European Child and Adolescent Psychiatry, Volume 1, Issue 2, April 1992, pp. 119–129 © Hogrefe & Huber Publishers

Development and Evaluation of a Diagnostic Algorithm for Depression in Childhood

I. Kolvin*, T. P. Berney**, L. M. Barrett*** and S. Bhate****

This paper describes the development and validation of a brief questionnaire intended for the rapid assessment of depression in childhood. It is primarily a clinical interview but there is an associated algorithm for the diagnosis of a syndrome of depression. It can equally be used for the clinical diagnosis of a depressive disorder. The current battery includes features representative both of "depressive cognitions" and "endogenous depression". However, it is less effective in the identification of a syndrome of endogenous depression than depressive cognitions.

Introduction

Much of the depression research in adulthood has taken into account the distinction between the symptoms and the syndrome of depression, the latter consisting of a number of symptoms which regularly occur together in a meaningful way. This has been complemented by the specification of criteria for depressive disorders by Feighner et al. (1972) and Spitzer et al. (1978) both of which have helped to define the sub-types of a major and minor depressive disorders as outlined in the DSM III.

Over the last decade, clinicians in the child and adolescent psychiatric field have examined the distinctions between *depressive* symptoms, syndromes and disorders. The term may be used to refer to a symptom of affect or mood, a syndrome consisting of a cluster of depressive symptoms, or a disorder reflecting a category of individuals with a depressive syndrome who are in one or other sense handicapped by the condition (Nurcombe et al., 1989). Screening instruments developed include those by Kovacs (1981). Syndromes have been delineated by the work of Weinberg (1973) and of Kolvin et al. (1984), Berney et al. (1981), and disorders have been defined by Kolvin et al. (1991) and Ryan et al. (1987).

The particular aim of this research was to develop further the KBB (Kolvin, I., Berney, T. P., Bhate, S.) questionnaire which was originally devised to discriminate between those school phobics who were depressed and those not depressed (Kolvin et al., 1984). More recent research into childhood depression provided an opportunity to revise this inventory geared to screen child populations for depressive syndromes and also to allow discrimination between depressed and non-depressed children attending a consulting child psychiatric practice (Kolvin et al., 1991). Revision was considered necessary because recent research (Ryan et al., 1987; Kolvin et al., 1991) has demonstrated that childhood depression is not a homogeneous phenomenon but includes at least two sub-types – "endogenous" and "negative cognition".

As part of this exercise, a number of specific themes were explored:

I To check on the validity of the KBB inventory using a fresh clinical sample.
II Does substitution of criteria modify the validity of the inventory?

Acknowledgements. This work was supported by the Health Promotion Trust. We are also grateful for the support from our other research colleagues on the depression research team – Dr. S. Tyrer, Dr. T. Fundudis and Dr. O. O. Famiyuwa. One of us is indebted to the late Dr. J. Puig-Antich for his generosity with ideas and making available the early and later versions of the Kiddie-SADS. And also Mrs. B. Slagel and Miss C. McNeil for secretarial help. Correspondence to Professor I. Kolvin.

* Royal Free Hospital, School of Medicine and The Tavistock Clinic, 120 Belsize Lane, London NW3 5BA, England.
** Prudhoe Hospital, Northumberland and Fleming Nuffield Unit, Newcastle upon Tyne, England.
*** Queen Elizabeth Hospital, Gateshead, Tyne and Wear, England.
**** Young People's Unit, Newcastle General Hospital and Fleming Nuffield Unit, Newcastle upon Tyne, England.

III Should dysphoric mood be a mandatory diagnostic criterion or merely one of the designated diagnostic criteria?
IV What is the best cut when applying a symptomatic algorithm?
V What is the relationship between an inventory diagnosis of depression and the sub-groups of depression identified by cluster analysis?
VI To delineate the pattern of symptoms occurring in different syndromes of depression and hopefully to specify criteria for the diagnosis of these syndromes.

Method

A. Sample Selection and Exclusions

As part of the Newcastle Depression Project, 316 children referred, between 9 and 16 years of age, were screened for depression using the shortened version of the Child Depression Inventory (CDI), (Carlson & Cantwell, 1980b; Kovacs, 1981) at their first appointment. The subjects of the survey had attended a university based Child Psychiatry Unit for the first time, either as an out-patient, day or in-patient. These constitute the usual type of cases attending a consulting clinical practice. Exclusions and losses are described elsewhere (Kolvin et al., 1991) and these reduced the sample to 275 cases.

A stratification procedure, based on high and low scorers on the CDI (Kolvin et al., 1991), was used to choose children for intensive interview, with the selection being made on the basis of probability of childhood depression in the population concerned. A score of less than nine on the short form of the CDI constituted a low score, a score of nine or more a high score; further details about method are available elsewhere (Fundudis et al., 1991). This reduced numbers to 93 for some analyses and 95 for others.

The mean age of the selected sample was 12.5 years; 27% were 9–10 years old; 26% were 11–12; 32% were 13–14; 14% were 15–16; 48% were male and 52% were female.

The current research is confined to comparing sub-groups of children in consulting clinical settings. As there are no general population controls the implications are that thresholds that emerge with patients seen in specialist psychiatric practice may be inappropriate for nonspecialist settings (Dohrenwend et al., 1978, 1980; Williams, 1986). Thus, scales that may be reliable in consulting practice samples may be much less so in general population samples.

