Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: a receiver-operating characteristic analysis

James J. Hudziak,1 William Copeland,1 Catherine Stanger,1 and Martha Wadsworth2
1University of Vermont, USA; 2University of Denver, USA

Background: This study examines the diagnostic accuracy of the CBCL syndrome AS scales for predicting DSM-IV Attention Deficit-Hyperactivity Disorder (ADHD) and Oppositional Defiant Disorder with or without Conduct Disorder (ODD/CD). Methods: The sample included 370 children (187 probands and 183 siblings) participating in a family genetic study of attention and aggressive behavior problems. Univariate and stepwise logistic regression analyses were used to derive models for predicting two diagnostic conditions: ADHD and ODD/CD. Results: The Attention Problems syndrome significantly predicted ADHD, and ODD/CD was significantly predicted by the Aggressive Behavior syndrome. Both scales demonstrated good diagnostic accuracy, as assessed through receiver operating characteristics analyses. Cut-point analyses confirmed the utility of low T-scores, 55 on the respective syndromes, for efficiently discriminating cases from noncases. Conclusions: CBCL syndromes display good diagnostic efficiency for assessing common externalizing disorders in children. Keywords: ADHD, ODD, attention problems, aggressive behavior.

Externalizing behavior problems account for over half of referrals for mental health services by children in the United States (Kazdin, 1995). Children with these problems are at increased risk for juvenile delinquency, academic failure, and social maladjustment, thus placing a tremendous strain on educational and mental health services (Frick, O’Brien, Wooton, & McBurnett, 1994; Hinshaw, 1992). These problems are represented in the DSM-IV (American Psychiatric Association 1994) by Oppositional Defiant Disorder (ODD), Conduct Disorder (CD) and Attention-Deficit/Hyperactivity Disorder (ADHD). Diagnosis of these conditions often requires structured clinical interviews or semi-structured interviews. These methods require administration by trained personnel, may be time-consuming, and thus, are quite costly. Current clinical practice and insurance guidelines put a premium on accurate and expeditious screening for the presence of these conditions. In addition, efficient screening expedites effective interventions. Behavior checklists have proven to be useful as quick, efficient, and importantly, low-cost, measures of broad constructs of psychopathology.

The Child Behavior Checklist (CBCL) is a widely used parent-report questionnaire which allows clinicians and researchers to assess for a wide range of behavior problems and competencies (Achenbach & Rescorla, 2001). A scored CBCL provides T-scores on eight empirically-derived syndromes and three competence scales, but does not directly assess for DSM diagnoses. Although the CBCL is used extensively in both clinical practice and research, the practical demands of both clinical work and research often require a categorical diagnosis. A number of studies have demonstrated convergence between the statistically derived syndromes of the CBCL and DSM-IV disorders (American Psychiatric Association, 1994) (Edelbrock & Costello, 1988; Kazdin & Heidish, 1984). In a clinical sample, the Aggressive Behavior syndrome was strongly related to Conduct Disorder and Oppositional Defiant Disorder (Edelbrock & Costello, 1988). A number of studies have supported strong relations between the Attention Problems syndrome and a DSM diagnosis of ADHD (e.g., Barkley, DuPaul, & McMurray, 1990; Steingard, Biederman, Doyle, & Sprich-Buckminster, 1992). A few studies have specifically assessed the diagnostic accuracy of the CBCL for predicting ADHD. Two studies have derived cut-points on CBCL syndromes that effectively discriminate cases from noncases (Chen, Faraone, Biederman, & Tsuang, 1994; Ostrander, Weinfurt, Yarnold, & August, 1998). Using exploratory logistic regression analyses, Chen and colleagues (1994) found the Attention Problems syndrome of the CBCL had the highest discriminating power for a DSM-III-R diagnosis of ADHD. Additional quality receiver operating characteristic analysis allowed comparison of various cut-points for discriminating ADHD cases from noncases. A cut-point of 55 to 60, depending on sample, on the Attention Problems syndrome of the CBCL had high specificity and relatively high sensitivity for predicting ADHD across various samples. In the proband group, a T-score of 55 on the Attention Problems syndrome was associated with a sensitivity of 88% and a specificity of 91%.

