British Journal of Psychiatry (1991), **159** (suppl. 11), 9–21

The Newcastle Child Depression Project
Diagnosis and Classification of Depression

I. KOLVIN, M. L. BARRETT, S. R. BHATE, T. P. BERNEY, O. O. FAMUYIWA, T. FUNDUDIS and S. TYRER

A total of 275 successive referrals to a university child psychiatry unit out-patient department were examined using the Child Depression Inventory. Of these, 95 children were examined further by a structured clinical interview, and the relationship between different instruments for the assessment of depression in childhood was investigated. Just over one-third of the children (35%) had significant depression, and it was found that depression may be missed unless children with other psychiatric diagnoses are examined closely. Multivariate analyses of the clinical data provided factorial validation of diagnoses when employing different clinical diagnostic schemas.

Our knowledge of depression and its aetiology in children was previously limited, but is increasing rapidly. Two fundamental issues in the current debate about depression in childhood are the validity of diagnostic criteria, and how best the classification of depression in children should proceed. With regard to the first issue, some consider it possible to diagnose major depression in childhood using criteria identical to those used in adults (Spitzer *et al*, 1978; Puig-Antich, 1980); others feel that different considerations should apply to children. This is reflected in alternative diagnostic schemas, including three from separate centres: those devised by Puig-Antich *et al* (1978), and by Weinberg *et al*, (1973) in the United States of America, and by Kolvin *et al* (1984) in Newcastle. How valid are each of these in the diagnosis of childhood depression and to what extent do these schemas agree or disagree with each other?

The second issue is how depression in childhood should be subclassified. In adult psychiatry, this is a notoriously complex area and even if childhood depression were similar, classification problems would be compounded by issues of child development. Many, therefore, consider it better to describe disorders in a systematic way as there is no consensus of the validity of any sophisticated subclassification. This can be done by giving an account; firstly, of the severity of depression and, secondly, by documenting any specific disorders associated with the illness. The wisdom of this approach is underlined by the fact that it is not yet proven whether depression in children is a homogeneous or a heterogeneous condition. If the latter is true, is depression missed because of concealment by other symptoms and, if so, what is the extent and nature of the association?

The Newcastle Depression Project was set up to examine these and related issues. The aims of this part of the study were fourfold:

(a) to explore the use of the Standard Psychiatric Interview (Goldberg *et al*, 1970) for diagnosing depression in pre-adolescents and adolescents
(b) to assess agreement between different schemas used for diagnosing major depression in childhood and 'validate' each against the other
(c) to explore the extent of mixed depression in childhood with particular reference to conduct, neurotic, and phobic disorders
(d) to examine multivariate approaches to classification of depression in childhood and adolescence.

There were two hypotheses.

(a) Validity of the different diagnostic schemas would vary. Closer agreement would be found between diagnoses that had been achieved using clinical interviews. The Weinberg criteria would perform more satisfactorily if sufficient allowance was made for severity of disorder.
(b) The proportions of mixed depression would vary according to the type of associated disorder.

Method

The method used for aims (c) and (d) above is described separately later in the paper.

Sample selection and exclusions

As part of the Newcastle Depression Project, 316 referrals between 9 and 16 years of age were screened for depression using the shortened version of the Child Depression Inventory (CDI) (Kovacs & Beck,

1977; Carlson & Cantwell, 1980*b*; Kovacs, 1981) at their first appointment. They had all attended the Nuffield Child Psychiatry Unit for the first time, either as an out-patient, day patient or in-patient. The following patients were excluded: those receiving pharmacological agents, those with any significant physical illness, those with diagnosed clinical seizures or other major neurological conditions, those with an intelligence quotient (IQ) of less than 70, and those with autism, schizophrenia or any substance-abuse disorder. Refusals and losses were as follows: in 22 cases (7%) the families did not attend the follow-up interview; six children (2%) were already on medication (for epilepsy or diabetes) and in two cases (0.6%) parental data were not available as the children were in care; a further 11 cases (3.5%) declined to be included in a parallel study which involved the dexamethasone suppression test – a test for disturbed endocrine function. This reduced the sample to 275 cases (90% of original number).

As there were insufficient resources to carry out in-depth interviews on all cases, it was decided to use a stratification procedure to choose children for intensive interview, the selection being made on the basis of probability of childhood depression. This was done according to CDI scores. A pilot study using this instrument, comprising 66 children in hospital and 13 in a social services residential assessment unit had shown that the mean score on the shortened version of the CDI was 6.9 with a standard deviation (s.d.) of 4.0. Scores of nine (half a s.d. above the mean) were called 'high scorers' and of eight or below were 'low scorers' in this subsequent clinical survey.

Approximately half of the high scorers (49 of 100) and one-quarter of the low scorers (46 of 175) were selected randomly giving roughly equal numbers of high and low scorers. The ratio of cases chosen was based on the principle of over-sampling from that proportion of the population which had the potential for an excess of childhood depression. Provided that the sampling fraction for each stratum selected is known, over-sampling allows estimates for the total population. Each stratum can be weighted appropriately (Robins, 1980). The mean age of the selected sample was 12.5 years (s.d. 2.0); 27% were nine or ten years old; 26% were 11 or 12; 32% were 13 or 14 and 14% were 15 or 16. Of the cases, 48% were male and 52% were female.

The current research is confined to comparing and contrasting subgroups of children in consulting settings. There are no general population controls and this is relevant for the interpretations of findings. The review by Williams (1986) highlights this. Psychiatric morbidity in the community is largely undifferentiated (Goldberg & Huxley, 1980) and "diagnostic criteria derived from, and for, patients seen in specialist psychiatric practice, may incorporate thresholds which are inappropriate for non-specialist settings" (Dohrenwend *et al*, 1978, 1980; Williams, 1986). Thus, scales which may be reliable in samples from consulting practices may be much less so in general population samples. This has crucial relevance in our paper concerning screen measures as criteria for diagnosis.

Psychiatric interviews: reliability and agreement

Three senior and experienced research child psychiatrists were involved in assessment of the children. As the assessment of depression was largely based on clinical assessment by the psychiatrists, their reliability was examined closely. Three different statistical techniques were used for this purpose, the kappa statistic, product-moment correlation coefficients and supplementation by analysis of variance and intraclass correlation coefficients where necessary.

Kappa is a statistic specifically developed to measure the interobserver reliability of categorical data and which corrects for chance or expected agreement (Cohen, 1960). The range of kappa is not necessarily the same as that of a correlation coefficient and thus maximum kappas have been calculated – the maximum values of agreement that can be attained in terms of the number of categories and frequency with which each rater or judge chooses that category (Cohen, 1960). Kappa was also used to measure agreement between schemas about the presence or absence of depression. As the cases were given overall ratings of severity on a number of different dimensions of disorder, such as depression and anxiety, another form of correlation proved necessary. With the limitation of resources we considered it more efficient to attempt to improve the agreement coefficients by examining ratings made independently by all three raters on a small group of children. For these purposes analysis of variance and intraclass correlation coefficients were performed. Intraclass correlations provide essentially an average intercorrelation.