Measures used

i) The Standard Psychiatric Interview

The Standard Psychiatric Interview (SPI – Goldberg et al., 1970) is a semi-structured schedule designed to study psychiatric disorder in adults in a community setting. It has a number of precise probes as well as clear-cut definitions of symptoms and can be used to give ratings on a range of clinical disorders.

For the purpose of this research with pre-adolescents and adolescents, an unstructured introductory interview was designed to precede the SPI. In addition, the interviewers (all senior research child psychiatrists) were allowed to use phraseology and concepts appropriate to the child's cognitive level and stage of psychological development, thus accommodating the different abilities of children to give accounts about themselves. Preliminary piloting indicated that, provided that the interviews were conducted with flexibility and sensitivity, little in the way of problems emerged when using the SPI with prepubertal children and adolescents. The instrument rates symptoms according to clinical judgement on a five-point severity scale. We utilized the concepts and definitions of disorders more usually geared to adults (as outlined and defined in the manual) which allows both clinical diagnosis and rating of severity of depressive disorder.

The reliability of the data was checked by employing multiple observers (Kolvin et al., 1991) on a small sample of cases. There were no discrepancies at all with regard to diagnosis, which is probably a reflection of the similar training, concepts and symptom definition of the three raters. This was checked by analysis of variance which revealed highly significant differences between those rated but not between raters. Intra-class correlation coefficients proved to be high for depressive disorder and anxiety disorder (0.96 and 0.92 respectively) (Kolvin et al., 1991). It was concluded that the Standard Psychiatric Interview schedule, appropriately modified and administered by experienced child psychiatrists, had satisfactory reliability in the assessment of the degree of overall

depression, overall anxiety and associated symptoms in the prepubertal period and adolescence.

ii) Kiddie-SADS

A commonly used and validated schedule in assessing depression in children is the Kiddie-SADS. It is essentially a modification of the Schedule for Affective Disorders and Schizophrenia (Spitzer et al., 1978) that has been employed with children between 6 and 17 years of age by Puig-Antich and colleagues (Chambers et al., 1985). It has been shown to be a reliable instrument for measuring symptoms of depression and conduct disorder, although the ratings of anxiety disorder have not been so consistent (Chambers et al., 1985). The recommended procedure is for the parents to be interviewed first and then the child; the interview of the child is guided by information provided by the parent.

Although the Kiddie-SADS was felt to be a very suitable instrument for assessing childhood depression, an important modification of the procedure was effected in the current study. Independent interviewing of parent and child by different interviewers was carried out which allowed independent ratings of the severity of the symptoms based on child interview and parent interview (Barrett et al., 1991).

Other modifications included the addition of a small number of symptoms using a format of probes and scales similar to that in the original Kiddie-SADS. These included paradoxical aggression, i. e. aggression which manifests itself in the home only (Berney et al., 1981), boredom, a sense of feeling unloved, a feeling of emptiness, self dislike, exaggerated illness behaviour and *deja/jamais vu*. Further, in our study a four-point severity scale was used on the Kiddie-SADS in place of the seven-point scale in the original.

iii) The Original KBB Questionnaire

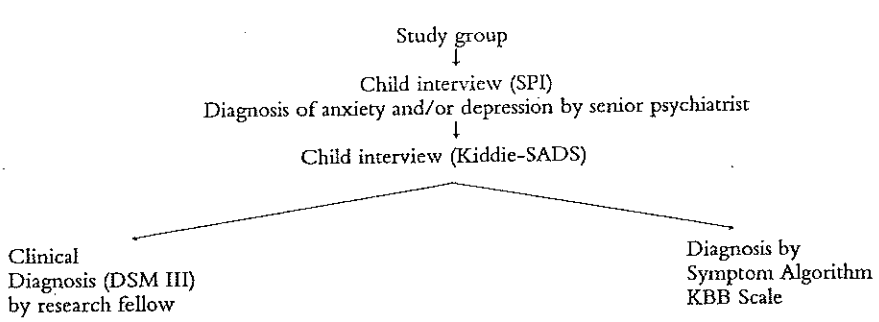
This is a set of eleven key diagnostic criteria (Kolvin et al., 1984, 1991) – dysphoric mood, weeping, life not worth living, sense of emptiness, unloved, loss of interest, loss of energy, initial insomnia, nocturnal restlessness, illness behaviour, and *deja ju/jamais vu*. Symptoms were rated on a four-point ordinal scale (1 2 3 4 – reflecting nil, some, moderate and marked) which was then re-coded to give a dichotomous (binary) scale indicating the presence or absence of a symptom (0 0 1 1). To achieve a diagnosis of depression and to allow evidence of severity of overall depression, a symptom algorithm was applied to the binary scale which delineates a cut-off. The presence of more than 5 symptoms (of which one must be dysphoric mood) was used to indicate the diagnosis of depression (Kolvin et al., 1984)

Sequence of Assessment and Diagnosis

It was intended that the child be interviewed first by a senior psychiatrist using the SPI followed by the Kiddie-SADS administered by a research fellow with the schedules being administered in sequence on the same day or within 48 hours of each other. However, lack of staff prevented the wide implementation of this intention, so that in practice in a proportion of cases the senior psychiatrist first administered the SPI and then the Kiddie-SADS.