Ostrander and colleagues (1998) derived classification trees for predicting ADHD from CBCL scales.
and syndromes by utilizing optimal discriminant classification tree analysis. A T-score of 54 on the Social Problems syndrome was the most effective discriminator of ADHD from non-ADHD. This study utilized a community sample of children with and without ADHD, whereas Chen et al. (1994) utilized a clinically referred sample of children with ADHD. Additional use of the Delinquent Behavior, Aggressive Behavior, and Attention Problems syndromes in the classification tree improved the classification of ADHD versus non-ADHD subjects. This method represents a statistically sophisticated method of combining CBCL syndromes and scales to hone diagnostic efficiency.

A number of statistical issues need to be considered in assessing the diagnostic accuracy of any measure. Multiple indices of the effectiveness of the scale should be examined to determine the usefulness of a scale for predicting a diagnosis. A scale must accurately identify positive cases, but also accurately identify noncases. For example, a scale may correctly identify all noncases (according to a ‘gold standard’), but may only correctly identify 25% of cases. In this situation the scale would have high specificity but low sensitivity. As these two constructs are reciprocally related, the most efficient scales would effectively balance both indices. It is also necessary to consider the base rates of the disorder within the population. If the disorder only occurs in 5% of the population then a test that always predicts the absence of the disorder would be accurate 95% of the time. The specificity for this test may be perfect and yet the sensitivity would be zero. This issue is particularly pertinent to the prediction of psychiatric diagnoses as these conditions often have low prevalence rates.

Receiver-operating characteristic (ROC) analyses, developed from signal-detection theory, allow visual analysis of the relative merits of continuous scales as predictors of dichotomous outcomes. ROC analyses have been used for diverse purposes ranging from predicting response to pharmacological interventions to determining predictors of violent recidivism (Rey, Morris-Yates, & Stanislaw, 1992; Rice & Harris, 1995). Used in conjunction with logistic regression, ROC analysis provides a means to assess the predictive value of a given test and to adjust cut-points for clinical or research purposes. On the ROC graph, the sensitivity of different cut-points on the test are graphed on the Y axis (true positive rate) along with 1 – the specificity of the cut-points on the X axis (false positive rate) to determine the ability of the test to optimize both measures for each point on the test. The higher the graph extends toward the upper left corner of the graph, the higher the discriminatory power of the test. The area under the curve statistic assesses the probability of correctly classifying a randomly selected pair of subjects in which one is a case and one is a noncase.

Previous studies assessing the diagnostic accuracy of the CBCL for predicting ADHD have supported the Social Problems and Attention Problems syndromes for discriminating children with ADHD from children without ADHD. This study attempted to replicate previous findings supporting the diagnostic utility of the CBCL in identifying ADHD, but also to extend previous work by evaluating the usefulness of the CBCL in identifying children with ODD and/or CD. CBCL syndromes and scales were included in logistic regression analyses to predict both diagnostic categories. Based on available research on the relations between CBCL syndromes and these DSM diagnoses, the Attention Problems syndrome was hypothesized to discriminate ADHD from non-ADHD cases, whereas the Aggressive Behavior syndrome should effectively discriminate children with ODD/CD from children without ODD/CD. Given the specific findings in regard to the Attention Problems syndrome, low cut-points, at a T-score of 55 to 60, are predicted to optimize classification of ADHD cases. The relations between the CBCL and the externalizing disorders were assessed in a sample of children with high levels of attention and aggressive behavior and a replication-sample of randomly-selected siblings of the probands.

Method

Participants

The sample was composed of 187 proband children aged 6 to 18 (114 boys, 73 girls) and 183 randomly-selected siblings (one sibling from each family; 101 boys, 82 girls) participating in a family study conducted in the Northeastern United States examining genetic and environmental correlates of attention and aggressive behavior problems. For families having more than one child, we randomly selected one child per family to be included in the sibling analyses. To randomly select children, we assigned each family a random number between 1 and 100 using the random number generator in SPSS. For families having 2 children in addition to the proband, if their random number was below 50, we selected the youngest child. A similar strategy was employed to randomly select one child from families having 4 or more children. Families were recruited from local pediatricians and psychiatrists, and via newspaper advertisements and posters placed throughout the county. The study had three demographic inclusion criteria for probands: (1) Proband child had to be between the ages of 6 and 18; (2) Proband lived with at least one biological parent; (3) Proband had at least one sibling between the ages of 6 and 18. In addition, subjects were excluded if the proband’s IQ fell at or below 70, as assessed by two sub-tests of the Wechsler Intelligence Scale for Children, Version III (WISC-III; Wechsler, 1991). In addition to these inclusion criteria, probands were recruited to fill four groups corresponding to various levels of externalizing behavior problems for genetic analyses. Each group was defined by the CBCL scores of the proband: (1) Children scoring
below a T-score of 60 on both scales were recruited into a control group; (2) Children with a T-score at or above 67 on the Attention Problems syndrome and a T-score below 60 on the Aggressive Behavior syndrome were recruited for an attention problems only group; (3) Children with a T-score at or above 67 on the Aggressive Behavior syndrome and a T-score below 60 on the Attention Problems syndrome were recruited for an aggressive behavior only group; (4) Children with a T-score at or above 67 on both syndromes were recruited for a comorbid group. Based on these group definitions, no proband had T-scores between 60 and 66 on the CBCL Attention Problems or Aggressive Behavior syndromes.