We have employed a number of different schemas to give ratings of depression. It was decided that data and ratings from the Standard Psychiatric Interview (SPI; Goldberg *et al*, 1970) should be the criterion for validating the other measures – as it was always totally dependent on the observations and clinical judgements of the senior research child psychiatrists, and clinical ratings were felt to be of paramount importance. However, some may view this as an arbitrary decision and argue for the obverse, namely that the DSM-III (American Psychiatric Association, 1980) diagnosis based on data gathered via the Kiddie-SADS interview (Chambers *et al*, 1985) should be the validating criterion for the other measures (SPI; Goldberg *et al*, 1970; Williams, 1986).

Measures used

The SPI

The SPI is a semistructured 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 give ratings on a range of clinical disorders. For 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) used 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 the interviews were conducted with flexibility and sensitivity, few problems emerged when using the SPI with prepubertal children and adolescents. The instrument rates symptoms according to clinical judgements on a five-point severity scale. We used concepts and definitions of disorders more usually geared to adults, as outlined and defined in the SPI manual. This allows both clinical diagnosis of depression and rating of severity of disorder.

The reliability of the data was checked by several observers whose agreement coefficients were pooled using the Spearman–Brown prophesy formula (Hartmann, 1977). Interviews on nine children were undertaken by a single interviewer (SB) and observed by video recording by two independent raters (TB & IK). Diagnoses of presence or absence of clinical depression and anxiety were made with the intention of undertaking statistical analysis for categorical data (Cohen, 1960). However, there were no discrepancies in diagnosis, which is probably a reflection of the similar training, concepts and symptom definition of the three raters.

Such categorical agreement does not necessarily provide information about the extent of agreement of severity on the dimensional scores of symptoms or on ratings of overall depression and anxiety. However, when the correlations of the interviewer ratings were pooled, an interobserver reliability of greater than 0.9 was found for overall severity of depression and anxiety disorder. The correlation coefficients on the 12 symptoms were lower, with the lowest agreements being on fatigue (0.75), irritability (0.84) and phobias (0.84). When applying the Spearman–Brown prophesy formula, these pooled correlations rise substantially. However, when using product–moment correlation coefficients, rater scores can differ by a constant across ratees and not be reflected in the coefficient. Hence, we also used intraclass correlation coefficients that take into account systematic error, which can serve to lower coefficients (Winer, 1971).

Analysis of variance revealed highly significant differences between ratees but not between raters. Intraclass correlation coefficients proved to be of a similar order to those obtained with the Spearman–Brown formula on ratings of depressive and anxiety disorders (0.96 and 0.92 respectively). It was concluded that the SPI 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 both the prepubertal period and adolescence.

The Kiddie–SADS

This is a commonly used and validated schedule in assessing depression in children. 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 (Orvaschel *et al*, 1982; Chambers *et al*, 1985). It is a reliable instrument for measuring symptoms of depression and conduct disorder, although ratings of anxiety disorder have not been so consistent (Chambers *et al*, 1985). The recommended procedure is for the parents to be interviewed, then the child; the child interview is guided by information from the parent. The examiner is required to make a clinical judgement of the presence or absence of the symptom under inquiry and, if present, its severity. The mode of onset and duration of symptoms contribute to ratings of severity. The rating is a synthesis of all the information available on the child and provides the basis for the diagnosis. A wide range of symptoms are covered and all have been found to be assessed with acceptable reliability, except for pathological guilt and impaired concentration (Chambers *et al*, 1985).

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 Newcastle study. Interviewing of parent and child by different interviewers was carried out, thus avoiding a halo effect in the rating of symptoms. Independent ratings of the severity of depression as reported by the child (by consultant) or parent only (by senior research fellow) were followed by separate ratings of clinical disorder.

Other modifications included the addition of a small number of symptoms using a format of probes and scales similar to that in the original. These included paradoxical aggression, i.e. aggression which manifests itself in the home only, boredom, a sense of being unloved, a feeling of emptiness, self-dislike, exaggerated illness behaviour and *déjà/jamais vu*. Further, in our study, a four-point severity scale was used in place of the seven-points in the original. Such a four-point scale can be converted easily from an original one reflecting severity, to a dichotomous one, reflecting significant presence or absence of symptoms; this is an important point, as discussed later. These two sets of data based on child and parent interviews were the basis of the analyses of an accompanying paper (Barrett *et al*, 1991).

The Newcastle Instrument

Our research group has also developed a brief set of key diagnostic criteria for depression (Kolvin *et al*, 1984, 1989). They differ from the criteria advanced by the North American research groups (Weinberg *et al*, 1973; Puig-Antich, 1980), but bear greater similarity to the Puig-Antich schema than the Weinberg variation. Ratings of symptoms are scored as an ordinal scale to reflect severity (1, 2, 3, 4) but this can be recoded as a binary scale (0, 0, 1, 1). Previously, it was found that the best cut was at six or over (summed scores) on the binary scale. To achieve a diagnosis of depression and to allow evidence of severity of overall depression, a symptom algorithm was employed consisting of a score of six or more on the binary scale being severe depression and five, moderate depression (Kolvin *et al*, 1984). These cuts were originally developed to discriminate between school phobics who were depressed and those who were not.

The Weinberg schema of 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). To be diagnosed as depressed, children must have the primary symptoms of dysphoric mood and self-deprecation plus two

more of a further eight symptoms. There are no guidelines as to the duration and severity of disturbances. Hence, a decision was made to allow for severity by scoring and recoding the Weinberg criteria in the same way as the Newcastle criteria. Similar operational decisions were made to score severe and moderate degrees of depression and an appropriate symptom algorithm was applied.

Evaluation of other symptoms

Depression in childhood may be a homogeneous entity although the evidence from adult psychiatry suggests that, apart from psychotic depression, the remaining depressive states are heterogeneous (Goldberg, 1983). To determine how far other symptoms and disorders were associated with depression, ratings were made for associated clinical disorders of sufficient severity to be given a diagnostic label in their own right. Patients were therefore given ratings on three dimensions of disorder, based entirely on the presence of the defined symptoms and irrespective of any other ratings.

