Classification and diagnosis of depression in school phobia.

I Kolvin, T P Berney and S R Bhate

BJP 1984, 145:347-357.
Access the most recent version at doi: 10.1192/bjp.145.4.347

British Journal of Psychiatry (1984), **145**, 347–357

Classification and Diagnosis of Depression in School Phobia

ISRAEL KOLVIN, THOMAS P. BERNEY and SURYA R. BHATE

Summary: Fifty-one school phobic children, aged nine to fourteen years, were assessed for psychiatric diagnosis; this revealed the presence of two clinically meaningful sub-groups—depressed and residual school phobic. A wide range of symptoms were studied to identify those which might prove useful in diagnosing adult-type depression in childhood and early adolescence, both in terms of frequency of symptoms in the depressed group and the extent of the distinction between the two groups. Eleven such key symptoms were identified and based on these, a formula for diagnosing adult-type depression was evolved. The validity of several different ways of classifying the above cases were explored; these covered: kind of disorder; type of onset; adolescence versus pre-adolescence; and sex of the child. However, on only one dichotomy—depression versus absence of depression—were there many significant discriminants. Affective symptomatology of more recent onset was contrasted with pre-morbid personality traits, usually associated with school phobia. There was no evidence to support the concept of 'masked depression' in childhood.

'School phobia' is a misnomer for a heterogeneous collection of mainly neurotic disorders, associated with a grave reluctance to attend school, and there are thus sound theoretical and practical reasons for attempting to identify the main sub-groups of this syndrome. Non-attendance at school has been divided into three main groups: those conditions in which there are no obvious social or psychological factors, such as stem from true physical illness; conditions where there are major social factors, as occur in truancy, or where the child is kept at home for domestic reasons; and thirdly, school refusal, where there are associated major psychological factors (Kolvin & Nicol, 1979). This paper is confined to the third group, and considers its clinical and statistical classification.

Aims

This study had three aims. The first was to ascertain if there were clinically meaningful sub-groups in a population of children with school phobia; in psychiatric practice, two broad approaches to classification are commonly used—clinical and statistical (Garside & Roth, 1978). Our second aim was to compare the depressive sub-type with the rest of the school phobic children, while the third was to construct a set of criteria for the diagnosis of depression.

Are there meaningful sub-groups?

The clinical approach: In essence, this consists of a careful study of the clinical phenomena to identify those characteristics which appear to hang together so well that they constitute a syndrome. However, the possibility that any groups identified were distinct clinical entities would be strengthened if they could be shown to have prognostic or predictive validity. Initially, these descriptions would not constitute more than clinical hypotheses which would have to be validated. The first essential step is to identify a set of cardinal diagnostic criteria, which ideally would then be ratified by statistical frequency analysis. A complementary step is the systematic empirical validation of the hypothesis by definition of the distinctions between the particular syndrome and other relevant syndromes or control groups. Such distinctions should go beyond the clinical phenomena to include historical, family, social, other aetiological, and (where possible) prognostic evidence.

The statistical approach

(i) By discriminant function analysis: The purpose is to test the distinctions within an existing classification of patients, using a set of clinically relevant measures. In addition, these measures are so weighted as to

maximise the distance between the groups; hence, the technique can be used to compare two populations of patients with different clinical diagnoses. In the same way, it can be used to check the validity of the dichotomisation of a population into mutually exclusive sub-groups (on the basis of defined criteria); furthermore, multiple ways of dichotomising the same population can be simultaneously studied. However, the technique is not without limitations, since for example, a variation of the set of items used or the set of subjects studied may give rise to different results. It is therefore essential to attempt to replicate the distinctions which have been identified on a new sample of cases, and with some variations of the measures used.

(ii) By factor and cluster analysis: These multivariate techniques are designed to identify symptoms (and perhaps other features) which hang together in characteristic patterns (factor analysis); this view is consistent with the hypothesis that there is one general factor (or dimension) underlying all the symptoms, as well as additional differentiating or bipolar factors. The techniques can also identify those sub-sets of features or individuals which cluster together, in the sense of having much in common within each cluster but less in common between the clusters (cluster analysis). These methods not only summarise relationships between features, but may also highlight a new set of harmonies or structures (Cattell, 1965; Garside & Roth, 1978).

However, they are not a classification panacea. The results of factor analysis may be considerably influenced by changes either in the size or composition of the subject sample or in the range of measures employed (Kendell, 1975). Another well-known limitation to the use of factor analysis for classification purposes is that in practice, only a minority of patients may be accommodated by the classification (Hewitt & Jenkins, 1946). Other problems are such questions as whether the data are normally distributed, and whether it is legitimate to include variables which are binary in nature (Garside & Roth, 1978). These qualifications do not apply in the case of principal component analysis, which therefore appears to be a more robust technique for analysing clinical data.

On the other hand, while cluster analysis theoretically appears to be the most useful way of identifying relatively homogeneous sub-sets of individuals, the validity and number of clusters identified are still open to debate. For instance, questions remain about how to establish the validity of the clusters generated; not only may differences emerge using different programmes of cluster analysis on the same population of patients, but different results are often obtained from a study of different samples of patients with the same diagnosis.

Despite such limitations, and provided that the pitfalls are constantly borne in mind, these techniques usefully complement the traditional clinical methods of identifying syndromes. In summary, multivariate techniques can either provide powerful statistical evidence in support of an already conceptualised classification or suggest new ways of categorising behaviour or classifying individuals.