Recruitment occurred through two sources. Subjects were recruited from a university-based outpatient clinic based on a review of clinic records. In addition, subjects were recruited from the community via posters and newspaper ads. Families that responded to advertisements were screened by telephone for demographic inclusion criteria. Families that met these criteria completed CBCLs on all eligible children to determine if any child in the family met the CBCL criteria for the four groups. If more than one child met the criteria for a study group, then the child identified by the parent as having problems with attention and/or aggressive behavior was designated as a proband. In some cases, the identified child failed to meet criteria for a group, but a sibling met the criteria and was thus designated as the proband. A total of 704 families were reviewed from the outpatient clinic: 346 of these families were ineligible and thus were not contacted further. Of the remaining 358 families, 156 refused to participate and 105 were unable to be contacted based upon the contact information available at the clinic, and another 3 families agreed to participate then failed to attend scheduled appointments. Thus, 75 of the 253 families contacted from the outpatient clinic participated (30% participation rate), and 75 of 358 eligible families participated (21% participation rate). A total of 467 families responded to community advertisements for the study: 295 families were ineligible due to their CBCL scores, 37 families were eligible based on CBCL and demographic information but refused to participate, and 4 families were eligible but failed to attend scheduled appointments. Thus, 131 of 172 eligible families participated in the study (76% participation rate). Of the 206 families participating in the study, 19 families were missing diagnostic data for the proband samples and 23 families were missing diagnostic data on the randomly-selected sibling.

Procedure

Families participated in a multi-stage data collection process. Families completed packets of questionnaires mailed to them prior to their appointment. Next, families attended a 2-hour data collection visit either at the research center or in their homes. During this visit, parents and children over the age of 11 were administered diagnostic interviews. All children were administered the Block Design and Vocabulary subtests of the WISC-III for a brief cognitive screen. All participating family members provided a sample of their saliva for genetic analyses.

Measures

A structured diagnostic interview of DSM-IV symptomatology, the Vermont Structured Diagnostic Interview, was completed with mothers of the probands and siblings. Diagnoses assessed included the disruptive behavior disorders of CD, ODD, and ADHD. Internalizing disorders assessed included generalized anxiety disorder, separation anxiety disorder, and major depressive disorder. Diagnostic status was assessed for current (within the past 6 months) and past (prior to past 6 months) diagnoses. Composite lifetime diagnostic status was also computed as positive if the diagnosis was present in either the past or present. Interviews were conducted by trained bachelor-level research assistants. Determination of diagnostic status was based on endorsement of symptomatology consistent with DSM-IV criteria. Test–retest reliability of the interview was completed on a sample of 22 parents recruited from an outpatient psychiatric clinic (7 other recruited parents refused participation). Parents were interviewed twice over a period of 7 to 21 days (mean = 16.23). The average age of the children was 10.84 (SD = 3.20) (12 boys; 10 girls). Parents were asked to report on the behavior of the child receiving services at the clinic. Kappa statistics for disruptive behavior disorders ranged from .433 for CD to .850 for ADHD (mean κ = .619), and for internalizing disorders ranged from .463 for separation anxiety disorder to .876 for generalized anxiety disorder (mean κ = .689). Intraclass correlation coefficients (ICCs) were computed for past and current symptom counts. ICCs for disruptive behavior disorders ranged from .738 for ODD to .878 for CD (mean ICC = .826) and for internalizing disorders ranged from .569 for major depressive disorder to .953 for separation anxiety disorder (mean ICC = .739). Test–retest reliability statistics are comparable to those reported for the DISC-2.3 (Schaffer, Fisher, Luca, Dulcan, & Schwab-Stone, 2000) and various versions of the K-SADS (Ambrosini, 2000). In the current analyses, lifetime diagnoses of ADHD, ODD, and CD were used to designate group status.

