

Improving estimation of the prognosis of childhood psychopathology; combination of DSM-III-R/DISC diagnoses and CBCL scores

Robert F. Ferdinand, Jeroen Heijmens Visser, Kirsten N. Hoogerheide, Jan van der Ende, Marianne C. Kasius, Hans M. Koot, and Frank C. Verhulst

Department of Child and Adolescent Psychiatry, Academic Hospital Rotterdam – Sophia/Erasmus University Rotterdam, Rotterdam, The Netherlands

Objective: To compare the predictive validity of the clinical-diagnostic and the empirical-quantitative approach to assessment of childhood psychopathology, and to investigate the usefulness of combining both approaches. **Method:** A referred sample ($N = 96$), aged 6 to 12 years at initial assessment, was followed up across – on average – a period of 3.2 years. It was assessed to what extent DISC/DSM-III-R diagnoses – representing the clinical-diagnostic approach, and CBCL scores – representing the empirical-quantitative approach, predicted the following signs of poor outcome: outpatient/inpatient treatment, or parents' wish for professional help for the child at follow-up, disciplinary problems in school, and police/judicial contacts. **Results:** Both diagnostic systems added significantly to the prediction of poor outcome, and neither of the two systems was superior. Use of both systems simultaneously provided the most accurate estimation of the prognosis, reflected by the occurrence of future poor outcome. Even diagnostic concepts that are generally regarded as relatively similar, such as ADHD (DSM) and attention problems (CBCL), or conduct disorder (DSM) and delinquent behavior (CBCL), appeared to differ in their ability to predict poor outcome. **Conclusions:** The present study supports the use of the empirical-quantitative approach and the clinical-diagnostic approach simultaneously, both in research and in clinical settings, to obtain a comprehensive view of the prognosis of psychopathology in children. **Keywords:** DISC, CBCL, prognosis, psychopathology, validity.

To assess psychopathology in children and adolescents, two main approaches can be followed: the *clinical-diagnostic* and the *empirical-quantitative* approach. The *Diagnostic and Statistical Manual of Mental Disorder* (American Psychiatric Association, 1980, 1987, 1994) is a representative of the clinical-diagnostic approach. DSM diagnostic criteria are mainly the result of consensus among experts. In this way, disorders are defined by a rather arbitrary set of criteria. Furthermore, for most disorders, the DSM system does not provide more information than that the disorder is 'present' or 'absent', which is not informative, for instance, regarding the severity of a disorder, or the number of symptoms. Mostly, DSM diagnoses are assessed via clinical interview.

In contrast, the empirical-quantitative approach uses quantitative procedures to determine – empirically – which symptoms tend to co-occur in syndromes. Multivariate statistical procedures, applied to data from large clinical samples, are used to find out which symptoms constitute a syndrome. Quantitative scores for psychopathology are typically derived via rating scales. By using quantitative – rather than categorical – scores, information can be obtained regarding the number and severity of problems, compared to other subjects in clinical or normative samples, instead of just indicating the presence or absence of problems. Hence, decisions on the number and severity of problems can be

based on actual distributions of scores in populations rather than on preset criteria.

The fact that the two approaches as described above exist together shows that neither of them fully satisfies the numerous questions with regard to assessment and diagnosis of child and adolescent psychopathology. The combination of both diagnostic paradigms may be valuable if they compensate for the weaknesses and augment the strengths for each approach. However, to date, no satisfying paradigm exists as to how to combine information from the two approaches, and in which cases to let one of the systems prevail.

To investigate the strengths and weaknesses of both systems, it is important to compare the validity of both approaches. Validity of a diagnostic system can be defined as the degree to which the system covers concepts that are relevant, given the aims of the system. Most important types of validity are: content validity, concurrent validity, divergent validity, criterion-related validity and predictive validity. The first four types of validity rely on cross-sectional information, while the latter uses longitudinal information.

Content validity refers to coverage. For instance, does an anxiety measure really cover anxiety items only, or does it also cover items aimed at other types of psychopathology? Concurrent validity indicates the association between a measure, for instance, aimed at ADHD, and other measures, also aimed at

ADHD. Divergent validity reflects the ability of a measure to be specific for the area of psychopathology at which it is aimed. For instance, a measure aimed at assessing delinquency should have a low correlation with a measure aimed at anxiety. Criterion-related validity can be assessed by investigating the association between scores on a measure and an external criterion. For instance, Verhulst et al. (1996) found that children and adolescents who were referred to mental health services scored significantly higher on 114 of the 120 problem items of the Child Behavior Checklist (CBCL; Achenbach, 1991b) than subjects from a normative sample from the general population. This indicated that CBCL items probably measure 'psychopathology'.

Although important, sole reliance on the described types of validity, based on cross-sectional data, can be hazardous. Achenbach and McConaughy (1997) compared the assessment of the concurrent validity of instruments or diagnostic systems with 'bootstrapping', based on the way Baron von Münchhausen lifted himself from a swamp by his own boots. In other words, although diagnostic systems may converge strongly, this still does not provide absolute guarantee that these systems provide valid information, because both systems might be invalid. This problem can – in a way – be circumvented, by assessing and comparing the predictive validity of diagnostic systems. Predictive validity of a measure reflects its ability to predict future events, related to the content of the measure. For instance, in a general population sample of 4- to 16-year-olds, Hofstra, van der Ende, Koot, and Verhulst (submitted) found that scores on the CBCL scale Delinquent Behavior predicted later police contacts across a 14-year period. Predictive validity is important for clinical practice because it is indicative of the prognosis of psychopathology.

