which ASD is usually diagnosed, unlike many of the ASD-specific screening instruments.

Rescorla (1988) used an early version of the CBCL (Achenbach, 1978) to rate case records in a clinic sample of 204 boys aged 3 to 5 years evaluated prior to the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed.; DSM-III): 24% “severe atypical,” 14% “mild atypical,” 40% “reactive,” and 12% “undiagnosed.” Cluster analysis of CBCL 8-factor profiles (including an autistic-like factor with the items *Confused/in a fog*, *Strange behavior*, *Withdrawn*) yielded three clusters. The first cluster included most of the children diagnosed “severe atypical” or “mild atypical” (i.e. essentially children with more severe and less severe autism), the second cluster included most of the “reactive” group (i.e. children with a mixture of internalizing and externalizing problems), and the third cluster included most of the children with low CBCL scores (i.e. the undiagnosed children).

Several years after this initial study, Sikora et al. (2008) used a T score cut point of ≥ 70 on the CBCL/1½–5–Withdrawn syndrome and DSM-PDP scale (Achenbach and Rescorla, 2000) to differentiate among 147 children referred to an autism clinic (aged 36–71 months). Based on results of the Autism Diagnostic Observation Schedule-Generic (ADOS-G), children were classified with “autism” ($n = 79$), “autism spectrum” ($n = 18$), or “non-spectrum disorder” ($n = 50$). SENS values were 80% for the DSM-PDP scale and 65% for the Withdrawn syndrome, whereas SPEC values were 42% for DSM-PDP scale and 62% for the Withdrawn syndrome. Specificities were probably low because the non-spectrum disorder group had other developmental disabilities (e.g. low intelligence quotients (IQs)).

Muratori et al. (2011) used the CBCL/1½–5 with three groups of Italian children (aged 24–60 months): 101 diagnosed with ASD (diagnosed by clinical evaluation plus the ADOS), 95 diagnosed with other psychiatric disorders (OPD), and 117 with TD. When Muratori et al. (2011) compared the ASD group with the TD group, using a cut point of $T \geq 65$, SENS/SPEC values were 85%/90% for the DSM-PDP scale and 89%/92% for the Withdrawn syndrome. When they compared the ASD group with the OPD group, also using a cut point of $T \geq 65$, SPEC was lower (60% for the DSM-PDP scale and 65% for the Withdrawn syndrome). The lower SPEC indicates that some children in the OPD group had high scores on these scales even though they did not have ASD but rather other diagnosable psychiatric problems. However, SENS was unchanged (85% and 89%, respectively), indicating that both scales identified most of the children who received an ASD diagnosis.

Rescorla et al. (2015) used the CBCL/1½–5 to screen Korean children (aged 18–71 months), obtaining the following groups: 46 with ASD, 111 with DD, 71 with OPD, and 228 non-referred (NR). The ASD group scored significantly higher than the other groups on the Withdrawn and DSM-PDP scales. With a $T \geq 65$ cut point on the DSM-PDP scale, SENS was 80% for identifying ASD relative to the other three groups, but SPEC varied across groups: NR = 87%, and OPD = 55%, DD = 60%, replicating results from Muratori et al. (2011). Results suggested that the CBCL/1½–5 performs best in Level 1 screening, namely, differentiating children with ASD from children in the general population.

Havdahl et al.'s (2016) findings for 161 children (aged 2–5 years) further suggested that the CBCL/1½–5 PDP scale performs better for Level 1 screening (“typical” comparison group) than for Level 2 screening (“clinical” comparison group). The sample included 104 young children with ASD and 57 children with non-ASD developmental and behavioral/emotional problems (58% with language delay and 25% with attention-deficit/hyperactivity disorder (ADHD)). Decision statistics, reported for the Withdrawn syndrome only, were SENS = 0.74, SPEC = 0.53 for a cut point of ≥ 62 and SENS = 0.63, SPEC = 0.65 for a cut point of ≥ 65 .

Despite some positive findings regarding the CBCL/1½–5 as a screening tool for ASD in young children, research is still quite limited in this area. Furthermore, to our knowledge, the CBCL/1½–5 has not been used previously in studies in which high-risk (HR) infants—known as “baby siblings” because their older sibling has an ASD diagnosis—are compared with low-risk (LR) infants with no known familial risk for ASD. In a recent review of studies using the baby sibling paradigm, Szatmari et al. (2016) noted that the “first generation” of screening tools for ASD in young children, which generally compared children with ASD to typically developing children and children with non-ASD DDs, reported sensitivities from 12 to 24 months of age that were generally low. The authors further noted that many studies also failed to report PPV, namely, the percentage of screened positive cases who really turn out to have ASD.

