

## The relevance of the multiple criterion screen to an adolescent population

MAURICE PLACE,<sup>1</sup> ISRAEL KOLVIN, ANGUS MACMILLAN<sup>2</sup> AND RORY NICOL

*From the Nuffield Child and Adolescent Psychiatry Unit, Fleming Hospital, Newcastle upon Tyne*

**SYNOPSIS** The 4th- and 5th-year pupils of four large comprehensive schools, which were broadly typical of the schools serving an urban population, were screened using a multiple criterion approach. The screen results were compared with those found in a younger population drawn from the same area, and a similar rate of efficiency was found. However, in the adolescent age group the screen was more accurate at detecting true negative cases (specificity) but less able to identify true positive cases (sensitivity). Overall, the multiple criterion screen proved more efficient than a variety of single screen criteria, and the findings also suggest that certain screen measures used in adolescence may be more effective at identifying disturbed girls while others are more effective at identifying disturbed boys. This confirms the principle that the choice of screening instruments must be made with reference to their proposed use.

### INTRODUCTION

Adolescence is a phase of development which has excited much study and interest. The original picture, which was based upon clinical experience, portrayed adolescence as a period of great stress and turmoil (Hall, 1904) and emotional chaos (Eissler, 1958). However, epidemiological surveys of adolescents (Douvan & Adelson, 1966; Offer, 1969; Rutter *et al.* 1976) have found little evidence to suggest that such major disruptions are widespread.

An important first step in the screening of an adolescent population for psychiatric disturbance is to seek instruments that are known to be reliable and valid for that age group. Many well-validated instruments exist for children and adults, but there have been few attempts to assess the validity of these instruments when used with adolescents. Instruments specifically designed for this age group tend to be lengthy and therefore inappropriate for screening purposes (Spivack *et al.* 1967; Offer, 1969, p. 224).

In previous study in Newcastle upon Tyne (MacMillan *et al.* 1980) a multiple criterion screen was developed to identify those children

starting secondary education who prove to have special needs because of psychiatric disorder, which manifests itself in the school situation. This screen consisted of reports from the teacher (Rutter B2 Scale), peers (sociometry), and the youth (Junior Eysenck Personality Inventory) which were weighted to produce an overall score. The foremost requirement of such a screen is that the false positives should be kept to a minimum: in other words, the selected children should be those who are unequivocally disturbed. This must be balanced, however, with not missing too many of the disturbed children, that is keeping false negatives to a minimum. MacMillan and his colleagues demonstrated the value of a multiple criterion screen as originally commended by Bower (1969). This early study was carried out on 1723 children with a mean age of 11.8 years, and therefore could be considered to be in their early adolescence.

By repeating this exercise with a parallel sample of children of approximately 15 years of age, it proved possible to compare the efficiency of this self-same screen when applied to a later stage of adolescence. As well as such methodological considerations, the study allowed estimates of prevalence to be made, but the relative performance of the screen in this older age group was the primary concern. If this screen were to prove equally valid, it would constitute an

<sup>1</sup> Address for correspondence: Dr M. Place, Child and Family Psychiatry Unit, Sunderland General Hospital, Sunderland, Tyne and Wear.

<sup>2</sup> Present address: Crichton Royal Hospital, Dumfries.

important step in defining disturbance across adolescence using comparable measures.

## METHOD

The 4th- and 5th-year pupils of four schools in the area were screened, using the method previously described by MacMillan *et al.* (1980). Sixty-five children who were on the school registers had not attended school for at least a term; they were therefore excluded. This gave a study population of 1805 pupils, with a mean age of 15.7 and a range of 14–16.6. There were equal numbers of boys and girls. The four large comprehensive schools were selected as being broadly representative of the local educational settings; it is therefore not surprising that the social class distribution of the sample was similar to that reported in a previous Newcastle community study (Neligan *et al.* 1975).

The assessments were undertaken by subject teachers during a double period in the school timetable, and then repeated with any absentees who attended class during the subsequent two weeks. Three hundred and fifty-nine youths did not complete the self-rating measure and so were assumed to be adolescents who were frequently absent from school. The characteristics of this group are reported in detail elsewhere (Place & Kolvin, 1985). The teachers were asked to complete a rating scale on each member of the class during this period, and all proformas were then collected for analysis.

In addition to the community sample, 30 adolescents resident in a medium-stay psychiatric unit were also assessed blind, which offered an evidently pathological group against which to validate the psychiatric interview.

### Screen measures

The instruments used explore three areas of functioning.

#### (a) Teacher's ratings

Teachers completed a Rutter B2 Scale (Rutter, 1967) for each youth in the study. The B version of the scale, which differs only slightly in wording from the B2 version, has been shown to have inter-rater reliability of the order of 0.72 in the 12-year age group, and a test-retest reliability of 0.89. As well as the total score, subscales of neurotic and antisocial behaviour

were used. These were the Newcastle Revision of Rutter's original scores, which increases the number of items contributing to the subscores, and has been shown to increase their reliability (MacMillan *et al.* 1980).