(a) *Neurotic disorder*: symptoms of significant anxiety, sensitivity, obsessive-compulsive phenomena, somatic symptoms, hypochondriasis or hysteria. The disorder, as defined here, overlaps broadly with category 300 of the ICD-9 (Rutter *et al*, 1975). This system has been used extensively in the past and the reliability of ratings established elsewhere (Kolvin *et al*, 1984; Wrate *et al*, 1985). The subcategories are not listed, but neurotic disorder overall is rated according to severity on a four-point scale. Ratings are based on clinical judgement where four indicates marked disorder, three represents moderate severity, two denotes dubious presence and one indicates no evidence of disorder. A score of three is taken to represent neurotic disorder at a clinical level irrespective of any other diagnosis. A score of two represents subclinical symptoms.

(b) *Conduct disorder*: a combination of lying, stealing, truanting, poor interpersonal relationships and various types of antisocial behaviour including aggression. The disorder, as defined here, overlaps broadly with category 312 of the ICD-9. Severity ratings were made according to the clinician's judgement, using the convention and scale described above. The clinical scale takes into consideration frequency, severity and duration of behaviour.

(c) *Phobic disorder*: symptoms of irrational fears of situations or people, accompanied by avoidance and a sense of social handicap. It includes school phobia. This overlaps broadly with category 300.2 of the ICD-9. Ratings of severity were made according to the clinician's judgement and scored as above.

Sequence of interviews and scoring of measures

The shortened version of the CDI was administered first; on the basis of their scores the children were interviewed by a senior psychiatrist who had no knowledge of the results of the screening. The parent interview (Kiddie-SADS) was always administered by one of the research fellows. This was usually undertaken simultaneously with, but independent of the child interview. The child was interviewed using the SPI followed by the Kiddie-SADS; these schedules were either administered in sequence on the same day or within 48 hours of each other. It had been intended that the SPI and the child Kiddie-SADS be given independently, the latter by the second research fellow or another senior psychiatrist, but a lack of staff time prevented the implementation of this intention in every case, and in about one-fifth of the cases the senior psychiatrist first administered the SPI and then the Kiddie-SADS. On the basis of the information contained in these extended Kiddie-SADS schedules, 'symptom' diagnoses of neurotic, conduct and phobic disorders, and their severity, were made. In addition, the extended Kiddie-SADS schedule was subjected to algorithms relating to defined symptoms listed in the Weinberg (Weinberg *et al*, 1973) and Newcastle (Kolvin *et al*, 1984) schemas, and this allowed relevant 'symptom' diagnoses to be achieved. In the case of the Weinberg, due allowance was made for severity of defined symptoms.

Following Puig-Antich (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 relatively discrete and associated with the defined number of symptoms of the depressive syndrome. The detailed definition and criteria were those applicable to DSM-III diagnostic categories of either major depression, single episode or recurrent episode. The annotated child Kiddie-SADS protocols were later reviewed blindly by a research fellow who made a diagnosis of depression and ratings of severity based on Kiddie-SADS data. The categories and ratings that emerged were used to check agreement between schemas.

On the SPI, categorisation 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 schemas. Composite ratings of severity of clinical disorder and diagnosis were established by one of the senior psychiatrists (IK, TB and SB) based on a synthesis of all the information, but these ratings have been little used.

Results

The estimated prevalence of depression in the out-patient sample, based on direct clinical assessment utilising the SPI, was found to be 35.3%. Clinical depression was subdivided into marked or severe depression, and moderate depression. Marked or severe depression was present in a quarter of the total sample irrespective of whether the proportions were based on direct assessment using the SPI (26%), case-note review applying DSM-III criteria (25%), the Newcastle symptom algorithm (25%), or the Weinberg criteria (23%). However, the rates for moderate depression varied. The Newcastle formula, when there were five inclusion criteria, including dysphoric mood, gave a rate of 18%. The percentage when using DSM-III was 21%, which was similar – again probably reflecting decisions about severity of inclusion criteria. The rate based on the SPI proved intermediate at 9% – this was considered the most accurate of our methods as it was based on both direct assessment

and clinical judgement which allowed finer discrimination on the five-point scale. Finally, as our earlier decisions on the Weinberg about the inclusion criteria for moderate levels of depression proved rather rigorous, few patients entered this intermediate category.

The next step was the comparison of the main instruments used in our study – the Kiddie–SADS (as a basis for interviewing) and DSM–III (for diagnosis), the Weinberg criteria and the Newcastle scale – were compared. This was to determine the relationship between different ways of categorising depression as compared with the SPI. Although kappa scores were moderate, intraclass correlations for all three measures were satisfactory and this indicates acceptable agreement. A comparison of agreement between these same three scales also showed satisfactory intraclass correlation coefficients (Table 1).

While depression was present in over one-third of the sample, it presented alone in only 11.2%. It should also be noted that the rates were marginally lower when diagnosis was based on clinical judgement using one source of information, namely the SPI (Goldberg *et al*, 1970). They are higher when a more detailed and intensive interview is used and are highest when they include information from two interview sources, namely the child and the mother. This is generally to be expected: the greater the information, the less symptoms of depression will be missed. The prevalence of depression was higher in in-patients than out-patients. Our estimated rates of depression of about one in three do not necessarily represent a main diagnosis and there is considerable overlap – for example one-sixth of the clinical sample proved phobic, one-third had a conduct disorder, and half had a neurotic disorder.

Examination of our data (Table 2) suggests possible explanations which could account for the high rates of depression in our clinic population. It would seem that a simple categorisation into any major category of child psychiatric disorder merely obscures the picture. Where the main diagnosis is of a conduct or phobic disorder, we may overlook a significant associated depressive disorder. Thus, about one-sixth of our patient population have a phobic disorder and of these about two-thirds are depressed; about one-tenth of our out-patient group are depressed and phobic. In contrast, one-third of our clinical population have a conduct disorder but only one-fifth of these are also depressed – thus one-in-four of the clinical sample have a conduct disorder without depression. Finally about half of the out-patient population have neurotic disorders and one-third of these are depressed; put another way, one in six have a neurotic disorder without depression. An alternative perspective is that about one-third of the depressed group are phobic, half have a neurotic disorder, and one-fifth, a conduct disorder.

Age and sex trends

One crucial theme is the analysis of the data to display age and sex trends. The sample size was too small for any fine-grain analysis, so the new out-patient group of cases was dichotomised at about the mean (12 years or less and 13 years or more) and according to sex. Only about a fifth (18%) of younger girls proved depressed but more than three times as many (68%) of the older group were. In contrast, two-fifths (39%) of the younger boys were depressed but in the older group this rate reduces by half (19%). These sex ratios roughly replicate those reported by Rutter (1986; based on Pearce's 1974 data). Pearce emphasises that the differences are likely to be in relation to sexual maturity rather than chronological age. Therefore we checked the dichotomy according to chronological age against our clinical assessments of puberty and found that age and pubertal stage were not in agreement in over one in five boys (22.7%) and about one in six girls (14.9%). Re-analysis of the data according to whether the cases were prepubertal or pubertal gave rise to a modest variation of the proportions of depression – 42% of prepubertal males compared with 30% of females, and 27% of pubertal males as opposed to 66% of females. These data are consistent with the notion that there is a shift in the sex ratio of depressed children from a male preponderance before puberty to a female preponderance after puberty (Rutter, 1986).