Because the HR baby sibling paradigm often involves comparing three groups of children—HR children who do receive an ASD diagnosis, HR children who do not receive an ASD diagnosis, and LR children—it is ideally designed for use of a continuous, quantitative measure such as the CBCL/1½–5. The CBCL/1½–5 DSM-PDP scale might show a pattern

of highest scores for the HR diagnosed children, intermediate scores for the HR non-diagnosed children (a subset of whom may display the “broader autism phenotype” or BAP), and lowest scores for the LR children. Thus, applying the CBCL to young children who are at familial HR for ASD allows us to characterize profiles of behavior that extend beyond categorical boundaries.

Goals of the study

In this study, we analyzed 24-month CBCL/1½–5 scores for 56 children at familial HR for ASD (HR) and 26 age-matched children with no first- or second-degree relatives having known or suspected ASD (LR). We had two primary goals in the study. Our first goal was to test for differences on 24-month CBCL/1½–5 scale scores between the HR and LR children and then between the HR children diagnosed with ASD (HR-ASD), the HR children not diagnosed (HR-NEG), and the LR children. Based on previous research, our primary focus was on the Withdrawn and DSM-PDP scales, but we also conducted one-way analyses of variance (ANOVAs) for the other CBCL scales. Our second goal was to determine the screening efficacy of the CBCL, namely, the ability of the Withdrawn and DSMPDP scales to distinguish young children with ASD from those without ASD as indicated by receiver operating characteristic (ROC) results and decision statistics.

Method

Participants

Participants included 56 children at HR for ASD (57% boys) and 26 age-matched toddlers with no familial history of ASD (77% boys). Most of the children in both the LR and HR groups were white: 20 (80%) and 47 (89%), respectively. Of the rest, two children in each group were mixed race, three were black, and one HR child was Asian (race was missing for four children). Most of the mothers in both groups reported having at least a college degree (81% of the LR and 77% of the HR groups). Participants for both groups included all participants seen at the Center for Autism Research (CAR) at the Children’s Hospital of Philadelphia (CHOP) site of the longitudinal Infant Brain Imaging Study (IBIS; NIH Autism Center for Excellence) Network (Estes et al., 2015) between 2010 and 2016 for whom CBCL scores and diagnoses were available at 24 months. Because the CBCL/1½–5 was not included as part of the standard 24-month assessment protocol for IBIS, this study sample was limited to 82 children recruited through CHOP.

In order to qualify as an HR infant, the infant’s older sibling needed to have a community diagnosis of ASD, meet the cutoff for ASD on the SCQ, and meet criteria for autism on the ADI-R (Rutter et al., 2003). Informed consent was received from the infants’ parents prior to the start of the study.

Procedure

The diagnostic protocol from the larger IBIS was used to determine whether an ASD diagnosis should be given at 24 months. Specifically, children in both groups were given the Autism Diagnostic Observation Schedule (ADOS-G) (Lord et al., 2000) and the ADI-R was administered to parents. In addition, the DSM-IV diagnostic criteria were utilized (autistic

disorder, Asperger's disorder, and pervasive developmental disorder not otherwise specified (PDD-NOS) were combined to form the ASD group), and these assessment tools together with expert clinical judgment determined ASD diagnostic status. The CBCL/1½–5 (Achenbach and Rescorla, 2000) was sent home for parents to complete and bring back at the time of their 24-month visit in order to obtain a standardized measure of behavioral/emotional functioning. The CBCL was not used to inform clinical diagnoses made at 24 months as this measure was not part of the diagnostic protocol for IBIS. Rather, CBCL forms were scored at a later date by other research personnel.

Measures

CBCL/1½–5—The CBCL/1½–5 (Achenbach and Rescorla, 2000) contains 99 problem items rated on 0 = *not true*, 1 = *somewhat or sometimes true*, 2 = *very true or often true* based on the previous 2 months. These 99 CBCL/1½–5 items yield scores on seven empirically based syndromes and five DSM-oriented scales. To be consistent with previous studies (Muratori et al., 2011; Rescorla et al., 2015), our primary focus was on the Withdrawn syndrome and the DSM-PDP scale. The CBCL/1½–5 is a widely used measure with strong psychometric properties, including high test–retest reliability (Achenbach and Rescorla, 2000). Internal consistency Cronbach's alphas were 0.75 for the 8-item Withdrawn syndrome and 0.80 for the 13-item PDP scale (Achenbach and Rescorla, 2000).

ADOS-G and ADI-R—The ADOS-G (Lord et al., 2000) is a semi-structured assessment comprising a series of activities (or presses) designed to allow the examiner to observe behaviors relevant to an ASD diagnosis, with cutoff scores to differentiate children with ASD from typically developing children and developmentally delayed children without ASD. The ADI-R (Rutter et al., 2003) is a semi-structured parent interview designed to determine history and current symptoms related to ASD. Both the ADOS-G and ADI-R were administered by research-reliable staff members, with both intra- and inter-site reliability maintained over time.

Language and cognitive skills—The MacArthur-Bates Communicative Development Inventory (CDI) (Fenson et al., 1993) Words and Gestures form, although designed for ages 8–18 months, was administered at 24 months, as per the IBIS protocol. Measures analyzed for this study were Words Understood and Words Produced. We used the Early Learning Composite (ELC) from the Mullen Scales of Early Learning (MSEL) (Mullen, 1995) to assess cognitive development.