Method

(a) *The population:* The population in this study has previously been described (Berney *et al*, 1981); it consists of 51 children suffering from school refusal, which was defined as the association of a neurotic disorder with a marked reluctance to attend school for at least four weeks. Clear-cut cases of truancy and psychotic disorders were excluded. The children were between their ninth and fifteenth birthdays on referral, with a slight preponderance of girls (55%); the social class distribution was similar to that of the local community. Although baseline data were obtained for all cases, complete follow-up data are available for only 47 cases.

(b) *Data collection:* At the beginning of the study, baseline data were obtained by both the psychiatrist in charge of the case and a second assessor, who separately and independently interviewed the patient and parents. Thereafter, both doctors carried out independent assessment at monthly intervals over the following 12 weeks, with the inter-rater reliabilities proving satisfactory (Berney *et al*, 1981).

Methods of classification

A. The clinical approach

From a study of the historical data and clinical information available at initial assessment, we classified our 51 cases into two groups—those with significant depression, and a residual group of school refusers without significant depressive features (the RSR group). In this latter group, the features were predominantly those of anxiety.

Depression was defined as a sustained sad, unhappy mood, associated with at least one of the following: an appearance of gloom or tearfulness; a lack of usual energy; a feeling of hopelessness or that life was not worth living. This state differed from the child's usual mood swings in severity, quality, and duration.

Findings: Having dichotomised the cases, we compared the two sub-groups on a variety of features, which fell roughly into three areas—psychiatric phenomenology, psychological data, and environmental data. Of the 51 cases originally studied, 22 (45%) were diagnosed as showing significant depression.

(i) *Psychiatric phenomenology*

Intensive assessment: These ratings were based on information deriving from focussed depth interviewing of the parents and the children. (For the sake of brevity, the definitions and scales used have not been included, but are available on request). Although 21 items were studied, there were significant differences on only ten of these, which are listed in Table I.

A core depression index was calculated by summing the scores of those seven symptoms considered to be most

TABLE I
Psychiatric phenomenology at baseline assessment

	Depressed (%)	RSR (%)	Significance (chi-square)
Intensive assessment			
Dysphoric mood	83	32	**
Tearfulness	48	14	*
Sense of being unloved/unwanted	61	4	**
Sense of emptiness/isolation	43	11	**
Self-denigration	30	4	**
Suicidal preoccupation	39	7	**
Loss of interest	52	7	**
Loss of energy	52	25	*
Initial insomnia	70	39	*
Exaggerated illness behaviour	35	11	*
Core Depressive Score (Best cut)	89	29	**
Total Global Depressive Score (Best cut)	70	14	**
Unreality Index	43	4	**
Obsessive Compulsive Index	52	25	*
Brief assessment (from parents at baseline)			
Dysphoric mood	74	32	**
Tearfulness	61	14	**
Suicidal preoccupation	43	7	**
Initial insomnia	74	46	*
Night awakening	39	14	*
Terminal insomnia	4	0	NS
Irritability	65	54	NS
Paradoxical aggression	43	18	*
Somatic pains	57	36	NS
Nausea	22	14	NS
Anorexia	22	29	NS
Panic attacks	26	29	NS
Inability to shop alone	30	36	NS
Profound reluctance to attend school	70	89	NS
Global score (Best cut)	52	7	**

diagnostic of depression: the severity of dysphoric mood; its persistence; its reactivity; bouts of weeping; diurnal variation; a sense of emptiness and isolation; and a sense of being unloved. We sought the best cut-off point, and an excess of cases in the depressed group scored above it, the differences proving highly significant. Nevertheless, 29% of the non-depressed cases still scored above this threshold and 11% of the depressed scored below; we concluded that the core global score was moderately sensitive (71%) and adequately specific (87%).

A second global index was calculated which included not only those seven items listed above, but another 14 items, suggested in the literature as being important. These included: self-denigration; suicidal pre-occupation; exaggerated illness behaviour; loss of energy; loss of interest; poor concentration; irritability; initial insomnia; restless sleep; early morning wakening; nightmares; secondary nocturnal enuresis; stealing; and aggression within the home. These items were summed and the best cut sought, but 30% of the 'depressed' groups scored below this threshold and 14% of the non-depressed cases scored above it. We concluded that this global score is as sensitive (80%) and specific (89%) as the core global score. However, neither the core nor the total global scores proved to be highly sensitive; we had wrongly anticipated that the addition of the further 14 items would increase the validity.

We were interested in those items in which there were great differences in frequency between the groups. Firstly, those for which there were low rates in the depressive group comprised early morning waking; diurnal variation of mood; enuresis; stealing; poor concentration; nightmares; responsiveness of the depression to environmental circumstances, and durability of depression. We also noted a number of items which, while being moderately frequent in the depressed groups, showed no significant differences between the groups; the most important of these were 'irritability' and 'aggression within the home'.