The Child Behavior Checklist (CBCL; Achenbach, 1991) was used to assess emotional and behavior problems in children. The CBCL is a standardized behavioral checklist for parents to report the frequency of 120 problem behaviors exhibited by their child in the past six months, and was used to test for symptoms of internalizing and externalizing psychopathology. Respondents rate each behavior or symptom on a three-point scale: not true (0), somewhat or sometimes true (1), or very true or often true (2). The results of principal components analysis of the 120 behavior problem items have yielded eight empirically-validated syndromes including Withdrawn, Somatic Complaints, Anxious/Depressed, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, and Aggressive Behavior. The CBCL also contains items about activities, social relationships, academic performance, chores, and hobbies, which are summarized in three competence scales: Activities, Social and School scales. This measure has established excellent reliability and validity. The eight behavioral syndromes and three competence scales were used in logistic regression
analyses. CBCL syndromes are scored to provide a raw sum score and $T$-scores normed by age and gender. While both $T$-scores and raw scores were available for scales and syndromes, raw scores were used in logistic regression analyses, and $T$-scores were utilized for ROC analyses. Raw scores are preferred for parametric statistical analyses due to greater variability in scores, whereas $T$-scores are appropriate for evaluating cut-points.

**Data analysis**

Analyses were conducted separately for the proband and sibling groups. The two categories to be predicted were all children who had ever carried a diagnosis of ADHD and all children who had ever carried a diagnosis of ODD or CD. The proband sample included 89 children diagnosed with ODD/CD and 95 diagnosed with ADHD. A total of 50 children had ODD without CD, and 39 had ODD and CD. None had CD without ODD. In the sibling sample, 66 children met diagnostic criteria for ADHD and 68 for ODD/CD. Of the 68 children in the ODD/CD group, 49 only met the criteria for ODD, and 19 met the criteria for both ODD and CD. None had ODD without CD. In all analyses, comparisons were made between children in the diagnostic group and all other children in the sample; thus, children with a diagnosis of ODD/CD but not ADHD would be included in the comparison to the ADHD group. As expected, significant comorbidity between groups was observed. For probands, 70 met the diagnostic criteria for both ADHD and ODD/CD, and 42 met both criteria in the sibling sample.

**Logistic regression analyses.** Univariate and stepwise logistic regression analyses were used to develop candidate models for predicting the two diagnostic conditions. First, all syndromes and scales from the CBCL were included individually in separate univariate logistic regression analyses to test the relative significance of each scale or syndrome for predicting the presence of the specific diagnosis (i.e., 11 regression equations for each diagnosis). A Bonferroni correction was made to the significance level in these analyses (i.e., $p < .0045$). The candidate predictors included the following scales and syndromes from the CBCL: Activities scale, Social scale, School scale, Withdrawn, Anxious/Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, Aggressive Behavior, and Delinquent Behavior. Next, significant scales and syndromes identified in the univariate analyses were used as predictors in stepwise logistic regression analyses to select the best predictor or combination of predictors of ADHD and ODD/CD. As these were exploratory analyses, the stepwise analyses were run with a probability of entry level of .05. Significant predictor models derived from the stepwise logistic regression analyses were then submitted to ROC analyses.

**Receiver-Operating Characteristic (ROC) analysis.** ROC analyses were conducted separately for both the proband and sibling samples. Models tested included scales or syndromes that were significant in logistic regression analyses. If more than one predictor was significant then the multiple predictors were tested both separately and at the same time to identify the amount of improvement for the combination of predictors relative to each predictor alone. ROC analyses were first developed in signal detection theory for assessing the predictive value of a test for a ‘gold standard.’ A cut-off that predicts all true positives has 1.0 sensitivity and a cut-off that predicts all true negatives has 1.0 specificity. Attempts to maximize either sensitivity or specificity, independently, will lead to inflated rates of false positives or false negatives, respectively, for imperfect tests. ROC can visually demonstrate the cut-points that efficiently maximize both sensitivity and specificity. The most common index of accuracy in ROC analysis is the area under the curve (AUC), which assesses the possibility of correctly classifying a randomly selected pair of subjects in which one is a case and one is noncase. AUC values range between .5, in which correct classification occurs in 50% of cases, to 1.0, in which correct classification occurs with every case (sensitivity and specificity = 1). Acceptable AUC values vary depending on the base rate of the diagnosis and other sample characteristics.