Analyses

To address the first goal of our study, we used ANOVAs to test for CBCL/1½–5 group differences. We first compared HR and LR groups and then compared HR-ASD, HR-NEG, and LR groups. Because we tested 15 scales in these two sets of initial analyses (seven syndromes, five DSM-oriented scales, and the three broadband scales of Internalizing, Externalizing, and Total Problems), we used Bonferroni correction (0.05/30) to generate $p = 0.002$. To test for group differences in CDI Words Produced and Words Understood as well as MSEL ELC scores (three measures \times two sets of ANOVAs), we used a separate Bonferroni correction of $0.05/6 = 0.008$. To address the second goal of our study, we

determined the ability of the Withdrawn and DSM-PDP scales to distinguish children with ASD and those without by comparing HR-ASD versus HR-NEG + LR groups using decision statistics/ROC analyses.

Results

Age 24 months ASD status

Based on the clinical assessment procedures described above, 13 of the 56 children in the HR group (11 boys, 2 girls) received an ASD diagnosis at 24 months (HR-ASD = 23%), and 43 did not (HR-NEG, 77%). None of the LR infants received an ASD diagnosis.

Language and cognitive findings

Comparisons between the LR and HR groups yielded no significant CDI differences at 24 months. However, one-way ANOVAs comparing the three groups (LR, HR-NEG, and HR-ASD) yielded a significant group difference on Words Understood, $F(2, 64) = 6.44$, $p = 0.003$, $\eta_p^2 = 0.17$, at the Bonferroni-adjusted alpha level of 0.008 used for the language and cognitive measures. Student–Newman–Keuls (SNK) post hoc tests indicated that the HR-ASD group scored significantly lower than the LR and HR-NEG groups, which did not differ from each other (see Table 2). There was also a group difference for CDI Words Produced, but only at $p < 0.02$.

The HR and LR groups did not differ on the Mullen ELC at 24 months. However, a one-way ANOVA comparing the three groups yielded a significant difference at 24 months, $F(2, 64) = 8.80$, $p < 0.001$, $\eta_p^2 = 0.22$. SNK post hoc tests indicated that the HR-ASD group scored significantly lower than the LR and HR-NEG, which did not differ from each other (see Table 2).

CBCL/1½–5 scale score findings at 24 months

Although the HR group generally obtained slightly higher scores than the LR group on the 15 CBCL/1½–5 scales tested, no differences were significant at $p < 0.002$. However, as shown in Table 2, one-way ANOVAs testing CBCL/1½–5 scores at 24 months for the 26 LR children, 43 HR-NEG children, and 13 HR-ASD children yielded 13 effects significant at $p < 0.002$, namely, for six syndromes, four DSM-oriented scales, and all three broadband scales. All of the scales with significant overall effects showed the same pattern according to SNK post hoc tests, namely, HR-ASD > HR-NEG = LR. Effect sizes as measured by η_p^2 were extremely large for the two CBCL/1½–5 scales of primary focus in this study: DSM-PDP, $F(2,79) = 46.77$, $p < 0.001$, $\eta_p^2 = 0.54$ and Withdrawn, $F(2,79) = 39.97$, $p < 0.001$, $\eta_p^2 = 0.50$.

As seen in Table 2, the LR and HR-NEG groups had mean T scores of 50–51 on the DSM-PDP scale, whereas the HR-D group had a mean T score of 65 (93rd percentile), a 1.5 SD difference. The difference was almost as large with the Withdrawn syndrome ($T = 64$ vs 50–51). It is noteworthy that the HR-ASD group did not have particularly high scores on most

CBCL/1½–5 scales. Aside from their mean scores of 64 on Withdrawn and 65 on DSM-PDP, their only elevated score was 62 on Attention Problems, with scores on all other scales within 1 *SD* of the mean *T* score of 50.

Decision statistics findings

After dichotomizing status at 24 months as ASD (the 13 HR-ASD children) or non-ASD (the 43 HR-NEG and 26 LR children), we did two ROC analyses, one with Withdrawn as a predictor and the second with DSM-PDP as a predictor. The area under the curve (AUC) was 0.89 for Withdrawn and 0.91 for DSM-PDP, indicating strong ability of both scales to differentiate between children with and without an ASD diagnosis at 24 months.

We also computed dichotomous decision statistics analyses using a $T \geq 60$ (84th percentile) cut point for deviance on each scale. Although the Withdrawn and DSM-PDP scales only share five items, they had the same SENS and NPV (77% and 96%), both identifying 10 out of the 13 children diagnosed with ASD. The three “missed” children had scores of 50–51. The DSM-PDP scale yielded one “false positive” (SPEC = 99%, PPV = 91%), but the Withdrawn syndrome yielded two (SPEC = 97%, PPV = 83%). Using a $T \geq 65$ score in this sample would have yielded poorer decision outcomes. For example, a cut point of $T \geq 65$ on the DSM-PDP scale would miss two additional children (for a total of 5 out of 13), thereby reducing SENS from 77% to 62% but with SPEC still at 99% due to one false positive.

