JOURNAL OF CHILD AND ADOLESCENT PSYCHOPHARMACOLOGY
Volume 19, Number 1, 2009

Pp. 23–30
DOI: 10.1089/cap.2008.067

The Child Behavior Checklist (CBCL) and the CBCL-Bipolar Phenotype Are Not Useful in Diagnosing Pediatric Bipolar Disorder

Rasim Somer Diler, M.D.,[1] Boris Birmaher, M.D.,[1] David Axelson, M.D.,[1] Ben Goldstein, M.D.,[1] MaryKay Gill, M.S.N.,[1] Michael Strober, Ph.D.,[2] David J. Kolko, M.D.,[1] Tina R. Goldstein, Ph.D.,[1] Jeffrey Hunt, M.D.,[3] Mei Yang, B.Sci.,[1] Neal D. Ryan, M.D.,[1] Satish Iyengar,[4] Ronald E. Dahl, M.D.,[1] Lorah D. Dorn, M.D.,[5] and Martin B. Keller, M.D.[3]

Abstract

Objectives: Previous studies have suggested that the sum of Attention, Aggression, and Anxious/Depressed subscales of Child Behavior Checklist (CBCL-PBD; pediatric bipolar disorder phenotype) may be specific to pediatric bipolar disorder (BP). The purpose of this study was to evaluate the usefulness of the CBCL and CBCL-PBD to identify BP in children <12 years old.

Methods: A sample of children with BP I, II, and not otherwise specified (NOS) ($n = 157$) ascertained through the Course and Outcome for Bipolar Disorder in Youth (COBY) study were compared with a group of children with major depressive/anxiety disorders (MDD/ANX; $n = 101$), disruptive behavior disorder (DBD) ($n = 127$), and healthy control (HC) ($n = 128$). The CBCL T-scores and area under the curve (AUC) scores were calculated and compared among the above-noted groups.

Results: Forty one percent of BP children did not have significantly elevated CBCL-PBD scores ($\geq$2 standard deviations [SD]). The sensitivity and specificity of CBCL-PBD ≥ 2 SD for diagnosis of BP was 57% and 70–77%, respectively, and the accuracy of CBCL-PBD for identifying a BP diagnosis was moderate (AUC = 0.72–0.78).

Conclusion: The CBCL and the CBCL-PBD showed that BP children have more severe psychopathology than HC and children with other psychopathology, but they were not useful as a proxy for *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV) diagnosis of BP.

Introduction

Pediatric bipolar disorder (BP) is now recognized as a significant public health problem, noteworthy for its severity and chronicity of symptoms (Pavuluri et al. 2005; Youngstrom et al. 2004; Birmaher et al. 2006). Because of their ease of administration, good psychometric properties, and cross-cultural validation, the Achenbach Rating Scales (e.g., parent-rated Child Behavior Checklist; CBCL) (Achenbach and Rescorla 2001) have been used widely in studies of youths with varying types of psychopathology, pediatric BP included (Biederman et al. 1995; Kahana et al. 2003; Mick et al. 2003; Faraone et al. 2005; Youngstrom et al. 2005).

Biederman and colleagues (Biederman et al. 1995) proposed that the sum of the Attention, Aggression, and Anxious/Depressed subscales of the CBCL, pediatric BP phenotype, (CBCL-PBD; alternatively referred to as the juvenile BP phenotype [CBCL-JBD]) may be specific for the diagnosis of pediatric BP. A meta-analysis of seven studies (Biederman et al. 1995; Biederman et al. 1996; Carlson and Kelly 1998; Carlson et al. 1998; Geller et al. 1998; Hazell et al. 1999; Dienes et al. 2002) showed that in comparison with children with attention-deficit/hyperactivity disorder (ADHD) children with BP had significantly higher scores on these three CBCL subscales (Mick et al. 2003). Efforts to establish the specificity of these scales in identifying children with BP have thus far met with mixed results.

[1]Department of Psychiatry, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania.
[2]Department of Psychiatry, UCLS Semel Institute, Los Angeles, California.
[3]Department of Psychiatry, Brown University, Providence, Rhode Island.
[4]Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania.
[5]Division of Adolescent Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio.

The projects described were supported by Grant Numbers 2R01MH059929-06A2 (Dr. Birmaher), P01MH41712 (Dr. Ryan), and 5R01MH057727 (Dr. Kolko) from the National Institute of Mental Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

The CBCL-PBD was reported as a highly efficient way of identifying children with BP in a clinical sample (Faraone et al. 2005) and was used as proxy research definition of BP in a nonclinical sample (Hudziak et al. 2005). However, others have reported that the CBCL poorly identified pediatric BP and suggested that the high scores in the CBCL-PBD might be due to the symptom severity, co-morbidity, or functional impairment in subjects with BP (Kahana et al. 2003; Youngstrom and Youngstrom 2005). Methodological limitations of previous studies include different approaches to clinical-diagnostic ascertainment, small sample sizes with a preponderance of male ADHD subjects (Kahana et al. 2003; Mick et al. 2003), lack of control for co-morbidities or other clinical and demographic variables, and the absence of the full spectrum of BP phenotypes. Two recent studies underscore these points in a large community (Volk and Todd 2007) and clinical (Holtmann et al. 2007) samples noting that whereas the CBCL-PBD profile was associated with suicidality and functional impairment and was moderately heritable, no children with high CBCL-PBD scores (T-scores ≥ 2 standard deviations [SD]) fulfilled diagnostic criteria for BP.

The goal of the present study was to evaluate the diagnostic performance of the CBCL subscales (e.g., Total, Externalizing, Attention, Aggression, and Anxious/Depressed) and the CBCL-PBD phenotype in a large group of clinically ascertained children diagnosed with BP spectrum disorders (BP I, II, and not otherwise specified [NOS]), unipolar major depressive/anxiety (MDD/ANX) disorders, disruptive behavior disorders (DBD), and a group of healthy control (HC) children <12 years old. We sought to determine first, whether or not children with BP have higher scores on the CBCL compared to MDD/ANX, DBD, and HC groups, and, second, if the CBCL-PBD phenotype is useful in identifying BP youth.