Discussion of rating scale results

Several authors have previously distinguished between symptoms and syndromes of depression in childhood and have asserted that they are not synonymous; sadness does not necessarily constitute depression (Hamilton, 1982; Kolvin *et al*, 1984).

Table 1
Diagnostic schemas

	Raw kappa values	Maximum kappa values	Intraclass correlation
Validation against SPI			
DSM–III	0.45	0.92	0.64
'Weinberg'	0.56	0.77	0.64
Newcastle (Kolvin *et al*)	0.55	0.94	0.63
Comparison			
DSM–III v. 'Weinberg'	0.50	0.75	0.71
DSM–III v. Newcastle	0.57	1.00	0.76
Newcastle v. 'Weinberg'	0.56	0.77	0.66

Table 2
Co-morbidity in a clinic sample with depression and associated disorders[1]

Associated disorder	Total: %	Plus depression: %	Minus depression: %	Depression minus the associated disorder[2]: %
Phobia	17.7	11.2	6.5	24.3
Neurotic	49.5	15.9	33.6	19.6
Conduct	31.6	6.5	25.2	29.0

1. Estimates of rates in total clinical sample based on calculation from proportions in stratified sample.
2. Column four represents the rate of depression where the listed associated disorder is excluded – for example depression without phobia.

However, while dysphoria is commonly viewed as a necessary characteristic of depression, it is not a sufficient criterion in itself. Our data suggest a variety of ways in which childhood depressive syndromes overlap with other childhood syndromes.

Technically, although at times we have described the different conditions that we have identified as 'disorders', it is more appropriate to view them as 'syndromes'. Carlson & Cantwell (1980*b*) take this one stage further by indicating that, to achieve the status of depressive disorders, such syndromes need to be distinguished from other disorders by differences in family factors, family patterns of illness, natural history, biological factors and response to treatment. Nurcombe *et al* (1989), tried to deal with the confusion regarding the concept of depression – pointing out that the term may be used to refer to a symptom of affect or mood, a syndrome cluster of depressive symptoms, a disorder reflecting a category of individuals with a depressive syndrome, or a disease consisting of a disorder which is physiologically or biologically based. Much current research is aimed at identifying clear depressive syndromes (Ryan *et al*, 1987; Seifer *et al*, 1989).

It is clear that, once the severity of diagnostic symptoms is taken into account, agreement between diagnostic schemas is substantial and, further, all are found to be equally valid in relation to an independent clinical criterion. This is of particular relevance to the Weinberg criteria, which previously did not take account of severity and consequently gave rise to rather high rates of depression in a learning-disorder population (Weinberg *et al*, 1973). Cantwell (1983) reports only moderate agreement between different sets of diagnostic criteria and our findings, on the surface, seem to support this. However, the same diagnostic criteria can be used to achieve quantitative variations in relation to qualitatively distinct disorders (Rutter, 1986). When the different diagnostic schemas were treated in this way and the emergent dimensions were compared using intraclass correlation coefficients, the agreements proved much more satisfactory. Data were not available for checking the reliability of the diagnoses of depression or the dimensional ratings except for the clinical SPI, which proved to have satisfactory reliability.

The association of depression with other psychiatric conditions

Mixed depression is a well known phenomenon in adult psychiatry; witness the affective basis of some phobic and obsessional states. In children also, as Rutter (1988) points out, it is not uncommon for depression to be associated with other psychiatric conditions. For instance, Puig-Antich & Gittelman (1982) reported that one-third of boys fitting research diagnostic criteria for major depressive disorders could also be diagnosed as conduct disordered, using DSM–III criteria; Hendren (1983) found that more than half of patients with anorexia nervosa also have depression, using these same criteria. Carlson & Cantwell (1980*b*) determined that 50% of children in their clinic survey presenting with affective disorder had another diagnosis with hyperactivity (25%) and anorexia (11%) proving most frequent.

The above evidence suggests a similarity between depression in children and that in older people (although the association with conduct disorder and hyperactivity may be seen as age-specific). An extension of this view would be that depression underlies the other disorder with which it is associated but that, once treated, both disorders will improve. This is a known scenario in adults, but the only evidence for this in children is the study by Puig-Antich (1982) of major depression in prepubertal boys. One-third of these also met the DSM–III criteria for conduct disorder and, of those 13 who had mixed disorder, 11 lost both disorders on appropriate antidepressant treatment. In a subsequent study, Puig-Antich *et al* (1985) found that four months after the depressive features had cleared with antidepressant treatment, residual symptoms remained. Similarly, Berney *et al* (1981) found that school-phobic children with associated depression (according to the Newcastle criteria) experienced rapid relief of their depressive symptoms with treatment, but showed considerable persistent residual neurotic disturbance. An alternative interpretation of the overlap is that mixed depression is the independent association of two separate disorders mostly with distinct aetiologies and outcomes. The degree of relatedness presumably varies with the type of associated disorder, being higher for neurotic/phobic disorders, and lower for conduct disorder, despite the results described by Puig-Antich (1982).

Adult/child links

The association between psychiatric illness in childhood and later adult illness is not close (Robins & Hill, 1966). However, Zeitlin (1986) has shown that in some individuals, depression in childhood is followed by adult depression. He investigated a group who had been under psychiatric care at the Maudsley Hospital in London, both as children and as adults. Zeitlin examined two control groups: one of children who had not re-attended as adults, and one of adults whose disorders had begun in adult life. He reported that much adult depression was not

preceded by any childhood disorder. Further, of those adults who had attended as children, only a small proportion had predominantly depressive disorders. In addition, of the children diagnosed as depressive using clearly stipulated criteria, most were not diagnosed as suffering from depression in adulthood. However, in a small number of cases where the diagnosis of depression was made on the basis of specified criteria in both childhood and adulthood, strong continuities between the child and adult group were found. It may have been that many of the depressed adults had presented with other symptoms in childhood and were not recognised as depressive.

Masked depression

The theme of one disorder masking or concealing another is not new in child psychiatry, and dates from the theories and studies of Cytryn & McKnew (1972), Cytryn *et al* (1980) and Glaser (1968). It would appear that these authors were correct in suggesting that depression was often hidden, but they concluded, erroneously, that depressive symptoms manifested in a different way.

Our findings suggest that depression is simply undetected because of inadequate techniques of clinical assessment. Other workers have demonstrated that children with so-called 'masked' depressive disorders (Rutter, 1988) do show overt depression (Carlson & Cantwell, 1980*a*,*b*). On the other hand, some research suggests that the clinical features of depression may be modified or may be more primitive in younger children (McConville *et al*, 1973; Zeitlin, 1986). This current work underlines the importance of assessing depression accurately, regardless of the presenting complaint. All associated symptoms should be recorded and as long as each syndrome is not viewed as discrete from disorders reflecting other psychiatric and behavioural difficulties, it should be easier to understand.