Studies that compared the validity of the clinical-diagnostic and the empirical-quantitative approach have mainly focused on the concurrent and divergent validity of the approaches, by assessing associations between results of both approaches in one sample (Edelbrock & Costello, 1988; Jensen et al., 1993; Gould et al., 1993; Steingard et al., 1992; Biederman et al., 1993; Kasius et al., 1997). For instance, Edelbrock and Costello (1988) assessed relationships between DSM-III diagnoses derived from one of the earliest versions of the parent DISC (NIMH Diagnostic Interview Schedule for Children; NIMH 1992), and pre-1991 CBCL scales (Achenbach & Edelbrock, 1983) which preceded the 1991 version of the CBCL scales (Achenbach, 1991a, b). There was considerable overlap between CBCL-scale scores and DISC diagnoses in their sample of 270 clinically referred children aged 6 to 16 years. The strongest relations were found between the CBCL scales Hyperactive, Delinquent, and Depressed and the DISC-derived diagnoses of Attention Deficit Disorder, Conduct Disorder, and Depression/Dysthymia,

respectively. Similarly, Kasius et al. (1997) assessed associations between DISC 2.3./DSM-III-R diagnoses (NIMH, 1992) and 1991-CBCL syndromes, in a sample of 231 six- to sixteen-year-olds who had been referred to an outpatient clinic. They found a strong specific convergence between CBCL scales Attention Problems, Delinquent Behavior and Aggressive Behavior and DISC/DSM diagnoses Attention Deficit Hyperactivity Disorder, Conduct Disorder and Oppositional Disorder, respectively. Associations between internalizing CBCL scales and DSM Anxiety or Affective disorders were moderate, on average, and did not reveal very specific associations between particular CBCL scales and specific DSM diagnoses.

Studies thus far have not compared the predictive validity of measures based on the clinical-diagnostic versus the empirical-quantitative approach. Furthermore, while previous studies provided information on convergence and divergence between systems, such studies cannot be used to investigate how information from both systems can be combined. For this purpose, studies investigating how information derived via both diagnostic systems can be combined to optimally predict the prognosis of psychopathology can be useful.

To test and compare the predictive validity of both diagnostic approaches, it is important that information reflecting the two approaches is derived from the same type of informant. For example, if we compare the predictive validity of parents' CBCL ratings with DSM diagnoses derived from clinical interviews with the child, we do not know if discrepancies reflect the difference in diagnostic approaches or the use of different informants (parent versus child). It is also important that the outcome variables that are used are clinically relevant, but – by their content – not more likely to be associated with one of the two systems in advance, because this would hamper optimal comparison.

In the present study, the predictive validity of CBCL scores, representing the empirical-quantitative approach, versus DISC-2.3/DSM-III-R diagnoses, representing the clinical-diagnostic approach, were compared. Furthermore, it was studied how results of both diagnostic systems can be combined, to obtain an optimal view on the prognosis of psychopathology. Initially 6- to 12-year-olds who visited a child psychiatric outpatient clinic were assessed with the CBCL and the DISC-2.3, and were reassessed, on average, three years later. At follow-up, poor outcome variables that were 'external' to initial measures of psychopathology were assessed.

Method

Measures

The Child Behavior Checklist (CBCL; Achenbach, 1991b) is a parent questionnaire for assessing problems in 4- to 18-year-olds. It consists of 20 competence

items and 120 items on behavioral or emotional problems in the past 6 months. Only data from the problem section were used in the present study. The response format is 0 = not true, 1 = somewhat or sometimes true, and 2 = very true or often true. The good reliability and validity of the CBCL (Achenbach, 1991b) were confirmed for the Dutch translation (Verhulst et al., 1996; De Groot et al., 1994).

Achenbach (1991a) constructed eight so-called 'narrow band' syndrome scales: 'Withdrawn', 'Somatic Complaints', 'Anxious/Depressed' (which together constitute the 'Internalizing' broad band syndrome), 'Delinquent Behavior' and 'Aggressive Behavior' (together constituting the 'Externalizing' grouping of syndromes), and the scales 'Social Problems', 'Thought Problems', and 'Attention Problems'. Furthermore, a Total Problem score can be computed, reflecting the sum of the total score of all problem items.

The NIMH Diagnostic Interview Schedule for Children, version 2.3, a highly structured interview to assess the more common DSM-III-R Axis 1 diagnoses (APA, 1987) in children and adolescents, was used. Unless otherwise specified, the timeframe of the DISC is the past 6 months. The DISC-2.3 covers the diagnostic categories listed in Table 1. The DISC has two parallel forms: DISC-C administered directly to the child or adolescent, and DISC-P administered to the parent or parent substitute. The DISC-P was used in the present study. The reliability and validity of the earlier versions of the DISC-2.3, the DISC-R and the DISC 2.1, were supported by Piacentini et al. (1993), Schwab-Stone et al. (1993), and Shaffer et al. (1993). The DISC was administered by two professionals. The first was one of the authors (MK), who had been trained at Columbia University/NIMH DISC training center, NY-USA. The second was a research psychologist, who was trained in the DISC by MK. These researchers were not acquainted with results of other assessment procedures or clinical data.

For all diagnoses, except for the Psychosis Screen, all DSM-III-R criteria had to be met to assign a DSM diagnosis. The Psychosis Screen does not represent a specific DSM diagnosis. In the DISC-2.3 25 questions are asked about delusions ($n = 13$), hallucinations

($n = 5$), disorganized speech ($n = 5$), catatonic behavior ($n = 1$), and alogia ($n = 1$). These features cover all 5 key criteria for schizophrenia. The emphasis is on delusions, hallucinations, and disorganized speech, probably on the assumption that catatonic behavior and alogia are less prevalent in children. All symptoms are rated absent or present by the interviewer, on the basis of a yes/no answer from the respondent. If a symptom is rated present, the interviewer is instructed to write a complete description of the behavior: what happened, when and where it happened, how many times, how long, etc. All descriptions were reviewed, and eventually recoded, by two experienced clinicians. If these clinicians were not convinced that the descriptions indicated symptoms of schizophrenia, such as described in DSM-III-R, positive scores were changed into negative. The Psychosis Screen was rated positive if at least one symptom of DSM-III-R Schizophrenia was present.