Methods

Clinical diagnoses

To evaluate whether the CBCL scores are specifically higher in children with BP, we compared these children with historical samples of children with non-BP and HC. The non-BP and HC children were recruited and evaluated using a methodology similar to the one used in our study, including *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV) diagnostic criteria (American Psychiatric Association 2000), the CBCL (Achenbach and Rescorla 2001), and the Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime Version (K-SADS-PL) (Kaufman et al. 1997). Moreover, two of the investigators (B.B. and D.A.) participated in these studies. The MDD/ANX group did not include children with BP and the DBD group did not include subjects with MDD or BP. Participants in the healthy control group were free of any lifetime episode of any major psychiatric disorder. The methods and instruments used to ascertain and evaluate the BP youth have been described in detailed elsewhere (Axelson et al. 2006).

An operationalized definition of BP NOS was used to diagnose children with this subtype of BP who did not meet the DSM-IV criteria for BP I or BP II, but had a distinct period of abnormally elevated, expansive, or irritable mood plus the following minimum criteria: (1) two DSM-IV manic symptoms (three if the mood is irritability only) that were clearly associated with the onset of abnormal mood; (2) a clear change in functioning; (3) mood and symptom duration of a minimum of 4 hours within a 24-hour period for a day to be considered meeting the diagnostic threshold; and (4) a minimum of 4 days (not necessarily consecutive) meeting the mood, symptom, duration, and functional change criteria over the subject's lifetime, which could be two 2-day episodes, four 1-day episodes, or another variation. Using these criteria, we have previously shown that children and adolescents meeting these more strictly defined BP-NOS criteria have similar, but less severe, clinical pictures, co-morbid disorders, family history, and longitudinal outcome than the BP I subjects (Axelson et al. 2006; Birmaher et al. 2006). Moreover, about 25% of these youth diagnosed with BP NOS converted into BP I or BP II (Birmaher et al. 2006).

Subjects

A subgroup of 157 children younger than 12 years old (9.4 ± 1.5) from the Course and Outcome of Bipolar Youth (COBY) was included in this study (Axelson et al. 2006; Birmaher et al. 2006). Seventy nine of these children had BP I, 3 had BP II, and 75 had BP NOS. About 85% of the subjects with BP had significant manic and/or depressive symptoms during the month before the assessment, but they were not required to meet the criteria for a manic, mixed, or hypomanic episode at the time of the study intake (Axelson et al. 2006; Birmaher et al. 2006). For this study, only children, and not the adolescents, with BP were included because existing CBCL/BP studies have mainly included children younger than 12 years old (Mick et al. 2003). The MDD/ANX ($n = 101$, 10.0 ± 1.2 years old), DBD ($n = 127$, 8.7 ± 1.6 years old), and HC ($n = 128$, 9.5 ± 1.6 years old) children were part of the Neurobehavioral Changes in Pediatric Affective Disorder (Birmaher et al. 2000) and the Resources to Enhance the Adjustment of Children-REACH (Kolko et al., 2008) studies at the University of Pittsburgh. The MDD/ANX group included 46 children with MDD, 33 with anxiety disorder without MDD (28 children with generalized anxiety disorder, 2 children with panic disorder, 1 child with separation anxiety disorder, 1 child with social phobia, and 1 child with both generalized anxiety disorder and separation anxiety disorder) and 22 with co-morbid MDD and anxiety disorders (9 children with generalized anxiety disorders, 9 children with separation anxiety disorder, 2 children with panic disorder, 2 children with generalized anxiety disorders and social phobia). The DBD group consisted of 24 children with conduct disorder (CD) and 103 children with oppositional defiant disorder (ODD). Informed consent was obtained from the study subject's parent or guardian before initiation of the assessment.

As shown in Table 1, there were significant between-group differences in age, sex, race, and family intactness. Children in the MDD/ANX group were the oldest and the DBD children were the youngest among four groups. There were more male subjects in all groups and its prevalence of male subjects was highest in the DBD group and lowest in the MDD/ANX group. Except for the DBD group, race was significantly nonwhite in the study groups. General level of psychosocial functioning as measured through the Global Assessment Scale for Children (C-GAS) (Shaffer et al. 1983) was significantly lower in the BP group when compared to the other three groups. The prevalence of an ADHD diagnosis was significantly higher in the BP and DBD groups as compared

Table 1. Demographic and Clinical Variables in Bipolar (BP), Major Depression/Anxious (MDD/ANX), Disruptive Behavior Disorder (DBD), and Health Control (HC) Subjects

	BP[a] (n = *157*)	*MDD/ANX*[a] (n = *101*)	*DBD*[a] (n = *127*)	*HC*[a] (n = *128*)	*Statistics*	p *Value*
Age	9.4 ± 1.5^{b}	10.0 ± 1.2^{c}	8.7 ± 1.6^{d}	9.5 ± 1.6^{b}	KW $\chi^2 = 4\ 0.4$	<0.001
Sex (% Male)	65.6^{b}	52.5^{c}	85.0^{d}	65.6^{b}	$\chi^2 = 28.8$	<0.001
Race (% White)	83.4^{b}	80.2^{b}	47.2^{c}	74.2^{b}	$\chi^2 = 51.9$	<0.001
SES	3.3 ± 1.2	N/A	3.1 ± 1.0	3.3 ± 1.0	KW $\chi^2 = 2.6$	0.27
Intactness of Family (%)	38.2^{b}	40.7^{b}	N/A	94.5^{c}	$\chi^2 = 55.3$	<0.001
C-GAS Current	56.3 ± 12.5^{b}	63.0 ± 10.3^{c}	72.0 ± 7.0^{d}	89.2 ± 5.6^{e}	KW $\chi^2 = 261.5$	<0.001
C-GAS Most Serious Lifetime	39.6 ± 10.1^{b}	60.6 ± 11.1^{c}	N/A	87.9 ± 9.8^{d}	F = 139.5	<0.001
Lifetime ADHD	72.6^{b}	18.8^{c}	76.4^{c}	0.0^{d}	$\chi^2 = 231.8$	<0.001
Lifetime CD	7.0^{b}	0.0^{c}	18.9^{d}	0.0^{c}	$\chi^2 = 45.9$	<0.001
Lifetime Anxiety	41.4^{b}	$32.7^{b,f}$	15.0^{c}	0.0^{d}	$\chi^2 = 77.6$	<0.001