Are these findings in accord with those features described in the literature as representative of depression and, in particular, do they support the notion of masked or hidden depression? The answer must be an emphatic 'No'. By contrast, a number of the features listed by Weinberg (1973) and his colleagues as being diagnostic show significant differences between our groups; this is not surprising, as

TABLE II

Psychiatric assessment of child – individual items at moderate and marked level of severity (includes one item elicited from parents)

	Baseline		Significance Level	One month		Significance Level
	RSR %	Depressed %		RSR %	Depressed %	
Dysphoric mood	27	80	**	15	19	NS
Tearfulness	11	52	**	15	14	NS
Suicidal preoccupations	8	62	**	0	14	NS
Initial insomnia	35	76	*	23	24	NS
Night waking	15	38	NS	15	38	NS
Terminal insomnia	4	19	NS	0	14	NS
Irritability	42	61	NS	23	24	NS
Somatic symptoms	38	48	NS	27	19	NS
Nausea	19	24	NS	19	24	NS
Appetite problems	31	24	NS	15	5	NS
Unable to go shopping alone	35	24	NS	23	10	NS
Panic attacks	23	47	NS	27	5	NS
Paradoxical aggression	18	43	*	15	14	NS
n =	28	23		26	21	

Legend: * significant at 5% level **significant at 1% level

many of the items would be considered characteristic of adult depression. Surveying our items, in addition to dysphoric mood, we identified other important features, defined as those where there are very significant differences associated with a steep gradient of high rates in the depressed group and low rates in the non-depressed group. These features included a sense of being unloved and a loss of interest in daily activities. Next in importance, characterised by a less steep gradient, were: self-denigration; sense of isolation/emptiness and suicidal preoccupations.

Brief assessment: In standard interviews with parents, information was gathered about the child on 14 items of behaviour, which were rated on four-point scales (from absence of disorder to marked disorder). These items were summed subsequently to give global scores. From a brief psychiatric examination of the child, ratings were obtained for a similar set of 13 items and these brief assessments were repeated monthly for three months.

In separate interviews of parent and child, the rate of occurrence of these symptoms seldom varied by more than 10%. In Table I we have therefore confined data to those deriving from parental interview. At baseline, there were significant differences on only six of the 14 symptoms. When these 14 items are summed and a best cut sought, it is notable that the brief assessment is only moderately sensitive (70%), but adequately specific (86%).

It is evident from the data that many symptoms described in the literature as characteristic of depression often occur in both groups, so that the differences between the groups prove modest and therefore such symptoms are unlikely to be useful in a diagnostic exercise. Those which seemed likely to be most useful are dysphoric mood, tearfulness, and suicidal preoccupations; those which are likely to be moderately useful are sleep disturbances and paradoxical aggression (i.e. which emerges in the 'safe' circumstances of the home, and not at school). The latter is contrary to expectation.

Table II contains data from the brief psychiatric assessment of the child. The pattern of differences at baseline is similar to those based on parental interview about the child, while the fact that just four weeks later, there are no longer any differences between the groups, is very important.

(ii) *Psychological data:* On cognitive assessment, both groups proved to be of average intelligence, but with reading quotients about eight points below the verbal IQ. On the Juvenile Eysenck Personality Inventory, the depressed group had a significantly higher neuroticism score than did the RSR group.

On the Highlands Dependency Questionnaire (Berg, 1974), the depressed group again obtained significantly higher scores on 'real assistance' and 'real affection'. As both of these groups have problems with school attendance, differences in parental dependency are of considerable interest: the higher scores suggest that the depressed children are currently doing very little for themselves, either because thay are apathetic, or because their parents have recognised their incapacity, or both.

That the depressed children were giving or receiving significantly more affection from their mothers is reminiscent of what occurs in a younger age-group. This may be a regressive phenomenon: when a child is substantially depressed, it may seek reassurance through closer physical contact with its parents.

(iii) *Environmental factors:* The groups were compared on about 35 background factors but, with this population size, few significant differences were found. Double the number of mothers in the depressed group (compared with the non-depressed) were known to have been treated for a depressive illness (44% v 21%), but this difference did not achieve statistical significance.

Data reflecting a history of cerebral insult in the first five years of life were based on two pieces of information: known head injury and any history of fits. Historical evidence for

TABLE III

Summary of discriminant function analysis (DFA)

Data Subset	Summary of statistics	Dichotomies				
		Depressed/ not depressed	Acute/ insidious	Adolescent pre-adolescent	Males/ females	Attends school/ does not attend school
A. Background features	a. DFA	P <.03	NS	NS	NS	NS
	b. correctly classified	82%	—	—	—	—
B. Intensive psychiatric assessment	a. DFA	P <.002	P <.02	NS	P <.01	NS
	b. % correctly classified	86%	84%	—	82%	—
C. Brief child assessment at base (Parental interview)	a. DFA	P <.03	NS	NS	NS	NS
	b. % correctly classified	80%	—	—	—	—
D. Brief child assessment at base (Child interview)	a. DFA	P <.003	NS	NS	NS	NS
	b. % correctly classified	88%	—	—	—	—
E. Repeat C. at one month	a. DFA	NS	NS	NS	NS	P <.03
	b. % correctly classified	—	—	—	—	81%
F. Repeat D. at one month	a. DFA	NS	P <.002	NS	NS	NS
	b. % correctly classified	—	87%	—	—	—
G. Repeat C. at two months	a. DFA	P <.001	NS	P <.05	NS	P <.03
	b. % correctly classified	89%	—	85%	—	83%
Ten subsets	No. significant	5 of 10	2 of 10	1 of 10	1 of 10	2 of 10

Note: I. The P value is only given when the discriminant function is significant.
II. Similarily % cases correctly classified only given when DFA is significant
III. Rest of analyses are not significant except on one occasion.

cerebral insult was found significantly more often in the depressed group (39%) than in the non-depressed (4%) (P <0.001).