Outcome measures were assessed with a parent questionnaire, containing items on (1) receiving outpatient, or (2) receiving inpatient mental health services at time 2, (3) parents' wish for professional help regarding problems of the child at time 2 (in case the child was still in treatment we respected the parents' wish for more or alternative treatment), (4) disciplinary problems in school during follow-up (defined as problems relating to school other than learning problems, e.g., being suspended or expelled from school, truancy, violent behavior, misbehavior, and social problems), and (5) police/judicial contacts during the follow-up period. Each outcome variable was scored 0 if absent and 1 if present.

Participants

At the first time of assessment (time 1), between April 1992 and April 1994, 246 children and adolescents (6- to 16-year-olds), who had been consecutively referred to the outpatient department of Child and Adolescent Psychiatry of Sophia Children's Academic Hospital in Rotterdam, and their parents, were asked to participate in a study, performed by Kasius (1997).

Table 1 Percentages of children with time 1 CBCL scores above the 95th percentile of a Dutch normative sample,¹ and rates of time 1 DISC/DSM-III-R diagnoses

CBCL syndrome	% > P95	DSM-III-R diagnosis	Rate (%)	DSM-III-R diagnosis	Rate (%)
Withdrawn	46.9	Simple Phobia	25.6	Attention Deficit Disorder	41.1
Somatic Complaints	37.5	Social Phobia	10.0	Oppositional Disorder	17.8
Anxious/Depressed	51.0	Agoraphobia	11.1	Conduct Disorder	6.7
Social Problems	54.2	Panic Disorder	6.7	Any Disruptive Disorder	45.6
Thought Problems	49.0	Separation Anxiety Disorder	12.2	Eating Disorder	.0
Attention Problems	54.2	Avoidant Disorder	4.4	Nocturnal Enuresis	10.0
Delinquent Behavior	38.5	Overanxious Disorder	17.8	Diurnal Enuresis	7.8
Aggressive Behavior	49.0	Generalized Anxiety Disorder	14.4	Encopresis	7.8
		Obsessive Compulsive Disorder	5.6	Tic Disorder	13.3
		Any Anxiety Disorder	53.3	Psychosis Screen	20.0
		Major Depressive Disorder	13.3	Substance use disorders	.0
		Dysthymia	16.7		
		Mania	4.4		
		Any Mood Disorder	23.3		

Note: ¹See Verhulst et al. (1996).

Exclusion criteria were: severe mental retardation of the child, and parents' lack of command of the Dutch language. One hundred and sixty-nine (68.7%) subjects and their parents participated at time 1.

To assess the representativeness of her sample, Kasius (1997) compared CBCL scores in her sample with those of a Dutch referred sample of 2004 children and adolescents. This sample, which was derived from a large number of Dutch mental health agencies, has been described in detail by Verhulst et al. (1996) in the manual for the Dutch CBCL/4-18. Kasius found that referred 4- to 11-year-old boys in her study sample were scored significantly higher on the following CBCL scales: Somatic Complaints ($t = 2.38$, $p < .05$), Anxious/Depressed ($t = 2.32$, $p < .05$), Social Problems ($t = 2.75$, $p < .01$), Thought Problems ($t = 5.56$, $p < .001$), Attention Problems ($t = 2.36$, $p < .05$), and Delinquent Behavior ($t = 2.35$, $p < .05$). CBCL problem scale scores of girls from her sample in the same age range did not differ from scores of girls from the large clinical sample. Hence, the present study's sample does probably contain boys with higher problem levels than in general mental health care settings. This might be explained by the fact that the sample originates from an academic clinic, with referrals of rather complex and severe cases which have often been unsuccessfully treated before in other mental health agencies.

The time 1 target sample scored as follows on a 6-point scale of parental occupation (1 = lowest SES, 6 = highest SES; Van Westerlaak et al., 1975): 1-16%, 2-19.5%, 3-25.4%, 4-9.2%, 5-9.7%, 6-20.0%. At time 1, there was no significant difference in mean SES between responders versus refusers ($t = 1.36$, $p > .05$).

In the present study, we report on 132 children who were aged 6 to 12 years at time 1. These participants were followed up across time (mean interval = 3.2 years; $sd = .62$ years; range = 1.8-4.5 years). At time 2, 96 (72.7%; 63 boys and 33 girls) cooperated. For those, time 1 CBCLs were available for all 96 participants, and time 1 DISC data for 90. Follow-up data were obtained from a larger follow-up study (Heijmens Visser et al., 1999).

To assess selective attrition we compared remainers and drop-outs with respect to age (mean age = 9.5 vs. 10.4 years; $t = 2.74$; $p < .01$), sex ($\chi^2 = .602$; $p = n.s.$), and mean socioeconomic status on a six-step scale of parental occupation (3.33 vs. 3.42; $t = .251$; $p = n.s.$). There were no significant differences for time 1 CBCL syndrome scores and Total Problem scores.

Of the 96 participants, 68 lived with their biological parents, 25 lived with their biological mother, with (10) or without (15) a partner, one lived with his/her biological father and his new partner, and 2 lived with adoptive parents. In 95% of the cases, the mother of the child filled out the CBCL. Ninety-one children (95%) were born in the Netherlands and had Dutch nationality (see also Heijmens Visser et al., 1999).

Between times 1 and 2, almost all children (95%) received some kind of treatment. In 15 cases treatment was day treatment or inpatient treatment. Sixty-seven (85%) of the 79 children who received outpatient treatment were treated at our outpatient department. Ninety percent of them received individual psychotherapy. Thirteen percent of the children received medication. Mean duration of treatment was 15.6 months ($sd = 15.2$

months; range 1 to 89 months). In 61% of the cases there were more than 10 therapeutic sessions. Seventy-seven percent of the parents received parental counseling.

CBCL scores and rates of DISC/DSM-III-R disorders in the sample are presented in Table 1. Frequencies of outcome variables are presented in Table 2.