[a]% Yes or mean ± SD.
[b–e]Different superscripts indicate pairwise $p \leq 0.05$.
[f]Anxiety disorder without MDD.
Abbreviations: BP = Bipolar disorder; MDD/ANX = major depressive/anxiety disorders; DBD = disruptive behavior disorders; HC = healthy controls; CD = conduct disorder; ADHD = attention-deficit/hyperactivity disorder; SES = Socioeconomic Status on Hollingshead Four Factor Index Scale; C-GAS = Global Assessment Scale for Children; N/A = data not available.

with the MDD/ANX and HC groups. The prevalence of anxiety disorders was significantly higher in the BP and MDD/ANX groups as compared with the DBD and HC groups.

Clinical measures

Research interviewers were trained to reliably administer the K-SADS-PL (Kaufmann et al. 1997). The results of each interview were reviewed by a child psychiatrist or psychologist for diagnostic consensus. As reported elsewhere, there was high reliability for differentiating BP from non-BP subjects (K = 0.90) and from the BP diagnostic subtypes (K = 0.79). For the nonmood disorders, values were 0.80 or higher (Axelson et al. 2006).

The CBCL has 11 subscales, including Delinquent Behavior, Aggressive Behaviors, Withdrawn, Somatic Complaints, Anxious/Depressed, Social Problems, Thought Problems, Attention Problems, Externalizing Problems (includes Delinquent and Aggressive Behaviors), Internalizing Problems (includes Withdrawn, Somatic Complaints, and Anxiety/Depressed Problems), and Total Problems (includes Externalizing, Internalizing, Social, Thought, and Attention Problems) (Achenbach and Rescorla 2001), and cut-off scores of 70 (2 SD above normal level) have been recommended as clinically meaningful thresholds for a deviation from age-and sex-matched healthy children. In accordance with previous studies (Mick et al. 2003; Faraone et al. 2005), the CBCL-PBD phenotype has been defined as the sum of Anxiety/Depression, Attention Problems, and Aggressive Behaviors subscales on the CBCL, and a cut-off score of 210 is 2 SD above normal level. Subjects included from the BP and DBD studies were administered the 2001 version of CBCL/6-18 (Achenbach and Rescorla 2001) and subjects from the MDD/ANX study were administered 1991 version of CBCL/4-18 (Achenbach 1991), both of which obtain ratings of 118 specific behavioral–emotional problem items. The data were modified to be in accordance with the 2001 version after consultation with the publisher.

The socioeconomic status for all subjects was measured using the Hollingshead four-factor scale (Hollingshead 1975).

Statistical analysis

Kruskal–Wallis and analysis of variance (ANOVA) tests for nonparametric and parametric data were used, respectively. This was followed by chi-squared and F tests to compare demographics, clinical variables (co-morbid lifetime ADHD, anxiety, and conduct disorders and C-GAS), and CBCL T scores including the CBCL-PBD phenotype between BP I, BP II, and BP NOS as well as between BP, MDD/ANX, DBD, and HC subjects. There were no significant differences in the CBCL and CBCL-PBD results among children with BP I, BP II, and BP NOS. Thus, for the comparison with other groups, these subtypes of BP are combined into one group. CBCL analyses were adjusted for significant demographic variables and global functioning. Logistic regression was used to calculate the odds ratios (ORs) reflecting how well each CBCL score (eight subscales and internalizing, externalizing, and total T scores) and CBCL-PBD predicted bipolar youth, while controlling for significant demographic variables and global functioning. Sensitivity and specificity of CBCL-PBD for BP diagnosis were analyzed using the receiver operating characteristics (ROC) analyses and area under the curve (AUC) scores. ROC analysis is a valuable tool to evaluate diagnostic tests and predictive models (Biederman et al. 2005; Zou et al. 2007). An ROC curve is a plot of sensitivity on the *y* axis against false positive rate (1-specificity) on the *x* axis (Park et al. 2004) and can be used to select the optimal cutting scores under a variety of clinical circumstances, balancing the tradeoffs between sensitivity and specificity (Zou et al. 2007).

The AUC score is an overall summary of diagnostic accuracy and provides useful information about how a test performs across different cutting scores (Park et al. 2004; Zou et al. 2007). The AUC equals 0.5 when the ROC curve corresponds to random chance and 1.0 for perfect accuracy (Zou et al. 2007). The AUC scores can also be interpreted as the probability that a patient with BP would score higher on the CBCL

Table 2. Child Behavior Checklist (CBCL) and CBCL–Pediatric Bipolar Disorder (CBCL-PBD) Profiles in Biopolar (BP), Major Depression/Anxious (MDD/ANX), Disruptive Behavior Disorder (DBD), and Healthy Control (HC) Subjects