#### (b) Sociometry

This technique gives two scores which represent an assessment of the index youth's standing in the eyes of his peers. Each pupil in a class is asked to nominate the three class-mates he most likes to be with, and the three he most likes to sit beside. He also nominates the three peers he least likes to spend time with or sit beside. The scores for each pupil are then collated. Those with only one nomination for being liked, or none at all, are considered to be isolated. Those with 12 or more nominations for not being liked are considered to be rejected. Such a sociometric assessment has been shown to be reasonably reliable in children (MacMillan *et al.* 1978), in whom test-retest reliability is 0.72 for isolation and 0.87 for rejection. The technique's usefulness in the adolescent age group and the appropriateness of the cut-off points in this age group have yet, however, to be established.

#### (c) Self-report

The Junior Eysenck Personality Inventory (JEPI) (Eysenck, 1965) was completed by all the youths who attended the nominated class at least once during the two weeks of the study. This inventory yields scores – introversion/extroversion, neuroticism and a lie scale. The neuroticism dimension is said to measure characteristics which are associated with instability and a low score is likely to predict a 'calm, carefree, easy-going' personality (Eysenck, 1965). The neuroticism scale was used in the screening exercise. Split-half reliabilities for neuroticism are in the region of 0.84, while test-retest reliability is reported as 0.77.

#### Establishing cut-offs and assigning weightings

As in the original study, it was decided that for each of the above screen measures an extreme score would be used as an indicator of potential disturbance. These scores were used to determine cut-offs which selected the highest-scoring 10–14% of the population on that measure. On the Rutter B2 Scale, for example, a total score of 9 was taken as the cut-off (Rutter *et al.* 1970).

These cut-offs were used to determine the weighting to be assigned to each case. In the case of the B2 Scale, a score of 9 was given a weighting of one and a total score of 15 or more was assigned a weighting of two. In addition, a score of 4 or more on the Newcastle revision of the neurotic subscore of this instrument was assigned a weighting of one, as was a score of 4 or more on the antisocial subscore.

For the sociometry technique, extreme scores on either scale were assigned a weighting score of one point. In the case of isolation, a score of 0 or 1 was assigned a score of one point and, for rejection, a score of 12 or more was allocated one point. Because of the nature of the instrument, scores were obtained for all of the youths on the class list, but only those for whom full data were available were assigned weighted scores in this part of the study. The neuroticism scale of the JEPI required a more complex weighting system. The norms for an adolescent population necessitate the use of a lower cut-off in boys. In the current study, a girl with a JEPI(N) score of 21-22 was assigned a weighting of two points, more extreme scores being assigned a weighting of three points. For boys, a score of 18-20 was given a weighting of two points, more extreme scores being assigned a weighting of three points. These cut-offs were chosen because they correspond to  $1\frac{1}{2}$  and 2 standard deviations above the mean for 15-year-olds (Eysenck, 1965).

A random sample of youths who had an extreme total score on at least one of the four screening instruments was selected for interview. The percentage levels used were varied to ensure the inclusion of at least 20 individuals on each major criterion. However, because of the overlap between the criteria, the numbers selected on each were approximately 30. In addition, 50 adolescents were selected at random from the youths who had no extreme scores on any of the measures. All these adolescents were then interviewed blind.

Two psychiatric interviews were used in conjunction, with the first being more geared to adults, namely the Standardized Psychiatric Interview (Goldberg *et al.* 1970), and the second being more geared to children, namely a semi-structured approach modelled on that originally described by Rutter & Graham (1968), but using a wider range of themes. This interview

examines specific items such as anxiety, fears, relationships which reflect neurotic behaviour and also themes which reflect antisocial behaviour. In addition, the youths' reactions during the course of the interviews were noted. The version devised for children has been used in previous studies (Neligan *et al.* 1976; Wrate *et al.* 1985), which found the average coefficient of agreement (Cohen, 1960), between two diagnosticians, to be in the region of 0.71 (Wrate *et al.* 1985). On this occasion, as additional items specific to adolescence had been added, the inter-rater agreement was checked in 20 cases. For these purposes, the psychiatrist rated the presence of psychiatric disorder on a 4-point scale, on three dimensions - conduct, neurotic and overall disturbance. The average coefficient was found to be 0.73. It is of note that the parents of the interviewed youths were visited and several measures were completed, but these measures are not reported on in any detail.

The above mentioned weightings were added together to give a summed weighted score for each subject in the community sample. Previously, a cut-off of 3 or more had been found to be the best predictor of disturbance in the 12-year-old population (MacMillan *et al.* 1980), and the optimum cut-off for the adolescent population is discussed later in this paper.

In addition to the community sample, 30 adolescents resident in a medium-stay psychiatric unit were also assessed blind, which offered an evidently pathological group against which to validate the psychiatric interview.

