Screening schoolchildren for high risk of emotional and educational disorder.
I Kolvin, R F Garside, A R Nicol, I Leitch and A Macmillan
Access the most recent version at doi: 10.1192/bjp.131.2.192

To obtain reprints or permission to reproduce material from this paper, please write to permissions@rcpsych.ac.uk

http://bjp.rcpsych.org/cgi/eletter-submit/131/2/192

http://bjp.rcpsych.org/ on September 5, 2011
Published by The Royal College of Psychiatrists

To subscribe to The British Journal of Psychiatry go to:
http://bjp.rcpsych.org/site/subscriptions/
Screening Schoolchildren for High Risk of Emotional and Educational Disorder

By I. KOLVIN, R. F. GARSIDE, A. R. NICOL, I. LEITCH and A. MACMILLAN

Junior schoolchildren were screened for high risk of emotional and educational disorder. The information was gathered from the school, using standard and objective tests. The multiple criterion screen employed, comprised: (a) classroom behaviour scale (Rutter B Scale), rated by teachers; (b) sociometric tests; choice of companions by classmates. From these, lack of positive choices was taken as a measure of isolation and a high rate of negative choices, as a measure of rejection; (c) Reading quotient of 75 or below on the Young Reading Test; (d) Absence from school for reasons considered by teachers to be trivial.

The number of cases identified by the screening was found to be 322 children per thousand. They may be seen as representing a high risk rate.

Using extreme scores as indicative of high risk, approximately 17 per cent of children were identified on the basis of the behaviour test; 12 per cent by the reading test; 9 per cent by the isolation test; 8 per cent by the rejection test and 3 per cent by the absenteeism test. Seventy per cent of the identified children were rated clinically as disturbed. Absenteeism identified the smallest percentage of cases and made the smallest independent contribution to identification. Isolation was not impressively related to neurotic or antisocial behaviour. The three important criteria, therefore, were behaviour, rejection and reading.

Corrected disturbance rates for our population of 7-8-year-old children, including those not identified by the screen, were 6-8 per cent markedly maladjusted and 33.7 per cent somewhat maladjusted.

Introduction

Although the home has traditionally been recognized as the most potent influence in child development, the impact of the school has recently attracted increased interest (Rutter et al., 1975; Power, Benn and Morris, 1972; Reynolds et al., 1976). This has added force to the argument that children are not necessarily 'born to fail' (Wedge, 1974), and that great advantages may accrue from helping children with special needs in ordinary school. This may include children who are at risk of developing psychiatric disorder, problems of social relationships, learning and other difficulties connected with school. This paper will present the results of a procedure for identifying such children as part of a research project aimed at evaluating the effectiveness of provision for children with special needs. First however, it is necessary to clarify some of the concepts involved in such a venture.

The currently accepted method of boosting educational provision to needy children is through the Educational Priority Area Concept.
(Plowden, 1967). In this, deprived neighbour-
hoods are allocated extra resources and staff
to deal with their problems. However, it is
uncertain whether the characteristics of a
neighbourhood alone provide a valid criterion
on which to base a prescription for individual
children: not all children in deprived neigh-
bourhoods are equally deprived, nor would the
most needy children best be able to benefit from
the extra help. Further, not all children in
need would be living in such neighbourhoods,
and some, therefore, would not receive help.
The assumption that a wide variety of problems
would respond to a single form of help is also
debatable. For these reasons, as well as for
diagnostic purposes, information was obtained
about individual children in need by gathering
information from teachers, from peers and from
the children themselves.

In attempting to provide special services on a
population-wide basis, there is the problem of
how to define the group in need. It was felt that
at the age of the present sample, seven and
eight years old, it would be unduly restrictive to
to consider only those with established handicaps,
whether emotional or educational in nature.
Caplan (1964) has described the importance of
early diagnosis of psychological disorders, at a
time when they are just detectable, as a prelude to
'secondary prevention'. This has commonly
involved the selection of individuals who
obtain statistically deviant scores on question-
naires to the teacher and parent (Shepherd
et al., 1971; Rutter et al., 1970) teacher and pupil
(Hallworth and Morrison, 1964; Roff et al., 1972)
or teacher, peer and self (Bower, 1969; Zax and
Cowen, 1969; Cowen et al., 1971). Rogers (1942)
used a 9-criterion screen.