Ethics

Each assessment phase of this study was approved by the Committee for Medical Ethics, Sophia Children's Hospital/Erasmus University Rotterdam. At each phase, informed consent was obtained from both parents and children.

Statistics

To examine the ability of CBCL and DISC results to predict poor outcome variables, logistic regression analyses were conducted.

For the CBCL, first, 5 sets of 11 univariate logistic regressions were performed, in which CBCL time 1 scores (all eight narrow band syndromes, Internalizing, Externalizing, and Total Problem score) were entered as separate candidate predictors, and - respectively - the five poor outcome measures as dependent variables (see Table 3). Similarly, 5 sets of 13 regression analyses were performed for DSM-III-R/DISC diagnoses that occurred in - for reasons of power - at least 10% of the time 1 sample, and 5 sets of 3 regression analyses were performed for the broader DSM-III-R/DISC categories 'Any Anxiety Disorder', 'Any Mood Disorder', and 'Any Disruptive Disorder' (Conduct Disorder, Oppositional Defiant Disorder, or Attention Deficit Hyperactivity Disorder; see Table 1). Diagnoses, coded as '0' if absent and '1' if present, were entered as candidate predictors, and outcome variables as dependent variables. For analyses using time 1 CBCL data, data from 96 participants were used, while for analyses regarding DISC/DSM data, data from 90 participants were used (see 'Participants').

Subsequently, a set of five forward stepwise univariate analyses was performed, for each poor outcome variable separately, in which all CBCL scale scores and DISC/DSM-III-R diagnoses that contributed significantly to the prediction of a specific poor outcome variable in the first logistic regression analyses were entered (see Table 4). These analyses indicated which CBCL scores or DISC/DSM-III-R diagnoses were the most significant predictors of a poor outcome variable, *independently* of other scores/diagnoses. In other words, suppose that one CBCL score and one DISC/DSM-III-R diagnosis both were statistically significant

Table 2 Rates of poor outcome variables at Time 2

Outcome variable	No.	%
Disciplinary problems in school	32	33.3
Police/judicial contacts	6	6.3
Parent's wish for help	39	40.6
Outpatient treatment	34	35.4
Inpatient treatment	11	11.5

Table 3 Prediction of signs of malfunctioning by CBCL scale scores and DISC/DSM-III-R diagnoses

Predictor ¹	Disciplinary problems in school			Outpatient treatment time 2			Inpatient treatment time 2			Police/judicial contacts			Wish for professional help time 2		
	Odds	chi sq	p	Odds	chi sq	p	Odds	chi sq	p	Odds	chi sq	p	Odds	chi sq	p
Social Problems	–	–	–	1.20	7.37	<.01	1.29	6.78	<.01	–	–	–	1.14	4.28	<.05
Attention Problems	–	–	–	1.11	5.63	<.05	1.16	4.91	<.05	–	–	–	–	–	–
Delinquent Behavior	1.20	7.49	<.01	1.15	4.83	<.05	–	–	–	–	–	–	1.21	7.95	<.005
Aggressive Behavior	–	–	–	1.06	5.88	<.05	–	–	–	–	–	–	–	–	–
Externalizing Problems	1.04	4.79	<.05	1.05	6.51	<.05	–	–	–	1.08	4.57	<.05	–	–	–
Total Problem Score	–	–	–	1.02	5.73	<.05	–	–	–	–	–	–	–	–	–
Simple Phobia	–	–	–	–	–	–	–	–	–	–	–	–	2.79	4.37	<.05
Generalized Anxiety Disorder	–	–	–	–	–	–	–	–	–	7.40	4.69	<.05	3.72	4.54	<.05
Agoraphobia	–	–	–	8.31	8.43	<.005	–	–	–	–	–	–	6.67	6.72	<.01
ADHD	–	–	–	2.67	4.91	<.05	–	–	–	–	–	–	–	–	–
Oppositional Def. Disorder	3.04	3.91	<.05	–	–	–	5.15	5.32	<.05	12.00	7.70	<.01	–	–	–
Conduct Disorder	11.15	6.56	<.05	–	–	–	–	–	–	–	–	–	–	–	–
Psychosis Screen	–	–	–	4.55	7.79	<.01	–	–	–	–	–	–	7.00	11.85	<.001
Any Disorder	–	–	–	4.83	5.20	<.05	–	–	–	–	–	–	–	–	–

Note: ¹Only predictors that contributed significantly ($p < .05$) to the prediction of an outcome variable were entered in the table.

predictors of a poor outcome variable in Table 3, they were entered in simultaneously in a new regression analysis. If the CBCL score and DISC/DSM-III-R diagnosis would mainly cover similar behavioral or emotional problems, only the predictor for which the highest association between the predictor and the outcome was found in Table 3, either the CBCL score or the DISC/DSM-III-R diagnosis, would appear from the new regression analysis as an independent predictor of the poor outcome variable. However, if the CBCL and DISC/DSM-III-R predictor that were significant in Table 3 would both constitute a risk factor for the poor outcome variable, but would, at least partially, cover different behavioral or emotional problems, both predictors might appear from the new analysis as independent predictors of the poor outcome variable.

Furthermore, we also allowed two-way interactions in the final regression analyses. Significant two-way interactions in Table 4 indicate that the risk for an outcome is determined, not by an increase in scores on a single predictor, but by an increase in scores on two predictors simultaneously. For instance, if a two-way

interaction between a CBCL score and a DISC/DSM-III-R diagnosis is statistically significant, this indicates that the combination of high CBCL scores plus the presence of a DISC/DSM-III-R diagnosis is a risk factor for future poor outcome. In other words, by using this analytic approach, we were able to investigate if, by combining information obtained via the CBCL and the DISC, outcome could be predicted more accurately.