On the basis of the above assessments a multiplicity of diagnoses of depression were achieved. First, on the SPI (Goldberg et al., 1970) categorization and ratings of severity of depressive disorder were made by the senior psychiatrists and these were used for calculating rates of depression in the total sample and also to check agreement with other schemes. This was considered to be the most

Flowchart of Assessment and Diagnosis



accurate of our methods as it was based on both direct assessment and clinical judgement, which allowed finer discrimination on a five point scale and in addition allowed categorisation of disorder rather than merely of a syndrome.

Second, the original KBB Questionnaire items were subject to algorithms (Kolvin et al., 1984) and this allowed a symptomatic diagnosis to be made.

Third, following Puig-Antich (1980) (Chambers et al., 1985) major depressive episodes were defined as periods of dysphoric mood or of pervasive loss of interest or pleasure and were categorised as major if both relatively discrete and were associated with the defined number of symptoms of the depressive syndrome. In addition, the annotated child Kiddie-SADS protocols were later reviewed blindly by one of the research fellows, who made a diagnosis of depressions, based on DSM III criteria, of major depression (single episode or recurrent episode) and, in addition, made ratings of the severity of depressive disorder. The categories and ratings that emerged were used to check agreement between schemes.

The Patterns of Symptoms in Different Syndromes and Disorders of Depression

a) Clinical Diagnosis

The three diagnostic schemes mentioned above allowed us to sub-classify our cases into depressed and not depressed (the KBB schema); or major depression – its presence or absence (DSM III diagnosis); finally, the SPI allowed us to sub-divide our clinic population according to four operational categories:

I Non-depressed group
II Anxiety states
III Depression without anxiety
IV Depression with anxiety

b) Multivariate "Classification"

We used an expanded symptom list, which overlapped with that used by Ryan et al. (1987), for the purpose of multivariate analysis (Kolvin et al., 1991). Principal component analysis (PCA) with varimax rotation was undertaken and components were retained following two criteria – an eigenvalue greater than one; and having at least three variables that loaded significantly (Kolvin et al., 1991). Like Ryan, we identified four major factors or principal components and as two of these are so important for the current paper we list them in Table 1. These two components accounted for 45 % of the variance.

Clusters were identified using Ward's (1963) method as outlined by Kolvin et al. (1991). Equivalents of the above two components re-emerged in our cluster analysis (hereafter called cluster 'a') and they consisted of:

I Negative depressive cognitions
II "Endogenous" depression
III An amalgam of conduct-neurotic anxiety features.

Another cluster analysis (hereafter called cluster 'b') based on data from the SPI but which also included the item "school non-attendance" gave a rather different picture which is best represented in graph form (Figure 1).

I Moderate depression without symptoms
II Severe depression including masquerade syndrome
III A residual maladjustment group (Kolvin et al., 1989) without depression

Hence the structure of the emergent clusters varied according to the content of the symptoms included in cluster analysis.

The DSM III and the KBB diagnoses were validated against the SPI, using Kappa (Cohen, 1960). Although the Kappa coefficients proved moderate, intra-class correlations of severity of DSM III diagnoses and KBB diagnoses proved satisfactory (Kolvin et al., 1991). Furthermore, a comparison of agreement between the KBB symptomatic diagnoses and DSM III diagnoses also showed acceptable intra-class correlation coefficients (Kolvin et al., 1991).

Exploring the Eleven Item KBB Questionnaire

This is an eleven item inventory and the original version included the item of *'deja vu'* (Kolvin et al., 1984). It had been a surprise that an item of inappropriate familiarity/strangeness (deja or jamais-vu) was found to be an important discriminator for depression (Kolvin et al., 1984). Gelder (1983) has pointed out, that *deja-vu* experiences may accompany depersonalisation which is encountered in a number of psychiatric conditions, including depression (and at times even in healthy people when