## RESULTS

Three hundred and fifty-nine youths, who were taken to be chronic absentees, did not complete the self-report instruments. Thirty per cent of these would have been screen positive even without a JEPI score and, when compared with the attending school population they show a significant excess of high scores on all the other instruments (Place & Kolvin, 1985). Thus the youths who have frequent absences from school constitute a distinct group.

Obviously, such a large group of absentees must make predictions about the total population very tentative. The absentees were not interviewed to validate the screening procedure in what is a distinct subgroup of the school-age population.

Table 1. Children scoring above the cut-off on each measure at different ages

	Cut-off used	12-year-olds (N = 1723) %	15-year-olds (N = 1446) %
Rutter B2 total score	≥ 9	12.0	13.6
Newcastle neurotic subscore	≥ 4	11.0	12.2
Newcastle antisocial subscore	≥ 4	14.0	15.5
JEPI neuroticism			
Boys	≥ 18	—	5.8
Girls	≥ 21	—	7.7
	≥ 20	17.0	—
Isolation	≤ 1	14.0	15.0
Rejection	≥ 12	14.0	8.0

Table 2. Frequency distribution of multiple criterion screen scores and projected rates of psychiatric disturbance at each level

Multiple criterion screen score	0	1	2	3	4+
Number of youths with screen score	911	233	125	63	114
Rate (%) of disturbance in interviewed sample	10	32	49	67	71
Estimated number of disturbed youths	96	74	61	42	81

Their absence is explained by the fact that in the last two terms of the academic year, many 5th-formers simply ceased to attend school on a regular basis. Discussion with the teachers revealed that this would account for the vast majority of absent youths. Consequently, this paper deals with the validity of the multiple criterion screen in that part of the adolescent population which attends school on a reasonably regular basis.

The percentage of subjects scoring above the cut-off on the various instruments are compared by age in Table 1. As can be seen, the Rutter B2 Scale yields slightly more cases in adolescence than at a younger age when a score of 9 or more is taken as the cut-off point. Similar increases are seen in the neurotic and antisocial trait scores. The rejection measure of sociometry selects fewer cases in the adolescent sample, as does the neuroticism scale of the JEPI. It is not clear why this cut-off should yield less than the predicted 10%, but perhaps it is a function of the high rate

of absenteeism in this sample which may be distorting the previously described patterns.

In order to estimate the most efficient cut-off point, the weighted scores were compared with the psychiatrist's assessment of disturbance. This assessment had been validated by comparing its performance in three different groups. In a residential psychiatric unit the overall rating of disturbance was 100%. In the group of youths who obtained a weighted score of two or more 58% were rated as disturbed and, in the group with a weighting of less than two, 19% were rated as showing disturbance. In the interviewed sample, 50 of the youths had no extreme scores and 26 had extreme scores on only one instrument. Considering the multiple criterion screen weightings, a cut-off point of 3 or more gave a Kappa coefficient (Cohen, 1960) of 0.26, and a cut-off of 2 or more gave a Kappa coefficient value of 0.04.

The method of selecting the interview sample did not permit direct inferences about the study population as a whole. Population estimates were calculated using the proportion of disturbance in the interviewed sample in relation to the total number of youths at each level of multiple criterion screen score (see Table 2). Using these data, a cut-off of 3 or more gave a Kappa coefficient of 0.35 and a cut-off of 2 or more gave a Kappa coefficient of 0.42.

The findings suggest that a cut-off of 2 or more is more efficient in an adolescent population and, to test this hypothesis, the screen's efficiency at the two levels was assessed using the reconstituted population.

#### Efficiency of the screen

It is possible to assess the efficiency of the screening procedure by other mathematical formulae. The sensitivity of the screen is its ability to identify true cases out of all possible true cases; likewise, specificity of the screen is its ability to select persons who are truly free of disorder. These are calculated by the following formulae:

$$\text{sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \times 100;$$

$$\text{specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}} \times 100.$$

Table 3. Comparing the efficiency of weighted score cut-offs at both ages

	15-year-olds* (N = 132)		12-year-olds† (N = 244)
	≥ 2	≥ 3	≥ 3
Total weighted score cut-off			
Specificity (%)	89	95	57
Sensitivity (%)	52	35	92
Overall efficiency (%)	80	80	79
Kappa coefficient	0.42	0.35	0.43
Maximum kappa within the border totals	0.89	0.59	0.61

\* Estimated for total sample using data in Table 2.

† From MacMillan *et al.* (1980).

In addition, an overall efficiency of the screen may be calculated:

$$\text{overall efficiency} = \frac{\text{true positives} + \text{true negatives}}{\text{total sample}} \times 100.$$

The results of these calculations for this adolescent population are presented in Table 3. As can be seen, the multiple criterion screen was somewhat more sensitive when a cut-off of 2 or more was used, and this is confirmed by the Kappa coefficients. When the adolescent screen was compared with the results obtained in the 12-year-old age group the efficiency was almost the same, but the rates of sensitivity and specificity were reversed.

