European Child & Adolescent Psychiatry
9:263–270 (2000) © Steinkopff Verlag 2000

ORIGINAL CONTRIBUTION

R. J. McGuire
I. Berg
I. McKenzie
B. Wright
D. Foreman
K. Chandiramani

Have the Cross-Informant Syndromes of the CBCL any practical value in identifying grouped ICD10 diagnoses?

Accepted: 17 April 2000

R. J. McGuire
Department of Psychiatry
University of Edinburgh
EH10 5HF, Scotland

I. Berg (✉)
Cramond Old Barn
Low Gate Farm
Sawley
Ripon
North Yorkshire
HG4 3EL, England

I. McKenzie · B. Wright
Academic Department
of Child Mental Health
University of Leeds
LS2 9LT, England

D. Foreman · K. Chandiramani
Department of Psychiatry
University of Keele
ST4 7QH, England

Abstract 120 children referred to a child and adolescent psychiatric service in a university clinic were studied with the aim of deriving predictors for grouped ICD10 diagnoses using the CBCL Cross-Informant-Syndromes (CISs). CIS7 (Delinquent Behaviour) and CIS8 (Aggressive Behaviour) were shown to significantly separate Disruptive Behaviour Disorders from all other disorders. As cross-validation, a separate clinical sample of 118 children from a similar service in another part of the country was used to see to what extent the CIS predictors from the first sample held up in the second sample. Positive and Negative Predictive Powers, all corrected for chance, confirmed that the Disruptive Behaviour Disorder group could be usefully separated from all other disorders using the Delinquent Behaviour and the Aggressive Behaviour Cross-Informant Syndrome scores. There was no good evidence that Emotional (Anxiety-Mood) Disturbance could be usefully separated in the same way using the Anxious-Depressed Syndrome (CIS3) or any other syndrome. Discriminant Function Analysis showed that there was no significant improvement in prediction when more elaborate linear combinations of the syndromes were used.

Key words CBCL – Cross-Informant Syndromes – ICD10 – predictors – evaluation studies

Introduction

The objective specification of childhood psychiatric disorders is an important exercise, since it helps clinicians, wherever they work, to know that they are dealing with similar sorts of problems. Two different systems (5) have been used to find out what kinds of psychiatric disturbances affect children and to identify and characterise them. One is the method of empirically based taxonomy (2), which involves the use of factor analysis, a statistical technique that indicates what sort of disturbances exist and provides dimensions to measure each of them. Items which tend to correlate or cluster are identified. The CBCL (Child Behaviour Checklist), is a scale which has been employed in this way (3). It consists of a questionnaire which is completed by parents. It contains 120 behavioural items scored 0, 1 or 2 in the direction of problematic behaviour. Similar companion scales (2) can be completed by either teachers or the children themselves. It was developed by studying clinical samples but has been extensively used, more generally, all over the world (14). Eight Cross-Informant Syndromes or Factors (CISs) were identified in clinical samples using 89 of the items. (Six items of disturbed behaviour contributed to more than one factor). As the name implies, the CISs apply whichever informant, parent, teacher or child, completes the appropriate scale. The CISs also apply whatever the age and sex of the child. High scores on a particular factor may be said to indicate the existence of the kind of disturbance represented by that dimension (1, 14).

European Child & Adolescent Psychiatry, Vol. 9, No. 4 (2000)

The other major system of identifying disturbances has applied the traditional medical approach of classifying disorders by specifying criteria to separate one disorder from another on the basis of clinical experience and research. Disorders are conceptualised as categories rather than dimensions. ICD10, the tenth version of the International Classification of Diseases (15), is used all over the world. It contains categories which have been considered to be generally acceptable to those who work with disturbed children. Unlike its main rival amongst categorical systems of classification, DSM IV, it prefers one category to be used rather than encouraging the employment of several. Both the CBCL and the ICD10 systems of specifying psychiatric disturbances are widely employed in clinical work and research. Without any preconceptions that one approach is better than the other, it would be helpful to know whether important groups of disorders categorised by ICD10 can be predicted using the CBCL CISs. If they are, then the preliminary use of the CBCL might have some advantages in grouping cases for separate study, assessment and treatment.