More of the depressed group had suffered serious illness immediately prior to the present disturbance, but the differences did not prove statistically significant. We also studied life events in the six months immediately preceding the current disorder: 35% of the depressed group had a high score, compared with 15% of the non-depressed group. Again, this difference fell short of statistical significance.

There were no significant differences in the following variables: age, sex of child, family size, social class, nervous problems in siblings or father. A number of suggestive trends emerged, but did not achieve statistical significance. For example, evidence of chronic or recurrent physical or emotional illness in the family occurred more frequently in the depressed group, as did the occurrence of slightly more recent bereavements.

B. Multivariate analyses

1. **Discriminant function analysis** *(Table III):* As described earlier, we sought ways of dichotomising our total sample of children with school refusal. On the basis of a review of the literature and from clinical hunches, we decided to explore the following dichotomies: depressed versus non-depressed; cases with an acute onset versus those with an insidious onset; pre-adolescent versus adolescent patients; boys versus girls; and ability to attend school at three months after the initiation of intensive therapy. The following ten sub-sets of data were used in the discriminant function analysis:

(i) Thirteen background features.
(ii) Fourteen features from intensive psychiatric examinations.
(iii) Fourteen child symptoms derived from brief interview with parents at baseline.
(iv) Thirteen child symptoms derived from brief interview with child at baseline.
(v)—(vii) A repeat of the parent interview as described at (iii) at one month, two months, and three months after the base assessment.
(viii)—(x) A repeat of the child interview as described under (iv) at one month, two months, and three months after the base assessment.

Findings

A summary of the significant findings is presented here. Further details are available on application to the authors.

Depression versus absence of depression (residual school refusal group): This is the only dichotomy where there is a large number of significant discriminants—five of the ten sub-sets discriminate between the groups. It is not only sub-sets of psychiatric data which discriminate between groups, as might be expected, but also a sub-set which is entirely composed of background data. Two of the background features that make an important contribution to this discrimination are evidence of living with only one biological parent (which occurs more frequently in the depressed sub-group) and organic features, which again occur more frequently in the depressed sub-group.

TABLE IV

Principal component analysis, social, family and developmental data

Pole A: Psychosocial adversity
0.63 Civil state of parents
0.60 Living with one biological parent
0.63 Mother, neurotic illness
0.56 Chronic illness in first-degree relatives
0.56 Life Event Index
0.52 Mother, depressive illness
0.41 Father, neurotic illness
0.40 Marital difficulties
0.38 Lower occupational class
0.32 Father, depressive illness
0.32 Organic features in first five years
Pole B: Perinatal adversity
−0.24 Pregnancy of short duration
−0.26 Male
−0.30 Large family size
−0.49 Neonatal breathing problems

Note: First Component Eigen value 3.4 % variance 14.7

Type of Onset—acute versus insidious: Two of the ten data sub-sets significantly discriminate the groups. One of those consists of psychiatric phenomenology, derived from intensive psychiatric assessment. However, it is to be expected that this type of data will differentiate the groups, as the symptomatology studied often indirectly reflects the more chronic nature of the symptoms.

Pre-adolescent versus adolescent patients: On only one occasion is significant discrimination between the sub-groups achieved, and even then it is only at the 5% level. This is an important negative finding, as it suggests that there is little difference between these groups.

Males versus females: Again, there was a significant difference with only one of the data sub-sets—using information derived from intensive psychiatric assessment. Subsequently, we studied an extended sub-set of such features, and found that the two groups could be significantly discriminated with only a 10% mis-classification rate. The features which proved most important in the discrimination were premorbid obsessional traits, a sense of unreality, and wide-spread fears, all of which occur more frequently in girls than in boys.

Eventual ability to attend school: On only two of the data sub-sets was there significant discrimination between the groups, i.e. when using follow-up interview data gathered one and two months respectively after the base assessment.

2. Principal Component Analysis

Principal Component Analyses were undertaken with two sets of data: firstly, social, family and developmental characteristics; secondly, psychiatric phenomenology derived from intensive psychiatric assessments. Where baseline information only was included, we were able to use 50 cases in the analysis, but where follow-up data were studied, mainly because of incomplete information, the number of cases was reduced to 47.

TABLE V

Principal component analysis. Child psychiatric phenomenology based on intensive interview

Features	Description of item or indices	Factor 1 loadings	Factor 2 loadings
Supplementary depressive symptom index	Self-denigration, suicidal pre-occupations, loss of interest, poor concentration, irritability, sleep problems, etc.	0.72	−0.32
Fears index	A composite of multiple fears covering 14 themes, 5 of which related to home, school or shops	0.68	0.22
Recent obsessionality	(Single item)	0.68	−0.15
Unreality index	Sum of items – depersonalisation, derealisation, inappropriate familiarity or strangeness	0.67	−0.13
Obsessional traits	Prior to illness – reoccupation with detail, cleanliness, obstinancy, thrift, etc.	0.62	0.34
Core depressive symptom index	Depressive symptoms, diurnal variation mood, emptiness, feeling unloved	0.61	−0.37
Anxiety	History of recently becoming more anxious and worried	0.57	−0.29
Appetite	Recent loss of appetite	0.54	−0.36
Dependence index	Dependence on close relatives or familiar environment prior to illness	0.41	0.77
Premorbid sociability	Prior to illness	0.35	0.66
Hysterical traits	Prior to illness – attention-seeking, exploitation of physical symptoms, marked possessiveness etc.	0.11	0.28
Type of onset	Insidious onset	−0.02	0.23
Situational phobias	(Single item)	0.1	0.24
Duration	Duration of illness in weeks	0.1	0.45