Because of overlap between diagnostic categories, interaction effects between Oppositional Defiant Disorder and Conduct Disorder were not allowed. For the same reason, broad categories of DSM diagnoses (Any Anxiety Disorder, Any Mood Disorder, Any Disruptive Disorder, and Any Disorder) were excluded from these analyses, as were broad band CBCL syndromes ‘Internalizing’ and ‘Externalizing’, and the Total Problem score. An exception was made for final analyses regarding the prediction of police/judicial contacts, in which Externalizing Problems was entered, because the narrow-band externalizing syndromes (Delinquent Behavior and Aggressive Behavior) did not predict this outcome in the first set of analyses (see Table 3).

Table 4 Prediction of signs of malfunctioning by CBCL scale scores and DISC/DSM-III-R diagnoses; results of composite analyses

Outcome	Predictors		Statistics			
	DISC/DSM-III-R diagnosis	CBCL scale	Odds	chi q	p	df ¹
Disciplinary problems in school	Conduct Disorder	² Delinquent Behavior	1.30	7.23	<.01	1
	Agoraphobia	–	6.90	22.37	<.0005	3
	Psychosis Screen	–	5.19	–	–	–
Outpatient treatment time 2	ADHD	² Delinquent Behavior	1.21	–	–	–
	Oppositional Defiant Disorder	–	4.33	10.41	<.01	2
	–	Social Problems ²	1.01	–	–	–
Police/judicial contacts	Generalized Anxiety Disorder ²	–	27.0	9.89	<.005	1
	Oppositional Defiant Disorder	–	–	–	–	–
Wish for professional help time 2	Psychosis Screen	–	7.57	19.78	<.0005	2
	–	Delinquent Behavior	1.21	–	–	–

Notes: ¹df = degrees of freedom for chi-square statistic final model. ²*indicates interaction effect.

For analyses concerning prediction of outpatient and inpatient treatment, Wald statistics were not statistically significant for, respectively two and one, predictors included in the final model. In other words, the predictive power of some separate predictor variables that constituted the final set of predictor variables that were identified by these regression analyses was not statistically significant. This is caused by the fact that different statistical tests are used to test the significance of an entire regression model, that is constituted by a set of predictors (chi-square tests), than for the significance of all predictors from the model separately (Wald tests). If Wald tests were not significant for a predictor, we excluded this variable. Hence, final analyses concerning the prediction of outpatient and inpatient treatment were repeated, excluding the least powerful non-significant predictor as a candidate predictor variable.

Results

Table 3 shows results of the first sets of logistic regression analyses, aimed at assessing the predictive power of all DISC/DSM-III-R disorders and CBCL scales. Non-significant predictors are not shown. For instance, police/judicial contacts were predicted by scores on the Externalizing Problems scale of the CBCL, and by DISC/DSM-III-R diagnoses Generalized Anxiety Disorder and Oppositional Defiant Disorder. The odds ratios indicate that the chance of having a specific outcome increases if a predictor is present. An odds ratio of 12.00, for instance, as found for Generalized Anxiety Disorder, indicates that the probability of police/judicial contacts increased 12-fold if a Generalized Anxiety Disorder was diagnosed at time 1.

Table 4 shows results of analyses in which predictors that were significant in the first analyses were combined. For instance, it is shown that scores on the CBCL scale Externalizing Problems, which predicted police/judicial contacts in the first set of analyses, did not predict this outcome independently of Generalized Anxiety Disorder or Oppositional Defiant Disorder. Table 4 also shows interaction effects. Interaction effects are indicated by *. Interaction effects occurred between DISC/DSM-III-R disorders and CBCL scales (e.g., DSM/ADHD * CBCL Delinquent Behavior predicted Outpatient treatment at time 2), among DSM diagnoses (e.g., Generalized Anxiety Disorder * Oppositional Defiant Disorder predicted police/judicial contacts), and among CBCL scales (e.g., Social Problems * Attention Problems predicted Inpatient treatment at time 2).

Corrections for chance findings were not applied, although multiple tests were performed. As argued by Rothman (1990, p. 530), adjustments for making multiple comparisons in large bodies of data are recommended to avoid rejecting the null hypothesis too readily. Unfortunately, reducing the type I error for null associations increases the type II error for those associations that are not null. Hence, possibly

important results would be missed if correction for multiple comparisons were applied too rigorously. Instead of adjusting testing for multiple comparisons, a better approach to reduce type I errors is to repeat the present study in other samples, and to test if the results are stable across studies, countries, samples, assessment procedures, etc.

Discussion

The present study compared the predictive validity of the clinical-diagnostic and the quantitative-empirical approaches, which represent two widely used taxonomic systems of psychiatric symptoms. Baseline assessment with the Child Behavior Checklist and the Diagnostic Interview Schedule for Children was followed by assessment of outcome, on average, after a 3-year period. Participants were 96 six- to twelve-year-olds who visited our child psychiatric outpatient clinic. By keeping informants similar for all assessments, informant biases were ruled out as a source of variation.

The outcome variables used in the present study all represented signs of poor outcome that were external to initial candidate predictors. That is, instead of being based on one of the two diagnostic paradigms, they indicated general measures of adverse outcome, like treatment not being finished at follow-up, disciplinary problems in school, parents' wish for (more) professional help, or police/judicial contacts. Table 3 shows that data based on both taxonomic paradigms, clinical-diagnostic and quantitative-empirical, contributed to the prediction of all five outcome variables. Hence, both diagnostic approaches provided information on the severity of problems at time 1, exemplified by persistence of problems across the follow-up period. Convergence of the two systems in this respect supported results of previous cross-sectional studies, which, generally, found moderate to strong concurrent validity (Edelbrock & Costello, 1988; Jensen et al., 1993; Gould et al., 1993; Steingard et al., 1992; Biederman et al., 1993; Kasius et al., 1997).