There is plenty of evidence to suggest that
such deviance has predictive utility. For
example, ratings by teachers have often proved
to be predictive of later delinquency (Mulligan
et al., 1963; Khrief, 1964), and so have sociome-
tric indices (Harper, 1965; Skaberne et al.,
1965). In a recent longitudinal study of 4,600
children, Roff, Sells and Golden (1972) found
that middle class delinquents had previously
been rejected by their peers, while working
class delinquents may have been either rejected
or popular. Sociometric ratings of disorderly

behaviour in primary school have been found
to correlate 0.50 with secondary school teachers'
later ratings of indiscipline (Gibson and Hansen,
1969) and may be more predictive of high
academic achievement or withdrawal from
school than self-rating personality tests (Ullman,
1957). Inadequate personal adjustment may also
be foreshadowed by earlier peer difficulties.
Thus, unpopular children are more likely to be
community-wide psychiatric register (Cowen
disproportionately represented later in life in a
et al., 1973). Among the indicators identified in
the Austin Longitudinal Research Project (Currie
et al., 1974) as precursors of personal adjustment
some nine years later, rejection by peers in the
first grade has been one of the most useful.

It appears, therefore, from previous work
that we can have some confidence that a 'high
risk' group can be identified. The prospect also
seems sound on a theoretical basis, for we are
concerned here with disorders which are
quantitative deviations from normal (Goldberg,
1972; Kolvin, Wolff et al., 1975). Qualitative
deviations such as childhood psychosis, are rare
in ordinary school and are not the subject of
this study. There is thus no clear dichotomy
between disorder and non-disorder, and the
problem is one of deciding the optimum cut-off
point. Indeed, the above quantitative distinc-
tion needs to be validated clinically. Minor
deviations may be expected, in some cases, to
progress to more severe and disabling deviations
over time. The marginally deviant individual
can thus be described as 'at risk' of developing
psychiatric or educational disorder.

So much for the rationale for screening and
the 'at risk' concept. We may expect from this
reasoning to discover three categories of
children from deviant scores on a screen
instrument:

(i) Those who at initial assessment are
already experiencing clear psychiatric inter-
personal or educational difficulties;

(ii) Those who will subsequently develop such
difficulties (who may be termed those who
will 'take');

(iii) Those who are neither experiencing, nor are
destined to develop any type of difficulty within
the defined period.
There can be some confidence that those cases which are shown to have a clinical disorder are at risk for persistence or progression. The problem is with the remaining high scorers. These seem likely to be at greater risk of developing psychiatric, social or educational disorder than low scorers. While skilled clinical evaluation may help to improve prediction, there will still remain a considerable uncertainty about which other children will develop disorders during the defined period. For this reason, differentiation between those children classified as 'at risk' who are destined to become handicapped and those who are misclassified because of the inaccuracy of our screening instruments or because they have mild or transient variations will have to await the results of follow-up studies.

The same problem of differentiation exists for those children who fall below the cut-off point. It is not possible to tell until follow-up which, and what proportion of such children are destined to develop difficulties, although the studies mentioned above suggest that the proportion is likely to be lower than for the screen-positive group. As some children—whether currently handicapped or destined to become so—are likely to recover over a defined period, we must be concerned with the period prevalence rate at either screen or follow-up.

The present paper will focus on the interrelationships between screen criteria and will test the validity of the screen in distinguishing cases that have established psychiatric disorder from those that have not (i.e. it will differentiate (i) in the previous paragraph from (ii) and (iii)). The differentiation of (ii) and (iii) will become possible when follow-up data has been collected and will be the subject of future reports.

**Method**

(a) *The population*

In our main study we screened 515 children who first entered junior school in the academic year beginning September 1973. The sample had a mean age of 7 years 9 months and attended six junior schools. There were almost equal numbers of boys and girls (52 and 48 per cent respectively). The schools chosen were broadly representative of the schools in this area; the social class distribution of the area is slightly below the national average (Neligan et al., 1974). The normal controls were subsequently selected at random from the pool of children not identified by any screen criterion, with the proviso that they should be drawn from all six schools and should reflect the sex ratio in the 'at risk' group, which was found to be 3 boys to 2 girls.