Rates of depression in child psychiatric populations vary widely. Weinberg *et al* (1973) reported that 57% of children aged 6–12 attending an educational diagnostic clinic had a depressive disorder. Petti (1978) using modified Weinberg criteria diagnosed depressive disorder in about 60% of in-patients. Lower rates (27%) were reported by Hudgens (1974) in adolescent in-patients and Kovacs (1977) found only 14% of clinic patients to be depressed. Carlson & Cantwell (1980*a*) reported 28% of the interview sample of 102 children attending the UCLA Neuropsychiatric Clinic had a clinically diagnosed depressive disorder and 49% a 'cross-sectional depressive syndrome'. The variation in rates is likely to reflect the concept of depression employed and definition of criteria and also the nature of the clinical sample studied. In our study, depressions diagnosed by the SPI are likely to represent a disorder, whereas those identified by other instruments or by symptom algorithms should be seen as syndromes, with a consequently higher prevalence (Carlson & Cantwell, 1980*a*). The relatively high frequency of depression found in our sample was likely in part to reflect the severity of disturbance in cases referred to a university-based consulting clinic. We do not believe that we have overdiagnosed depression in our sample, although this is a possibility in any study where a particular disorder is being sought for in detail. It is more likely that there has been underdiagnosis in those studies where co-morbidity has been overlooked. This being said, because of the selection factors operating in referral to our clinic, it is not relevant to extrapolate rates of depression found in our study to the general child population.

Clinical and multivariate approaches

In the United Kingdom, multivariate exploration of adult depression was pioneered by the Newcastle School of Psychiatry (Kiloh & Garside, 1963; Carney *et al*, 1965). The use of these adult diagnostic criteria for describing depressive disorders in school-age children and adolescents has been promoted in the UK by the work of Berney *et al* (1981), Kolvin *et al* (1984) and this current research. In the USA, Puig-Antich & Chambers (1978) have been the main architects with the development of semistructured interview assessments. These latter developments have facilitated multivariate approaches to classification of childhood depression.

In this section we describe principal component and cluster analysis on all children and adolescents on whom full Kiddie–SADS data were available.

Method

In exploratory factor analysis it is hypothesised that a set of variates exist which substantially account for any interrelationships (Maxwell, 1977). Thus, theoretically, exploratory factor analysis is capable of identifying symptoms (and perhaps other features) which hang together in characteristic patterns; this view is consistent with the hypothesis that there is one general factor underlying all the symptoms, as well as additional differentiating or bipolar factors. Varimax rotation may be of help in identifying those symptoms which are especially related to each other, but in the process their variances may be greatly inflated by error (Maxwell, 1977).

Principal component analysis (PCA) is a widely used technique for exploring a covariance structure. Hopefully the derived components will represent more basic variability in the data than in the observed variates (Maxwell, 1977). Where PCA is used as an alternative to exploratory factor analysis, the first few rotated components "often give a robust identification of major trends in the data" (Taylor, 1979). The mathematical basis of the PCA approach makes it a useful technique in non-standard situations "where the data may not fit classical assumptions of multinormality" (Taylor, 1979).

Many of the qualifications about factor analysis, especially those relating to multinormality, do not apply in the case of PCA, which therefore appears to be a more robust technique for analysing clinical data. (As much of the data collected in this research was found to not always conform to the conditions defined above, PCA was chosen in preference to factor analysis.)

The purpose of cluster analysis is to identify those subsets of features or individuals which cluster together, i.e. have much in common within each cluster, but little in common between the clusters (Aldenderfer & Blashfield, 1984). These methods not only summarise relationships between features, but may also highlight new sets of harmonies or structures (Cattel, 1965; Garside & Roth, 1978).

Cluster analysis theoretically appears to be the most useful way of identifying relatively homogeneous subsets of individuals, the validity and number of clusters identified are still open to debate. For instance, questions arise about how to establish the validity of the clusters generated: not only may differences emerge using different programs 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.

However, the multivariate techniques are not a classification panacea. The results of factor analysis and PCA may be considerably influenced by changes either in the size or composition of the subject sample, or in the range of measures employed (Kendell, 1968, 1975). Another well known limitation when using factor analysis for classification purposes is that in practice only a minority of patients may be accommodated by the emergent classification (Hewitt & Jenkins, 1946).

Provided that the pitfalls are constantly borne in mind, these techniques usefully complement the traditional clinical methods of identifying syndromes. 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.

In selecting items for multivariate analysis, the criteria defined by Ryan *et al* (1987) were followed. To avoid the confusing effects of unreliability and lack of variance, only those items from the Kiddie-SADS were included for which there were both test-retest reliability coefficients above 0.50 and a good range of scores for the sample population tested. However, some additional items which we thought merited consideration on theoretical grounds were also included. In all, 22 items were selected which overlapped extensively but not completely with the 22 items included by Ryan *et al* in their analysis. This constitutes the *first set*: within this, three parallel subsets of data were examined, the first deriving from parent interview, the second from child interview and the third consisting of summed data derived from parent and child interviews. A *second set* of data was established in which school refusal, and aggression confined to the home were included instead of the final two items in set one listed in Table 3.

PCA with varimax rotation was undertaken and components were retained following the two criteria: an eigenvalue greater than one; and having at least three variables that loaded significantly on it. This gave rise to a smaller number of components than selected by Ryan *et al* (1987) for their larger patient population.

Clusters were identified using Ward's (1963) method: in this the coefficients are of the dissimilarity type, with iterative relocation with hierarchic fusion. The method tends to derive clusters of relatively equal sizes. Cluster numbers are determined by examining the fusion coefficients for sharp changes in their values, the number of clusters before these sharp changes being accepted as the most probable solution (Aldenderfer & Blashfield, 1984). Some view these changes as rather subjective and hence other methods have been advocated, such as, the procedure developed by Mojena (1977) in an attempt to obtain a better definition of a significant jump. However, most procedures have to be checked for validity, and clusters need to be judged in terms of interpretability and predictive ability (Everitt, 1977).

Clusters found by Ward's method can be ordered in terms of their overall organisation, that is, the elevation of their profile, which determines the selection of the clusters.

Principal-component analyses

Each of the three PCAs of set one produced four similar clinically interpretable components. Together these accounted for 58% of the variance using child interview data; 53% using mother interview data; and 60% using combined data.

The four comprised:

(a) depressed, anhedonia, increased fatigue, psychomotor retardation component which is similar to an adult endogenous-type depression factor
(b) negative cognitions component accompanied by thoughts of suicide
(c) anxiety
(d) anger–agitation component.