As was stated earlier, the initial instruments cut-offs were chosen as those that might be predicted to yield the highest scoring 10% of the

population. Since rejection did not select 10% of the adolescent population, it is possible that this may have reduced the efficiency of the screening process. The data were thus reanalysed using cut-offs which yielded 10% on each of the measures. The new system reclassified 3 false negatives as true positives, but 3 true negative cases became false positives. The Kappa coefficient for this reclassification did not alter, and thus these changes did not improve the efficiency of the screen which in this population is optimally effective when adolescents with a total weighted score of two or more are designated as screen positive.

#### The relative contribution of the screen measures

So far, we have explored the validity of the multiple criterion screen when compared with an independent clinical assessment of disturbance. It is obviously important, however, to explore the relationship that the individual instruments have with each other, and the contribution made by each to the weighted score. As can be seen in Table 4, the correlations of the different individual measures with each other tended to be rather low, compared with their correlation with the total weighted score. As expected from previous research (MacMillan *et al.* 1980), the Rutter subscales showed a high correlation with the total scale score, and the intercorrelation of the neurotic and antisocial subscales with each other was 0.47. All the screen measures correlated significantly with the total weighted score – and, as all the correlations were moderate, the high level of significance must be a reflection of the large number of cases included in the

Table 4. Correlation of screen criteria (N = 1446)

	Rutter B Scale	Newcastle neurotic subscale	Newcastle antisocial subscale	JEPI neuroticism	Isolation	Rejection
Total weighted score	0.55*	0.44*	0.51*	0.24*	0.28*	0.19*
Rejection	0.20*	0.18*	0.16*	0.01	0.29*	—
Isolation	0.21*	0.25*	0.17*	0.03	—	—
JEPI neuroticism	0.01	0.11*	0.02	—	—	—
Newcastle antisocial subscale	0.89*	0.47*	—	—	—	—
Newcastle neurotic subscale	0.77*	—	—	—	—	—

\* Pearson correlation:  $P < 0.0001$ .

analyses. The scores which appeared to make the most important contribution to the total weighted score were the Rutter B2 Scale's total score and subscores, whereas the JEPI and sociometry made a lesser contribution. The rejection score, in particular, showed a reduction in comparison with that found at 12 years of age, a correlation 0.19 compared with 0.55 in the younger age group. Furthermore, the rejection score did not correlate well with the other screen measures.

Since fewer extreme scores were obtained in the adolescent sample (8% of the adolescents had a high rejection score compared with 14% of the 12-year-olds), the instrument cut-off was varied, but this did not substantially alter the scale's correlation with the total weighted score. This perhaps suggests that rejection, as measured by sociometry, is a less powerful social force in adolescence than at a younger age. However, there is one caveat, namely the high rate of absenteeism which complicates the interpretation of sociometry. This is examined elsewhere (Place & Kolvin, 1985) but, for example, if all the absentees had been present there would have been slightly fewer 'isolated' children and slightly more 'rejected' ones, simply because more choices were being made. However, we question whether this could totally account for the considerable differences in the rate of rejection found at the different ages. Nevertheless, absentees will be less salient in the minds of those making choices than will regular school attenders. It thus seems that sociometry is more difficult to interpret at this phase of development than it was at a younger age.

The neuroticism subscale of the Rutter B2 Scale correlated significantly with the neuroticism scale of the JEPI, but the correlation was low and its statistical significance again reflects the large number of cases studied. There was a moderate correlation between the neurotic subscale of the Rutter Scale and other items of the screen, but no similar correlation was found for the JEPI neuroticism scale. This is not surprising, as the Rutter neurotic subscale measures the neurotic behaviour of the school child and the early adolescent; the neuroticism dimension of the JEPI, on the other hand, has been designed to elicit the more internalized forms of neuroticism, as found in adults. This may well render the JEPI the more relevant instrument for screening an adolescent population for the neurotic traits.

#### Efficiency of the screen measures

If each of the screen measures makes an additional contribution to the multiple criterion score then, theoretically, this would prove to be a more efficient method of identifying disturbed youths (as validated by clinical interview) than any individual instrument. This was tested empirically by examining the individual instrument's ability to select potentially disturbed youths in the reconstituted population. All the measures showed an overall efficiency of around 75%, but the sensitivity of the instruments ranged from 17% for rejection to 32% for the Rutter B2 total score. The Kappa coefficients ranged from 0.15 to 0.29, confirming that the multiple criterion screen was a more reliable predictor of potential disturbance than any single criterion. When the sexes were considered separately the Rutter B2 Scale was more efficient at identifying potentially disturbed boys, and the neuroticism dimension of the JEPI was more efficient in girls.