As a supplemental analysis, we redid the ROC and decision statistic analyses comparing the HR-ASD group with the HR-NEG group only ($N = 56$). AUCs were 0.90 for DSM-PDP and 0.89 for Withdrawn, very similar to those using both LR + HR-NEG as the non-ASD comparison group. SENS was still 0.77 for both scales, with the same three children missed. Because one of the two false positives for the Withdrawn syndrome had been an LR child, the decision statistics for the DSM-PDP and Withdrawn scales were identical when using the HR group alone: PPV = 0.91, SPEC = 0.93, and NP = 0.93. SPEC and NP were very high but slightly lower than with the full sample because the denominator was 82 rather than 56.

Discussion

In this preliminary study of the CBCL/1½–5 in children at high familial risk for ASD, we found that children diagnosed with ASD at 24 months had, on average, much higher scores on the DSM-PDP and Withdrawn scales than children in the LR and HR-NEG groups, with a mean difference of 1.5 *SDs* for the PDP scale and >1 *SD* for the Withdrawn scale. Furthermore, a significant proportion of the variance in DSM-PDP and Withdrawn scales was accounted for by ASD group status. ROC analyses yielded very strong AUCs (0.91 for DSM-PDP and 0.89 for Withdrawn). A cut point of $T \geq 60$ on the DSM-PDP scale yielded quite good discrimination between ASD and non-ASD groups (i.e. SENS = 77%, SPEC = 99%, PPV = 91%, NPV = 96%). In contrast with most previous studies using the DSM-PDP scale, our SENS was somewhat lower but our SPEC and PPV were substantially higher. Overall, the results we obtained using the baby sibling paradigm indicated that the CBCL DSM-PDP scale, and to a somewhat lesser extent, the Withdrawn syndrome, did a good job of differentiating children diagnosed with ASD at 24 months from those not diagnosed.

Of the 13 HR-ASD children, three were “missed” by the cut point of $T \geq 60$ on the DSM-PDP scale and the Withdrawn syndrome. Although a cut point of 60 seems to be the lowest reasonable cut point to use, given that it represents 1 *SD* above the mean; use of an even lower cut point would not have identified these three children. This is because they had DSM-PDP and Withdrawn scores of 50 or 51, which are in the normal range and comparable to scores of the LR and HR-NEG groups. One of these three children appeared to have a significant cognitive and expressive language delay, but the other two did not. Research with additional baby sibling samples is needed to determine how common it is that 24-month-olds receiving an ASD diagnosis have scores this low on the DSM-PDP and Withdrawn scales and what factors might explain such anomalous results.

There was one false positive using the cut point of $T \geq 60$ on the DSM-PDP scale, a child with a score of 72 on the DSM-PDP scale (and 67 on Withdrawn). This HR-NEG boy had scores of 60 or above on 10 of the 15 CBCL scales, with a Total Problems score of 65. This Total Problems score was higher than that obtained by any other LR or HR-NEG child and seven points higher than scores of 58 obtained by the next two highest scoring HR-NEG children. This finding is consistent with Havdahl et al. (2016), who reported that having elevated scores (T scores ≥ 70) on either the Aggressive Behavior syndrome or the DSM-Affective Problems scale was associated with more false positives in school-age children. The second false positive, based on a score of 65 on Withdrawn, did not show this same pattern. This was an LR boy with CBCL scores of 44 to 56 (normal range) on all CBCL scales except for Withdrawn.

Because two of the HR-ASD children had DSM-PDP scores of 63, a T score v of 60 (84th percentile) achieved better SENS than the more traditionally used cut point of 65 (93rd percentile). Lowering a cut point to raise SENS typically reduces SPEC and PPV (due to more false positives), but it did not do so in this baby sibling sample, as indicated by our high SPEC and PPV values. Raising the cut point to 65 would also not have eliminated the single false positive on the DSM-PDP scale, who had a score of 72. Although $T \geq 65$ (1.5 *SDs*, 93rd percentile) has been the most widely used cut point for ASD screening using the CBCL/1½–5, Sikora et al. (2008) used the more extreme cut point of 70 (2 *SDs*, >97th percentile), whereas Havdahl et al. (2016) used both 62 and 65. A possible reason that a T of 60 worked better than a cut point of 65 in our sample is that the HR-ASD children were diagnosed as part of a baby sibling study, rather than being clinic referrals, which could have led to lower CBCL/1½–5 scores. For example, because the HR parents already had a child with an ASD diagnosis, they may have rated behavior differently than a parent without this first-hand knowledge of ASD. Another possible reason is that the children were only 24-months-old, which may have led to fewer or less severe ASD symptoms as measured by the CBCL/1½–5 in some of the diagnosed children. Additionally, unknown idiosyncratic factors leading to somewhat lower CBCL/1½–5 scores than found in other ASD screening studies may explain our findings. Further research with other 24-month-old baby sibling samples, other 24-month non-baby sibling samples, and older baby sibling samples is needed to test these alternative hypotheses. However, our results raise the possibility that a lower DSM-PDP cut point may be optimal when screening for ASD at 24 months than is optimal at 3 to 5 years when children with ASD may manifest more numerous or more severe symptoms.