(b) *Designing the multiple criterion screen*

A child must make three major adjustments at school. First, there are the formal educational and academic demands (where failure leads to educational handicap); second, there are the demands of emotional adjustment and behavioural control (where failure leads to emotional handicap); and third, merging with both the others, there are the demands of social relationships with peers. We proposed that signs of failure in any of these areas would constitute early evidence of developing handicap, and that failings in one area could have repercussions in either of the others. We felt, therefore, that a single screen measure was inadequate for our purposes, and that the multi-screen model of Bower (1960–69) was more appropriate. We scanned the literature for screen measures which could be applied in schools, and which were likely to be reliable, valid and reasonably efficient predictors of disorder.

In an initial pilot study we used five screen criteria in three junior classes involving 90 children. In the event, all these were later used in the main study. They include:

1. *Sociometry*: Each child in a class was asked to complete a form which asked him to choose which three of his class-mates he would like to sit beside in class and play with at play-time; and conversely, which he would not like to sit beside or play with. This procedure yielded two scores: *isolation*, which is defined as a lack of positive choices; and *rejection*, which is defined as receipt of a large number of negative choices. In a class of approximately 30 children, the child was considered to be isolated if he scored 0 or 1 positive choice and to be rejected if he
received 14 or more negative choices. Each of these cut-off scores was intended to yield about 6 per cent of the population, judged by the pilot study data, but in practice we found that the yield was higher. The sociometric criteria are described more fully elsewhere (Macmillan et al., 1976).

2. Reading: In the Isle of Wight study (Rutter et al., 1970), the definition of educational backwardness was a reading accuracy or comprehension 20 months below the child's chronological age, and this included 7.9 per cent of the population. Some of the children in our study (i.e. those who were just seven years old) were too young for a similar definition to be used with any degree of confidence, so we selected a cut-off of a reading quotient (RQ) of 75 or less on the Young (1968) Reading Test. This meant that, at this age, the selected children were non-readers on this test. We appreciated that this would produce a higher yield than the Isle of Wight Study, but we considered this to be justified in view of the greater degree of economic deprivation in the North East. In fact it leads to an inclusion of 12.2 per cent of the population. However, it is to be noted that due to illness or absenteeism we only managed to administer the reading test to 512 of the children during the screening period.

3. Behaviour—The Rutter Teacher Scale B. Rutter (1967) and Rutter et al. (1970) found that a cut-off of 9 or more had discriminative value. It selected about 10 per cent of the boys and 4 per cent of the girls, averages 7 per cent. This cut-off produced nearly 30 per cent of children in our pilot study and over 20 per cent of the children studied overall. We decided therefore, to use a slightly more rigorous criterion, and raised the cut-off to 10. At this level 17 per cent of the population were included, more than double the rate reported by Rutter. The number of cases and the percentages selected is set out in Table I and Fig 1.

4. Absenteeism: The Rutter Teacher Scale B contains the item 'Tends to be absent from school for trivial reasons'. We decided that any child who qualified for a clear-cut affirmative answer to this question should be included in the 'at risk' category. We found that 90.5 per cent of the children were rated as 'does not apply', 6.9 per cent were rated as 'applies somewhat', and 2.5 per cent of the children were rated as 'certainly applies'.

(c) Reliability

The Rutter Teacher Scale B has already been shown to have an adequate reliability of 0.72 between raters and a retest reliability, using the same teacher, of 0.89. The reliability of the absentee item of the Teacher Scale is likely to be less than this, as on any scale the constituent items are usually less reliable than the score based on a summation of the items, but reliability is assumed because of the continued inclusion of the item in the scale. The Young Reading Test has been shown to have a reliability of 0.95 (1968). For Sociometric criteria, we found that the test-retest reliability with a one month gap was 0.88 for rejection and 0.63 for isolation.

Table I

<p>| Proportion (per cent) of cases identified by each criterion selected by that criterion only |
|---------------------------------|--------------------------------|----------------|----------------|----------------|</p>
<table>
<thead>
<tr>
<th>Selected by</th>
<th>Behaviour criterion</th>
<th>Reading criterion</th>
<th>Rejection criterion</th>
<th>Isolation criterion</th>
<th>Absent criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific percentage</td>
<td>...</td>
<td>42</td>
<td>36</td>
<td>24</td>
<td>53</td>
</tr>
<tr>
<td>Overlap percentage</td>
<td>...</td>
<td>58</td>
<td>64</td>
<td>76</td>
<td>47</td>
</tr>
<tr>
<td>No. cases identified (= 100%)</td>
<td>...</td>
<td>88</td>
<td>63</td>
<td>40</td>
<td>48</td>
</tr>
<tr>
<td>Total cases—515</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig 1.—Percentage of total population identified by each screen criterion. Shaded area represents percentage selected only by that criterion.