Some success has already been achieved in attempting to link CBCL factor scores and the categories of the DSM system of classification (8). Edelbrock and Costello (6) used a particular version of the CBCL (3) in relation to DSM III diagnoses, with some success. Biederman et al. (5) also reported some ability of individual CBCL factors to predict DSM IIIR diagnoses. Steinhausen et al. (12) found satisfactory differentiation of groups of children with either Anxiety Disorder or ADHD or with a combination of these two DSM IIIR diagnoses, the differentiation was judged from patterns of differences in mean scores: there was no attempt to assign cases using these patterns or any other allocation procedure. Other attempts are referred to by Kasius et al. (9) who reported their own study of the CBCL in relation to DSM IIIR diagnoses. Individual categories were collapsed into broader groups. Some support was found for a useful correspondence. As far as we know, no studies have been reported so far using CBCL CISs with ICD10.

If the CBCL is going to be helpful in predicting diagnostic groupings, it should be able to distinguish between the broadest categories. It was therefore decided to look at two common types of problem coming to clinics, Disruptive Disorders (F90 and F91) and Emotional Disturbances (F93, F32, F40, F41, F42, and F43). Because of its high incidence, a third type of problem, incontinence, comprising non-organic encopresis and enuresis (F98) was initially considered as worth looking at, but the CBCL scores of this group were found to be not significantly different from those of categories other than Disruptive Behaviour Disorders and Emotional Disturbances, and this diagnosis was therefore included with the miscellaneous group in the main analyses.

Methods

Samples

Leeds cases consisted of 120 children referred to the child and adolescent psychiatric service of the Leeds General Infirmary, a university hospital. They were any referral aged 6–15 years inclusively, where there was informed parental consent to participate in the research. As far as practical considerations allowed, they were unselected consecutive referrals. The Keele cases, 118 children, were also an unselected series of referrals to a clinic for childhood psychiatric disorders associated with the University of Keele, in Staffordshire. In their case, the CBCL had been routinely used as part of clinical assessment, just before parents and child were seen.

Diagnostic groups

In Leeds, mothers were administered the CBCL by computer (4) an hour or so before child and parent(s) were interviewed by clinician A for clinical assessment to whom CBCL scores were not known by him when this assessment was carried out. The case notes were subsequently gone through by clinician B who made a principal ICD10 diagnosis (15). He also did not know the CBCL scores. In Keele, CBCL scores were known when clinician C carried out his clinical assessment. However, Clinician D, who subsequently made an ICD10 diagnosis from the case notes, did not know them.

In the main analyses the ICD10 diagnoses were grouped as follows: 1. *Disruptive Behaviour Disorders* (*DBD*), including Conduct Disorder (F91.0, F91.1, F91.2, F91.8 and F91.9), Oppositional Defiant Disorder (F91.3) and Hyperkinetic Disorder (F90.9) 2. *Emotional Disturbance* (*ED*), Anxiety and Mood Disorders, including F93.0, F93.1, F93.2 and F93.8, F32.0, F40.2, F41.1 and F41.3, F42.1 and F42.2, F43.21 and F43.23 and F43. 3. *Alternative Diagnoses* (*ALTD*), any other category, including both Non-organic Enuresis (F98.0) and Non-organic Encopresis (F98.1) (*NOEE*).

Cross-Informant Syndromes (CISs)

The CISs are labelled as follows (number of items): CIS1: Factor I Withdrawn (9), CIS2: Factor II Somatic Complaints (9), CIS3: Factor III Anxious-Depressed (14), CIS4: Factor IV Social Problems (8), CIS5: Factor V Thought Problems (11), CIS6: Factor VI Attention Problems (11), CIS7: Factor VII Delinquent Behaviour (13) and CIS8: Factor VIII Aggressive Behaviour (20). CIS scores are produced by the simple summation of appropriate item scores (0, 1 or 2) for each factor.

Inter-rater reliability of the ICD10 diagnoses

Rater I independently diagnosed a random sample of 74 Leeds cases to make an estimate of reliability by comparison with rater 2's diagnoses, which were used in the study. For individual ICD10 categories the overall Kappa statistic was 0.9 which establishes that the categorisation was statistically reliable (10).

Statistical analyses

All the statistics were undertaken using the SPSS package for Windows. Measures of predictive efficacy were calculated using the formulae given in Kraemer's List of Statistical Notations (10). One-way analyses of variance (ANOVAs) were used initially, with cases in both samples combined, to compare the ICD10 diagnostic groups on their mean scores for each of the Cross-Informant Syndromes. Graphs of cumulative frequency distributions (i.e. numbers of children within each of the diagnostic groups having scores on some Cross-Informant Syndromes at or below a particular level), were produced for the Leeds cases. This used only those Cross-Informant Syndromes which had shown significant differences between diagnostic groups on the one-way ANOVAs.