At a first glance, DISC/DSM-III-R diagnoses may seem to have stronger predictive power than CBCL scales, because of average higher odds ratios, up to 12.0, while the highest odds ratio for CBCL scales was 1.29. However, odds ratios reflect the relative increase in the probability that an event (in case of the present study, an outcome variable) will take place. For DISC/DSM-III-R diagnosis, only one increase in probability can take place, because a diagnosis can be absent ('0') or present ('1'). For CBCL scales, multiple increments can take place; with every step on a scale, a rise in scores of one point, the probability that an outcome will take place may be multiplied by the amount indicated by the odds ratio. For instance, if an odds ratio is 1.10, the probability that an event will take place may be

multiplied by 1.10^{10} (=2.59) if a scale score increases by 10 points.

Concordance in predictions by different diagnostic systems may support the validity of both systems. For instance, disciplinary problems in school were predicted by DISC/DSM-III-R diagnoses of Oppositional Disorder and Conduct Disorder, and by CBCL scales Delinquent Behavior and Externalizing Problems. Hence, the strong association between a DISC/DSM-III-R diagnosis of Conduct Disorder and scores on the Delinquent Behavior scale of the CBCL found in previous studies (e.g., Edelbrock & Costello, 1988; Kasius et al., 1997) was supported in a longitudinal fashion in the present study. However, divergence in predictions of diagnostic concepts that are generally regarded as similar may also provide valuable information. Remarkably, despite high concurrent validity between a DISC/DSM-III-R diagnosis of Oppositional Defiant Disorder (ODD) and the CBCL scale Aggressive Behavior (Kasius et al. 1997), ODD contributed to the prediction of disciplinary problems in school, while Aggressive Behavior did not. This indicated that the two diagnostic concepts diverged with respect to prediction of that part of the prognosis that concerned these problems in school.

Divergence also occurred regarding internalizing symptoms assessed with the CBCL, versus the DISC. For DSM-III-R anxiety disorders, 5 significant odds ratios were found, reflecting prediction of outpatient treatment at follow-up (probably indicating treatment resistance), police/judicial contacts, and parents' wish for professional help at time 2. Conversely, CBCL scales aimed at assessing internalizing problems did not predict poor outcome. Hence, the present study indicates that assessment of DSM-III-R anxiety disorders, compared to sole reliance on the CBCL, may improve estimation of the prognosis of psychopathology in children who visit an outpatient clinic.

Two major factors may be responsible for these discrepancies concerning internalizing symptoms. The first factor is a difference in the structure of the two diagnostic systems. A major difference between the structure of the quantitative-empirical system and that of the clinical-diagnostic approach is the presence of a number of different anxiety syndromes in the DSM system (see Table 1), but not in the quantitative-empirical system. Spence (1998) supported the need for a diagnostic system providing the ability to differentiate between different anxiety disorders in children, instead of using just one general measure of anxiety, by assessing different DSM-IV dimensions of anxiety in children and adolescents, using a factor analytic approach. Furthermore, the CBCL does not differentiate between anxiety and depressive syndromes, while, despite considerable comorbidity between anxiety disorders and depressive disorders (i.e., Last et al., 1987, 1992; Cole et al., 1998; Kashani et al., 1987), some studies indicate that distinct qualities exist that differentiate these

disorders from each other. For instance, Barrett and Turner (2001) and Lowry-Webster et al. (2001) found differences with respect to age of onset, duration, and associated features. The fact that, in the present study, several 'pure' DSM-III-R anxiety disorders, but none of the depressive disorders, predicted poor outcome may indicate that the predictive power of CBCL scores was affected by combination of symptoms of anxiety and depression in one scale.

Multiple testing may constitute a second factor that may be responsible for the predictive power of DSM anxiety disorders, but not of internalizing scales of the CBCL. Ten DISC/DSM-III-R anxiety disorders were assessed, while the CBCL covers only three narrow-band internalizing syndromes. By testing the predictive validity of a greater number of DSM anxiety disorders than CBCL internalizing scales, the probability that findings concerning the DSM system were merely chance findings increased.

Knowledge about interaction of the two diagnostic systems with respect to prediction of outcome may be useful in clinical practice. An interaction was found between a DSM-III-R diagnosis of Conduct Disorder and scores on the Delinquent Behavior scale. Children with a combination of Conduct Disorder plus high scores on the Delinquent Behavior scale were especially at risk for antisocial behaviors resulting in police/judicial contacts. Generally, a strong association is found between the diagnostic concepts Conduct Disorder and Delinquent Behavior. However, the results of the present study suggest that by assessing just one of these concepts in children, valuable information regarding the prognosis would be lacking. This finding may indicate that a diagnosis of Conduct Disorder is not sufficient to determine the part of the prognosis regarding future antisocial behavior, because information regarding the severity – or the number – of problems is missing. Scores on the – dimensional – Delinquent Behavior scale might make up for this lack. Another explanation might be that the two different syndromes partially cover different concepts. However, this is unlikely, given their very strong association (e.g., Kasius et al. 1997) and the similarity of their content, suggesting similar content validity.

Interaction between DSM disorder and CBCL information was also relevant for the prediction of outpatient treatment at follow-up, probably a sign of treatment resistance. While the first set of analyses (Table 3) indicated that children with ADHD were likely to receive treatment at follow-up, it was subsequently shown (Table 4) that those with high scores on the CBCL scale Delinquent Behavior were especially at risk for this outcome. This interaction, between attention problems/hyperactivity (assessed with the DISC) and delinquent behaviors (assessed with the CBCL), would not have been found if time 1 assessment had relied solely on DISC/DSM-III-R diagnoses. DSM Disruptive Behavior Disorders,

which may also be regarded as indicative of delinquent behaviors, were not predictive of outpatient treatment at time 2, and did not interact with a DISC/DSM-III-R diagnosis of ADHD, with respect to the prediction of outpatient treatment.