Note: All the indices are composites of features as listed above

(i) *Social, family and developmental data (Table IV):* Twenty-three variables were used in this analysis. This is a rather large number, in relation to the size of the sample studied, and hence we confine ourselves to discussing only the first component and to those items where there is a component loading greater than ±0.2. The first component proved bipolar and took up 14.7% of the variance. It is not unexpected when dealing with this type of data that there will be no general component underlying all the symptoms. The one pole of this first component can be best described as *perinatal adversity,* particularly in boys; the other pole has high loadings on variables representing adverse social and family experiences and psychiatric illness—especially neurotic and depressive illness in the mother—and a life event index reflecting stressful life events over the previous six months. The latter component was therefore described as one of *psychosocial and psychiatric adversity.*

(ii) *Child psychiatric phenomenology (Table V):* Fourteen variables were included in this analysis. The first component takes up 25.6% and the second 14.8% of the variance. The first is general, and appears to represent general psychiatric disturbance, as reflected by wide-spread depressive and neurotic symptoms, together with premorbid personality features, etc.

The second component is bipolar: at one pole, are features representative of premorbid traits, especially a dependence on close relatives and poor sociability, but it also includes loadings on a premorbid obsessionality index. In addition, there are loadings on features representative of fears and phobias and also of less acute onset and longer duration. These features in many ways resemble the characteristics so commonly reported in children suffering from the school refusal syndrome. At the other pole, there are high loadings on features which have a recent and abrupt onset and are usually considered representative of affective states; of interest are the two features of recent obsessionality and sense of unreality.

One could interpret these findings as suggesting the presence of a general psychiatric disturbance component (first component), together with a differentiating bipolar component. The latter contrasts affective symptomology of more recent onset at the one pole, with the premorbid personality traits often associated with school refusal at the other.

C. Depression — selection of diagnostic criteria

There is a clear need to identify a set of questions which will be helpful in diagnosing depression in childhood. A series of items were obtained during the comprehensive assessment of the 51 cases of 'school phobia'; it is well known that a substantial proportion of school phobics are depressed, and indeed 45% or our 51 cases were rated as being unequivocally depressed. This provided the opportunity to identify items which discriminated well between those phobics who were depressed and those who were not, as well as to identify a subset of such items which maximised this discrimination. There is considerable advantage in comparing the two groups: items that discriminated between the groups would probably be reasonably specific to depression. However, as there was no normal control group, it is possible that the items identified would not necessarily discriminate between those who were well and those mildly depressed. In the following analysis, we employed methods more usually employed in screening populations, using questionnaires to detect psychiatric disturbance.

Item analysis: We were able to compare the two groups on each individual item, scored on four-point scales, and the results indicate that a large number of items discriminate well between the groups. A 'good' item should be scored as positive in as few as possible of the pure school phobics and in as many as possible of those with depression. Thus, the gradient of difference should be relatively steep, but never as steep as would occur if, in addition, there had been a normal control group. In deciding on the gradient, we applied the Likert-type procedure, with weights assigned to each position on our four-point scale (1, 2, 3, 4), or recorded as a bi-modal scale, so that only the clear-cut presence of an item is scored; i.e. there are zero scores in columns 1 and 2, and unity scores on columns 3 and 4 (i.e. 0, 0, 1, 1). The latter is not only a simple procedure, but also eliminates errors of end-users and middle-users (Goldberg, 1972).

The data are presented here in this binary form and the gradient is therefore reflected by the difference in percentage of cases with positive scores for the two groups. We decided to exclude all items with a gradient of less than 25%; where gradients were higher than this, the differences between the two groups were usually statistically significant as well. If we had had a normal control group, we would have excluded items in which there were positive scores in more than, say, 20% of the controls, but as we were comparing two abnormal groups, we relied chiefly on gradients. Further reduction of items was achieved by the following principles: items were likely to be retained when there were highly significant differences between the groups on them, and if they had emerged as important in component analysis and discriminant function analysis.

Subsequently, we examined the cutting scores that give optimum discrimination between the non-depressed (the 'residual phobic') and the 'depressed' groups. Using Likert scoring, the best cuts with various composite scores still gave rather a high mis-classification rate.

We looked at the data to see how we could obtain a better distinction of the symptoms by weighting which can be clinically or statistically determined. A weighting taking those items which proved most discriminating on simple comparison again did not greatly improve the distinction; i.e. the best of these cuts did not provide a better distinction than that obtained by simple addition of the clinical scores. This was because a loading of minor symptomatology on a range of features gave rise in practice, to high scores, whereas maximum scores (indicating considerable severity) on a circumscribed number of features specifically related to depression did not necessarily give rise to high scores. We therefore dichotomised the scores on each item into Absent/Present—where 'Present' reflected a deviant rating on any particular item. The 11 key items which best distinguished the groups and which showed a sharp gradient (see Table VI) were included.