Values for specificity and sensitivity were calibrated to the base rates of the disorder within the sample through a quality ROC analysis (QROC). This procedure allows evaluation of possible cut-points for making categorical classification decisions. As discussed previously, optimal cut-points rely upon one’s valuation of sensitivity and specificity. For the CBCL syndromes, we assessed cut-points consistent with .5, 1.0, 1.5, and 2.0 standard deviations above the mean for the relevant syndromes (i.e., $T$-score = 55, 60, 65, 70). Cut-point analyses provide rates of the disorder within the sample, sensitivity and specificity, positive predictive power and negative predictive power (the probability of having either a negative or positive diagnosis among those with a positive or negative test, respectively), and efficiency (derived from QROC analyses, an index of the probability that the test result and the diagnosis agree after sensitivity and specificity are recalibrated using base rates). Cut-points were only derived for the sibling sample as the CBCL syndrome scores for the proband sample were truncated in the recruitment process.

**Results**

CBCL scores for each of the diagnostic groups in the proband and sibling samples are displayed in Table 1. Both samples scored, on average, above the mean on most syndromes of the CBCL, with the proband sample scoring about one standard deviation above the mean and the sibling sample scoring about a half standard deviation above the mean.

**Logistic regression analyses**

For probands, each syndrome and competence scale significantly predicted each respective diagnostic condition in the univariate logistic regression analyses, except the Somatic Complaints syndrome and the Activities scale. In the sibling sample, all scales
Table 1 Means and standard deviations of scores on the CBCL syndromes and scales for probands and siblings

<table>
<thead>
<tr>
<th></th>
<th>Probands (n = 187)</th>
<th>Siblings (n = 183)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Dx</td>
<td>ADHD</td>
</tr>
<tr>
<td></td>
<td>(N = 82)</td>
<td>(N = 95)</td>
</tr>
<tr>
<td>Age</td>
<td>10.87 (3.16)</td>
<td>10.88 (2.56)</td>
</tr>
<tr>
<td>Competence scales</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>46.90 (8.14)</td>
<td>39.28 (9.02)</td>
</tr>
<tr>
<td>Activities</td>
<td>49.00 (6.26)</td>
<td>45.93 (6.93)</td>
</tr>
<tr>
<td>School</td>
<td>44.51 (9.83)</td>
<td>34.68 (7.87)</td>
</tr>
<tr>
<td>Behavioral syndromes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Withdrawn</td>
<td>55.51 (8.36)</td>
<td>62.08 (9.53)</td>
</tr>
<tr>
<td>Somatic Complaints</td>
<td>56.37 (8.50)</td>
<td>59.96 (8.97)</td>
</tr>
<tr>
<td>Anxious/Depressed</td>
<td>56.63 (9.74)</td>
<td>63.32 (10.96)</td>
</tr>
<tr>
<td>Social Problems</td>
<td>54.56 (7.60)</td>
<td>65.27 (11.12)</td>
</tr>
<tr>
<td>Thought Problems</td>
<td>54.09 (8.00)</td>
<td>61.49 (10.58)</td>
</tr>
<tr>
<td>Attention Problems</td>
<td>56.05 (9.26)</td>
<td>69.52 (9.98)</td>
</tr>
<tr>
<td>Delinquent Behavior</td>
<td>55.50 (7.53)</td>
<td>62.31 (9.46)</td>
</tr>
<tr>
<td>Aggressive Behavior</td>
<td>55.34 (9.85)</td>
<td>65.63 (13.24)</td>
</tr>
</tbody>
</table>

Note: All CBCL scores are T-scores.

Table 2 Results of stepwise logistic regression analysis

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE B</th>
<th>Wald T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Models for probands</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADHD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Problems</td>
<td>.18</td>
<td>.08</td>
<td>4.85</td>
</tr>
<tr>
<td>Attention Problems</td>
<td>.23</td>
<td>.05</td>
<td>19.24</td>
</tr>
<tr>
<td>ODD/CD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Scale</td>
<td>-.23</td>
<td>.10</td>
<td>4.63</td>
</tr>
<tr>
<td>School Scale</td>
<td>-.32</td>
<td>.14</td>
<td>4.91</td>
</tr>
<tr>
<td>Aggressive Behavior</td>
<td>.09</td>
<td>.02</td>
<td>15.77</td>
</tr>
<tr>
<td>Models for siblings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADHD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention Problems</td>
<td>.51</td>
<td>.08</td>
<td>42.54</td>
</tr>
<tr>
<td>ODD/CD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggressive Behavior</td>
<td>.19</td>
<td>.03</td>
<td>37.71</td>
</tr>
</tbody>
</table>

Note: B = unstandardized regression coefficient; SE B = standard error of regression coefficient.

and syndromes significantly predicted the diagnostic conditions except that the Activities scale did not predict ODD/CD.