Findings

1. Results of the screening procedures

(a) Percentage of the population identified by each criterion

As Fig 1 shows, the multiple criterion screen yielded about a third of the total population. The percentage of cases identified by each screen criterion was 17.1 per cent behaviour; 12.2 per cent reading; 9.3 per cent isolation; 7.8 per cent rejection; and 2.5 per cent absenteeism. The overall yield was 32.2 per cent, but the yield varied from school to school, the lowest being approximately 20 per cent and the highest 50 per cent. As one might expect, the highest yield was in the poorest neighbourhood.

Our multiple screen thus has the potential of identifying not only individuals at risk but also those schools and neighbourhoods with a high concentration of problems. It is therefore a powerful and economic instrument which could be applied more widely.

(b) Overlap of cases selected by the different criteria

Inevitably, some cases will be selected by a single criterion and others by more than one: this overlap can be indicated in several ways. Fig 1 and Table I show that 42 per cent of the cases selected by the behaviour criterion are picked out by that criterion alone; the corresponding percentage for isolation is 53 per cent, for reading is 36 per cent, for rejection is 24 per cent and for absenteeism is only 11 per cent.

A second way to consider this is simply in terms of the number of criteria by which each case is selected (Table II). Most are selected by one criterion, less than half of that by two criteria, and half again by three or more criteria. While the highest percentage of cases is selected by the behaviour criterion, and the lowest by the absenteeism criterion, nearly all the cases selected by the latter are selected by other criteria as well. As absenteeism selects so few cases, and as its independent contribution is so small, we have excluded it from most of the subsequent analyses. The behaviour criterion not only makes the greatest contribution but makes a high independent contribution as well. The other three criteria select a roughly equal number of cases, but isolation makes the highest independent contribution.

A third way of studying overlap is to examine the joint contribution of any two criteria, such
as behaviour and reading, to see which two have the best yield. As one might expect, behaviour combined with any other criterion selects approximately 70 per cent of the cases identified (Table III). Rejection combined with isolation selects the lowest percentage of those identified.

The extent of agreement, or concordance, between screen criteria also can be measured by using a statistic, kappa (Cohen, 1960; Kendell, 1975). In simple terms kappa constitutes a form of correlation coefficient. The kappas in Table IV are statistically significantly larger than zero, but all are small. This suggests that all of the four measures thus analysed are making a useful contribution to the variety of expressions of being at risk, or potential maladjustment. The highest kappas are between behaviour and reading, and between behaviour and rejection; the lowest include the isolation criterion.

The Teacher Scales can provide two subscores as well as the total score presented above (Rutter et al., 1970). It was considered that the inclusion of additional items would maximize the stability of the subscales and hence, from an examination of Rutter's own data, we selected additional items which appeared to have discriminative value. Thus the items for the antisocial subscale were A, B, D, E, O, P, S, T and Z, and for the neuroticism subscale were G, H, J, K, N, Q, R, V and W, in each case on Scale B.

The salient findings (Table V) are that on the total behaviour scale the group identified by the isolation criterion obtains a low mean score. It is also evident that, compared to their neuroticism means, the reading and rejection groups have high antisocial scores. The absentee group have the highest mean neurotic score, but only a relatively high antisocial score. Isolation is the one characteristic therefore which is not impressively related to neurotic or antisocial behaviour as rated by the teachers.

2. Clinical ratings of disturbance

Effectiveness of the screen measures—any criterion

Rutter and colleagues (1970) based their analysis on a comparison of true and false positives as determined by a clinical judgement based on interviews with the parents and children as well as additional information from the teachers and the screen data. In the present study, the parents only were interviewed by a social worker; this was because interviews with children themselves, even at the age of 11 years, have been found to yield little additional

---

**Table II**

<table>
<thead>
<tr>
<th>Number of criteria</th>
<th>Percentage of identified cases (N = 166)</th>
<th>Percentage of total population (N = 515)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One criterion</td>
<td>.62</td>
<td>20</td>
</tr>
<tr>
<td>Two criteria</td>
<td>.25</td>
<td>8</td>
</tr>
<tr>
<td>Three or more</td>
<td>.13</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>.100</td>
<td>32</td>
</tr>
</tbody>
</table>