The distinction between groups was then explored—we were seeking the minimum positive criteria which would give the best distinction between the groups. It is to be noted (Table VII) that the best cuts on both the core global and the

TABLE VI

The key items

	Marked	Moderate	Slight	Nil
Dysphoric mood	4	3	2	1
Weeping	4	3	2	1
Sense of emptiness/isolation	4	3	2	1
Sense of being unloved	4	3	2	1
Exaggerated illness behaviour	4	3	2	1
Loss of interest	4	3	2	1
Loss of energy	4	3	2	1
Initial insomnia	4	3	2	1
Nocturnal restlessness	4	3	2	1
Feeling life is not worth living	4	3	2	1
Feeling of *déjà/jamais vu*	4	3	2	1

TABLE VII

Validity

	Mis-classification	Sensitivity	Specificity
Core global score	22%	71%	87%
Total global score	22%	80%	89%
11 items – cut-off 5+	16%	89%	81%
11 items – cut-off 6+	14%	100%	80%

total scores gave rise to rather high mis-classification rates (using the formula $\frac{FP + FN}{Total}$ (FP = False positive FN = False negative)) despite reasonable levels of sensitivity and specificity on the total global score. However, where a deviant rating was found on any five of the 11 key items, similar levels of sensitivity and specificity were achieved, and the mis-classification rate was marginally lower. The best discrimination was achieved when any six of the 11 key items had deviant ratings. The use of a larger number of items than those 11 did not improve the ratings.

Four of the cases clinically diagnosed as depressed did not have high scores on the item of dysphoric mood. Nevertheless, to be sure that one is dealing with depression, one of the five or six diagnostic items should be that of dysphoric mood.

Comment

Inevitably, certain cases will fall into the indeterminate area between depression being clearly present or absent, and such diagnostic problems are likely to give rise to some degree of mis-classification. Nevertheless, we believe that our criteria provide a useful and valid means of diagnosing depression in older children and adolescents. Thus the clinician could usefully employ the following 11 items, and rate them on four-point scales, the object being to ensure that he bears in mind the distinction between substantial and minor symptomatology. In the past, a problem has been the tendency to give equal importance to a large number of symptoms, some of which may be vague and of uncertain diagnostic significance, rather than giving appropriate attention to a small number of symptoms which are present to a severe degree.

For example, a child scoring two on each of our 11 key items will have total score of 22, while a child scoring three on the first six items and one on the rest will have a total of 23. It is evident that the scoring system and cut-offs suggested above are more likely to give rise to a valid discrimination of depression.

We were particularly surprised to discover that the item of sense of inappropriate familiarity/strangeness (*déjà* or *jamais-vu*) provided an important discriminator between the two groups. Of the 11 key items, the only other one which would be unusual in adult depression is that of exaggerated illness behaviour.

Another way of looking at the relationship of each of the 11 items with the clinical ratings of depression is by using appropriate correlative analyses. All these items correlated positively with clinical severity, but the correlations prove significant in only eight cases. The highest set of correlations related to sense of being unloved, dysphoric mood, sense of inappropriate familiarity/strangeness, loss of interest and life not worth living (ranging from 0.43 to 0.62); the next highest set included sense of emptiness, weeping and initial insomnia (ranging from 0.30 to 0.41), and the lowest, loss of energy, restlessness and attention-seeking behaviour (ranging from 0.17 to 0.28). Yet another approach is to see how well each item correlates with the summed scores of the eleven items and for this purpose, we dichotomised each item according to presence or absence of abnormality. All the coefficients proved highly significant, ranging from 0.39 to 0.69, so that

each of the items makes an important contribution to a summed depressive score.

We also looked at the kinds of agreement obtained from data arising from brief interviews of the parents and the child respectively; these were not independent interviews, and there was the possibility of contamination from one to the other. We noted that the agreement on these items (rated on four-point scales) was moderate, ranging from 0.48 (dysphoric mood) right through to 0.72 (for weeping). For all the items which we suggest be included in a depressive scale, the correlations were significant at the level $P < 0.001$. Finally, we correlated each of the items with its own global score and found the following correlations: 0.78 for sad mood; 0.61 for weeping; 0.62 for life not worth living; 0.65 for initial insomnia; and 0.57 for night waking.

The above were based on parental interview; when we repeated this for child interview, the correlations were similar: 0.61 for sad mood; 0.68 for weeping; 0.57 for initial insomnia; 0.67 for life not worth living; and 0.43 for night waking.

We also inter-correlated the total scores obtained from parent and child interviews respectively, and attained a correlation of 0.76.

The above leads us to believe that there is reasonable agreement for the data obtained respectively from interviewing children and their parents. However, though there is substantial agreement with some items, with other items there is less. Therefore, after an initial screen to identify the likely candidates for a depressive disorder, a subsequent diagnosis needs to be dependent on information both from parents and from children, in order to obtain reasonably accurate ratings of severity on items which will contribute towards the diagnosis of depression.

We conclude that the use of a binary scale reduces the misclassification rate, despite using only a small number of diagnostic items; the reliability of the items has been discussed previously (Berney *et al*, 1981). However, the subject of validity of the eleven-item scale is more complex. Content validity is ensured by the selection from a wide range of items which had been suggested as relevant both to adult and child-type depression, and its factorial validity has been demonstrated. As the diagnosis of depression was checked by independent raters, this does in a sense also validate the diagnosis. Some further indirect evidence of validity can be gleaned from our own data determining the number of depressed patients that the eleven-item scale mis-diagnoses (its sensitivity) and the number of non-depressed patients that it mis-diagnoses (its specificity). It can therefore be concluded that the eleven-item scale, each item of which reflects severity, is highly sensitive and adequately specific. Finally, we need to pose questions about the relatively high misclassification rate, but we are not greatly deterred by this, as it is inevitable that there will be some overlap at the margin between severe school phobia with depression and depression in school phobia.