#### Prevalence of disorder

Using the data on screen positive selections and the frequency with which this agreed with the psychiatric assessment of severity, it is possible to estimate the prevalence of disorder that would be found in the total sample. It must be remembered that 359 cases did not complete the self-report elements of the screening procedure, and thus the prediction of prevalence of disorder is necessarily confined to those youths who proved regular school attenders.

As was seen in Table 2, there were estimated to be 354 potentially disturbed adolescents among the regular school attenders, giving a rate of disorder of 25%. This corresponds closely with the rate of 24.8% found in the 12-year-old age group, and such consistency between age groups is compatible with the findings of other surveys of inner-city populations (Rutter *et al.* 1976). If one assumes that three-quarters of the chronic absentees would prove to be screen positive, the estimated prevalence of psychiatric disorder rises to 29%.

The period of adolescence is one in which the nature of psychiatric disorder changes. The high incidence of problems among males, which is seen in childhood, changes to the preponderance of problems in females, which is found in adult populations. If the present study sample is

Table 5. The prevalence of psychiatric disorder in boys and girls

Multiple criterion screen score	0	1	2	3	4+
Number of boys with screen score	460	122	72	31	60
Rate (%) of disturbance in interviewed sample	8	31	42	63	67
Estimated number of disturbed boys	37	37	30	20	40
Number of girls with screen score	451	111	53	32	54
Rate (%) of disturbance in interviewed sample	13	33	58	67	75
Estimated number of disturbed girls	59	37	31	22	41

Estimated rate of disorder in boys =  $\frac{147}{652} \times 100 = 22.0\%$ .

Estimated rate of disorder in girls =  $\frac{161}{593} \times 100 = 27.1\%$ .

broken down by sex (Table 5), it can be seen that the estimated rate of disorder in girls is greater than that in boys.

## DISCUSSION

### Efficiency of the screening procedure

In theory, the more searching the assessment, the more disorder one is likely to uncover. Richman (1980) points out that, for screening purposes, estimates of prevalence are improved if more than one source of information is available, a finding echoed by others (Rutter *et al.* 1970, 1975; MacMillan *et al.* 1980). This is recognized in clinical practice, where it is customary for psychiatric disorder to be diagnosed on the basis of information obtained from multiple sources, most particularly parental interview, child interview and, hopefully, information from the school. To use such multiple sources in a research project naturally enlarges the scale of even the most modestly conceived study. Consequently, the value of combining several measures must be closely scrutinized to ensure that it is efficient. Home-based research entails a great deal of effort, and returns are usually not as good as those obtained from school-based projects. For this reason, we were looking for a school-based screen that was economical and efficient in an adolescent population. There are problems, however, in using only teacher ratings. For instance, girls and boys may be perceived differently (Wilson, 1974), passive children with

problems which do not attract the attention of the teachers may be overlooked (Garner & Bing, 1973), and there may be individual differences in the teachers' perception of problem behaviour (Davies, 1976). In addition, teachers may not be sufficiently aware of the interpersonal difficulties experienced by some children, and these elements would thus be missed (MacMillan *et al.* 1980). The inclusion of self-rating scales and sociometry to complement teacher scales offers the opportunity to detect personal unhappiness and discomfort that adults might not have recognized. A multiple criterion screen made up of these elements proved commendably efficient in the early adolescent age group, with 78% of the screen positives exhibiting psychiatric disorder at interview (MacMillan *et al.* 1980). This compares well with the Inner-London Borough Study of 10-year-olds, in which 43% of screen positives exhibited psychiatric disorder (Rutter *et al.* 1975). In the present adolescent survey 69% of the screen positives were estimated to have psychiatric disorder. Another way of comparing the efficiency of such screens is by studying the Kappas: in the first study mentioned above the Kappa was 0.43; in the second it was 0.20, and in this study Kappa was 0.42. Unfortunately, we could not find data to allow a calculation of Kappas for the Isle of Wight Adolescent Study.

Varying the cut-off of the total weighted score obviously affects the screen's efficiency. If we take the higher score, that is a more extreme cut, we are sacrificing our chance of including a larger percentage of true positives in order to avoid an excess of false negatives. When using the lower cut-off, we are likely to detect more true positive cases but, at the same time, a lower percentage of youths identified by the screen will be true positives. Thus, in the adolescent age group, the multiple criterion screen has achieved similar rates of efficiency to those found at a younger age, but the number of true cases in the screen positive group is reduced. The merit of this type of screening which is school-based is the simplicity and economy of its administration and the high yield of cases for minimal professional effort. The high rate of absenteeism of adolescents in the inner-city may have reduced these benefits, although Gray *et al.* (1980) did not find high rates of maladjustment in youths who had not attended school regularly in their final year.