The associations we found between the CBCL/1½–5 DSM-PDP and Withdrawn scales and ASD diagnosis are consistent with Sikora et al. (2008), Muratori et al. (2011), Myers et al. (2014), and Rescorla et al. (2015). There are several possible factors that may account for the CBCL/1½–5’s good screening performance across several studies. The items in the CBCL/1½–5, which tap many kinds of behavioral/emotional problems, were selected based on extensive pilot testing over many decades. Items that did not significantly discriminate between referred and NR children were eliminated. Another factor may be that CBCL/1½–5 0–1-2 item ratings are summed to yield continuous quantitative scores on scales that were normed on a large, nationally representative sample, a process that yielded robust standardized scores. Another factor that may be relevant in some studies is that the CBCL/1½–5 is not labeled as an ASD screener and contains many kinds of problems, so a parent’s proclivity to endorse or deny features of ASD may be less likely to influence ratings than might be the case on an ASD-specific instrument. This last factor probably did not contribute to results in our study, however, because it was conducted in a center for ASD research.

We did not find evidence for the BAP in this study, in that the HR-NEG children in our study were indistinguishable from the LR children on the CBCL, as well as on the CDI and the MSEL. While this is consistent with findings from the larger IBIS at the age of 2 years (e.g. Estes et al., 2015), other groups have reported evidence of the BAP among HR-NEG children measured at the age of 3 years (e.g. Ozonoff et al., 2014).

The positive results reported in this and other studies suggest that the CBCL/1½–5 has potential for use as an ASD screener, such as in pediatric practices. No professional time is needed to administer it, the paper forms are inexpensive and can be completed by a parent in a waiting room in less than 15 min, it can be scored in <5 min by clerical staff, and it can also be administered online. These features suggest that it could be efficiently incorporated into a 24-month pediatric checkup. These features also make it less costly in terms of professional time than the M-CHAT with the 15-min follow-up interview, which was added to the original M-CHAT because decision statistics for the M-CHAT alone were less than optimal. Additionally, because the scored profiles for the CBCL/1½–5 indicate scores on many syndromes and scales, it is easy for a practitioner to determine not only whether the child is at risk for ASD but also to see whether scores are elevated on other scales, such as Attention Problems or Aggressive Behavior. Given that the M-CHAT is already well established in pediatric practice use, it might be useful to administer both the M-CHAT and the CBCL/1½–5 at 24 or 30 months and then only conduct the M-CHAT follow-up interview if both instruments suggest ASD.

Limitations

The main limitation of this study is the somewhat small sample size. Additionally, it is possible that our sample of HR-ASD children was idiosyncratic in some way and thus not representative of ASD in general. This may account for the fact that three of the 13 diagnosed children had such low DSM-PDP scores and two others had scores of only 63. These limitations might be resolved by extending our analysis to a larger sample of LR and HR children. Although the differences between the HR-ASD versus the LR and HR-NEG

groups were very large, a larger sample might also provide greater power to detect evidence of the BAP. Furthermore, a larger baby sibling sample might indicate whether our findings of better SENS and no reduction in SPEC with a cut point of 60 versus 65 are unique to our sample or characteristic of 24-month-old baby sibling samples more broadly. As reducing a cut point to increase SENS typically reduces SPEC in population samples, it would be good to see whether our finding to the contrary would be replicated in larger baby sibling samples of 24-month-old children. Finally, it should also be noted that after the current data were collected and scored using the 13-item DSM-PDP scale based on DSM-IV, the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM-V; American Psychiatric Association, 2013) was published. Based on new expert ratings, one of the PDP scale items—*Afraid to try new things*—was removed, yielding a 12-item DSM-5 scale that was renamed the DSM-ASD scale, to be consistent with DSM-5 nomenclature (Achenbach, 2014). However, as the two versions of the scale share 12 items, it is unlikely that results would have differed markedly.

Future directions and implications

Because our sample size is somewhat small, we consider our findings preliminary rather than definitive with respect to how the CBCL/1½–5 might perform in a larger study of young children at familial risk for ASD. Future research should examine CBCL scores in larger baby sibling samples to determine the generalizability of our findings. Additionally, future research with larger baby sibling samples might explore the cognitive/behavioral profiles of any false-positive or false-negative cases to determine for whom the CBCL/1½–5 scales are most sensitive and specific. It might also be productive to examine item-level data to understand how components of the DSM-PDP scale differentially contribute to an estimate of risk. For example, do DSM-PDP problem items that are more common in the general population predict less well than problem items that are rarer? Furthermore, one could examine whether items that reflect social communication/interaction problems are better predictors than items that reflect restricted/repetitive interests and behavior.