---

**Table IV**

<table>
<thead>
<tr>
<th>Measure of agreement (Kappa) between screen criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behaviour</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

* = significant at 5 per cent level
** = significant at 1 per cent level
*** = significant at 0.1 per cent level

---

**Table III**

<table>
<thead>
<tr>
<th>Screen criteria</th>
<th>Rutter</th>
<th>behaviour</th>
<th>Reading</th>
<th>Rejection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>.79</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Rejection</td>
<td>.79</td>
<td>54</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Isolation</td>
<td>.76</td>
<td>54</td>
<td>48</td>
<td>—</td>
</tr>
</tbody>
</table>

Total identified cases—166

As the Teacher Scale identifies both the greatest overall number of cases and the greatest overlap with other scales, it is of interest to compare the score on this scale for those children selected as deviant on each of the other scales.
information. From the 166 cases identified, 139 were used for this part of the study as we did not have the resources to interview more. The remainder were made up of cases that were omitted on a random basis, uncooperative families, families who had moved away and families we were unable to contact. In addition, further data were gathered from the teacher, using the Devereux Elementary behaviour rating scale (Spivack and Swift, 1967). Using behavioural data available from all these sources in addition to the screen data an assessment of diagnosis and severity of disorder was made by one of the psychiatrists in the research team (I.K. or A.R.N.). The procedure has been found to be of satisfactory reliability (Wrate et al, 1977). Following assessment, three groups emerged:

(a) Those identified by the screen and who at global clinical assessment proved to be moderately (somewhat) or markedly maladjusted. These can be considered to be true positives (98 cases).

(b) Those identified by the screen and who at global clinical assessment showed only slight evidence of disturbance, or none. Some might call these false positives, but, as we have argued in the Introduction, they may hypothetically be considered to be a high risk group with a greater likelihood of 'taking' than have the group of children not identified by the screen (41 cases).

(c) A random sample of the residual children not identified by the screen (53 cases). Of these, those who at global clinical assessment were rated as being moderately or markedly disturbed must be considered to be false negatives.

At the time of the assessment the clinician had no knowledge of which of these groups the child was in, nor could he guess, as only behavioural data were available and the child could have been selected on educational grounds.

Table VIa shows that, while 4 per cent of the controls proved to be markedly disturbed, three quarters fell into the 'not disturbed' category. On the other hand, only about 30 per cent of the 'at risk' group proved not to be disturbed. In interpreting these findings, it should be remembered that 14 per cent of the screen positive group were identified as educationally retarded only, and that this might tend to dilute the discriminating power of the screen in identifying maladjustment. The lower part of the Table indicates the discriminating power of each screen measure separately. It can be seen that the smallest percentages of false negatives are gained from the Rutter Teacher Scale, the Sociometry rejection and the absenteeism scales. However, the last are so few in number that they contribute little to the screen. All criterion groups generate significantly more cases than does the control group.

Table VIb indicates the proportion of screen-positive cases which were used for further study.
Note: (1) Some cases selected by more than one criterion
(2) All criterion group significantly different from controls at $P < .001$

Table VIa
Global clinical assessments, per cent of total cases

<table>
<thead>
<tr>
<th>Controls</th>
<th>Not disturbed</th>
<th>Some-what</th>
<th>Marked</th>
<th>Total cases (= 100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>73</td>
<td>23</td>
<td>4</td>
<td>53</td>
</tr>
<tr>
<td>At risk total</td>
<td>29</td>
<td>58</td>
<td>13</td>
<td>139</td>
</tr>
<tr>
<td>Rutter B scale</td>
<td>20</td>
<td>66</td>
<td>14</td>
<td>84</td>
</tr>
<tr>
<td>Reading quotient</td>
<td>34</td>
<td>47</td>
<td>19</td>
<td>47</td>
</tr>
<tr>
<td>Sociometry</td>
<td>Isolation</td>
<td>32</td>
<td>62</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Rejection</td>
<td>21</td>
<td>71</td>
<td>8</td>
</tr>
<tr>
<td>Absenteeism</td>
<td></td>
<td>11</td>
<td>66</td>
<td>22</td>
</tr>
</tbody>
</table>