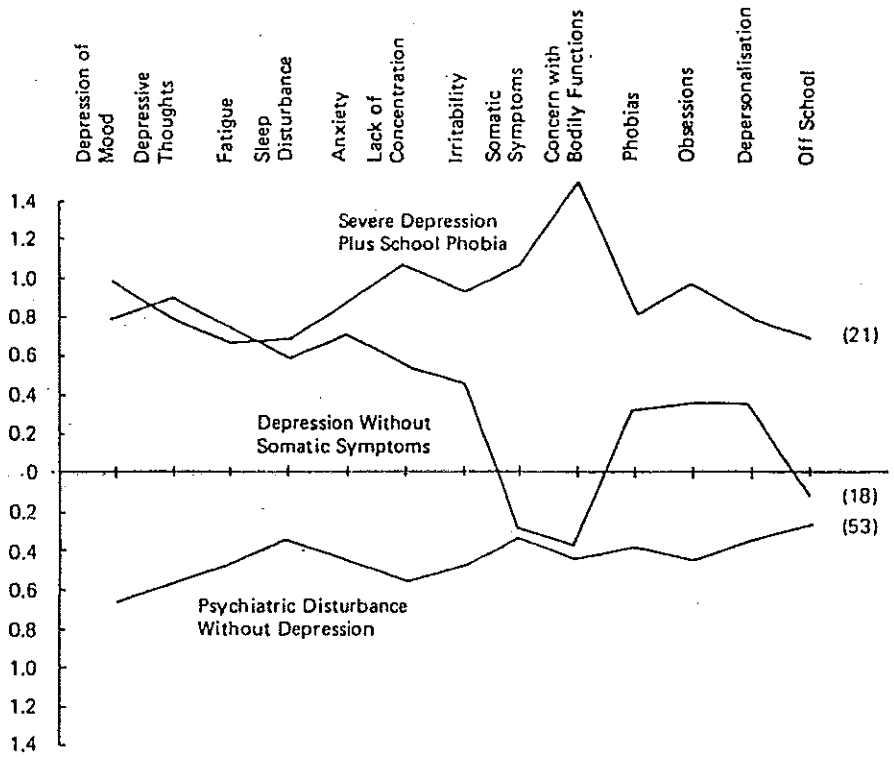


Figure 1: Three cluster solution (Kolvin et al., 1989)

tired). Hence, it seemed appropriate to explore the utility of depersonalisation as a diagnostic criterion as an alternative to *deja vu*. This was looked at in a number of ways. Firstly, the eleven items including *deja vu* were summed and the summed score was correlated with a clinical rating of the severity of depression on the SPI, and also with the overall rating of severity of depression based on DSM III criteria from the Kiddie-SADS interview. The correlations were 0.72 and 0.75 ($n=93$; $p<.001$) respectively. When depersonalisation was substituted for *deja vu*, the correlations were 0.75 and 0.77 ($n=93$; $p<.001$) respectively.

Second, the KBB inventory was viewed as a screen measure and its measurement characteristics were explored using the concepts of sensitivity, specificity and misclassification which derive from epidemiology. Sensitivity is the ability of a measure to select individuals who actually have a disorder. Specificity is the ability to select individuals who are truly free of the disorder. If the data used to check sensitivity and specificity were not from the whole sample but from two stratified random samples, this could give rise to some distortion; hence whole sample rates were estimated by recalculating back to the original sample of 275. The criterion used to validate the measure was the diagnosis of depression based on the SPI scheme. The patterns of sensitivity, specificity and also the overall mis-classification rate

Table 1. Principal component analysis based on Kiddie-SADS Data. Depression Components with Symptom Loadings

	"Endogenous"	Negative Cognitions
Anhedonia	66	–
Dysphoric Mood	58	61
Loss of Appetite	80	–
Lack of Energy	63	–
Somatic Complaints	47	–
Irritability	47	–
Self Denigration	–	72
Hopelessness	–	63
Guilt	–	48
Sense of Failure	–	83
Suicidal Ideation	–	69
Brooding	–	58
Withdrawal	52	48
Slowing Thoughts	–	53
Anger	–	55
Eigen Value	1.6	8.2
Proportion Variance	7.4	37.4

Items with lower loadings have been excluded from the table. This is based on data reported elsewhere (Kolvin et al., 1991)

proved very similar when de-personalisation was substituted for *deja vu*, as one of the eleven criteria. This proved true, irrespective of whether the validating criterion was severe depression or severe plus moderate depression.

The next question was whether the KBB diagnostic schema should include dysphoric mood. This is an important theoretical issue as the concept of masked depression suggests that depression can be diagnosed without the presence of dysphoria (Carlson & Cantwell, 1980a). Dysphoria was both included and excluded as one of the mandatory diagnostic criteria. Again, it was noted that the patterns of sensitivity and specificity and misclassification hardly differ irrespective of whether the symptom algorithm includes dysphoria as mandatory or whether dysphoria is merely one of the diagnostic items.

Further Development of the KBB Scale

The current research provided an opportunity for optimising the diagnostic utility of the KBB scale. The rationale was the identification of a set of symptom items which discriminated well between depressed and non-depressed groups within a clinical sample as it was likely that such items would prove to be reasonably *specific* to depression. Further, despite the absence of a normal control group, on theoretical grounds it was to be expected that items discriminating well between the above clinical groups would prove to be even more powerful differentiators between samples of depressed and non-depressed children in the community.

Item Analysis

Following the model of our previous work (Kolvin et al., 1984) we sought operational criteria for discriminating between groups of depressed and non-depressed children. A decision was made that a "good item" should be scored as "positive" in as few as possible of the non-depressed group and in as many as possible of the depressed group. Hence the gradient of percentage differences between the groups should be relatively steep, but never as steep as would occur if, in addition, there had been a normal control group (Kolvin et al, 1984). As in our previous research, we re-scored the items on binary scales so that only the presence of clear cut deviance is scored as positive. This simple procedure eliminates potential errors of end-users and middle-users (Goldberg, 1970).