CBCL T scores (n = *513*)	*BP* (n = *157*)	*MDD/ANX* (n = *101*)	*DBD* (n = *127*)	*HC* (n = *128*)	*Statistics*	p *Value*
Total problems	72.9 ± 5.6^a	65.3 ± 8.9^b	68.7 ± 7.9^b	43.8 ± 9.3^c	$KW\chi^2 = 291.7$	<0.001
Internalizing	69.5 ± 8.7^a	69.9 ± 9.1^a	61.0 ± 10.2^b	45.3 ± 8.5^c	$F = 201.3$	<0.001
Externalizing	72.5 ± 6.5^a	62.7 ± 10.2^b	71.1 ± 7.5^a	45.3 ± 9.0^c	$KW\chi^2 = 290.0$	<0.001
Anxious/Depressed	69.6 ± 10.2^a	65.6 ± 10.2^a	61.9 ± 9.6^b	54.9 ± 9.3^c	$F = 55.5$	<0.001
Attention Problems	68.0 ± 9.3^a	58.2 ± 7.6^b	67.4 ± 9.1^a	56.6 ± 9.4^c	$F = 58.6$	<0.001
Aggressive Behaviors	78.2 ± 10.1^a	65.0 ± 11.6^b	73.4 ± 9.9^c	58.0 ± 11.4^b	$F = 95.8$	<0.001
Withdrawn	65.7 ± 9.1^a	67.7 ± 8.9^a	59.2 ± 9.0^b	55.1 ± 7.4^b	$KW\chi^2 = 150.4$	<0.001
Somatic complaints	64.2 ± 9.0^a	65.6 ± 9.8^a	58.7 ± 8.9^b	54.4 ± 6.6^c	$KW\chi^2 = 118.5$	<0.001
Social Problems	68.5 ± 9.2^a	62.1 ± 8.9^b	63.0 ± 9.7^b	55.2 ± 7.8^c	$KW\chi^2 = 138.0$	<0.001
Thought Problems	70.3 ± 8.5^a	59.2 ± 7.7^b	61.9 ± 8.8^b	54.2 ± 7.3^c	$KW\chi^2 = 196.2$	<0.001
Rule-breaking behavior	65.1 ± 8.1^a	57.3 ± 7.7^b	67.8 ± 8.3^c	56.7 ± 9.2^a	$F = 55.9$	<0.001
CBCL-PBD	215.9 ± 21.6^a	188.7 ± 24.0^b	202.7 ± 23.1^c	169.4 ± 26.8^d	$F = 96.1$	<0.001

[a–d]Different superscripts indicate pairwise $p \leq 0.05$.

Abbreviations: BP = Bipolar disorder; MDD/ANX = major depressive/anxiety disorders; DBD = disruptive behavior disorders; HC = healthy controls; CBCL = Child Behavior Checklist; CBCL-PBD (Child Behavior Checklist–Pediatric Bipolar Phenotype) = Anxious/Depressed + Attention Problems + Aggressive Behaviors.

than would a patient without BP, when both are randomly selected from the sample. In accordance with the literature, AUC scores were interpreted as excellent (1.00–0.90), good (0.80–0.90), moderate (0.70–0.80), or poor (0.60–0.70) accuracy of the measure (CBCL) for predicting the (BP) diagnosis (Swets 1988). To evaluate how the CBCL would function as a screening test in a clinical population, we calculated optimal cut-off scores that yield the maximum value of the equation of the "sensitivity + specificity − 1." Depending on the instrument used to ascertain diagnoses, the referral source, and the clinic's specialization, the prevalence of BP in clinical populations is between 0.6 and 15% (Pavuluri et al. 2005). On the basis of this information, we estimated a BP prevalence rate of 8% for a psychiatric outpatient clinical population and calculated negative predictive values (NPV) and positive predictive values (PPV) of CBCL subscales. All p values are based on two-tailed tests with $\alpha 2 = 0.05$.

Results

Subscales of the CBCL

Without any adjustments for between-group differences in demographics and C-GAS, the BP group had significantly higher CBCL Total Problem, CBCL-PBD, and Anxious/Depressed, Aggressive Behaviors, Withdrawn, Social Problems, and Thought Problems compared with the MDD/ANX, DBD, and HC groups (all p values < 0.001). The CBCL Rule-breaking and Withdrawn Problems were the only two subscales significantly higher in the DBD and MDD/ANX groups, respectively, when compared with the BP group. The CBCL Attention and Externalizing Problems were significantly higher in the BP and DBD groups as compared with the MDD/ANX and HC groups and the CBCL Somatic and Internalizing Problems were significantly higher in the BP and MDD/ANX groups compared with the DBD and HC groups (all p values < 0.001). After adjusting for between-group differences in demographic variables and C-GAS, the analysis showed that the CBCL Anxious/Depressed (previously higher in the BP group) and Withdrawn Problems (previously higher in the MDD/ANX group) were not different between the BP and MDD/ANX groups (Table 2).

CBCL-PBD phenotype

There were significantly more BP children (58.6%) that had elevated CBCL-PBD scores ≥ 2 SD compared with the MDD/ANX (22.8%), DBD (37.0%), and HC (10.9%) children. However, 41.4% of the BP children did not have significantly elevated CBCL-PBD scores. As depicted in Table 3, when only subjects with psychopathology were analyzed (BP + MDD/ANX + DBD), using the cut off of ≥ 2 SD, the sensitivity was 57% and the specificity was 70% for the CBCL-PBD. When healthy controls were also included in the above-noted analysis, sensitivity and specificity of CBCL-PBD scores ≥ 2 SD were 57% and 77%, respectively. In subjects with psychopathology, the PPV was 14% and NPV was 95% for the CBCL-PBD scores ≥ 2 SD. The optimal cut off score for CBCL-PBD was ≥ 1.5 SD (≥ 195), and in comparison to scores ≥ 2 SD it provided higher sensitivity (83% vs. 57%, respectively) and NPV (97% vs. 95%, respectively) but lower specificity (50% vs. 70%, respectively) and PPV (13% vs. 14%, respectively). The optimal cut-off scores were lower for several CBCL subscales (e.g., Total, Internalizing, Externalizing, Anxious/Depressed, Aggressive Behaviors, Withdrawn, and Social Problems) when HC group was included in the analysis.