On all three analyses the loadings of three of the four identified components overlapped substantially with those described by Ryan *et al* (1987) (see Table 3). On the negative cognitions component using child data, the overlap was in the order of five of the five variables; with parent data three of five; and with combined data five of five. With the endogenous component using child data, the overlap of such loadings is four of the nine variables, with parent data it is five of nine; with combined data it is six of the nine. However, on the anxiety component the overlap of loadings using child data is only one of the seven; with parents two of seven; and with combined data it is only two of the seven. In contrast to Ryan, neither an appetite–weight nor conduct components

Table 3
Principal component analysis derivation of the main components with symptom loadings set one (with three subsets)

	Components child data				Components parent data				Components both			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
Anhedonia		66X			77X					72		
Self-denigration	72X					85X			76X			
Hopelessness	63X					64X			69X			
Guilt	48			59	59	63			60			
Withdrawal	48	52X								69X		
Dysphoric mood	61X	58X							58X	66X		
Brooding	58X		49			52			72X			
Diurnal variation												
Terminal insomnia					63X					54X		
Loss of appetite		80X			77X					72X		
Lack of energy		63X			62X					82X		
Irritability		47		48			83X		49			47X
Suicidal ideation	69X					53X	43X		80X			
Somatic complaints		47		42						61		
Slowing thoughts	53				51X					59		
Sense of failure	83					79			76			
Separation anxiety			83X					81X			84X	
Conduct problems			−41		−50						−46	
Anxiety			71					55X	41		66X	
Fear			67					83			82	
Anger	55			40			74X		44			67X
Agitation				75								77X
Eigenvalue	8.2	1.6	1.6	1.4	5.8	2.3	1.8	1.6	8.0	2.2	1.8	1.3
Proportion variance	37.4	7.4	7.1	6.2	26.4	10.7	8.1	7.4	36.4	9.9	8.0	6.1
Interpretation	Negative cognition	Endogenous type	Anxiety	Anger–agitation	Endogenous type	Negative cognition	Anger–agitation	Anxiety	Negative cognition	Endogenous type	Anxiety	Anger–agitation irritation

X represents 'overlap' with loadings reported by Ryan *et al* (1987). For purposes of presentation, loadings below 40 have been excluded.

were identified. However, on all three analyses an anger–irritability component emerged.

An important finding is that the stuctures of both the negative cognitions and endogenous component differed according to whether the data was derived from child or parent. Notably the item dysphoric mood did not load highly on the parent-derived negative cognitions component. On the endogenous component based on parent interview data, dysphoric mood, somatic complaints and social withdrawal do not load highly; on the child-derived endogenous component, insomnia and slowing of thoughts do not load highly.

In the circumstances, a better definition of the major components probably emerges from the analysis of the combined data. The endogenous components consist of: anhedonia; lack of energy; loss of appetite; withdrawal; dysphoric mood; somatic complaints; slowing of thoughts; terminal insomnia. The negative cognitions components consist of: self-denigration and sense of failure and hopelessness; brooding/worrying, guilt; dysphoric mood; suicidal ideation; irritability; anxiety and anger.

In the separate analysis of the data (set two), school refusal and aggression confined to the home were included and again in relation to the parallel subsets. Again, on all three analyses, three of the identified components overlap substantially with those described by Ryan *et al* (1987) but the fourth did not prove to be clinically meaningful. The pattern of loadings in the other three were broadly similar to those described in set one. However, when employing child-interview data or combined data, poor school attendance loaded highly on the endogenous component, but only moderately on the anxiety component. The conclusion must be that changing the data set gives rise to some important variations.

Cluster analysis

The same two data sets, with the data standardised, were subject to cluster analyses which were repeated for child data, parent data and combinations of the data. In the first set, irrespective of whether child, parent or combined data was employed, three broad clusters emerged. The numbers of

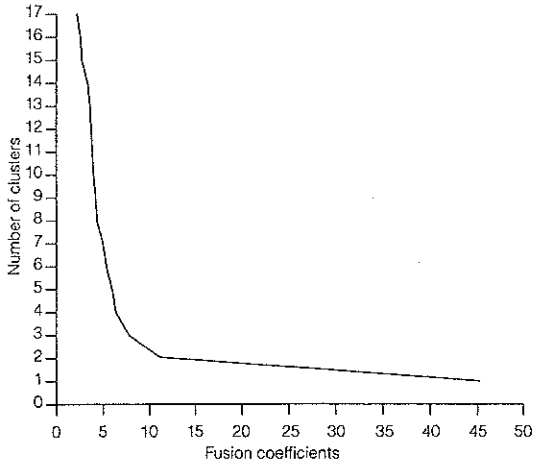

Fig. 1 Number of clusters using child, parent or combined data v. fusion coefficients.

Table 4
Comparison of clinical and cluster diagnoses based on selected items from standardised psychiatric interview (SPI)

	Cluster diagnosis			
	Depression	Severe depression plus 'masquerade' symptoms	Mixed maladjustment cluster	Total
Clinical diagnosis based on SPI				
Neither anxious nor depressed	0 0%	0 0%	21 100%	21
Anxiety category	2 8%	1 4%	22 88%	25
Depression with or without anxiety	16 35%	20 43%	10 22%	46
Total	18 20%	21 23%	53 58%	92

clusters were plotted against the fusion coefficients (Fig. 1). It is noted that the flattening to the curve begins at the three-cluster point suggesting three clusters are present in the data. The clusters consisted of negative cognitions, endogenous depression and an amalgam of conduct–neurotic–anxiety features. The latter was always the largest cluster and included 42 children when using child data, 44 children when using parent data and 51 when employing combined data. The negative cognitions cluster ranged from 22 children when using child data to 30 children when analysing parent data, reducing to 29 when using combined data. The structure of these clusters varied according to whether the data derived from the children or parents, and while a negative-cognition cluster was easily recognised when analysing combined data there was only an overlap of five out of 11 items when analysing parent data, as compared to the picture that emerged when analysing child data. The endogenous cluster gave eight identifying peaks from child data and six of these re-emerged when using the parent data. However, the endogenous cluster consisted of 30 children when using child data; reducing to 17 with parent data and 10 with combined data.

On the analysis of set two, the picture that emerges is not highly consistent. Cluster analysis deriving from child data gives rise to an endogenous cluster, a mixed conduct–anxiety cluster, and finally, a mixed endogenous negative cognition and school refusal cluster. With the parent data the syndromes that emerge are more distinct and circumscribed: an endogenous cluster which includes school refusal, negative cognitions, and again a mixed conduct neurotic anxiety cluster. On this occasion, the endogenous cluster consisted of 22 children when using child data, and 30 children when using parent data, reducing to only 14 children when using combined data.