Table 2. Items analysis according to categorisation the KBB items

	SPI Depressed %	Not %	D	F
I INCLUSIONS:				
Dysphoric Mood	90	15	75	5.0
Unloved	60	9	51	5.7
Weeping	62	9	53	5.9
Loss Energy	70	21	49	2.3
Loss Interest	55	17	38	2.2
Emptiness/Isolation	43	13	30	2.3
Depersonalisation	49	9	40	4.4
Suicidal Ideation	37	15	22	1.5
II DELETIONS				
Nocturnal Restlessness	49	34	15	0.4
Initial Insomnia	66	30	36	1.2
Deja/Jamais Vu	40	23	17	0.7
Exaggerated Illness Behaviour	23	16	7	0.4
III ADDITIONS				
Anhedonia	58	15	43	2.9
Depressive thoughts	49	4	45	11.1
Lack of concentration	38	0	38	38.0
Hopelessness	47	13	34	2.6
Loss of Appetite	40	11	29	2.6

Legend D =difference in rates between depressed and not depressed; $F = \frac{D}{\%\text{ not depressed}}$

The data in Table 2 are presented in this binary form. A "good item" formula was developed which utilized data in the binary form and is represented by a ratio 'F'. The gradient consists of the differences (D) in percentage of subjects with positive scores for the two groups; and the ratio 'F' represented by such differences between the groups divided by the base rate in the non-depressed group. Thus the higher the 'F' the better the gradient: For example, if 90 % of the subjects in the depressed group score positive but only 15 % in the non-depressed group, the ratio 'F' will be 5.

Of the 12 items studied (including depersonalisation), only 8 showed good discrimination: these were retained and the other four were discarded. The wider symptom list already referred to (Kolvin et al., 1991) was studied in the same way. By using the above criteria an additional five items were selected which proved good discriminators and these were added to the KBB scale. This allowed a better selection of items based on the item list used by Ryan et al. (1987). It is to be noted that the sense of "feeling unloved" which was one of the eleven original items and which was thought to constitute the equivalent of a depressive delusion

Table 3. Measurement characteristics of the revised KBB scale

Thereshold	At least moderate depression			Severe depression		
	Sensitivity	Specificity	Misclassification	Sensitivity	Specificity	Misclassification
Cut-off 4/5	81	85	17	93	80	16
Cut-off 5/6	75	96	13	85	89	12

Note – the above are estimates derived by recalculating back to the base poputation.

of adulthood (Kolvin et al., 1984), continue to be a good discriminator in this fresh sample.

This expanded 13 item scale was then examined further in an attempt to identify the best cut in relation to the external criterion of ratings of depression on the SPI. For these purposes the three related indices of sensitivity, specificity and misclassification were again used. Again, in order to obtain a better representation of the efficiency of the KBB extended scale, the rates on the above indices were not limited to the sub-sample of children who were randomly selected for interview in depth. Rather these rates were determined by recalculating back to the original sample of 275 children. The original stratified sampling procedure was utilized because of the anticipated low prevalence of depressive disorder (Goldberg, 1986) and thus it was essential to weight the data back to the original sample, otherwise estimates of specificity will be too low and sensitivity too high. Figure 2 provides graphical presentation of values for weighted estimates of sensitivity, specificity and misclassification so as to apply to the original sample for a variety of scores on the KBB inventory. The intersection point of the graphs for sensitivity and specificity is used as the optimum cut-off score for diagnostic confidence and it is noted that this is 5 or more in relation to marked level of depression on the SPI. Further, in brief, in relation to a moderate level of depression, the best cut is 4 or more. Using these thresholds the estimates of sensitivity, specificity and misclassification are given in Table 3.

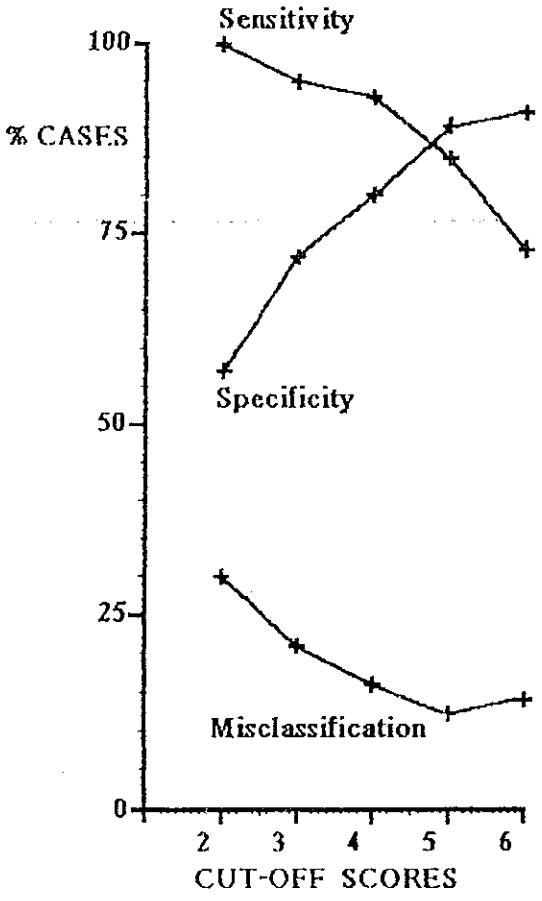


Figure 2: Percent rates of sensitivity, specificity and missclassification in relation in different cut-off scores on the KBB (R) Scale

Checking for Validity

The subject of validity is complex. Content validity is ensured by selection from a wide range of items which are considered relevant both to adult and child type depression (Ryan et al., 1987; Kolvin et al., 1984). The original eleven item scale had some discriminant validity in relation to depression and school phobia. While ideally a new or modified inventory should be tested on fresh clinical and community samples, multiple checks have been undertaken so as to minimise the possibility of invalidity and these are as follows:

FIRST, the cut of 5 was compared with different levels of severity of depression when using the SPI, as follows:

a) Clinically marked level of depression on the SPI Fewer children were classified as depressed, with overall, rates of misclassification 12%, sensitivity 85% and specificity 89%.