Logistic regression was used to analyze how well each CBCL scored, and the CBCL-PBD predicted BP youth in subjects with psychopathology and in HC plus subjects with psychopathology. After controlling for between-group differences in demographic variables and C-GAS, only Aggressive Behaviors (OR = 1.04, 95% confidence interval [CI] = 1.00–1.08, $p = 0.03$) and Thought Problems (OR = 1.12, 95% CI = 1.08–1.17, $p < 0.001$) were statistically significant in our study groups (BP + MDD/ANX + DBD and BP + MDD/ANX + DBD + HC) when the eight CBCL subscales were analyzed. In subjects with psychopathology, the CBCL-PBD (OR = 1.03, 95% CI = 1.01–1.04, $p = 0.04$) and Thought

Table 3. Child Behavior Checklist (CBCL) and Area under the Curve (AUC) Scores for Predicting BP Diagnosis in All Subjects and in Subjects with Psychopathology

	Healthy controls + subjects with psychopathology							*Subjects with psychopathology*						
CBCL n = *513*	*T scores*		*Sensitivity*	*Specificity*	*PPV*	*NPV*	*AUC*	*T scores*		*Sensitivity*	*Specificity*	*PPV*	*NPV*	*AUC*
Total problems	≥2 SD		77	68	17	97	**0.81**	≥2 SD		77	51	12	96	**0.71**
	Optimal cut off	**68**	85	64	17	98		Optimal cut off	**74**	52	79	18	95	
Internalizing	≥2 SD		50	76	15	95	**0.74**	≥2 SD		50	62	10	93	**0.62**
	Optimal cut off	**62**	86	59	15	98		Optimal cut off	**64**	79	46	11	96	
Externalizing	≥2 SD		72	68	16	97	**0.78**	≥2 SD		72	51	11	95	**0.66**
	Optimal cut off	**62**	94	49	14	99		Optimal cut off	**68**	74	44	10	95	
Anxious/ depressed[a]	≥2 SD		47	80	17	95	**0.74**	≥2 SD		47	74	14	94	**0.67**
	Optimal cut off	**59**	92	46	13	99		Optimal cut off	**64**	68	59	13	96	
Attention Problems[a]	≥2 SD		31	81	12	93	**0.70**	≥2 SD		31	76	10	93	**0.63**
	Optimal cut off	**59**	90	48	13	98		Optimal cut off	**58**	90	36	11	98	
Aggressive Behaviors[a]	≥2 SD		77	62	15	97	**0.78**	≥2 SD		77	50	12	96	**0.71**
	Optimal cut off	**66**	90	52	14	98		Optimal cut off	**78**	51	78	17	95	
Thought Problems	≥2 SD		58	88	30	96	**0.84**	≥2 SD		57	85	25	96	**0.79**
	Optimal cut off	**66**	76	76	22	97		Optimal cut off	**66**	76	69	18	97	
CBCL-PBD	≥2 SD[b]	**≥210**	57	77	18	95	**0.78**	≥2 SD[b]	≥210	57	70	14	95	**0.72**
	Optimal cut off	**195**	84	61	16	98		Optimal cut off	196	83	50	13	97	

[a]CBCL-PBD (Child Behavior Checklist–Pediatric Bipolar Profile) = Anxious/Depressed + Attention Problems + Aggressive Behaviors.
[b]T scores ≥ 2 SD is ≥ 210 for CBCL-PBD and ≥ 70 for all other subscales.
All Subjects = BP + MDD/ANX + DBD + HC Groups; Subjects with Psychopathology = BP + MDD/ANX + DBD Groups.
Abbreviations: CBCL = Child Behavior Checklist; MDD/ANX = major depressive/anxiety disorders; DBD = disruptive behavior disorders; BP = bipolar disorder; HC = Healthy Controls; PPV = positive predictive value, NPV = negative predictive value (NPV and PPV were calculated using an estimated prevalence of 8% BP in an outpatient psychiatric population).

Problems (OR = 1.12, 95% CI = 1.07–1.17, $p < 0.001$) were the only significant scores when the CBCL-PBD and the five remaining CBCL subscales (Delinquent Behavior, Withdrawn, Somatic Complaints, Social Problems, Thought Problems) were included in the analysis. The OR was 1.04 (95% CI = 1.02–1.05, $p < 0.001$) when the sum of Thought Problems and the CBCL-PBD scores were included in the analysis. When the CBCL Externalizing, Internalizing, and Total Scores were analyzed, only Total Scores significantly predicted the bipolar cases (OR = 1.14, 95% C = 1.00–1.30, $p = 0.04$).

When only subjects with psychopathology (BP + MDD/ANX + DBD) were included in the ROC analysis, the accuracy of predicting BP spectrum disorders was in the moderate range for the CBCL Thought Problems (AUC = 0.79), CBCL-PBD (AUC = 0.72), Total Problems (AUC = 0.71), and Aggressive Behaviors (AUC = 0.71) (Table 3). The AUC scores were higher when healthy subjects were included in the analysis (Fig. 1) and the accuracy of predicting BP spectrum disorders was in good range for the CBCL Thought Problems (AUC = 0.84) and Total Problems (AUC = 0.81) (Table 3).

Discussion

The main goal of this study was to evaluate the usefulness of the CBCL and more specifically the CBCL-PBD as a diagnostic tool for identifying pediatric BP. Our study showed that there were no significant differences in the scores of the CBCL and the CBCL-PBD between the BP I, II, and NOS groups and that BP children have more severe psychopathology as measured by the CBCL or the CBCL-PBD than healthy controls and children with other psychopathology. Aggressive Behaviors, Thought Problems, Total Scores, and the CBCL-PBD were more common in children with BP than other disorders, but ORs for these disorders were not higher than 1.14. Moreover, using the ROC analyses, the CBCL and the CBCL-PBD did not reliably distinguish between children with BP and those with MDD/ANX, DBD, and HC. That is, the specificity of CBCL-PBD was 70–77% for scores ≥ 2 SD and 50–61% for scores ≥ 1.5 SD (optimal cut off) for the BP diagnosis and the accuracy of CBCL-PBD for identifying BP diagnosis was moderate (AUC = 0.72–0.78). Thus, the CBCL-PBD phenotype was more frequently present, but not specific to BP children. In addition, when only children with psychopathology were included in the analyses, the AUC scores for the CBCL-PBD were even lower. No CBCL subscales were in the good range for identifying BP diagnosis among children with psychopathology, and only the CBCL Thought and Total Problems were in the good range when healthy children were included in the analysis.