The score at which two frequency distributions are equal, causing their curves to cross, also corresponds to the score at which two cumulative frequency curves have maximum vertical separation and therefore represents an optimal cut-off for distinguishing between those two groups. For three groups, as here, ideally one would hope that the frequency distributions of the groups would show clear separation with one group having mainly low scores, another group bunched in the middle and the third group with mainly high scores. Such data would lead to one of the CF curves rising steeply from quite a low score so that at a score less than or equal to, say, 15 almost all of that diagnosis would have been included, at which point one of the other diagnoses would start to rise so that by less than or equal to, say, 25 it too would be completely covered and finally the last diagnosis would begin to accumulate its cases. It would then be a simple matter to note the two points near 15 and 25 where the vertical distances between the CF curves are greatest and use them as cut-off points for the selection of three samples which would largely allocate each subject to his/her appropriate diagnosis. At the other extreme, a scale which had no discriminating power would show great overlap among the diagnoses and the CF curves would follow very similar paths with no points at which there was any useful vertical distance between them.

In this study, the appropriate cut-off points for each relevant CIS scale were derived from the CF curves and then used to calculate Sensitivities, Specificities, Positive Predictive Powers, Negative Predictive Powers and Overall Efficiencies, which are the standard measures used to evaluate the usefulness of a scale to select one particular group of subjects from a heterogeneous sample. These measures are defined in Kraemer's List of Statistical Notations (10). These statistics were corrected to help guard against spurious results arising by chance (see below). They were calculated for the Leeds cases and then for the Keele cases, as a test of cross-validation, after the Leeds selection procedures had been applied to them.

Discriminant Function (DF) analyses were next used so that the relationship between the best combination of several CISs and the diagnostic groups could be explored. DF equations derived from Leeds cases were then applied to the Keele sample, again as a test of cross-validation.

When a CIS scale selected a group as positive for a specific diagnosis, only some were clinically positive constituting the *true positives* (TP), but others were clinically negative, the *false positives* (FP). Of those rejected by the scale for the diagnosis, some were clinically negative, constituting the *true negatives* (TN), while others were clinically positive, the *false negatives* (FN).

There are two main ways of measuring the success of any *selection* procedure, Sensitivity (SENS) which states what proportion of the clinical cases were TP and Positive Power of Prediction (PPP) which states what proportion of the cases selected by the scale were TP. Similarly, there are two ways of measuring the success of the *rejection* procedure, Specificity (SPEC) which measures what proportion of the clinical non-cases are TN and Negative Power of Prediction (NPP) which measures what proportion of those rejected by the scale were TN. Overall efficiency (EFF) is the proportion of the whole sample who have been correctly assigned, i.e. the true positives plus the true negatives. It should be noted that a cut-off point can be arbitrarily moved so as to increase the number of true positives, but that will almost inevitably decrease the number of true negatives with the effect usually of increasing SENS and PPP whilst decreasing SPEC and NPP. Which measure is 'best' to use depends on the purpose of the selection which may depend on such things as the importance of identifying as many cases as possible versus the cost of treating non-cases.

Since even random allocation will inevitably produce a certain number of true positives and true negatives, it is advisable to estimate this effect of chance and to correct for it. The number of TP and TN expected by chance can be easily calculated from the marginal totals of each two by two table of scale selection ($+/-$) versus clinical diagnosis ($+/-$). The expected value is then subtracted from the observed value and also from the

value of the group from which they have been selected. This gives (10) Kappa(0,0) as the identical corrected values of both SPEC and PPP, Kappa(1,0) as the identical corrected values of both SENS and NPP and Kappa(0.5,0) as the corrected value of EFF. This last is also known as Cohen's Kappa and is arithmetically the harmonic mean of Kappa(0,0) and Kappa(1,0) so there are really only two independent measures when corrected for chance. Chi^2 is equal to K(0,0) times K(1,0) times N, the total number in the sample, which gives both a useful check on one's calculations and a measure of statistical significance. If the cases are exactly half the total sample (as is the case for DBD in Keele) then K(1,0) is the fraction by which TP exceeds that expected by chance, equally so for K(0,0) if selection/rejection splits the sample in half and in both cases it is so for K(0.5,0) in respect of (TP + TN). Unfortunately there is no easy interpretation of the Kappas in other situations.

Results

(Statistical significance is at the $p < 0.5$ level unless otherwise stated).

Comparing Leeds and Keele cases

There was a significant difference between the mean ages of all the Leeds (9.7 years) and Keele (8.4 years) cases and also for each of the four diagnostic groups. There was no significant difference in the proportion of boys and girls in the two samples. Social class was not recorded. About half of all the diagnoses in both samples were DBD, a quarter ED and a quarter ALTD.