Table 4. Cluster diagnoses and item frequency

	CLUSTER SET A			CLUSTER SET B		
	Depression	Severe Depression "Masquerade" Group	Residual Maladjustment	Neg. Cognitions	Endogenous	Residual Maladjustment Group
	%	%	%	%	%	%
INCLUSIONS						
Dysphoric Mood	89	90	26	90	82	9
Unloved	56	76	9	86	18	5
Weeping	80	62	11	57	59	7
Loss of Energy	67	91	23	80	59	14
Loss of Interest	61	67	15	60	46	14
Emptiness	56	52	9	63	18	7
Depersonalisation	33	43	4	40	18	2
Suicidal Ideation	56	33	9	33	36	9
ADDITIONS						
Anhedonia	67	76	11	70	51	5
Depressive Thoughts	56	67	2	60	27	2
Lack of concentration	28	62	0	57	32	7
Hopelessness	51	48	11	70	18	7
Loss of Appetite	39	38	13	30	68	0

b) Clinically moderate plus marked levels of depression on the SPI. More children were now classified as depressed with rates of misclassification 13%, sensitivity 75% and specificity 96% (again based on recalculation back to the original sample).

SECOND, an item analysis was undertaken for the 13 items in relation to the cluster groupings previously specified (Table 4). The gradient of differences proved to be as good as when the groups were defined clinically; this is to be expected because of the nature of the statistical derivation of the clusters. Note: in a previous clinical presentation the clusters became inverted – this has now been corrected.

THIRD, for each cluster the rates of depression were calculated using the extended KBB scale (see Table 5). Again, the fit proved tight for the first but not the second cluster, with higher levels on the negative cognitions cluster.

FOURTH, on the basis of the SPI assessment, all the cases were classified according to our four operational categories previously described (Kolvin et al., 1991) and the percentage that scored 'positive' on the extended KBB scale was noted. A close relationship was found according to the type of disturbance and hits 'positive' on the KBB scale, with the percentages varying marginally according

Table 5. Rates of depression in the different clusters identified by the extended KBB scale diagnostic cut-offs

FIRST CLUSTER SET	
Depression	94%
Severe Depression/Masquerade	95%
General Maladjustment	12%
SECOND CLUSTERSET	
Depressive Cognitions	93%
Depressive/Endogenous	59%
General Maladjustment	0%

to whether different definitions of dysphoric mood are used and also according to whether the percentages are based on the sample studied or calculated back to the original 275 cases.

I Not depressed: 0 to 5% score positive on the KBB scale
II Anxious: 8 to 15% score positive on the KBB scale
III Depressed: 57 to 64% score positive on the KBB scale
IV Depressed plus: 77 to 84% score positive on the KBB scale Anxious

FIFTH, the cases with depression in the cluster sets were compared with the KBB diagnosis. In the first set the misclassification rate was under 12% (Table 5). We examined those that had been misclassified and found that about half of those categorised as depressed had low scores on the KBB scale and this probably related to mild levels of

Table 6. The KBB scale and first cluster set

	Cluster Diagnosis	
	All Depressed	All Not Depressed
KBB depressed	41	0
KBB not depressed	11	43

depression. However, on the *second cluster set*, 5 of the 22 cases with "endogenous depression" had a marginal score of 3 or 4 but only 2 of 30 cases of "negative cognitions" had such scores and 9 of the 42 residual cases. Thus if the cut were reduced to a lower level, then substantially fewer of the endogenous depression cluster would be misclassified by the KBB scale, somewhat fewer of the negative cognitions cluster, but many more of the residual maladjusted cluster would become identified as depressed. This suggests that the optimum cut-off on the scale for identifying marked levels of depression is less effective at identifying children with endogenous depression as the latter is represented by more moderate levels of depression.

SIXTH, discriminant function analysis was undertaken with the extended KBB scale items used as independent variables in relation to the clinical diagnosis of depression on the SPI. Discrimination proved highly significant, with the best discriminants being:

I *For moderate plus marked depression:*
dysphoric mood, weeping, feeling unloved, impaired concentration.

II *For marked depression:*
weeping, depressive thoughts, loss of interest, unloved, anhedonia.

SEVENTH, in relation to ordinal scaling of items, the question arises of the utility of summating the ordinal scores? The mean score of those not depressed on the SPI was 1.6, and those markedly depressed 7.3. The mean scores of the operational categories were: not depressed 1.1; anxious 2.0; depressed 5.9; and depressed and anxious 7.2. The two cluster analyses give interesting results. In the first, severe depression with masquerade has a mean of 7.8; depression of 7.5 and non-depressed of 1.8. In the second, the negative cognitions cluster has a mean of 7.7; endogenous depression of 5.4; and non-depressed of 1.2. Thus the evident gradient with this cluster set is more pronounced than with the first. Previously Kolvin et al. (1991) found a clear distinction between endogenous depression and negative cognition clusters. However, the data in this section and in Table 4, suggests the distinction is not only one of type but also of severity.

EIGHTH, sex differences.