Unstandardized regression coefficients, standard errors of the unstandardized regression coefficients, and Wald T-test scores for the multivariate stepwise regression analysis are displayed in Table 2. For ADHD, the final proband model included both the Attention Problems and Social Problems syndromes. A model that included the Aggressive Behavior syndrome and the Social and School scales significantly predicted ODD/CD among probands. Results with the sibling sample were similar, albeit more parsimonious. ADHD was significantly predicted by the Attention Problems syndrome and ODD/CD was significantly predicted by the Aggressive Behavior syndrome.

Receiver-Operating Characteristic analysis

Based on results from the logistic regression analysis, models were evaluated with ROC analyses. For ADHD, ROC analyses tested three models for probands: (1) Attention Problems syndrome alone; (2) Social Problems syndrome alone; (3) a combination of both syndromes. Only the Attention Problems syndrome was tested in the sibling sample. In the proband sample, four models were tested for ODD/CD including the Aggressive Behavior syndrome, the Social scale, and the School scale, and the combination of all three scales. The Aggressive Behavior syndrome was tested in the sibling sample. AUC statistics determined the models with the highest discriminatory power. Z-score transformations of AUC scores were used to test differences between models using a formula correcting for the nonindependence of the scores (Hanley & McNeil, 1983).

In the proband sample, the Attention Problems syndrome alone displayed a higher AUC (AUC = .841) than the Social Problems alone (AUC = .791), and Social problems in addition to Attention Problems displayed the highest AUC (AUC = .850). Comparisons between models indicated no significant difference between the Attention Problems syndrome alone versus the Social Problems syndrome alone (z = 1.79, p = ns) or versus the combination of the syndromes (z = .29, p = ns), or the combination of syndromes versus the Social Problems syndrome alone (z = 1.52, p = ns). Thus, results support all three models in the proband sample. The Attention Problems syndrome had significant discriminatory power for predicting ADHD in the sibling sample (AUC = .904). All models were significantly different from a random predictor (AUC = .500).

For the ODD/CD condition, ROC analyses indicated an adequate AUC for the Aggressive Behavior syndrome for probands (AUC = .803). The AUC for the School scale was significantly lower than the Aggressive Behavior syndrome (AUC = .693; z = 2.99, p < .05), whereas the AUC for the Social scale
was not significantly lower than the AUC for the Aggressive Behavior syndrome (AUC = .756; $z = 1.43$, $p = \text{ns}$). The combination of all three scales resulted in the highest AUC (AUC = .829), although this was not significantly greater than the AUC for the Aggressive Behavior syndrome alone ($z = .84$, $p = \text{ns}$). The combined model was significantly better than the Social scale alone ($z = 2.09$, $p < .05$). For siblings, the Aggressive Behavior syndrome displayed high discrimination of ODD/CD (AUC = .838). For ADHD, all models were significantly different from a random predictor. ROC graphs for Aggressive Behavior predicting ODD/CD and Attention Problems predicting ADHD for probands and siblings are displayed in Figures 1 and 2.

Cut-point analysis

Table 3 presents the results of the cut-point analysis for the sibling sample. Cut-point analysis of the proband sample is not included as the predictor syndromes (i.e., Attention Problems and Aggressive Behavior) were used in the recruitment of the sample, resulting in a truncated range of possible scores. The cut-point analyses used the single predictor syndrome with the highest AUC for each diagnostic category: The Attention Problems syndrome was used as the predictor of ADHD, and the Aggressive Behavior syndrome was used as the predictor of ODD/CD. As noted earlier, these cut-points can be evaluated using a variety of criteria in different situations (i.e., research versus clinical work). Overall, optimal cut-points for the siblings were consistent with cutpoints derived in previous analyses. The optimal cut-point for predicting ADHD from the Attention Problems syndrome was a T-score of 55, and the optimal cut-point for predicting ODD/CD from the Aggressive Behavior syndrome was also a T-score of 55.