Our baby sibling study did not involve population screening in a large unselected sample, given that children in the HR group were recruited because of their elevated risk for ASD. However, our study simulated Level 1 screening in that both the LR children and the HR-NEG children turned out to be typically developing on the CBCL/1½–5. Our study thus provides an important addition to the literature on use of the CBCL/1½–5 as an ASD screener by demonstrating good discriminative validity (such as very high AUCs) when used in a baby sibling study in which the HR-NEG children manifest few problems. It should be reiterated, however, that previous studies (Havdahl et al., 2016; Muratori et al., 2011; Rescorla et al., 2015) suggest that the CBCL/1½–5 is less successful when the comparison group comprises children with developmental and emotional/behavioral problems, as generally found in Level 2 screening.

We identified only one false positive with the DSMPDP scale and two with the Withdrawn scale because most of the LR and HR-NEG children in our sample had mean *T* scores close to 50 on all CBCL/1½–5 scales. Thus, an important implication of this and other screening studies of ASD is that the strength of the screening prediction is strongly dependent on the

nature of the non-ASD comparison groups. As with other screening instruments used to identify young children with ASD, the CBCL/1½–5 has better SPEC and PPV when the comparison group comprises typically developing children than when it comprises children with other developmental or psychiatric problems. As shown in Sikora et al. (2008) and Myers et al. (2014), false positives are common when differentiating children with ASD from non-ASD children with language or cognitive delays. Additionally, although SENS may remain high with a comparison group containing children with OPD, SPEC and PPV may decline due to false positives, as found by Muratori et al. (2011) and Rescorla et al. (2015).

Although false positives reduce overall screening efficiency, the CBCL/1½–5's standardized scores on 15 scales provide extensive documentation of the kinds of behavioral/emotional problems false-positive cases have or do not have, hence providing valuable clinical information. Specifically, scores on the seven empirically based syndromes can indicate possible target areas for intervention, scores on the five DSM-oriented scales can suggest possible diagnoses to consider (such as ADHD, oppositional defiant disorder (ODD), or anxiety), and scores on the broadband scales indicate whether the child has primarily internalizing problems, primarily externalizing problems, both kinds of problems, or neither kind of problem.

This study adds to a small body of literature suggesting that the CBCL/1½–5 shows promise as a Level 1 screener for young children with ASD, while also providing additional information about other kinds of behavioral/emotional problems the children assessed may have that require attention. However, to definitely establish the CBCL/1½–5 as a Level 1 ASD screening tool, a larger and more comprehensive study is needed. Such a study should be epidemiological in scope and should involve large samples of rigorously diagnosed young children with ASD (e.g. >300) and at least two large control groups—one, children from the general population, and the other, children with developmental or behavioral/emotional problems. Such a study would provide definitive data about the ability of the CBCL/1½–5's DSM-PDP and Withdrawn scales to identify children with ASD with reference to both a “typical” and a “clinical” comparison group.

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by an NIH Autism Center of Excellence grant (NIMH and NICHD #HD055741 and HD055741-S1) to J.P.

Declaration conflicting of interests

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The first author (L.A.R.) receives royalties from the University of Vermont Research Center for Children Youth, and Families, which publishes the CBCL/1½–5.

References

Achenbach TM (1978) The child behavior profile: I. Boys aged 6–11. *Journal of Consulting and Clinical Psychology* 46(3): 478–488. [PubMed: 670491]