Clinical depression is commoner in girls than boys (1.3 : 1); but the ration is even greater with cluster analysis (namely 1.5 : 1 for the first cluster and 1.8 : 1 for the second). What about severity? Girls who have a severe depression clinically have a mean score on the KBB of 7.5 and boys 6.9. In addition, on the second cluster set, the mean scores for girls and boys with negative cognitions is similar but there is a difference on endogenous depression with the girls being 5.8 and the boys 4.3.

Checking the intersection of sensitivity and specificity on graphs, the best cut for girls for severe depression is 5/6 and for moderate depression it is 4/5. For boys the respective levels are 4/5 and 3/4.

It has already been noted that a cut of 4/5 on the KBB scale misses 41% of the children with endogenous depression. If the cut is reduced to 3/4 then most of the girls with endogenous depression are captured but there is only a marginal difference in relation to boys. Similarly this reduced cut captures most of the girls with moderate or severe depression (96%) but is less effective for boys (76%) overall. The conclusion is that the revised version of the KBB is much more effective than the previous version in identifying depression with negative cognitions.

Discussion

The KBB questionnaire battery was initially developed as an attempt to discriminate between separation anxiety syndrome as represented by school refusal and depression as found in a group of school phobics (Berney et al., 1981; Kolvin et al., 1984). Subsequently, it has been used as a research instrument intended to provide a reasonably rapid assessment as the basis for diagnosing depression with a satisfactory degree of reliability and validity. The questionnaire is intended to be primarily a clinical interview instrument: but it also provides algorithms for the diagnosis of depressive syndromes of varying severity. The questionnaire also allows clinicians to utilise the information obtained during an interview as the basis of making a diagnosis of a depressive disorder.

Evidence is advanced to demonstrate that dysphoria is not a mandatory criterion for childhood depression; this replicates our previous findings (Kolvin et al., 1984) and supports the view that, although dysphoria is commonly viewed as a necessary characteristic of depression, it is not a sufficient criterion in itself (Hamilton, 1982; Rutter, 1988). Another important issue is how much interchange of one of the specified criteria by a similar or allied feature affects the psychometric qualities of the questionnaire – this proved to be minimal.

The extended KBB scale is only slightly longer than the original but it has many advantages. First, the items constitute a better representation of the different factors or clusters than the original scale; this is especially true of those reflecting "depressive cognitions" on the one hand and "endogenous" depression on the other. Second, there is good evidence of validity – although it could be argued that the validation levels might have been inflated by the method used to select items. Nevertheless, this brief scale is likely to make a useful contribution to either symptomatic diagnosis utilising the diagnostic algorithm described above when used by a less experienced clinician, or the clinical diagnosis of disorder by a more experienced clinician. However, despite including more items representing negative depressive cognitions, the scale is only moderately effective in identifying children with the syndrome of endogenous depression. Reducing the cut-off by one point will include more of the misclassified endogenous depression subjects (i.e. reduce the false negative rate) but will simultaneously increase the false positive rate. Thus the inventory can be used for the symptomative diagnosis of marked depression with a cut-off of 5 on the binary scale; or for moderate depression including endogenous depression with a cut-off of 4.

Some comment is merited concerning the ordinal scaling of items – the object is to ensure that the clinician bears in mind the distinction between substantial and marginal symptomatology so as to avoid giving equal importance to a large number of symptoms some of which may be of uncertain diagnostic significance. For instance, a subject scoring two on each of our 13 items will score 26 and is unlikely to be depressed; whilst a child scoring four on 4 items will have a total of 25 and could well be depressed. Thus while usually the summation of scores when using an ordinal scale will provide a better representation of overall severity, it could well be misleading diagnostically and could contribute to over-diagnosis especially when self-report measures are utilized (Weissmann, 1990). Hence a binary system of recoding is recommended as an essential supplement to the use of an ordinal scale as it is likely to contribute to the valid discrimination of depression. Finally, another possible source of over-diagnosis is the widespread use of clinical questionnaires by those who do not have sufficient experience to enable them to make valid judgments about clinical impairment (Weissmann, 1990).

Résumé

Ce travail décrit le développement et la calidation d'un bref questionnaire ayant pour but l'évaluation rapide de la dépression dans l'enfance. Il est en premier un interview clinique, mais il y a un algorythme associé pour le diagnostic d'un syndrome de dépression. Il peut également être utilisé pour le diagnostic clinique d'un trouble dépressif. La batterie courante comprend des traits représentatifs à la fois des "cognitions dépressives" et de "la dépression endogène". Cependant, il est moins efficace dans l'identification d'un syndrome de dépression endogène que dans le cas des cognitions dépressives.

Zusammenfassung

Wir stellen die Entwicklung und Validierung eines kurzen Fragebogens vor, der zur raschen Feststellung einer Depression im Kindesalter dienen soll. Es handelt sich primär um ein klinisches Interview, aber es gibt zusätzlich einen Algorithmus für die Diagnosestellung eines depressiven Syndroms. Er kann ebenso für die klinische Diagnose einer depressiven Störung benutzt werden. Die vorliegende Batterie umfaßt Merkmale, die sowohl für "depressive Kognitionen" als auch für die "endogene Depression" repräsentativ sind. Jedoch können erstere hiermit besser identifiziert werden als das Syndrom einer endogenen Depression.

References

Barrett, L.M., Berney, T.P., Bhate, S., Famiyuwa, O.O., Fundudis, T., Kolvin, I. & Tyrer, S. (1991). Diagnosing Childhood Depression. Who should be

interviewed – parent or child? *British Journal of Psychiatry, 159 (Suppl. 11),* 22–27.