The situation specificity of behaviour (Mischel, 1968; Kolvin *et al.* 1977) points to the

inevitability of a teacher-completed screen missing children whose disturbance manifests predominantly outside the classroom. The Isle of Wight research workers (Rutter *et al.* 1975, 1976) have turned to a combination of home and school screens supplemented by data from the youths themselves using the Malaise Inventory, whereas the Newcastle group have attempted to allow for this by seeking information not only from the children themselves but also from their peers. Perhaps this screen would have been rendered more efficient by using information gathered from the home. To address this question we examined the Rutter A2 Parent Scale, which provided information about the youths who were psychiatrically interviewed. While this is a screening instrument and is not intended for diagnostic purposes, it allowed a further comparison to be made. Using the formula for describing potential cases suggested by Rutter and his colleagues (1970), half the cases diagnosed as disturbed at psychiatric interview scored below the Rutter A2 cut-off of 13 or more, and only 3 cases, who were not rated as disturbed, scored above this cut-off. Our impression was that this information gave rise to only small changes in diagnostic categorization which indicated that the school-based measures had been effective in detecting potential psychopathology. This is not a totally novel finding: Berg & Fielding (1979) reported that their study could show 'no relationship between child interview and parental or teacher assessments of psychiatric disorder', and Rutter *et al.* (1976) concluded that 'many youngsters who appear normal to parents and teachers are diagnosed as showing disorder on the adolescent interview'. These findings seem to point to the primacy of the extended interview with the adolescent. However, there is one caveat: our figures are estimates based on a relatively small number of interviews in relation to a relatively large study population, and so there remains a margin for error.

#### Relationship between screen measures

This relationship has been studied previously in 12-year-olds and so can be compared with the current findings. On this occasion, the correlation of the total weighted score with its constituent measures proved to be lower than that found at a younger age range, which suggests a greater

heterogeneity of psychiatric disturbance in adolescence. The pattern of intercorrelation of the individual measures is broadly similar at both age ranges except that, at the lower age, the two Rutter teacher subscales show a much lower correlation (0.21 at 12 years and 0.47 at 15 years), suggesting less in the way of independence of these measures in adolescence. The pattern of non-significant correlations of the JEPI's dimension of neuroticism with the other measures in an adolescent population suggest that it may not be valid for the identification of disturbance; nevertheless, it has been shown to be more specifically helpful in identifying disturbed girls. This interpretation is consistent with the conclusion previously reached by MacMillan *et al.* (1980), that the JEPI appears to be a valuable adjunct to peer and teacher measures.

Again, there is a significant but not high correspondence between teacher ratings of disturbance and peer generated sociometry, which suggests that teachers are conscious of the importance of social acceptability in adjustment (Semler, 1960; MacMillan *et al.* 1980) and, to a certain extent, reflect situation-specificity. On the other hand, self-ratings are less tied to a particular setting and such measures may therefore make a unique contribution.

#### Prevalence

Previous epidemiological research in Australia (Krupinski *et al.* 1967) and Blackburn, UK (Leslie, 1974) has suggested that the rate of psychiatric disorder in adolescence was about 16-17%. Rutter *et al.* (1976), in their Isle of Wight study, demonstrate that prevalence rates are dependent on how searching the instruments are: for instance, when the source of information is parental interview, the rate is 13.2% for males and 12.5% for females; when based on direct interview with the adolescent himself, an overall rate of 16.3% is obtained. However, the school screening procedures apparently miss quite a few children with psychiatric disorder and, if correction is made for these, the rate rises to 21%. This is based on data from the Isle of Wight, and the same group of workers has demonstrated that the rate of psychiatric disorder (based on parental interviews) in 10-year-olds was twice as high in an Inner-London Borough as on the Isle of Wight. A similar increase in prevalence is reported to exist in the



adolescent population of the Inner-London Boroughs (Rutter, 1980). Lavik (1977) also reports considerable urban-rural differences in the prevalence of adolescent disorder. The rate for 16-year-old boys is 23% and that for girls is 16%, which gives an overall prevalence for urban Oslo of 19%. The rates for rural areas are much lower. Elsewhere, it is argued (MacMillan *et al.* 1980) that the industrialized urban area of Newcastle is similar to the Inner-London Boroughs and also that a remarkable correspondence of prevalence rates of around 25% for 10- to 12-year-old children exists. In mid-adolescence we now report similar urban rates, with an estimated prevalence of 22% for boys and of 27% for girls, giving an overall estimated prevalence of 25% for the regular school population.

Poor school attendance has been reported to be most common in the fifth and final year of compulsory schooling (Galloway, 1976; Rutter *et al.* 1979). Further, Rutter and colleagues report considerable differences between schools ranging from over 85% attendance to about 65% in the 12 schools they studied. These latter workers also report that it was the pupils of low average intellectual ability or from families of low occupational status who were most likely to be poor attenders. They also found that, when schools were ranked according to pupil attendance, pupil behaviour and exam success, these rankings proved to be fairly highly correlated (from 0.68 to 0.77). They also noted a sharp drop in school attendance in the last school year, the stage when national exams loom high for many children. The rates of chronic absenteeism in our schools was 20%, and allowance needs to be made for this subgroup when estimating prevalence.