Note: (1) Some cases selected by more than one criterion
(2) All criterion group significantly different from controls at $P < .001$

Table VIb
Proportion of identified cases which were used for further study

<table>
<thead>
<tr>
<th>Rutter Teacher scale</th>
<th>Identified</th>
<th>Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>(scoring 10+)</td>
<td>88</td>
<td>84 (95%)</td>
</tr>
<tr>
<td>Reading quotient</td>
<td>63</td>
<td>47 (75%)</td>
</tr>
<tr>
<td>Isolation</td>
<td>48</td>
<td>34 (71%)</td>
</tr>
<tr>
<td>Rejection</td>
<td>40</td>
<td>38 (95%)</td>
</tr>
<tr>
<td>Absenteeism</td>
<td>13</td>
<td>9 (69%)</td>
</tr>
</tbody>
</table>

It can be seen that, by chance, the proportional representation of the different criteria is somewhat variable.

Table VII re-analyses the data according to the number of screen criteria by which a particular child is identified, and shows that the rate of clinically perceived moderate maladjustment increases progressively with the number of criteria by which the child is identified. For instance, 52 per cent of those identified by one criterion, 63 per cent of those identified by two criteria, and 73 per cent of those identified by three criteria were shown subsequently on clinical assessment to be at least moderately maladjusted.

It is also possible to estimate the disturbance rate for the total population we have screened by extrapolating from the random sample of screen-negative cases who were fully assessed and considering this in conjunction with the screen-positive cases. The result of this calculation is shown in Table VIII. From this we can calculate that 63 per cent of those assessed clinically as markedly disturbed, 54 per cent of those as moderately disturbed but only 16 per cent of those assessed as normal were identified by the screen.

The clinical assessment of maladjustment, as described above, can only be of very limited value in estimating validity. The reason for this is that the measure is grossly contaminated with those measures for which independent validation is sought. An independent source of information is from the parents, and as part of the study the parents completed questionnaires, the Rutter A Scale (Rutter, 1970), about their children's behaviour in the screen-positive group and the random control group. It was found that the means of the total behaviour scores of the criterion groups were all higher than those of controls and significant for all except the absentee and isolation group (Table IX). These patterns were repeated for the means of the behaviour antisocial scores. However, for the neurotic scores, only the mean of the behaviour group differed significantly from the controls, despite the fact that the means of the reading, isolation and absentee groups were higher than that of the behaviour group (this lack of significance is probably related to the size of the samples and also the standard deviation). Such findings are particularly impressive in view of the well-known situation specificity of disturbed behaviour (Mischel, 1968; Rutter et al, 1970). Such specificity emphasizes the importance of the family and social environment, both in the genesis and in the manifestation of disturbed behaviour.
TABLE VIII
Comparison of screen and clinical assessment of maladjustment

<table>
<thead>
<tr>
<th>Screen assessment</th>
<th>Clinical assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Marked maladjustment</td>
</tr>
<tr>
<td>Identified by screen</td>
<td>4.3%</td>
</tr>
<tr>
<td>Not identified by screen</td>
<td>2.5%</td>
</tr>
<tr>
<td>Total</td>
<td>6.8%</td>
</tr>
</tbody>
</table>

Note—Percentages were calculated for the total sample of 515 from a weighted combination of percentages for 53 screen negatives and 159 screen positives.

Discussion
Before launching into an intervention programme based on identification of individuals at high risk, certain questions have to be posed. Different authors emphasize different questions (McKeown, 1968; Smith, 1972), but the main ones are as follows:

(i) Can a high risk group specifically be distinguished?
(ii) Can the screen procedure be applied economically and comprehensively?
(iii) What evidence is available about the natural history of the disorder?
(iv) Does the screen validity select those who subsequently develop a disorder?
(v) Is there reasonable evidence that intervention will be of benefit?
(vi) Can the costs be justified?
(vii) Will the tests used be acceptable to the public?
(viii) Will those identified be willing to accept help?
(ix) Is the disorder important?