One-way analyses of variance (ANOVA) of the CIS scores by diagnostic group

As a preliminary survey, using all 238 cases, a one-way analysis of variance was carried out for each of the CIS scales categorised into four diagnostic groups namely: DBD, ED, NOEE and MISC. Where the F-ratio was statistically significant, Tukey's Honestly Significant Difference (HSD) test was applied to discover which groups had significantly different mean scores from each other (Table 1).

For each of CIS1, CIS2, CIS4 and CIS5 there was no statistically significant difference between the mean scores of the four groups. For CIS3 (Anxious-Depressed) the F-ratio was significant ($p < 0.01$) and HSD showed that the ED group had a higher mean score than the other three groups. For CIS6 (Attention Problems) the F-ratio had $p < 0.001$ with the mean score of the DBD group greater than those of the ED and NOEE diagnoses. For CIS7 (Delinquent Behaviour) and CIS8 (Aggressive Behaviour) the F-ratios had each a $p < 0.001$ and the mean score of the DBD group was much higher than those for any of the other groups.

Since none of the CIS scales showed the mean score of the NOEE group to be significantly different from the MISC group and since there are no questions in the syndrome scales directly referring to soiling, those two groups were combined into one diagnostic group ALTD for all subsequent analyses.

Cumulative frequency curves

Using only the Leeds data, for each of the four CIS scales which ANOVA had shown to be significant, a cumulative frequency (CF) graph was drawn with separate curves for each of the three diagnostic groups.

Table 1 Means (M), Standard Deviations (sd) and One-way Analyses of Variance (ANOVA), for each Cross-Informant Syndrome Score of the CBCL (CISs), for four ICD10 grouped diagnoses: DBD, ED, NOEE and MISC for 238 referrals to Child and Adolescent Psychiatric Clinics in Leeds and Keele

Group numbers CIS scales	DBD 113		ED 60		NOEE 38		MISC 27		All cases 238		ANOVA 3/234	
	M	sd	M	sd	M	sd	M	sd	M	sd	F	p
1 Withdrawn	5.3	3.3	6.5	4.2	5.8	4.4	5.4	4.4	5.7	3.8	1.32	0.27
2 Somatic Complaints	3.2	3.0	4.3	3.6	3.2	2.4	4.0	3.8	3.5	3.2	1.82	0.14
3 Anxious Depressed	8.9	5.5	11.5	5.8	7.8	4.9	8.3	7.1	9.3	5.8	4.40	0.01
4 Social Problems	6.0	3.1	4.7	3.0	5.6	3.6	4.9	3.8	5.5	3.3	2.25	0.08
5 Thought Problems	3.2	2.8	2.9	2.3	2.2	2.1	3.6	3.1	3.0	2.7	1.79	0.15
6 Attention Problems	12.6	4.7	8.9	4.9	10.0	6.0	10.6	6.8	11.1	5.4	7.44	0.00
7 Delinquent Behaviour	9.4	4.5	3.1	2.5	4.2	3.3	3.3	3.7	6.3	4.9	48.71	0.00
8 Aggressive Behaviour	26.2	6.2	14.7	8.3	15.6	7.9	14.4	9.8	20.2	9.4	44.43	0.00

All the abbreviations are explained in the text. DBD Disruptive Behaviour Disorders, ED Emotional Disturbances, NOEE Non-organic Enuresis & Encopresis, MISC Other Miscellaneous Diagnoses

For CIS3 (Anxious-Depressed) the CF curves did not deviate far from each other (Fig. 1). At the point of optimum separation (less than or equal to 10), 23 out of 54 (43%) of the DBD group had been placed in the upper group along with 10 out of 31 (32%) of the ALTD group and 24 out of 35 (69%) of the ED group. Although this had led to a greater concentration of EDs it is clear that they were still outnumbered by the other two groups 24 against 33. The PPP for selecting the ED group was 0.42 (corrected for chance 0.19), NPP was 0.83 (corrected for chance 0.40), EFF was 0.63 (corrected for chance 0.25) and $Chi^2 = 9.12$, $p < 0.01$.

For CIS6 (Attention Problems), again the curves were not greatly separate, so that even at the optimum score (less than or equal to 15), while 32 out of 35 (91%) of the ED group were selected by the scale so were 39 out of 54 (72%) of the DBD group and 18 out of 31 (58%) of the ALTD group with the result that the ED cases were only 36% of those picked out which was no great improvement on their concentration of 29% in the original group.