Again, when using combined child–parent data, clear distinctions emerge. Three clusters are obtained: depressive cognitions, endogenous depression, and mixed conduct disturbance/neurotic disorder. Depressive cognitions consists of self-denigration, hopelessness and guilt, irritability, suicide and a sense of anxiety, fear and anger. The endogenous depression cluster is represented by anhedonia, hopelessness, withdrawal, dysphoric mood and worrying, insomnia, loss of appetite, lack of energy, suicidal ideation, somatic complaints, slowing of thought, anxiety and school refusal. Finally, the mixed conduct disturbance neurotic disorder cluster consists of diurnal variation of mood, insomnia, loss of appetite, somatic complaints, separation anxiety, conduct problems, fear, paradoxical aggression, and to a lesser degree, school refusal.

A fundamental question is the extent to which clinical and cluster diagnoses agree. Such comparisons proved feasible as we had previously undertaken cluster analysis of data derived from the SPI which gave rise to three clusters of 'depression', severe depression plus 'masquerade' symptoms and a mixed-maladjustment cluster (Kolvin *et al*, 1989). The cases were located in these three cluster categories. This was complemented by clinical categorisation of the cases based (again) on the SPI technique. The clinical groupings employed were maladjustment unassociated with anxiety or depression, anxiety without depression, and depression with or without anxiety (Table 4). The clinical category of maladjustment without major anxiety or depression did not overlap with the cluster categories of depression. Nor does the clinical category of pure anxiety overlap very much with the depressive clusters. But about four in five of those clinically diagnosed as depressed fall into the two cluster categories of depression. The same data can be reorganised to show a surprisingly tight fit between clinical diagnosis of depression and cluster diagnosis.

The next stage consisted of a comparison of clinical categories based on the SPI interview and allocation of cases into categories based on cluster diagnosis using data from the Kiddie–SADS interview. The clusters were derived from the reduced Kiddie–SADS list (Ryan *et al*, 1987). The picture proved similar (Table 5). There is a reasonably tight fit

Table 5
Comparison of clinical diagnosis (SPI) and cluster diagnosis using data from the Kiddie-SADS

	Cluster diagnosis			
	Endogenous	Negative cognitions	Remainder	Total
Clinical diagnosis				
Neither anxious nor depressed	2 (8%)	0 (0%)	19 (80%)	21
Anxiety category	3 (11%)	4 (15%)	20 (64%)	27
Depression without anxiety	7 (50%)	5 (36%)	2 (14%)	14
Depression with anxiety	10 (30%)	21 (63%)	2 (6%)	33
Total	22	30	43	95

between clinical diagnosis based on SPI interview and cluster diagnosis using the Kiddie-SADS data which constitutes factorial validation of the SPI assessments.

Finally, the identified clusters were checked with the diagnosis derived from the Newcastle schema (Berney *et al*, 1981; Kolvin *et al*, 1984), and again the fit was reasonably close, which provides factorial validation of this schema.

Age and sex analyses

The sample size was too small for partitioning to allow an examination of the structure of the principal components for groups of boys as compared with girls, and younger as compared with older children. However, it was possible to submit the emergent cluster groups to simple statistical analysis – we confined ourselves to the cluster analysis of the Kiddie-SADS data from child interview.

The 'endogenous' cluster consisted of 45% boys and 55% girls; the negative cognitions cluster of 33% boys and 66% girls; while the residual maladjustment cluster showed a reversal of this pattern with 59% boys and 41% girls. Although boys and girls had the same mean age (12.0 years) in the residual maladjustment group, girls were older in the two depressive subclusters with a mean age of 13.9 in girls as opposed to 12.4 years in boys in the endogenous group; and 13.0 as to 11.8 in the negative cognitions group.

Discussion of the multivariate results

Ryan *et al* (1987) in Pittsburgh undertook PCA of psychiatric symptoms on 296 children and adolescents aged six to 18 years. In terms of simple demographic data his sample was broadly similar to ours. The meaningful components that emerged proved similar for both children and adolescents and included 'endogenous' and 'anxious' factors found in many studies of adult depression. Three others were 'negative cognitions'; 'appetite-weight changes'; and a 'conduct' factor. The five components accounted for 55% of the variance.

In the current Newcastle research, similar component structures were identified consisting of negative cognitions/suicide; a depressed/anhedonic retardation component; an anxiety component; and anger/agitation and irritability. The four components that emerge in the Newcastle research overlapped with only three of the five components which emerged in the Pittsburgh research. Further, the Newcastle data did not generate a conduct component but rather an anger–agitation–irritability one. The loadings on the Pittsburgh components were examined to ascertain any similarities to the components identified in the Newcastle research.

The endogenous component overlapped on five items – anhedonia, fatigue, psychomotor retardation, depressed mood, anorexia – but in Newcastle three additional items had high loadings – insomnia, somatic complaints and anger–irritability. These similarities proved true (with some minor variations) irrespective of whether the mother or the child was the informant.

The negative cognitions component overlapped almost completely on negative self-image, sense of hopelessness, suicidal ideation and brooding (worrying), again irrespective of whether the mother or the child was the informant. The association between hopelessness and suicidal behaviour parallels the findings on Pittsburgh with their younger population and also the findings in adults by Beck *et al* (1975). It may be that our series was not old enough to allow an association to emerge of impulsiveness, irritability and antisocial behaviour, as it did in Pittsburgh (Ryan *et al*, 1987). The other factors did not emerge in Newcastle.

It is notable how the loadings on the components from the child and the mother in the Newcastle work paralleled the two main Pittsburgh factors. It must be concluded that the variations produced by separating out the mother from the child data do not give rise to fundamental component differences.

The first two components found in the Pittsburgh and Newcastle studies resembled those reported in a review of studies of adult depression (Nelson & Charney, 1981). The smaller size of the Newcastle sample did not allow separate analysis of pre- and post-pubertal groups.

The largest cluster consists of largely non-depressed individuals and can be broadly described as a category of mixed maladjustment; there is a moderate-sized group with negative cognitions and a smaller cluster of children with endogenous depression associated with school refusal. It is not surprising, in an analysis of the elements that make up depression, that the two themes of depressive cognitions and endogenous depression appear as separate clusters when cluster analysis is applied.

It is notable that when the final two items of the data subset are replaced by aggression within the home (which we have described as paradoxical) and school refusal, the structure of the cluster analysis changes. While the depressive cognition cluster overlaps moderately in the two analyses, the overlap is much more extensive in the endogenous cluster and covers anhedonia, hopelessness, withdrawal, dysphoric mood, terminal insomnia, loss of appetite, lack of energy, somatic complaints, slowing of thoughts and anxiety.