The Psychosis Screen of the DISC remained a significant predictor of poor outcome; outpatient treatment at time 2 and parents' wish for (more) help. This screen consists of a number of questions regarding psychotic symptoms, based on features of Schizophrenic Disorder. Apparently, psychotic features constituted an important risk factor for persistence of problems. It is important to note that the Thought Problems scale of the CBCL, which also covers some questions regarding psychotic features, did not predict poor outcome. This was probably due to the fact that this scale covers several other problems which are not indicative of psychosis at all. The fact that the DISC covers a relatively large number of questions regarding psychosis, or the use of clinical judgment regarding answers to these questions, may constitute other explanations. Regardless of the explanation, however, the present study indicates that assessment of psychotic features according to DSM criteria may add significantly to estimation of the prognosis of childhood psychopathology. Sole reliance on the CBCL scale Thought Problems would probably result in a lack of information.

Interaction between CBCL scales Social Problems and Attention Problems – besides DSM Oppositional Defiant Disorder – predicted inpatient treatment at follow-up. Again, this finding underscored the need for the simultaneous use of both diagnostic systems. First, the Social Problems scale of the CBCL does not even have a clear counterpart in DSM taxonomy (APA 1987, 1994), and still contributed to the prediction of inpatient treatment. Second, DSM diagnosis of ADHD did not predict inpatient treatment in the first set of analyses, despite its close association with scores on the CBCL scale Attention Problems (i.e., Edelbrock & Costello, 1988; Jensen et al., 1993; Biederman et al., 1993; Kasius et al., 1997). Hence, to determine the prognosis, information derived with the CBCL, in addition to information regarding DSM diagnoses, was very useful.

Limitations

Most limitations of this study are associated with the generalizability of its findings. The sample size was small, and all participants were derived from the same outpatient clinic. Furthermore, treatment was unstandardized, and some syndrome-incongruent results seem to cast doubt on the representativeness of the study.

One seemingly incongruent result concerned the lack of association between psychotic problems and inpatient treatment. While it can be expected that

severe psychotic problems in adolescents predict inpatient treatment, the category psychotic problems did not necessarily indicate severe problems (see Methods section), and concerned 6- to 12-year-olds, in whom psychotic symptoms may less likely result in inpatient treatment, because children with psychotic symptoms may more easily be handled by their environment than adolescents with psychotic features. Furthermore, it is known that, even in adults, most individuals who suffer from psychotic features will never be referred to mental health services (van Os et al., 2000).

Difficulties with the law were not associated with delinquency or aggression, but were associated with ODD and GAD. This casts doubt on the external validity of the study, which can also be questioned because of the disparity between the rate of ADHD (41%) and the rate of pharmacotherapy (13%). The finding that problems with the law were not associated with scores on the Delinquent Behavior scale of the CBCL, or with the DSM-III-R diagnosis Conduct Disorder, may indicate that these diagnostic constructs did not cover problem areas that were relevant for the prediction of police/judicial problems. Different results might have been found if the prediction of poor outcome had been assessed in other samples. In the present sample, the rate of police/judicial contacts during the follow-up period was only 6.3% ($n = 6$), which indicates the low power of the present study to detect indicators of police contacts, and relatively poor generalizability of these indicators. The discrepancy between high rates of ADHD versus low rates of pharmacotherapy may indicate a difference between DISC and clinical DSM diagnoses. Consulting clinicians were not acquainted with DISC/DSM-III-R diagnoses, while recent evidence suggests that considerable discrepancies may exist between DISC and clinical DSM diagnoses (Jensen & Weisz, 2002). This discrepancy may result from rate bias between parents and clinicians, who may weigh the presence and severity of symptoms in different ways, but also from differences in the structure of diagnostic procedures. The DISC is a highly structured interview, while clinical assessment is often fairly unstructured. Furthermore, the consulting clinician was acquainted with diagnostic information from parents (but not with the DISC results), teachers, and children themselves, while parents were the only information source for the DSM-III-R/DISC diagnoses.

Another possible problem is that multiple testing may have resulted in chance findings. This, indeed, is a limitation of the present study. Some might argue that, to reduce the number of chance findings, corrections for multiple testing are needed. However, a better approach to test the value of the present study's findings is to investigate if similar findings are obtained from other samples.

In summary, it is not clear to what extent findings from the present study may be generalized to other

settings, with other patients and other types of treatment. However, despite these limitations, this study yields valuable information regarding the validity of two major diagnostic approaches, as well as on the way information from both approaches may be combined. The limitations do not negate the primary conclusions made.

Conclusions

The present study compared the predictive power of two main diagnostic approaches for child and adolescent psychopathology, the clinical-diagnostic approach and the empirical-quantitative approach. The study found that both diagnostic approaches provide valuable information regarding the outcome of psychopathology in children, and that the way symptoms are grouped by both systems results in strengths and weaknesses. But more importantly, it was found that, for optimal estimation of the prognosis, defined as the risk for a number of adverse outcome variables, neither system was superior to the other. Instead, the risk for poor outcome was best estimated by combining information from both diagnostic approaches. Even diagnostic concepts such as ADHD (DSM) and Attention Problems (CBCL), or Conduct Disorder (DSM) and Delinquent Behavior (CBCL), appeared to differ in their ability to predict poor outcome. Therefore, the present study supports the use of both systems simultaneously, in both research and clinical settings. Because of the small sample size, and the use of participants who were referred to mental health services, future research replicating the study in large non-selected samples is needed.