For CIS7 (Delinquent Behaviour), Fig. 2 shows that the curves were more encouraging since the DBD group stayed low while the curves for the other two diagnoses climbed quite early. At a cut-off point of less than or equal to 3, the upper group (on the right-hand side of the cut-off point) consisted of 49 out of 54 (91%) of DBD, 8 out of 35 (23%) of ED and 9 out of 31 (29%) of the ALTD group. For the DBD group therefore their selection from the others gave PPP = 0.74 (corrected for chance 0.53), NPP = 0.91 (corrected for chance 0.79), EFF = 0.82 (corrected for chance 0.64) and $Chi^2 = 50.24$, $p < 0.001$. At no point was there a useful separation of the ED from the ALTD curve.

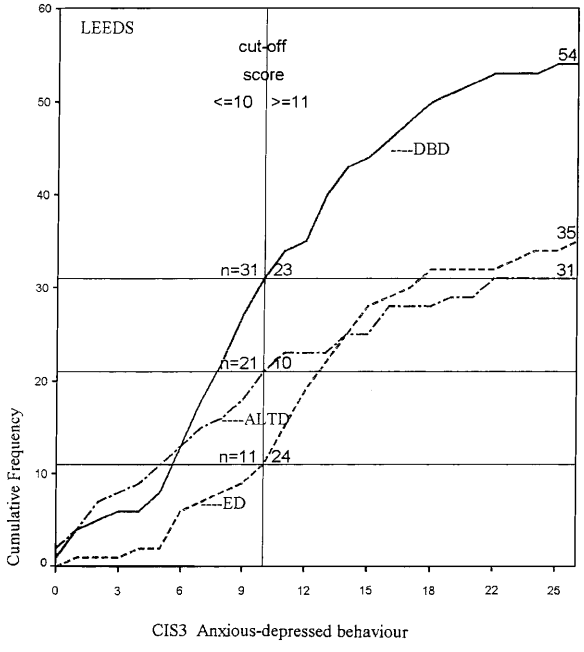

Fig. 1 Cumulative frequency curves for scores on CIS3 (Anxious-Depressed) of the Leeds diagnostic groups DBD, ED and ALTD. The optimum cut-off point (≤10) for separating the ED group from all others divides each group into numbers shown

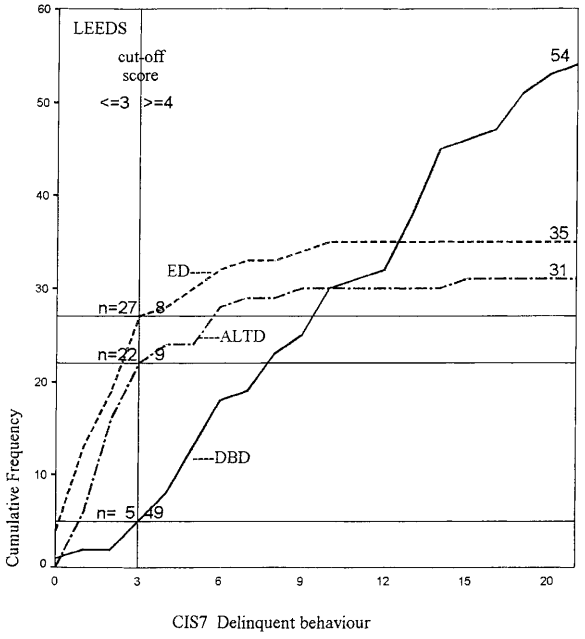

Fig. 2 Cumulative frequency curves for scores on CIS7 (Delinquent Behaviour) of the Leeds diagnostic groups DBD, ED and ALTD. The optimum cut-off point (≤3) for separating DBD from all others divides each group into the numbers shown

CIS8 (Aggressive Behaviour) gave a similar set of curves. Above the score of less than or equal to 20, those selected by the scale consisted of 44 out of 54 (81%) DBD, 6 out of 35 (17%) ED and 9 out of 31 (29%) of the ALTD group. For the separation of the DBD group from the others therefore PPP = 0.75 (corrected for chance 0.52), NPP = 0.89 (corrected for chance 0.64), EFF = 0.79 (corrected for chance 0.58) and $Chi^2 = 39.94$, $p < 0.001$.

The use of any single CIS score therefore gave adequate separation of only DBD cases from all other cases and this was done most effectively by using CIS7 (Delinquent Behaviour).