It is also interesting to note the higher rate of disturbance which we found in adolescent girls, a finding at odds with the equal sex ratio found in the Isle of Wight 14-year-olds (Graham & Rutter, 1973). This may be a valid phenomenon, or may help to confirm the hypothesis that there are sex differences in the prevalence of disorder in the inner-city areas of the UK, irrespective of age. Previous research by Rutter and his colleagues (1975) supports this latter hypothesis.

## CONCLUSIONS

Several tentative conclusions may be drawn from these findings. First, a degree of chronic absenteeism was found among the 4th- and 5th-year pupils which had not been present in the younger age group, and this obviously complicates analysis and interpretation of the data. Secondly, the rate of disorder in a mid-adolescent population of Tyneside is in the range of 25-29%, which is only slightly higher than the rate found in the early adolescent population (24.8%), giving rise to the conclusion that evident distress and psychiatric disorder do not substantially increase over the course of adolescence. However, the suggestion that adolescent distress may simply be hidden from the adults with whom they come into contact (Rutter *et al.* 1976) is given support from our sex-specific data, in which girls are more likely to be identified by self-report measures, and may be missed if teacher-completed questionnaires are the only ones used. Thirdly, a multiple criterion screen with a weighting system has been shown to be more specific, but less sensitive and with a similar overall efficiency when used in mid-adolescence as compared with the younger age. As previously reported (Tarnopolsky *et al.* 1979; MacMillan *et al.* 1980), the cut-off used in a screening procedure must be adjusted to obtain maximum accuracy in case identification in any given population. It does seem, however, that information provided by the youth during a psychiatric interview is more comprehensive and extensive than is available from screen data or interview with parents. In particular, it seems to facilitate the emergence of neurotic symptomatology, a finding not altogether surprising in view of the poor overlap between the self-reporting of neuroticism and information from the school.

One further question is whether a multiple criterion screen is more efficient than a single criterion method. The available evidence suggests that the overall efficiency of the multiple criterion screen is greater than any of the single criterion screen measures. There are variations in relation to the individual scales, and questionnaires designed for younger adolescents may be less helpful in identifying disturbance when used in a mid-adolescent population. When the sexes are considered separately, it is evident that the Rutter B2 Scale tends to be more efficient in

detecting potentially disturbed boys and the JEPI neuroticism measure more efficient in detecting potentially disturbed girls. Thus, instruments for screening purposes need to be chosen with careful reference to the aims and objectives if the population is an adolescent one but, together with other authors (Berg & Fielding, 1979), we believe that the psychiatric interview is the most valid way of collecting and categorizing psychiatric data.

We would like to acknowledge the valuable advice and constructive criticism that was given during this study by the late, and much missed, R. F. Garside. The research was funded by grants from the Department of Education and Science and the Newcastle Health Authority.