- Achenbach TM (2014) DSM-Oriented Guide for the Achenbach System of Empirically Based Assessment (ASEBA). Burlington, VT: Research Center for Children, Youth, & Families, University of Vermont.
- Achenbach TM and Rescorla L (2000) Manual for the ASEBA Preschool Forms & Profiles. Burlington, VT: Research Center for Children, Youth, and Families, University of Vermont.
- American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders. 4th ed Washington, DC: American Psychiatric Association.
- American Psychiatric Association (2013) Diagnostic and Statistical Manual of Mental Disorders. 5th ed Washington, DC: American Psychiatric Association.
- Baird G, Charman T, Baron-Cohen S, et al. (2000) A screening instrument for autism at 18 months of age: a 6-year follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry* 39(6): 694–702. [PubMed: 10846303]
- Baron-Cohen S, Allen J and Gillberg C (1992) Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *The British Journal of Psychiatry* 161(6): 839–843. [PubMed: 1483172]
- Barton ML, Dumont-Mathieu T and Fein D (2012) Screening young children for autism spectrum disorders in primary practice. *Journal of Autism and Developmental Disorders* 42(6): 1165–1174. [PubMed: 21842325]
- Berument SK, Rutter M, Lord C, et al. (1999) Autism screening questionnaire: diagnostic validity. *The British Journal of Psychiatry* 175: 444–451. [PubMed: 10789276]
- Bracken BA and Keith LK (2004) CAB, Clinical Assessment of Behavior: Professional Manual. Lutz, FL: Psychological Assessment Resources (PAR).
- Bradstreet LE, Juechter JI, Kamphaus RW, et al. (2017) Using the BASC-2 parent rating scales to screen for autism spectrum disorder in toddlers and preschool-aged children. *Journal of Abnormal Child Psychology* 45: 359–370. [PubMed: 27177744]
- Bryson SE, Zwaigenbaum L, McDermott C, et al. (2008) The Autism observation scale for infants: scale development and reliability data. *Journal of Autism and Developmental Disorders* 38: 731–738. [PubMed: 17874180]
- Chlebowski C, Robins DL, Barton ML, et al. (2013) Large-scale use of the modified checklist for Autism in low-risk toddlers. *Pediatrics* 131(4): e1121–e1127. [PubMed: 23530174]
- Estes A, Zwaigenbaum L, Gu H, et al. (2015) Behavioral, cognitive, and adaptive development in infants with autism spectrum disorder in the first 2 years of life. *Journal of Neurodevelopmental Disorders* 7(1): 24. [PubMed: 26203305]
- Fenson L, Dale PS, Reznick JS, et al. (1993) MacArthur Communicative Development Inventories: User's Guide and Technical Manual. San Diego, CA: Singular.
- Gilliam JE (2006) Gilliam Autism Rating Scale-2. 2nd ed Austin, TX: Pro-Ed.
- Havdahl KA, Von Tetzchner S, Huerta M, et al. (2016) Utility of the child behavior checklist as a screener for autism spectrum disorder. *Autism Research* 9(1): 33–42. [PubMed: 26140652]
- Lord C and Rutter M (2012) Autism Diagnostic Observation Schedule. 2nd ed Los Angeles, CA: Western Psychological Services.
- Lord C, Rutter M, DiLavore PC, et al. (2000) Autism Diagnostic Observation Schedule-Generic. Los Angeles, CA: Western Psychological Services.
- Mullen EM (1995) Mullen Scales of Early Learning. Circle Pines, MN: AGS, pp.58–64.
- Muratori F, Narzisi A, Tancredi R, et al. (2011) The CBCL 1.5–5 in a sample of children with autism in Italy. *Epidemiology and Psychiatric Sciences* 20: 329–338. [PubMed: 22201210]
- Myers CL, Gross AD and McReynolds BM (2014) Broadband behavior rating scales as screeners for autism? *Journal of Autism and Developmental Disorders* 44(6): 1403–1413. [PubMed: 24287879]
- Norris M and Lecavalier L (2010) Screening accuracy of Level 2 autism spectrum disorder rating scales: a review of selected instruments. *Autism* 14(4): 263–284. [PubMed: 20591956]
- Ozonoff S, Young GS, Belding A, et al. (2014) The broader autism phenotype in infancy: when does it emerge? *Journal of the American Academy of Child and Adolescent Psychiatry* 53: 398–407. [PubMed: 24655649]
- Rescorla L (1988) Cluster analytic identification of autistic preschoolers. *Journal of Autism and Developmental Disorders* 18: 475–492. [PubMed: 3215877]

- Rescorla L, Kim YA and Oh KJ (2015) Screening for ASD with the Korean CBCL/1½–5. *Journal of Autism and Developmental Disorders* 45: 4039–4050. [PubMed: 25239178]
- Reynolds CR and Kamphaus RW (2004) *Behavior Assessment for Children (BASC-2)*. Circle Pines, MN: American Guidance Service.
- Robins DL, Casagrande K, Barton M, et al. (2014) Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics* 133(1): 37–45. [PubMed: 24366990]
- Robins DL, Fein D and Barton M (2009) Modified checklist for Autism in toddlers, revised, with follow-up (M-CHAT-R/F). *Pediatrics* 133: 37–45.
- Robins DL, Fein D, Barton ML, et al. (2001) The modified checklist for Autism in toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders* 31: 131–144. [PubMed: 11450812]
- Rutter M, Le Couteur A and Lord C (2003) *ADI-R: The Autism Diagnostic Interview-Revised*. Los Angeles, CA: Western Psychological Services.
- Sikora DM, Hall TA, Hartley SL, et al. (2008) Does parent report of behavior differ across ADOS-G classifications: analysis of scores from the CBCL and GARS. *Journal of Autism and Developmental Disorders* 38: 440–448. [PubMed: 17619131]
- Szatmari P, Chawarska K, Dawson G, et al. (2016) Prospective longitudinal studies of infant siblings of children with autism: lessons learned and future directions. *Journal of the American Academy of Child and Adolescent Psychiatry* 55: 179–187. [PubMed: 26903251]

Table 1.

CBCL/1½–5 Withdrawn syndrome DSM-PDP scale items.