Discussion

Multifactorial origins: The study of environmental factors provides major support for a theory of multifactorial origin of depression in childhood. Despite the small numbers in each of the groups, there are both some important differences, and some factors in which significance is almost achieved. Top of this list is *parental history of depressive illness,* and second is the incidence of *recent life events,* which occur in one-third of the depressed group. A completely unexpected finding is the frequency of *organic factors in the early years of life* in the depressed group, which suggests that cerebral insult makes the child more vulnerable to depressive breakdown. Indirect support for the importance of organic factors comes with the clear emergence of a perinatal adversity pole on factor analysis of social, family, and developmental features. Rigorously conducted studies of family histories of depressed children reveal a significant *prevalence of affective disorders in biological relatives* (Carlson & Cantwell, 1979, 1980; Ling *et al*, 1970). This suggests a link between child and adult depressive disorders—a view supported by our own preliminary data, which suggest that there is likely to be a higher genetic loading for depression where depression is a feature of neurotic disorder in children.

Sub-groups in 'school phobia': Simple and multivariate analyses of psychiatric phenomenology found in our group of school phobics suggest that there are at least two major sub-groups—one of 'depression' and the other of a "residual school refusal syndrome", the latter having features characteristic of the school phobic syndrome described in the literature. The evidence for this classification is relatively strong, as in the first instance it derives from the traditional clinical approach, and in the second, it is supported by findings from both principal and discriminant function analyses.

The simple statistical comparisons between these two sub-groups provide some findings which are within expectation and some which are quite new. By definition, it was to be expected that the depressed group would show an excess of suicidal thoughts, loss of energy, and dysphoric mood. However, it is of interest that the depressed group also shows a higher rate of initial insomnia, though not of early morning wakening, than the non-depressed group. Indeed, the depressed group shows excess scores on a wide pattern of what can be considered depressive and anxiety features. This gives rise to speculation that instead of two separate groups, we may be dealing with a spectrum of severity, with phobic anxiety states giving rise to lower scores and depression to higher scores. The significantly higher neuroticism score of the depressed group on the JEPI tends to support this hypothesis.

A finding of considerable interest is the frequency of 'feeling unloved' in the depressed group, which may constitute the equivalent of depressive delusions of adulthood.

Classification: In a discriminant function analysis, Maudsley Hospital child psychiatry data (Pearce, 1977) have shown that depressive symptoms cluster together as a syndrome of depression. Achenbach (1966, 1980) applied multivariate (factor analytical) techniques, using extensive behaviour check-lists, with child psychiatric populations. Like other investigators, he demonstrated a robust depression factor in pre-puberty, but surprisingly, this proved not so clear-cut in adolescence. We have demonstrated here the presence of an adult-type depression in a higher percentage of school phobics, and there is now little room for doubt about the presence of a depressive disorder in childhood, similar to the type occurring in adults. Its presence in school phobics has been clearly shown by our research, both on clinical and multivariate grounds. This should be no surprise, since Frommer suggested in 1968 that phobic depression was one of the sub-categories of depression in childhood, but this was based on clinical impression, rather than on statistical analysis of features deriving from systematic psychiatric assessment.

Criteria for diagnosis: Weinberg *et al* (1973) have offered a list of primary and secondary symptoms, based on the criteria for diagnosis of depression in adulthood (Feighner *et al,* 1972). In order to be diagnosed as depressed, children need to have the primary symptom of dysphoric mood, to express self-deprecatory ideas, and also to have at least two of the following eight symptoms: aggressive behaviour, sleep disturbance, falling-off of school performance, social withdrawal, somatic complaints, loss of usual energy, appetite changes and change in attitude to school. Subsequently, Puig-Antich (1980) advanced the hypothesis that there was a depressive disorder in childhood, similar to adult depressive disorder, which could be diagnosed using unmodified Research Diagnostic Criteria (Spitzer *et al,* 1978). For diagnosis, the child had to have depressed mood or pervasive anhedonia, and also three of the following: guilt, anhedonia, energy changes, appetite changes, sleep problems, poor concentration, psychomotor retardation, and suicidal thoughts. Although we started with a different and simpler set of operational criteria from Weinberg for diagnosing depression, it is notable that there were differences between our groups on both primary symptoms and three of the eight secondary symptoms: loss of usual energy; sleep disturbance of the initial variety and nocturnal sleeplessness, but not terminal insomnia; and paradoxical aggression (i.e. aggressive outbursts within the home but not outside). Finally, there was an excess of somatic complaints in the depressed group, but the differences did not reach statistical significance.

With the Research Diagnostic Criteria, there were significant differences on both the primary ones and the following secondary criteria: loss of energy, suicidal ideation, sleep disorders as described above, and some presenting symptoms which appear to reflect anhedonia.

The picture which emerges is closer to the concept of the Research Diagnostic Criteria than to that of Weinberg, the only unusual characteristics being an excess of obsessive compulsive symptoms and a sense of unfamiliarity. Three interesting features were a sense of emptiness or isolation, a sense of being unloved or unwanted, and exaggerated illness behaviour. Some of these latter features have been proposed as being important by other authors (McConville *et al,* 1973; Glaser, 1968). However, even in the extended list of symptoms studied, there was no evidence of an excess of the so-called masked symptoms of depression (Glaser, 1968; Cytryn & McKnew, 1974).