Discriminant Function (DF) analysis

DF analysis is generally the most effective procedure for assigning subjects to their appropriate diagnoses by combining their scores from several scales. In this study, DF analysis of the Leeds children by a stepwise combination of the CIS scores that made a significant contribution was used to match, as far as possible, their

three diagnostic groupings, namely DBD, ED and ALTD. It produced two independent dimensions (as in principal component factor scores) in which each case's two scores were determined by two DF linear equations which used CIS7 (Delinquent Behaviour), CIS8 (Aggressive Behaviour), CIS3 (Anxious Depressed Behaviour) and CIS6 (Attention Problems), those same CIS scales which had been significant in the one-way ANOVA for this data. The equations are derived so as to maximise the distances between groups and yet to keep all subjects in each group as close together as possible round their centre of gravity, referred to as their centroid.

In this study the equations allowing cases to be assigned with statistical confidence to the three desired groups had the formulae:

$$DF1 = 0.61Cz7 + 0.59Cz8 - 0.11Cz3 - 0.10Cz6$$
$$DF2 = 1.06Cz3 - 0.96Cz6 + 0.23C27 + 0.12Cz8 \ ,$$

where Cz represents standard scores on the respective scales. One-way ANOVA for DF1 by diagnostic groups produced $F = 56.51$, $p < 0.001$, HSD separated the DBD group from the other two and ANOVA for DF2 had $F = 9.10$, $p < 0.001$ and HSD separated the ED group from each of the other two groups.

Figure 3 shows the centroid position (x) of each of the three diagnostic groups of the Leeds children plotted against the two derived functions. The boundaries which separate two groups, in terms of the probability of being included in one group or the other, are also shown. It is clear that DF1 which mainly consisted of CIS7 (Delinquent Behaviour) and CIS8 (Aggressive Behaviour) separated the DBD children from the children in the other groups, while DF2, mainly consisting of CIS3 (Anxious Depressed Behaviour) and a negative score on CIS6 (Attention Problems), separated the ED children from the ALTD children. As was the case with the HSD results of the original ANOVAs in Table 1.

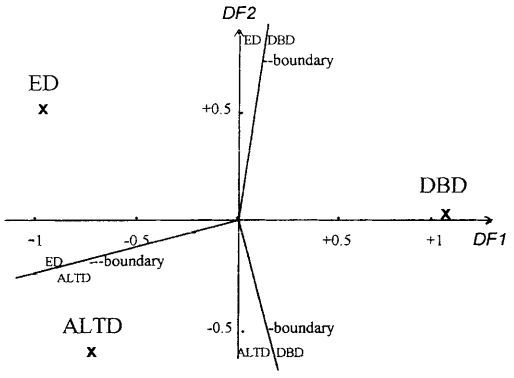

Fig. 3 Centroid position (x) and boundaries for the probability of belonging to each diagnostic group DBD, ED and ALTD of Leeds children plotted on the bivariate distribution of DF1 and DF2

Needless to say, in any realistic situation the equations will not be able to assign every case to its original diagnostic group, even though one is being wise after the event. On the Leeds cases from which they were derived, the DF equations correctly assigned 45 out of 54 (83%) DBD cases, 21 out of 28 (75%) ED cases and 21 out of 38 (55%) ALTD cases. The overall efficiency for all three groups was 87 out of 120 (0.73, 0.58 corrected for chance).

As cross-validation exactly the same formulae were applied to the Keele cases. One-way ANOVA of DF1 scores by diagnostic groups gave $F = 42.00$, $p < 0.001$ and, as with the Leeds cases, HSD separated DBD from both ED and ALTD. For DF2, ANOVA gave $F = 7.24$, $p < 0.01$ and HSD separated ED from both DBD and ALTD groups. Applying both functions correctly assigned 45 of 59 (76%) DBD cases, but only 13 out of 25 (52%) ED cases and 14 out of 34 (41%) ALTD cases The overall efficiency for all three assignments was 72 out of 118 cases (0.61, 0.38 corrected for chance).

Table 2 Comparison of different procedures used to separate the DBD children from the rest using: a) CIS7 (Delinquent Behaviour) with optimum cut-off score of greater or equal to 4, b) CIS8 (Aggressive Behaviour) with optimum cut-off score of greater or equal to 21 and c) DFA2 In each case the resulting allocation data are given and standard measures of predictive efficacy are given both for the Leeds sample from whom the predictors were derived and for the Keele children to whom they were applied

		Data				Measures of predictive efficacy							
		TP	FP	FN	TN	PPP	NPP	SENS	SPEC	EFF	K(0.5, 0)	K(1, 0)	K(0, 0)
CIS7	Leeds	49	17	5	49	0.74	0.91	0.91	0.74	0.82	0.64	0.79	0.53
	Keele	57	27	2	32	0.68	0.94	0.97	0.54	0.75	0.51	0.88	0.36
CIS8	Leeds	44	15	10	51	0.75	0.84	0.81	0.77	0.79	0.58	0.64	0.52
	Keele	51	17	8	42	0.68	0.84	0.86	0.71	0.79	0.58	0.68	0.50
DFA2	Leeds	43	8	11	58	0.84	0.84	0.80	0.88	0.84	0.68	0.65	0.71
	Keele	50	13	9	46	0.79	0.84	0.85	0.78	0.81	0.63	0.67	0.59