Discussion

This study replicates previous studies testing the diagnostic accuracy of the CBCL for assessing ADHD and also tested the diagnostic accuracy of the CBCL for assessing ODD with or without CD. Results suggested that the CBCL is an efficient screener for commonly-occurring externalizing disorders in a sample of children with high rates of externalizing problems and in a sibling cross-validation sample. For both diagnostic conditions, single scales proved as effective for predicting the condition as multivariate models. Cut-points derived from QROC...
analyses indicated a T-score of 55 on the Attention Problems and the Aggressive Behavior syndromes effectively minimizes false positives and false negatives in the assessment of ADHD and ODD/CD, respectively.

Past research supports the use of the CBCL as a screener for ADHD, although results have varied with respect to specific syndromes associated with ADHD. Consistent with our expectations, the Attention Problems syndrome effectively discriminated ADHD cases from noncases in both sibling and proband samples. In the proband sample, additional scales improved prediction of ADHD in logistic regression analyses, although not to a significant extent in ROC analyses. Similar to these results, Rey et al. (1992) found the Attention Problems syndrome to have a high AUC for predicting ADHD (AUC = .84) in a clinically referred sample of boys. Chen et al. (1994) found the Attention Problems syndrome predicted ADHD significantly, with an AUC of .961 for probands and .87 for siblings. In that study, all children in the proband sample had been referred for mental health services. Overall, the Attention Problems syndrome has consistently predicted the presence of ADHD in various studies.

The relation between the Attention Problems syndrome and ADHD may seem relatively intuitive to researchers and clinicians familiar with the CBCL. Indeed, a number of the 11 items on the Attention Problems syndrome overlap with DSM-IV criteria for ADHD including: can’t sit still; is restless; hyperactive; and poor schoolwork. Nevertheless, the Attention Problems syndrome differs from the DSM-IV definition of ADHD in a number of important ways. First, a number of items on the syndrome have little to no overlap with DSM-IV criteria for ADHD. For example, the Attention Problems syndrome includes an item on whether the child is nervous, highstrung, or tense. The Attention Problems syndrome also includes only one item assessing impulsivity, whereas the DSM-IV criteria include three items. In addition, the parental directions for the CBCL indicate that responses should only apply to the past 6 months, whereas the DSM-IV requires presence of symptomatology before the age of 7. Thus, the Attention Problems syndrome differs from DSM-IV ADHD in a number of important ways. Most importantly, DSM-IV disorders reflect a top-down approach to classification of psychopathology, whereas the CBCL syndromes reflect the statistical covariance of various behaviors and emotions.

ODD with or without CD was significantly predicted by the Aggressive Behavior syndrome of the CBCL in both proband and sibling samples. AUC values were slightly lower than for ADHD, although these values cannot be directly compared. In the proband sample, two competence scales, social and school competence, added to the discrimination of ODD/CD. Addition of these scales may support CBCL competence scales as a proxy for impairment criteria. Both ODD and CD are associated with impaired social and school functioning (e.g., Hinshaw, 1992). A number of studies have supported a strong relation between ODD and the Aggressive Behavior syndrome and CD and the Delinquent Behavior syndrome (Biederman et al., 1993; Edelbrock & Costello, 1988; Jensen, Salzburg, Richters, & Watanabe, 1993), although studies have not identified specific cut-points on these scales. The Aggressive Behavior syndrome contains quite a few items similar to some of the diagnostic criteria for ODD, but also a number of overt aggressive items such as fighting and bullying that are related to CD criteria. The distinction between CBCL syndromes of Aggressive Behavior and Delinquent Behavior corresponds closely to the overt versus covert aggressive behavior distinction (Frick et al., 1993). The slightly decreased AUC values for the ODD/CD condition versus the ADHD condition may reflect the combination of those two highly-related but distinct conditions. In addition, the Delinquent Behavior syndrome was not a significant predictor of the ODD/CD group indicating that this group may represent overt conduct problems.

Overall, AUC values from all analyses compare favorably to other commonly-used medical tests including the dexamethosone suppression test for predicting MDD which has an AUC of .79 (Mossman & Sonozia, 1989). AUCs are higher than findings for models constructed to predict violent behavior (AUCs = .73 to .76) from the Violence Risk Appraisal Guide (Rice & Harris, 1995). Analyses replicate the diagnostic accuracy of the Attention Problems syndrome for predicting ADHD and provide support for the Aggressive Behavior syndrome in predicting ODD/CD.

Consistent with previous findings, optimal cut-points on the Attention Problems and Aggressive Behavior syndromes in predicting the respective DSM diagnoses ranged from .5 to 1 standard deviation above the mean. Optimal cut-points refer to those scores that minimize false positives and false negatives, rather than one or the other. In practice, the relative cost of a false positive versus a false negative may not be equal. For example, screening in a clinical setting can tolerate a high false positive rate, if all true positive cases are identified. Clinicians can then utilize more time-intensive methods to identify true positive cases. In that setting, the screening instrument precludes administration of time-intensive assessment devices for individuals with a low possibility of having a specific diagnosis.