Correspondence to

Robert F. Ferdinand, Department of Child and Adolescent Psychiatry, University Hospital Rotterdam – Sophia, Dr. Molewaterplein 60, 3015 GJ Rotterdam, The Netherlands; Email: r.f.ferdinand@erasmusmc.nl

References

- Achenbach, T.M., & Edelbrock, C.S. (1983). *Manual for the revised Child Behavior Checklist and profile*. Burlington, VT: University Associates in Psychiatry.
- Achenbach, T.M., & McConaughy, S.H. (1997). *Empirically based assessment of child and adolescent psychopathology: Practical applications*. Thousand Oaks, CA: Sage Publications.
- Achenbach, T.M. (1991a). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles*. Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T.M. (1991b). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: University of Vermont Department of Psychiatry.
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd edn). Washington, DC: American Psychiatric Association.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (rev. 3rd edn). Washington, DC: American Psychiatric Association.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). Washington, DC: American Psychiatric Association.
- Barrett, P.M., & Turner, C. (2001). Prevention of anxiety symptoms in primary school children: Preliminary results from a universal school-based trial. *British Journal of Clinical Psychology*, 40, 1–12.
- Biederman, J., Faraone, S.V., Doyle, A., Krifcher Lehman, B., Kraus, I., Perrin, J., & Tsuang, M.T. (1993). Convergence of the Child Behavior Checklist with structured interview-based psychiatric diagnoses of ADHD with and without comorbidity. *Journal of Child Psychology and Psychiatry*, 34, 1241–1251.
- Cole, D.A., Peeke, L.G., Martin, J.M., Truglio, R., & Seroczynski, A.D. (1998). A longitudinal look at the relation between depression and anxiety in children and adolescents. *Journal of Consulting and Clinical Psychology*, 66, 451–460.
- Edelbrock, G., & Costello, A.J. (1988). Convergence between statistically derived behavior problem syndromes and child psychiatric diagnoses. *Journal of Abnormal Child Psychology*, 16, 219–231.
- De Groot, A., Koot, H.M., & Verhulst, F.C. (1994). Cross-cultural generalizability of the Child Behavior Checklist cross-informant syndromes. *Psychological Assessment*, 6, 225–230.
- Gould, M.S., Bird, H., & Staghezza Jaramillo, B. (1993). Correspondence between statistically derived problem syndromes and child psychiatric diagnoses in a community sample. *Journal of Abnormal Child Psychology*, 21, 287–313.
- Heijmens Visser, J., der Ende, J., Koot, H.M., & Verhulst, F.C. (1999). Continuity of psychopathology in youths referred to mental health services. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 1560–1568.
- Hofstra, M., van der Ende, J., & Verhulst, F.C. (submitted). 14-Year prediction of poor outcome and impaired social functioning.
- Jensen, P.S., Salzberg, A.D., Richters, J.E., & Watanabe, H.K. (1993). Scales, diagnoses and child psychopathology. I. CBCL and DISC relationships. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 397–406.
- Jensen, A.L., & Weisz, J.R. (2002). Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *Journal of Consulting and Clinical Psychology*, 70, 158–168.
- Kashani, J.J., Carlson, G.A., Beck, N.C., Hooper, E.W., Corcoran, C.M., McAllister, J.A., Fallahi, C., Rosenberg, T.K., & Reid, J.C. (1987). Depression, depressive symptoms, and depressed mood among a community sample of adolescents. *American Journal of Psychiatry*, 144, 931–934.
- Kasius, M. (1997). *Interviewing children. Development of the Dutch version of the Semistructured Clinical Interview for Children and Adolescents (SCICA) and*

- testing of the psychometric properties*. Academic dissertation, Erasmus University, Rotterdam, The Netherlands.
- Kasius, M.C., Ferdinand, R.F., van den Berg, H., & Verhulst, F.C. (1997). Associations between different diagnostic approaches for child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, 38, 625–632.
- Last, C.G., Strauss C.C., & Francis, G. (1987). Comorbidity among childhood anxiety disorders. *Journal of Nervous and Mental Disease*, 175, 726–730.
- Last, C.G., Perrin, S., Hersen, M., & Kazdin, A.E. (1992). DSM-III-R anxiety disorders in children: Sociodemographic and clinical characteristics. *Journal of the American Academy of Child and Adolescent Psychiatry*, 31, 1070–1076.
- Lowry-Webster, H.M., Barrett, P.M., & Dadds, M.R. (2001). A universal prevention trial of anxiety and depressive symptomatology in childhood: Preliminary data from an Australian study. *Behaviour Change*, 18, 1–14.
- NIMH (National Institute of Mental Health). (1992). *Diagnostic interview schedule for children*. New York: New York State Psychiatric Institute Division of Child and Adolescent Psychiatry.
- Piacentini, J., Shaffer, D., Fischer, P., Schwab-Stone, M., Davies, M., & Giola, P. (1993). The Diagnostic Interview Schedule for Children-revised version (DISC-R). III. Concurrent criterion validity. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 658–665.
- Rothman, K.J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology*, 1, 43–46.
- Schwab-Stone, M., Fisher, P., Piacentini, J., Shaffer, D., Davies, M., & Briggs, M. (1993). The Diagnostic Interview Schedule for Children-revised version (DISC-R), II. Test-retest reliability. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 651–657.
- Shaffer, D., Schwab-Stone, M., Fisher, P., Cohen, P., Piacentini, J., Davies, M., Conners, C.K., & Regier, D. (1993). The Diagnostic Interview Schedule for Children-revised version (DISC-R). I. Preparation, field testing, interrater reliability, and acceptability. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 643–650.
- Spence, S. (1998). A measure of anxiety symptoms among children. *Behavior Research and Therapy*, 36, 545–566.
- Steingard, R., Biederman, J., Doyle, A., & Sprich-Buckminster, S. (1992). Psychiatric comorbidity in attention deficit disorder: Impact on the interpretation of child behavior checklist results. *Journal of the American Academy of Child and Adolescent Psychiatry*, 31, 449–461.
- van Os, J., Hanssen, M., Bijl, R.V., & Ravelli, A. (2000). Strauss (1969) revisited: A psychosis continuum in the general population? *Schizophrenia Research*, 29, 11–20.
- Verhulst, F.C., van der Ende, J., & Koot, H.M. (1996). *Handleiding voor de CBCL/4-18 [Manual for the CBCL/4-18]*. Rotterdam: Afdeling Kinder- en Jeugdpsychiatrie, Sophia Kinderziekenhuis / Academisch Ziekenhuis Rotterdam, Erasmus Universiteit Rotterdam.
- Van Westerlaak, J.H., Kropman, J.A., & Collaris, J.W.M. (1975). *Beroepenklapper*. Nijmegen, The Netherlands: Instituut voor toegepaste sociologie.