Berney, T., 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,* 110–118.

Carlson, C.A. & Cantwell, D.P. (1980a). Unmasking masked depression in children and adolescents. *American Journal of Psychiatry, 137,* 445–449.

Carlson, C.A. & Cantwell, D.P. (1980b). A survey of depressive symptoms and disorder in a child psychiatric population. *Journal of Child Psychology and Psychiatry, 21,* 19–25.

Chambers, W.J., Puig-Antich, J., Hirsch, M., Paej, P., Ambrosini, P.J., Tabrizi, M.A. & Davies, M. (1985). The assessment of affective disorders in children and adolescents by semi-structured interview. *Archives of General Psychiatry, 42,* 696–702.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Dohrenwend, B.P., Shrout P.E., Egri, G. & Mendelsohn, F.S. (1980). Non specific psychological distress and other dimensions of psychopathology. *Archives of General Psychiatry, 37,* 1229–1236.

Dohrenwend, B.P., Yager, T.S., Egri, G. & Mendelsohn, F.S (1978). The psychiatric status schedule as a measure of dimensions of psychopathology in the general population. *Archives of General Psychiatry, 35,* 731–737.

Feighner, J.P., Robins, E., Guze, S.B., Woodruff, R., Winokur, G. & Munof, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry, 26,* 57–63.

Fundudis, T., Berney, T.P., Kolvin, I., Famiyuwa, O.O., Barrett, L.M., Bhate, S. & Tyrer, S. (1991). Reliability and validity of two self-rating scales in the assessment of childhood depression. *British Journal of Psychiatry, 159 (Suppl. 11),* 36–40.

Gelder, M.G. (1983). Anxiety and phobic disorders, depersonalisation and derealisation. In: *Handbook of Psychiatry, Vol. 4 (Ed. Russel, G.F.M. and Hersov, L.A); Chap. 5.21;* pp. 212–228. Cambridge: Cambridge University Press.

Goldberg, D.P., Cooper, B., Eastwood, M.R., Kedward, H.B. & Shepherd, M. (1970). A standardised psychiatric interview suitable for use in community surveys. *British Journal of Preventative Social Medicine, 24,* 18–27.

Goldberg, D.P. (1986). Use of the General Health Questionnaire in clinical work. *British Medical Journal, 293,* 1188–1189.

Hamilton, M. (1982). Symptoms and Assessment of Depression. In Paykel, E.S. (Ed.), *Handbook of Affective Disorders,* (pp. 3–11). Edinburgh: Churchill Livingstone.

Kolvin, I., Barrett, L.M., Bhate, S.R., Berney, T.P., Famiyuwa, O.O., Fundudis, T. & Tyrer, S. (1991). The Newcastle Child Depression Project: Diagnosis and Classification. *British Journal of Psychiatry, 159 (Suppl. 11),* 9–21.

Kolvin, I., Berney, T.P. & Bhate, S. (1984). Classification and diagnosis of depression in school phobia. *British Journal of Psychiatry, 145,* 347–357.

Kolvin, I., Barrett, L., Berney, T.P., Bhate, S.R., Famiyuwa, O.O., Fundudis, T. & Tyrer, S. (1989). The Newcastle child depression project: studies in the diagnosis and classification of childhood depression. In K. Davison and A. Kerr (Eds.) *Contemporary Themes in Psychiatry – A tribute to Sir Martin Roth,* 149–155. London: Gaskell.

Kovacs, M. (1981). Rating scales to assess depression in school aged children. *Acta Paedopsychiatrica, 46,* 305–315.

Nurcombe, B., Seifer, R. Scioli, A., Tramontana, M.G., Grapentine W.L. & Beauchesne, H.C. (1989). Is major depressive disorder in adolescence a distinct diagnostic entity? *Journal of the American Academy of Child and Adolescent Psychiatry, 22,* 333–342.

Puig-Antich, J. (1980). Affective disorders in childhood *Psychiatric Clinics of North America,* 3, 403–423.

Rutter, M. (1988). Depressive disorders. In Rutter, M.; Tuma, A.H; Lann, I.S., (Eds.). *Assessment and Diagnosis in Child Psychopathology.* (pp. 347–376).

Ryan, N.D., Puig-Antich, J., Ambrosine, P., Rabinovich, H., Robinson, D., Nelson, B., Iyengar, S. & Twomey, J. (1987). The clinical picture of major depression in children and adolescents. *Archives of General Psychiatry, 44,* 854–861.

Spitzer, T., Endicott, J. & Robins, E. (1978). Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry, 35,* 773–782.

Ward, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58,* 236–2544.

Weinberg, W.A., Rutman, J., Sullivan, L., Penick, E. & Dietz S. (1973). Depression in children referred to an educational diagnostic centre: diagnosis and treatment. *Preliminary report Journal of Pediatrics, 83,* 1065–1072.

Weissmann, M.M. (1990). Applying Impairment Criteria to Children's Psychiatric Diagnosis. *Journal of the American Academy of Child and Adolescent Psychiatry, 29,* 789–795.

Williams, P. (1986). The problem of diagnosis in psychiatric research. *Journal of the Royal Society of Medicine, 79,* 1–2.