Before discussing the above-noted results, the following limitations of the study need to be considered. The results reported only apply to children younger than 12 years old and

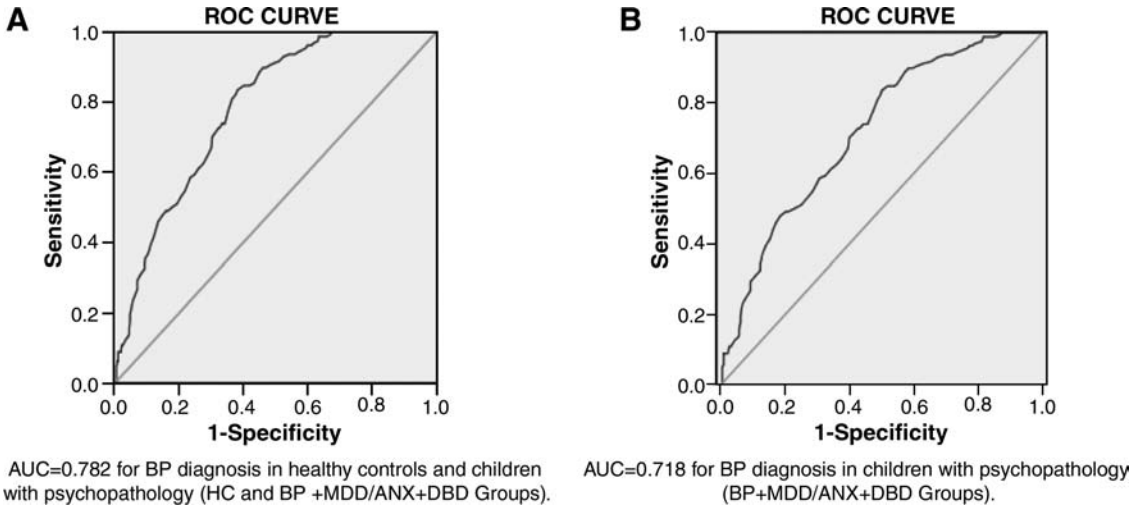

FIG. 1. Child Behavior Checklist–Pediatric Bipolar Disorder (CBCL-PBD) phenotype and area under the curve (AUC) scores for identifying bipolar diagnosis in all subjects and in subjects with psychopathology. All Subjects = BP + MDD/ANX + DBD + HC Groups, Subjects with Psychopathology = BP + MDD/ANX + DBD Groups.
Abbreviations: ROC = Receiver operating characteristics; BP = bipolar disorder; MDD/ANX = major depressive/anxiety disorders; DBD = disruptive behavior disorders; HC = healthy controls.

the data were obtained retrospectively. Because the study was carried out with a clinical sample, our results may not apply to children in community settings. Although the sizes of the BP I and BP NOS samples were relatively large, we did not have robust statistical power to detect small between-BP-subgroup differences (Axelson et al. 2006). Although nearly all subjects had significant manic and/or depressive symptoms during the month before intake, a small proportion of the subjects were euthymic at intake. This may have identified a cohort of subjects whose illness was less chronic or severe than bipolar subjects during a full-criteria episode. On the other hand, the CBCL identifies psychopathology for the past 6 months. However, the different time frames used by CBCL and lifetime clinical diagnosis of BP in identifying psychopathology have been identical to all reported CBCL and BP studies (Kahana et al. 2003; Mick et al. 2003; Faraone et al. 2005; Youngstrom et al. 2005; Diler et al. 2007).

Similar to the other studies, we found higher Total Problems, Anxiety/Depressed, Aggressive Behaviors, Social Problems, Thought Problems, and CBCL-PBD T scores in BP children compared to the depressed/anxious, disruptive behavior disorder, and healthy control children (Table 2). Faraone and colleagues reported that, for the lifetime diagnosis of BP ($n = 22$) in ADHD children ($n = 471$), the sensitivity of the CBCL-PBD ≥ 2 SD was 64% and the specificity was 93% (Faraone et al. 2005). In the same study, the sensitivity was 14% and the specificity was 98% for the lifetime diagnosis of BP ($n = 14$) in the siblings ($n = 410$) of ADHD children. In our study (157/513 children with BP), the sensitivity of the CBCL-PBD scores of ≥ 2 SD was 57%. In comparison with Faraone and colleagues (Faraone et al. 2005), we found lower specificity (93–98% vs. 70–77%, respectively), PPV (20–31% vs. 14–18%, respectively), and NPV (97–98% vs. 95%, respectively) of the CBCL-PBD scores ≥ 2 SD for the diagnosis of BP. Moreover, 41.4% of BP children did not have significantly higher (≥ 2 SD) CBCL-PBD scores and ORs for the CBCL and CBCL-PBD were not higher than 1.14. In this study, the above results do not support the use of the CBCL-BPD as a proxy for the BP diagnosis.

Kahana and colleagues suggested the possibility of high sensitivity and high negative predictive value of the CBCL when screening the BP diagnosis (Kahana et al. 2003). Similarly, there were more BP children in our study who had significant CBCL-PBD scores (≥ 2 SD) compared to the MDD/ANX, DBD, and HC children and the sensitivity of the CBCL-PBD for the BP diagnosis in our study was 83–84% for scores ≥ 1.5 SD (optimal cut-off score) and 57% for scores ≥ 2 SD. The sensitivity and NPV of the CBCL subscales' optimal T-scores were as high as 90% and 99%, respectively, suggesting that CBCL may be useful for ruling out BP diagnosis (Table 3). However, the NPV is good only in populations where the prevalence of BP is low (where the pretest rate of nonbipolarity is already very high) and the high NPV comes at a high price of false positives and poor accuracy. Because the specificity and sensitivity of a test can change according to the population studied and other factors, such as severity of the illness and the rate of the disorder (Youngstrom and Youngstrom 2005), the expected rate of BP in a given sample is very important to consider when interpreting the clinical use of CBCL. Others have suggested that higher rates of clinical diagnoses lower the specificity of the test, because they would increase the number of nonbipolar cases accidentally scoring high on screening tests (Youngstrom et al. 2004). Similarly, we found lower specificity and PPV for BP diagnosis in the group of children that had the higher rate of BP diagnosis (children with psychopathology) as compared to the group of children that had the lower rate of BP diagnosis (healthy children and children with psychopathology).