While PCA highlights components, these are not easily convertible into syndromes represented by a subgroup of individuals. Some of these components only account for a small proportion of the variance and it is not to be expected that they will be represented by individual subjects. Cluster analysis, by its very nature, is designed to maximise the similarity or dissimilarity in the group of individuals investigated. Despite this, we achieved a clear-cut distinction only in relation to negative cognitions and endogenous depression.

In the Pittsburgh research the results of cluster analysis were rather inconclusive, with the clusters formed not indicating a precise pattern, whereas in Newcastle the identified symptom clusters proved distinct. Some studies have not demonstrated the presence of major depressive disorder in childhood using either multivariate techniques or adult-type diagnostic criteria (Seifer *et al*, 1989). It is possible that the basis of the discrepancy resides in the methods they employed. These authors appear to question the clinical approach which lists diagnostic criteria for depression on two main grounds – the possibility of flaws due to confirmatory bias on behalf of the investigator or of the "identification of spurious syndromes in polysymptomatic children". However, our work shows that there is a substantial degree of concordance between clinical diagnoses and multivariate approaches. We argue that there is a potential flaw in some studies which inhibits concordance, consisting of a lack of appreciation of the presence of co-morbidity – depression may coexist as a clinical syndrome with other syndromes but may not necessarily be the main syndrome.

Similarities with adult studies

Adult factor or cluster analysis studies consistently reported an 'endogenous' factor or cluster (Nelson & Charney, 1981; Rassaby & Paykel, 1979) and other clusters. For instance, in research with depressed patients, Andreasen & Grove (1982) identified four clusters which roughly correspond to severe depression with endogenous features, less severe depression, bipolar depression, and psychotic depression. These factors and clusters proved to be stable irrespective of patient status, type of patient population or of the instruments used.

In the multivariate studies of child and adolescent data some additional factors have been reported, namely negative cognitions and irritability. In the Pittsburgh research, suicidal thoughts loaded on both of the last two factors and the association between a sense of hopelessness and suicidal behaviour parallels findings in adults by Beck *et al* (1975). Ryan *et al* (1987) argue that the loading of suicidal ideation and suicidal acts on these two separate factors is consistent with the view that there are two distinct sorts of suicidal children and adolescents: those who present with a sense of hopelessness (as part of a depressive disorder), and those who are impulsive, irritable or antisocial (but not necessarily as part of a depressive illness). In our research there are no loadings of suicidal ideation on the endogenous component, but there are significant such loadings on the negative cognitions component irrespective of whether the data derives from the child or the parents. Suicidal ideation also loaded on the anger–irritability component when using data deriving from the parents.

The traditional view of the development of negative cognitions in depression states that the altered thoughts in depressed individuals are consequent upon their depressed feelings. When the process of depression is ameliorated then the cognitions return to normal. The cognitive therapists take an opposite view. They believe that the patient's negative cognitions lead to a negative appraisal of the individual, the future and the world (Beck, 1970). If the depressive thoughts are found to be illogical when examined by the individual, there is the basis for improvement in depression overall. The cognitive model has impressive research evidence to back its assertions (Williams, 1984). However, there is a group of sufferers from depression with high genetic loadings who are less likely to respond to cognitive treatment (Rush *et al*, 1984). These individuals largely have endogenous symptoms of depression when they become depressed. Both these groups of patients have negative cognitions.

The separation of the negative cognitions and endogenous clusters in our analysis suggests that depression in childhood may differ from that in adults. We need to know more about the origins, nature and duration of these symptom patterns in our child population. A number of questions arise. Are the negative cognitions in our children enduring traits or are they consequent upon a depressive

process? Do the children with endogenous symptoms go through a stage of having negative cognitions before they develop their severe depressive symptoms?

The Newcastle school (Kiloh & Garside, 1963; Carney *et al*, 1965), on the basis of multivariate analyses of adult data, asserted that two separate syndromes of depression could be distinguished – neurotic and psychotic – but Kendell (1968) did not replicate these findings. The Newcastle child data suggest clustering on the basis of severity – a severe depression cluster similar to endogenous depression in adults. The less severe syndrome of depressive cognitions is difficult to relate to classification with adult psychiatric patients – it does not correspond entirely with the less severe category of Andreasen & Grove (1982).

Finally, our findings illustrate the truism that what is obtained from multivariate analysis reflects the data that are analysed. When using the SPI, a cluster of severe depression, anxiety and school phobia emerged and another of depression without somatic symptoms (Kolvin *et al*, 1989). When using the Kiddie-SADS symptom items the distinction is between a severe endogenous depression and depressive cognitions with suicidal thoughts. Nevertheless, the multivariate clusters provide good validation of the traditional clinical approach to classification. Further, as cases have been assigned to groups or clusters by mathematical means this has allowed independent confirmation of various clinical schemes for diagnosing depression in adolescence, namely the SPI, the Kiddie-SADS (DSM-III) and the Newcastle scheme.

Acknowledgements

During the period of this study IK received support from the Health Promotion Research Trust. We are also indebted to Mrs M. Fleeting, computer scientist. Our thanks to Mrs M. Blackburn for administrative help, and to Mrs K. McEnemy, Mrs B. Slayel and Miss C. McNeil for secretarial help. We are all indebted to the late Dr J. Puig-Antich for his generosity in making available to us the early and later versions of the Kiddie-SADS, and allowing modifications for the purpose of our project.

*Israel Kolvin, BA, MD, FRCPsych, DipPsych, *Professor of Child and Family Mental Health, Royal Free Hospital School of Medicine and the Tavistock Clinic, 120 Belsize Lane, London NW3 5BA, formerly Professor of Child and Adolescent Psychiatry, Director of the Human Development Unit, University of Newcastle, The Fleming Nuffield Unit, Burdon Terrace, Newcastle upon Tyne NE2 3AE*; Lynn M. Barrett MB, BS, MRCPsych, *Consultant Child Psychiatrist, Queen Elizabeth Hospital, Gateshead, Tyne & Wear formerly Research Senior Registrar, Fleming Nuffield Unit*; Thomas P. Berney, MB, ChB, DPM, FRCPsych, *Consultant Psychiatrist, Fleming Nuffield Unit, Newcastle upon Tyne*; S. Bhate, MB, BS, FRCPsych, *Consultant Child and Adolescent Psychiatrist, Young People's Unit, Newcastle General Hospital and Fleming Nuffield Unit, Newcastle upon Tyne*; O. O. Famuyiwa, BM, MRCPsych, *formerly Research Fellow, Fleming Nuffield Unit, Newcastle upon Tyne*; Trian Fundudis, PhD, MA, DipPsychother, CPsychol, FBPsS, *Top Grade Psychologist, Fleming Nuffield Unit, Newcastle upon Tyne*; Stephen P. Tyrer, MB, BChir, FRCPsych, *Consultant Psychiatrist, the Royal Victoria Infirmary, Newcastle*

*Correspondence