All the abbreviations are explained in the text. TP True Positive, FP False Positive, FN False Negative, TN True Negative, PPP Positive Power of Prediction, NPP Negative Power of Prediction, SENS Sensitivity, SPEC Specificity, EFF Overall Effectiveness, K(0.5, 0), K(1, 0), K(0, 0) are Kappas which correct for chance respectively EFF, SENS (& NPP), SPEC (& PPP)

A second DF analysis (DFA2) was done on the Leeds cases using the diagnostic groupings of DBD against all other cases so that a fair comparison could be made with the CF procedure which had proved capable of only that level of discrimination. The results for Leeds and Keele are given in Table 2.

Standard measures of the discriminating powers of the selection procedures used in the study

For comparison purposes, Table 2 shows the TP, FP, FN and TN resulting after using the three procedures described to select DBD cases from all the cases in each clinic sample. Eight indices assess the predictive values of: a) CIS7 (Delinquent Behaviour), b) CIS8 (Aggressive Behaviour) each used on their own and c) DFA2 analysis using linear combinations of all the statistically significant CIS scales. The figures are given for both the 120 Leeds cases from which the best cut-off points and DFA2 equations were derived and for the 118 Keele cases to which they were applied as cross-validation.

Discussion

Despite the fact that the mean raw CIS scores of the English children in DBD and ED (Table 1) were statistically significantly different from those of Swiss children (12) with diagnoses of ADHD and Anxiety Disorder, profiles of mean T scores for the eight CIS scales for English DBD and ED cases, derived from Table 1, correlated significantly (0.78, $p < 0.05$ and 0.88, $p < 0.01$ respectively) with the mean T score profiles of the ADHD and Anxiety Disorder Swiss groups. Differentiation of the profiles was also significant with DBD and ED correlating negatively (−0.81, $p < 0.05$ and −0.87, $p < 0.01$ respectively) with Anxiety Disorder and ADHD in the Swiss profiles. This suggests that these two pairs of diagnostic groups show corresponding patterns of CIS scores despite the fact that the diagnoses were not identical and were determined by the differing systems of ICD10 and DSM IIIR.

Although the mean ages of the Leeds and Keele samples were different, this applied across all diagnostic groups and would not be expected to bias the results since the CIS scales were designed to be unaffected by age or sex. There was no reason to suppose that the social class distribution of referrals to university clinics, in the two parts of the country characterised by industrial conurbations, would be substantially different.

The fact that the Leeds and Keele samples of cases showed similar features suggests that the findings of the study should be applicable to other clinical samples of the same sort. The two groups of cases were studied very differently. In Leeds, there was a carefully designed research project with attention paid to diagnostic reliability and independence of assessments. The CBCL was administered by computer, which is a reliable way of giving it, and has the advantage of making it impossible to ignore questions (4). Diagnoses were made independently and there was no bias due to knowledge of CBCL scores.

The situation in Keele was very different in that the CBCL had been routinely administered in the course of day-to-day clinical work. The CBCL scores may well have influenced what was put into the clinical records which were later used for making the ICD10 diagnoses. Since this would tend to increase the correspondence between CBCL scales and diagnoses, it may have undermined the positive findings linking the CBCL and Disruptive Behaviour Disorders. However, it would tend to strengthen the lack of findings in respect of other disorders, including Anxiety Mood Disturbances.

The justification for combining the various ICD10 disorders into groups was not entirely due to the need to keep the number of children in any cell at a reasonable level for statistical analysis, it was also done for practical purposes. ICD10 considers the various Disruptive Behaviour Disorders together. There has been a lot of concern expressed about the category of oppositional defiant disorder (7, 11). On the one hand, it can be conceptionalised as virtually normal childhood behaviour and, on the other, as a possible precursor of later conduct disorder. Likewise the strong association between conduct disorder and hyperkinetic disorder also leads to some blurring between the two (13). Anxiety and mood disorders are frequently associated, so it seems reasonable to consider them together. One of the problems with some of the earlier attempts to link the CBCL with a DSM diagnosis was the failure to lump similar disorders to create larger groups when this would have been reasonable. From a practical point of view there is much to be said for identifying disruptive and emotional disorders separately from each other, and from other disorders, to facilitate assessment and treatment.