As described above, the ROC provides the opportunity of plotting the sensitivity of a test as a function of its specificity, moving across all possible thresholds (Park et al. 2004; Zou et al. 2007). Few studies have used this method to evaluate the diagnostic efficacy of the CBCL and the CBCL-PBD scores. Kahana and colleagues studied the CBCL scores in a clinical sample of children with BP, DBD, MDD, and other diagnoses (Kahana et al. 2003). The ROC analysis suggested that the CBCL Aggression and Withdrawn scores had a high level of

predictive validity for BP (AUCs from 0.81 to 0.84). These authors did not analyze CBCL-PBD; however, similar to our results, they concluded that the sensitivity and specificity of prediction was too low for clinical use (Kahana et al. 2003). In earlier studies, Biederman and colleagues (Biederman et al. 2005) and Youngstrom and colleagues (Youngstrom and Youngstrom 2005; Youngstrom et al. 2005) studied the diagnostic performance of the CBCL scores for the diagnosis of BP in youth. They also did not analyze the CBCL-PBD phenotype, but the AUC scores for the CBCL subscales (Attention, Aggression, and Anxious/Depressed, which form the CBCL-PBD) were reported to be in the poor-to-moderate range for BP diagnosis ($n = 7$) in a referred pediatric ADHD sample ($n = 121$). Also, the CBCL externalizing scores (AUC $= 0.82$) were less specific for BP diagnoses compared to the other scales for the BP diagnosis (Youngstrom et al. 2004). In COBY's large sample of BP children ($n = 157$), our AUC scores for CBCL subscales were higher than those reported by Biederman and colleagues (Biederman et al. 2005) and similar to those found by Youngstrom and colleagues (Youngstrom et al. 2004) and by Kahana and colleagues (Kahana et al. 2003).

Using data from two family studies of ADHD, Faraone and colleagues analyzed the CBCL-PBD and reported that the CBCL-PBD score showed an AUC of 0.97 for the ADHD children and 0.82 for their siblings for current diagnosis of BP and of 0.89 for the ADHD children and 0.85 for their siblings for lifetime diagnosis of BP, suggesting that the CBCL-PBD provides a highly efficient way of identifying bipolar subjects in that sample (Faraone et al. 2005). We found lower AUC CBCL-PBD scores for BP (0.70–0.77 vs. 0.82–0.97, respectively) than Faraone and colleagues. The ROC analysis in our study suggested that the accuracy of predicting BP diagnosis was in the moderate range for CBCL-PBD. The accuracy of predicting BP diagnosis was in the poor-to-moderate range for the CBCL subscales among children with psychopathology and was in good range only for the CBCL Thought Problems (AUC $= 0.84$) and Total Problems (AUC $= 0.81$) among healthy children and children with psychopathology (Table 3). As expected, the ROC analysis in our study suggested that the CBCL scores did a better job in identifying BP cases when healthy children were included in the analysis (Fig. 1), suggesting that the CBCL may be more helpful in the community as compared to the clinical samples. Of particular attention, the Thought Problems in our study had the highest AUC score for the BP diagnosis and the PPV was as high as 30%. Similar to our study, the CBCL Thought Problems were suggested to identify bipolar phenotype in a community sample of prepubertal children (Diler et al. 2008). Future clinical and community bipolar studies in children that employ CBCL may consider analyzing Thought Problems for its clinical use.

Conclusions

In summary, our findings do not support the use of the CBCL or the CBCL-PBD as a proxy for DSM-IV diagnosis of BP in clinically ascertained children. On the other hand, lower scores in the CBCL subscales may help clinicians to reconsider a BP diagnosis, especially in samples such as the community samples that may have lower rates of BP. Moreover, any child with significantly higher scores in the CBCL and the CBCL-PBD should be carefully assessed for the presence of severe psychopathology, including BP. In our opinion, assessment of dimensional presentation of pediatric BP is important; however, particularly for BP that can be difficult to diagnose in children, no screening test can substitute for the judgment of well-trained clinicians in making an accurate BP diagnosis.

Disclosures

Bipolar subjects were recruited from the Course and Outcome for Bipolar Disorder in Youth (COBY) study, major depressive disorder/anxiety disorders subjects were recruited from the Psychobiology of Childhood Anxiety and Depression study, and disruptive behavior disorder subjects were included from the Resources to Enhance the Adjustment of Children (REACH) study.

Dr. Keller has been a consultant and received honoraria from the following companies: Collegium, Cypress Bioscience, Cyberonics, Eli Lilly, Forest Laboratories, Janssen, Organon, Otsuka, Pfizer, Pharmastar, Sepracor, Vela Pharmaceuticals, and Wyeth Pharmaceuticals. He has received grants for his research from Eli Lilly, Pfizer, and Wyeth Pharmaceuticals. He has served on the advisory boards for Abbott Laboratories, Bristol-Myers Squibb, Cyberonics, Cypress Bioscience, Eli Lilly, Forest Laboratories, GlaxoSmithKline, Janssen, Novartis, Organon, Pfizer, Sepracor, and Wyeth Pharmaceuticals.

References

Achenbach TM: Manual for the Child Behavior Checklist/4–18 and 1991 Child Profile. Burlington (Vermont): University of Vermont Department of Psychiatry, 1991.

Achenbach TM, Rescorla LA: Manual for the ASEBA School-Age Forms and Profiles. Burlington (Vermont): University of Vermont Research Center for Children, Youth, and Families, 2001.

American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 4th ed. Text Revision (DSM-IV-TR). Washington (DC): American Psychiatric Association, 2000.

Axelson D, Birmaher B, Strober M, Gill MK, Valeri S, Chiappetta L, Ryan N, Leonard H, Hunt J, Iyengar S, Bridge J, Keller M: Phenomenology of children and adolescents with bipolar spectrum disorders. Arch Gen Psychiatry 63:1139–1148, 2006.

Biederman J, Wozniak J, Kiely K, Ablon S, Faraone S, Mick E, Mundy E, Kraus I: CBCL clinical scales discriminate prepubertal children with structured interview-derived diagnosis of mania from those with ADHD. J Am Acad Child Adolesc Psychiatry 34:464–471, 1995.