Our research has also given rise to a set of key diagnostic criteria. They differ from the criteria advanced by the North American Research Groups (Weinberg, 1973; Puig-Antich, 1980), but there are more similarities than differences in relation to the Research Diagnostic Criteria. Further work needs to be undertaken to establish whether the formula we have developed is specific to separating school phobics into those with and without depression, or alternatively, whether it will have a wider application in the diagnosis of pre-pubertal and pubertal depression.

Acknowledgements

We are indebted to the late Dr R. F. Garside, who helped to plan this research, and also to Dr L. Scarth, Dr J. Jeans, and Mrs. B. Kay.

References

ACHENBACH, T. M. (1966) The classification of children's psychiatric symptoms: a factor analytic study. *Psychological Monographs,* **80.** (No. 615).

—— (1980) What is child psychiatry epidemiology the epidemiology of? In *Studies of Children* (ed. F. Earls) New York: Prodist.

BERG, I. (1974) A self-administered dependency questionnaire (SADQ) for use with the mothers of school children. *British Journal of Psychiatry,* **124,** 1–9.

BERNEY, T. B., KOLVIN, I., BHATE, S. R., GARSIDE, R. F., JEANS, J., KAY, B. & SCARTH, L. (1981) School Phobia: a therapeutic trial with clomipramine and short-term outcome. *British Journal of Psychiatry,* **138,** 110–18.

CARLSON, G. A. & CANTWELL, D. A. (1979) A survey of depressive symptoms in a child and adult psychiatric population. *Journal of the American Academy of Child Psychiatry,* **18,** 587–99.

—— & —— (1980) Unmasking masked depression in children and adolescents. *American Journal of Psychiatry,* **137,** 445–9.

CATTELL, R. B. (1965) The role of factor analysis in research. *Biometrics,* **21,** 405–35.

CYTRYN, L. & MCKNEW, D. H. (1974) Factors influencing the changing clinical expression of the depressive process in children. *American Journal of Psychiatry,* **131,** 879–81.

EYSENCK, S. B. G. (1965) *Junior Eysenck Personality Inventory Manual.* London: University of London Press.

FEIGHNER, J. P., ROBINS, E., GUZE, S. B., WOODRUFF, R. A., WINOKUR, G. & MUNOZ, R. (1972) Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, **26**, 57–73.

FROMMER, E. A. (1968) Depressive illness in childhood. *British Journal of Psychiatry*, special publication no 2, 117–36.

GARSIDE, R. F. & ROTH, M. (1978) Multivariate statistical methods and problems of classification in psychiatry. *British Journal of Psychiatry*, **133**, 53–67.

GLASER, K. (1968) Masked depression in children and adolescents. *Annual Progress in Child Psychiatry and Child Development*, **1**, 345–55.

GOLDBERG, D. P. (1972) *The Detection of Psychiatric Illness by Questionnaire*. Maudsley Monograph No. 21 London: Oxford University Press.

HEWITT, L. E. & JENKINS, R. L. (1946) Fundamental patterns of maladjustment: the dynamics of their origin. In *Education, Health and Behaviour*, Michigan Child Guidance Institute.

KENDELL, R. G. (1975) *The Role of Diagnosis in Psychiatry*. Oxford: Blackwell Scientific Publications.

KOLVIN, I. & NICOL, A. R. (1979) Child Psychiatry. In *Recent Advances in Clinical Psychiatry*, 3rd ed. (ed. K. Granville-Grossman) London: Churchill Livingstone.

LING, W., OFTEDAL, G. & WEINBERG, W. A. (1970) Depressive illness in childhood presenting as severe headache. *American Journal of Disturbed Children*, **120**, 122–4.

MCCONVILLE, B. J., BOAG, L. C. & PUROHIT, A. P. (1973) Three types of childhood depression. *Canadian Psychiatric Association Journal*, **18**, 133–8.

PEARCE, J. (1977) Depressive disorder in childhood. *Journal of Child Psychology and Psychiatry*, **18**, 79–82.

PUIG-ANTICH, J. (1980) Affective Disorders in Childhood. A Review and Perspective. *Psychiatric Clinics of North America*, **3**, No. 3, 1980.

SPITZER, R. L., ENDICOTT, J. & ROBINS, E. (1978) Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry*, **35**, 773–82.

WEINBERG, W. A., RUTMAN, J., SULLIVAN, L., PENICK, E. C. & DIETZ, S. G. (1973) Depression in children referred to an educational diagnostic centre. *Journal of Paediatrics*, **83**, 1065–72.

*Israel, Kolvin, BA, MD, FRCPsych, *Professor of Child Psychiatry, University of Newcastle, Nuffield Psychology and Psychiatry Unit for Children and Young People, Fleming Memorial Hospital, Great North Road, Newcastle upon Tyne, NE2 3AX*

Thomas P. Berney, MB, ChB, MRCPsych, *Consultant Psychiatrist, Prudhoe Hospital Northumberland and Nuffield Psychology and Psychiatry Unit, Newcastle upon Tyne*

Surya R. Bhate, MB, BS, MRCPsych, *Consultant Psychiatrist, Queen Elizabeth Hospital, Gateshead and Nuffield Psychology and Psychiatry Unit, Newcastle upon Tyne*

*Correspondence

(Received 19 December, 1983)