The fact, as shown by ANOVA, that the mean scores for the ED group on CIS3 (Anxious-Depressed) and CIS6 (Attention Problems) were different from those of the other two groups unfortunately did not help in differentiation since the frequency distributions of all three groups were so similar with a lot of overlap. This finding may nevertheless be useful to clinicians in labelling typical and atypical cases with the ED diagnosis. A general lesson to be learned from this study is that even highly significant differences in mean scores between different diagnostic groups does not demonstrate that these scores can usefully assign individuals to their appropriate diagnostic groups: only a cross-validation trial of actual discrimination can test that.

Corrections for chance should also be employed as in this and recent studies where there was a similar task to perform (9, 10).

For separating out DBD cases from all others, CIS7 could be routinely used to produce two groups with respectively high and low concentrations of that diagnostic group.

Although Discriminant Function Analysis was good at separating the total sample into three diagnostic groups and increased the overall hit rate to 72 as compared to the 44 that would have been expected by chance (a 64% improvement), it nevertheless still assigned only about half of the ED and ALTD children to their appropriate diagnosis. Even for just separating the DBD group from the others, it was hardly better than using just CIS7, a much simpler decision process.

The correspondence between the scales CIS7 and CIS8 and the Disruptive Behaviour Disorders for both clinics and for different statistical procedures indicates that it is a finding of some substance. Nevertheless, it remains to be seen whether administering the CBCL is the best way of identifying disruptive behaviour, when a smaller number of focussed questions might suffice for that task.

References

1. Achenbach TM (1991) Integrative Guide for the 1991 CBCL/4-18, YSR and TRF Profiles. University of Vermont, Burlington
2. Achenbach TM (1995) Empirically based assessment and taxonomy. Applications to clinical research. Psychological Assessment 7:261–274
3. Achenbach TM, Edelbrock C (1983) Manual for the Child Behaviour Checklist and Revised Child Behaviour Profile. Burlington: University of Vermont, Department of Psychiatry
4. Berg I, Lucas C, McGuire RJ (1992) Measurement of behaviour difficulties in children using standard scales administered to mothers by computer: reliability and validity. European Child and Adolescent Psychiatry 1:14–23
5. Biederman J, et al (1993) Convergence of the child behaviour checklist and structured interview-based psychiatric diagnosis of ADHD children with and without comorbidity. Journal of Child Psychology and Psychiatry 34:1241–1251
6. Edelbrock C, Costello A (1988) Convergence between statistically derived behaviour problem systems and child psychiatric diagnosis. Journal of Abnormal Child Psychology 16:219–231
7. Frick PJ, Lahey BB, Aplegate B, Kerdyck L, Ollendick T, Hynd GW, Garfinkel B, Greenhill L, Biederman J, Barkley RA (1994) DSMIV field trials for the disruptive behaviour disorders: symptom utility estimates. Journal of the American Academy of Child and Adolescent Psychiatry 33:529–539
8. Jensen PS, Salzberg AD, Richters JE, Watanabe HK (1993) Scales, diagnoses, and child psychopathology: I CBCL and DISC relationships. Journal of the American Academy of Child and Adolescent Psychiatry 24:151–168
9. Kasius MC, Ferdinand RF, van den Berg H, Verhulst FC (1997) Associations between different diagnostic approaches for child and adolescent psychopathology. Journal of Child Psychology and Psychiatry 38:625–632
10. Kraemer H (1992) Evaluation of Medical Tests X. Newtonbury Park, California Sage publications, Newtonbury Park, CA
11. Loeber R, Keenan K, Lahey BB, Green SM, Thomas C (1993) Evidence for developmentally based diagnoses of oppositional defiant disorder and conduct disorder. Journal of Abnormal Child Psychology 21:377–410
12. Steinhausen HC, Winkler Metzke C, Meier M, Kannenberg R (1997) Behavioural and emotional problems reported by parents for ages 6 to 17 in a Swiss epidemiological study. European Child and Adolescent Psychiatry 6:136–141
13. Taylor E (1994) Syndromes of attention deficit and overactivity. In: Rutter M, Taylor E, Hersov L (eds) Child and Adolescent Psychiatry. Modern Approaches, 3rd edn. Blackwell Scientific Publications, Oxford, pp. 285–387
14. Verhulst F, Achenbach T (1995) Empirically based assessment and taxonomy of psychopathology: cross-cultural applications. European Child and Adolescent Psychiatry 4:61–76
15. World Health Organisation (1992) The ICD10 Classification of Mental and Behavioural Disorders. WHO, Geneva