## REFERENCES

- Berg, I. & Fielding, D. (1979). An interview with a child to assess psychiatric disturbance: a note on its reliability and validity. *Journal of Abnormal Child Psychology* 7, 83-89.
- Bower, E. (1969). *Early Identification of Emotionally Handicapped Children in School*. C. C. Thomas: Springfield, Ill.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37-46.
- Davies, B. (1976). Piggies in the middle or 'Who, Sir? No, not me, Sir'. In *The Disruptive Pupil in the Secondary School* (ed. C. Jones-Davies and R. G. Cave), pp. 26-34. Ward Lock Educational: London.
- Douvan, E. & Adelson, J. (1966). *The Adolescent Experience*. Wiley: New York.
- Eissler, K. R. (1958). Notes on problems of technique in the psychoanalytic treatment of adolescents. *Psychoanalytical Study of the Child* 13, 223-254.
- Eysenck, S. B. G. (1965). *Manual of the Junior Eysenck Personality Inventory*. University of London Press: London.
- Galloway, D. (1976). Persistent unjustified absence from school. *Trends in Education* 30, 22-27.
- Garner, J. & Bing, M. (1973). The elusiveness of Pygmalion and differences in teacher-pupil contracts. *Interchange* 4, 34-42.
- Goldberg, D. P., Cooper, B., Eastwood, M. R., Kedward, H. B. & Shepherd, M. (1970). A standardised psychiatric interview for use in community surveys. *British Journal of Preventive and Social Medicine* 24, 18-23.
- Graham, P. & Rutter, M. (1973). Psychiatric disorder in the young adolescent: a follow-up study. *Proceedings of the Royal Society of Medicine* 66, 1226-1229.
- Gray, G., Smith, A. & Rutter, M. (1980). School attendance and the first year of employment. In *Out of School: Modern Perspectives on Truancy and School Refusal* (ed. L. Hersov and I. Berg), pp. 343-370. Wiley: Chichester.
- Hall, G. S. (1904). *Adolescence; its Psychology and its Relations to Physiology, Anthropology, Sociology, Sex, Crime, Religion and Education*. Appleton: New York.
- Kolvin, I., Garside, R. F., Nicol, A. R., Leitch, I. M. & MacMillan, A. (1977). Screening schoolchildren for high risk of emotional and educational disorder. *British Journal of Psychiatry* 131, 192-206.
- Krupinski, J., Baikie, A. G., Stoller, A., Graves, J., O'Day, D. M. & Polke, P. (1967). A community mental health survey of Heyfield, Victoria. *Medical Journal of Australia* i, 1204-1211.
- Lavik, N. J. (1977). Urban-rural differences in rates of disorder. A comparative psychiatric population study of Norwegian adolescents. In *Epidemiological Approaches in Child Psychiatry* (ed. P. J. Graham), pp. 223-251. Academic Press: London.
- Leslie, S. A. (1974). Psychiatric disorder in the young adolescents of an industrial town. *British Journal of Psychiatry* 125, 113-124.
- MacMillan, A., Walker, L., Garside, R. F., Kolvin, I., Leitch, I. M. & Nicol, A. R. (1978). The development and application of sociometric techniques for the identification of isolated and rejected children. *Journal of the Association of Workers for Maladjusted Children* 6, 58-74.
- MacMillan, A., Kolvin, I., Garside, R. F., Nicol, A. R. & Leitch, I. M. (1980). A multiple criterion screen for identifying secondary schoolchildren with psychiatric disorder. *Psychological Medicine* 10, 265-276.
- Mischel, W. (1968). *Personality and Assessment*. Wiley: London.
- Neligan, G. A., Prudham, D. & Steiner, H. (1975). *The Formative Years: Birth, Family and Development in Newcastle upon Tyne*. Oxford University Press: London.
- Neligan, G. A., Kolvin, I., Scott, D. McI. & Garside, R. F. (1976). *Born Too Soon or Born Too Small*. Spastics International Medical Publications/Heinemann Medical Books: London.
- Offer, D. (1969). *The Psychological World of the Teenager*. Basic Books: New York.
- Place, M. & Kolvin, I. (1985). Adolescents who are frequently absent from school. (In preparation.)
- Richman, N. (1980). Epidemiological studies of children: problems of screening. In *Studies of Children* (ed. F. Earls), pp. 117-125. Prodist: New York.
- Rutter, M. (1977). A children's behaviour questionnaire for completion by teachers: preliminary findings. *Journal of Child Psychology and Psychiatry* 8, 1-11.
- Rutter, M. (1977). Surveys to answer questions: some methodological considerations. In *Epidemiological Approaches in Child Psychiatry* (ed. P. J. Graham), pp. 1-30. Academic Press: London.
- Rutter, M. (1980). *Changing Youth in a Changing Society*. Harvard University Press: Cambridge, Mass.
- Rutter, M. & Graham, P. (1968). The reliability and validity of the psychiatric assessment of the child. 1. Interview with the child. *British Journal of Psychiatry* 114, 563.
- Rutter, M., Tizard, J. & Whitmore, K. (1970). *Education, Health and Behaviour*. Longman: London.
- Rutter, M., Cox, A., Tupling, C., Berger, M. & Yule, W. (1975). Attainment and adjustment in two geographical areas. 1. The prevalence of psychiatric disorder. *British Journal of Psychiatry* 126, 493-509.
- Rutter, M., Graham, P., Chadwick, O. F. D. & Yule, W. (1976). Adolescent turmoil: fact or fiction? *Journal of Child Psychology and Psychiatry* 17, 35-56.
- Rutter, M., Maughan, B., Mortimore, P. & Ouston, J. (1979). *Fifteen Thousand Hours*. Open Books: London.
- Semler, I. J. (1960). Relationships among several measures of pupil adjustment. *Journal of Educational Psychology* 51, 60-64.
- Spivack, G., Haines, P. E. & Spotts, J. (1967). *Devereux Adolescent Behaviour Rating Scale Manual*. The Devereux Foundation: Penn.
- Tarnopolsky, A., Hand, D. J., McLean, E. K., Roberts, H. & Wiggins, R. D. (1979). Validity and uses of a screening questionnaire (GHQ) in the community. *British Journal of Psychiatry* 134, 508-515.
- Wilson, J. A. (1974). Adjustment in the classroom: 2. Patterns of adaptation. *Research in Education* 11, 17-29.
- Wrate, R., Kolvin, I., Garside, R. F., Wolstenholme, F., Hulbert, C. & Leitch, I. M. (1985). Helping seriously disturbed children. In *Longitudinal Studies in Child Care and Child Psychiatry - Practical Lessons from Research Experiences* (ed. A. R. Nicol). Wiley: Chichester (in the press).