For basic research, studies often tolerate a high level of false negatives, if the resulting sample is free from false positives.

Sensitivity and specificity of a given test vary across samples based on the base rate of the disorder within the population and referred vs. nonreferred status (Kraemer, 1992). In this study, we included a proband and sibling sample. In the
proband group, ADHD participants included children referred for clinical attention and children recruited through community advertisements, who might or might not be receiving mental health services. Children with ADHD in the sibling sample served as a sample of children not included based on their referral status. As referral to clinical services is often associated with greater impairment, ADHD subjects in the sibling sample might be expected to display less severe symptomatology, although siblings of children with ADHD are at increased risk for psychopathology. In the current sample, children with ADHD in the proband sample displayed higher scores on the Attention Problems syndrome than children in the sibling sample with ADHD. Thus, differences in results across groups are likely attributable to differing sample characteristics. The similarity of ROC results across proband and sibling samples provides preliminary validation for these results, despite these sampling differences. In addition, our findings are consistent with a previous ROC analysis of the CBCL for predicting ADHD (Chen et al., 1994).

Interestingly, cut-points identified on the 1991 CBCL form for the borderline (T-score = 67) and clinical range (T-score = 70), although not developed for the purpose of diagnosing either ADHD or OD/CD, were substantially higher than those identified by the current study. The CBCL cut-points were so designated for identifying the 95th and 97th percentile and predicting children that are referred for mental health services (Achenbach, 1991). The recent version of the CBCL has revised the borderline cut-point down to a T-score of 65 to increase the number of impaired children identified as needing help (Achenbach & Rescorla, 2001). Nevertheless, any rote tendency to utilize such cut-points for identifying ADHD or ODD/CD may result in an unacceptable level of false negative cases; thus, CBCL cut-points may underdiagnose children that meet DSM-IV criteria for ADHD or ODD/CD, respectively.

A number of limitations in terms of the generalizability of the results are noteworthy. The proband sample was recruited to include four groups of children with various levels of attention and aggressive behavior for genetic analyses unrelated to the current analysis. Thus, an artificial bimodal distribution on two CBCL syndromes was created. This range limitation precluded the ability to evaluate various cut-points on these two syndromes with the proband sample. Thus, cut-points derived in our analyses were limited to those derived from the sibling sample. Next, a very low percentage of individuals recruited from an outpatient psychiatric clinic successfully completed the study resulting in a low overall participation rate. Two factors may have contributed to this low participation rate. First, many families were contacted from the clinic that were not currently receiving services through the clinic. A number of these families were not even able to be contacted based on antiquated contact information. Second, many of the families from the clinic refused to participate in the study based on the need for involvement from the whole family and the need to provide a saliva sample for genetic analyses. Thus, the families that agreed to participate in the study were making a substantial commitment and participation rates would be expected to be quite low. Next, both probands and siblings display a higher rate of comorbidity between ADHD and ODD/CD than would be observed in a community sample, perhaps related to recruitment of a significant portion of the sample from an outpatient psychiatric clinic. The high rates of comorbidity are also related to the inclusion criteria pertaining to CBCL T-scores.

This study supports the use of the CBCL as a screening instrument for commonly-occurring externalizing disorders. In particular, the Attention Problems syndrome is an effective screener for a DSM-IV diagnosis of ADHD, and the Aggressive Behavior syndrome is an effective screener for the diagnoses of ODD and/or CD. The CBCL is a time and cost-efficient means to assess for various behavioral and emotional concerns. Completion time of the CBCL involves between 10 and 15 minutes. The efficiency of CBCL administration and accuracy of the CBCL for predicting DSM-IV diagnoses also supports the use of this measure in settings in which mental health concerns may be secondary, such as schools and pediatric clinics.

Author note
James J. Hudziak, Department of Psychiatry; William Copeland, Department of Psychology; Catherine Stanger, Department of Psychiatry; Martha Wadsworth, Department of Psychiatry.

Acknowledgement
This work was supported by NIH grant MH01265.

Correspondence to
James Hudziak, University of Vermont, Given Building, Room B229, Burlington, VT 05405. Email: james.hudziak@zoo.uvm.edu

References


Manuscript accepted 26 November 2003