Biederman J, Faraone S, Mick E, Moore P, Lelon E: Child Behavior Checklist findings further support comorbidity between ADHD and major depression in a referred sample. J Am Acad Child Adolesc Psychiatry 35:734–742, 1996.

Biederman J, Monuteaux MC, Kendrick E, Klein KL, Faraone SV: The CBCL as a screen for psychiatric comorbidity in paediatric patients with ADHD. Arch Dis Child 90:1010–1015, 2005.

Birmaher B, Dahl RE, Williamson DE, Perel JM, Brent DA, Axelson DA, Kaufman J, Dorn LD, Stull S, Rao U, Ryan ND: Growth hormone secretion in children and adolescents at high risk for major depressive disorder. Arch Gen Psychiatry 57:867–872, 2000.

Birmaher B, Axelson D, Strober M, Gill MK, Valeri S, Chiappetta L, Ryan N, Leonard H, Hunt J, Iyengar S, Keller M: Clinical course of children and adolescents with bipolar spectrum disorders. Arch Gen Psychiatry 63:175–183, 2006.

Carlson GA, Kelly KL: Manic symptoms in psychiatrically hospitalized children—what do they mean? J Affect Disord 51: 123–135, 1998.

Carlson GA, Loney J, Salisbury H, Volpe RJ: Young referred boys with DICA-P manic symptoms vs. two comparison groups. J Affect Disord 51:113–121, 1998.

Dienes KA, Chang KD, Blasey CM, Adleman NE, Steiner H: Characterization of children of bipolar parents by parent report CBCL. J Psychiatr Res 36:337–345, 2002.

Diler RS, Uguz S, Seydaoglu G, Erol N, Avci A: Differentiating bipolar disorder in Turkish prepubertal children with attention-deficit hyperactivity disorder. Bipolar Disord 19:243–251, 2007.

Diler RS, Uguz S, Seydaoglu G,Avci A: Mania profile in a community sample of prepubertal children in Turkey. Bipolar Disord 10:546–553, 2008.

Faraone SV, Althoff RR, Hudziak JJ, Monuteaux M, Biederman J: The CBCL predicts DSM bipolar disorder in children: A receiver operating characteristic curve analysis. Bipolar Disord 7:518–524, 2005.

Geller B, Warner K, Williams M, Zimerman B: Prepubertal and young adolescent bipolarity versus ADHD: Assessment and validity using the WASH-U-KSADS, CBCL and TRF. J Affect Disord 51:93–100, 1998.

Hazell PL, Lewin TJ, Carr VJ: Confirmation that Child Behavior Checklist clinical scales discriminate juvenile mania from attention deficit hyperactivity disorder. J Paediatr Child Health 5:199–203, 1999.

Hollingshead AB: Four-Factor Index of Social Status. New Haven (Connnecticut): Sociology Department, Yale University, 1975.

Holtmann M, Bolte S, Goth K, Dopfner M, Pluck J, Huss M, Fegert JM, Lehmkuhl G, Schmeck K, Poustka F: Prevalence of the Child Behavior Checklist-pediatric bipolar disorder phenotype in a German general population sample. Bipolar Disord 9:895–900, 2007.

Hudziak JJ, Althoff RR, Derks EM, Faraone S, Boomsma D: Prevalence and genetic architecture of Child Behavior Checklist-juvenile bipolar disorder. Biol Psychiatry 58:562–568, 2005.

Kahana SY, Youngstrom EA, Findling RL, Calabrese JR: Employing parent, teacher, and youth self-report checklists in identifying pediatric bipolar spectrum disorders: An examination of diagnostic accuracy and clinical utility. J Child Adolesc Psychopharmacol 13:471–488, 2003.

Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N: Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. J Am Acad Child Adolesc Psychiatry 36:980–988, 1997.

Kolko DJ, Dorn LD, Bukstein OG, Burke JD: Clinically Referred ODD children with or without CD and healthy controls: Comparisons across contextual domains. J Child Fam Stud 17:714–734, 2008.

Mick E, Biederman J, Pandina G, Faraone SV: A preliminary meta-analysis of the child behavior checklist in pediatric bipolar disorder. Biol Psychiatry 53:1021–1027, 2003.

Park SH, Goo JM, Jo CH: Receiver operating characteristic (ROC) curve: Practical review for radiologists. Korean J Radiol 5:11–18, 2004.

Pavuluri MN, Birmaher B, Naylor MW: Pediatric bipolar disorder: A review of the past 10 years. Am Acad Child Adolesc Psychiatry 44:846–871, 2005.

Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, Aluwahlia S: A children's global assessment scale (CGAS). Arch Gen Psychiatry 40:1228–1231, 1983.

Swets K: Measuring the accuracy of diagnostic systems. Science 240:1285–1293, 1988.

Volk HE, Todd RD: Does the Child Behavior Checklist juvenile bipolar disorder phenotype identify bipolar disorder? Biol Psychiatry 62:115–120, 2007.

Youngstrom E, Youngstrom J: Evidence-based assessment of pediatric bipolar disorder, Part II: Incorporating information from behavior checklists. Am Acad Child Adolesc Psychiatry 4:823–828, 2005.

Youngstrom EA, Findling RL, Calabrese JR, Gracious BL, Demeter C, Bedoya DD, Price M: Comparing the diagnostic accuracy of six potential screening instruments for bipolar disorder in youths aged 5 to 17 years. Am Acad Child Adolesc Psychiatry 43:847–858, 2004.

Youngstrom E, Youngstrom J, Starr M: Bipolar diagnoses in community mental health: Achenbach Child Behavior Checklist profiles and patterns of comorbidity. Biol Psychiatry 58:562–568, 2005.

Zou KH, O'Malley AJ, Mauri L: Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation 115:654–657, 2007.

Address reprint requests to:
Dr. Rasim Somer Diler
Department of Child Psychiatry
Western Psychiatric Institute and Clinic
3811 O'Hara Street
Pittsburgh, PA 15213

E-mail: dilerrs@yahoo.com