Withdrawn syndrome	DSM-PDP scale
Item 1. <i>Acts young for age</i>	
Item 3. <i>Afraid to try new things</i>	Item 3. <i>Afraid to try new things</i>
Item 4. <i>Avoids looking others in the eye</i>	Item 4. <i>Avoids looking others in the eye</i>
Item 7. <i>Can't stand things out of place</i>	Item 7. <i>Can't stand things out of place</i>
Item 21. <i>Disturbed by any change in routine</i>	Item 21. <i>Disturbed by any change in routine</i>
Item 23. <i>Doesn't answer when people talk to him/her</i>	Item 23. <i>Doesn't answer when people talk to him/her</i>
Item 25. <i>Doesn't get along with other children</i>	Item 25. <i>Doesn't get along with other children</i>
Item 62. <i>Refuses to play active games</i>	
Item 63. <i>Repeatedly rocks head or body</i>	Item 63. <i>Repeatedly rocks head or body</i>
Item 67. <i>Seems unresponsive to affection</i>	Item 67. <i>Seems unresponsive to affection</i>
Item 70. <i>Shows little affection toward people</i>	Item 70. <i>Shows little affection toward people</i>
Item 71. <i>Shows little interest in things around him/her</i>	
Item 76. <i>Speech problem</i>	Item 76. <i>Speech problem</i>
Item 80. <i>Strange behavior</i>	Item 80. <i>Strange behavior</i>
Item 92. <i>Upset by new people or situations</i>	Item 92. <i>Upset by new people or situations</i>
Item 98. <i>Withdrawn, doesn't get involved with others</i>	Item 98. <i>Withdrawn, doesn't get involved with others</i>

Table 2.

Means (*SDs*) by group for CBCL/1½-5 scales by group.

Scale	HR	LR	HR-NEG	HR-ASD
Language and cognitive scales				
CDI-REC	262.6 (109.2)	316.1 (64.9) ^a	287.5 (94.8) ^a	181.1 (117.7) ^b
CDI-EXP	165.2 (117.3)	214.1 (99.1)	187.9 (114.6)	90.9 (97.2)
Mullen ELC	180.4 (18.1)	109.4 (10.7) ^a	104.8 (14.8) ^a	84.4 (20.4) ^b
Syndromes				
Emotionally Reactive	52.9 (4.9)	50.0 (0.2) ^a	51.0 (3.0) ^a	54.8 (7.0) ^b
Anxious/Depressed	51.9 (3.8)	50.5 (1.6) ^a	51.2 (2.7) ^a	54.4 (5.6) ^b
Somatic Complaints	51.6 (4.2)	50.6 (1.2) ^a	50.5 (2.4) ^a	55.5 (6.5) ^b
Withdrawn	53.6 (7.5)	51.4 (3.2) ^a	50.7 (2.7) ^a	63.5 (9.8) ^b
Sleep Problems	54.6 (7.8)	51.0 (2.2)	53.9 (7.4)	57.1 (8.8)
Attention Problems	54.4 (6.8)	51.3 (2.9) ^a	52.0 (3.9) ^a	61.6 (9.2) ^b
Aggressive Behavior	53.3 (5.6)	50.2 (1.5) ^a	52.1 (3.9) ^a	57.2 (8.1) ^b
DSM-oriented scales				
DSM-Affective	54.0 (5.8)	51.0 (2.0) ^a	52.5 (3.8) ^a	59.1 (8.2) ^b
DSM-Anxiety	52.4 (4.8)	50.4 (1.4) ^a	51.5 (3.7) ^a	55.5 (6.5) ^b
DSM-ADH	52.6 (4.4)	50.7 (1.7)	51.8 (3.6)	55.1 (6.0)
DSM-PDP	54.7 (8.2)	50.8 (1.7) ^a	51.4 (3.9) ^a	65.4 (9.8) ^b
DSM-ODP	52.7 (4.4)	50.8 (1.8) ^a	51.8 (3.1) ^a	55.9 (6.4) ^b
Broadband scales				
Internalizing	41.4 (12.2)	36.4 (6.4) ^a	38.1 (9.3) ^a	52.6 (14.3) ^b
Externalizing	46.2 (11.8)	38.3 (9.0) ^a	43.7 (10.1) ^a	54.5 (13.6) ^b
Total Problems	44.7 (12.3)	35.6 (6.7) ^a	41.3 (9.6) ^a	55.9 (14.1) ^b

CBCL: Child Behavior Checklist; HR: high risk (*n* = 56); LR: low risk (*n* = 26); HR-NEG: high-risk children not diagnosed with ASD (*n* = 43); HR-ASD: high-risk children diagnosed with ASD (*n* = 13); ADH: attention deficit hyperactivity; PDP: pervasive developmental problems; ODP: oppositional defiant problems; ELC: Early Learning Composite; DSM: *Diagnostic and Statistical Manual of Mental Disorders*.

Mean values with different superscripts are significantly different in one-way ANOVA post hoc comparisons.