In this paper we address ourselves mainly to questions (i) and (ii). We hope that the results of the main research project will provide answers to the remaining questions. Some workers assert that question (iii) is of crucial importance and that screening is not justifiable as more than a research procedure if the natural history of the disease is not known (McKeown, 1968). In our opinion, many medical conditions, let alone behaviour disorders, would not conform to this stringent criterion. Nevertheless, we do have some hard facts on which to base our research: there is evidence that antisocial disorders often continue as personality disorders in adult life, whereas neurotic disorders are unlikely to do so (Robins, 1966, 1970). Tests of intelligence, especially those administered in the infant school stage only broadly predict attainment at later stages, as they cannot take subsequent modifying factors into account. Furthermore, emotional and educational disorders interact in ways that are not fully understood (Rutter et al, 1970).

The concept of ‘at risk’ was initially received with interest and enthusiasm by paediatricians. It was seen as an aid to effective diagnosis and treatment of physical disorders at an early stage, even before the handicaps were manifest clinically (Forfar, 1968). Subsequently, ‘at risk’ registers were introduced (Sheridan, 1962). In the late sixties, such registers began to be viewed more critically (Rogers, 1967; Richards and Roberts, 1967). The most pertinent criticism is that of Rogers, who points out that there is a tendency to disregard the possibility of those children not on the register developing a handicap, and that the false negative rate reported in paediatrics is too high for justification of the register.

(a) Practical problems
We can start by tackling some of the practical problems. The technique of screening described here was completed in about four weeks by one psychologist, the latter part of this period being
devoted to follow-up of absentees from initial testing, so that the population in the selected schools was covered completely. We should emphasize that the screen measures which required child involvement—the reading and sociometric assessments—were designed for group administration and therefore could be applied to existing class groups without causing more than the minimum of disruption within the school. The tests used were found acceptable to staff in ordinary schools, and while there were very occasional queries from parents there were no reports that children were upset. Indeed, in various parts of our work we have now used the screens in 16 schools, and only on a single occasion did the project staff have to meet a parent to allay expressed anxiety about the testing programme.

(b) Variability in risk rate

We are not in a position at this stage to comment on absolute risk rate. Our estimate of 32.2 per cent is based on considered judgements as to what is a meaningful degree of statistical deviance on the various screen instruments used. Of interest, however, is the fact that in the schools studied the lowest 'risk' rate was approximately 20 per cent and the highest about 50 per cent. We had tried by consultation with the local services to ensure that the schools used were broadly representative of the cities in which we were working, but some of the schools used served very poor areas; it may be, therefore, that our high risk rate is not truly representative. However, selection could not have affected the rate to any great extent, for if we reduced the rate in each of the three schools with the highest rates by 5 per cent, we would bring about a reduction of the overall risk rate only to about 30 per cent.

(c) Efficacy of the screen

Our screen is intended to identify children who already have psychiatric disorder and also to predict future disorder by statistical deviations on the various screen measures. If we consider the first of these two aims in isolation, we can compare our findings with those of other studies whose purpose was to identify current disorder alone. In this, our screen identified some 69 per cent of the children who on more comprehensive clinical assessment proved to be markedly disturbed and 54 per cent of those with some maladjustment. The price paid for this was a 'false positive' rate of 16 per cent of the non-maladjusted population.

In the Rutter et al (1970) study, 46 per cent of the children with psychiatric disorder were identified by the school screen at the cost of 5 per cent of the non-maladjusted population.

It should be remembered, however, that our 'false positive' rate was inflated by children who were in educational difficulties only, so that our screen may have been somewhat more accurate than Rutter's (1970) school screen (carried out with one instrument) or Mulligan's school screen (Tizard, 1968) where the identified group were 85 per cent false positives. This may relate to higher prevalence of disorder, but it seems possible that the multiple screening technique may itself have made up to some extent for the lack of data from the children's parents at the screen stage. Other workers have shown that a multiple screen has much greater effectiveness and predictive utility than has each of its component parts (Rogers, 1942). Bower (1969) has pointed out that multiple classroom screen criteria, when combined, give a clear, comprehensive and economic picture which is the most predictive of maladjustment. This view is supported by Smith (1958), who concluded that not all the tests discriminated adequately when used alone.

(d) Sociometry

Some comment is needed on our finding that, of the two sociometric criteria used rejection is the one that relates most closely to other screen measures and to the presence of psychiatric disorder.

The relationship between adjustment and sociometric measures has been the subject of many studies. These vary widely in their definition of 'adjustment'. Although this is not specified in many cases, it is commonly taken to be synonymous with such indices as being a member of a sports team or a faculty leader or staying on at school past the 12th grade (Gronlund 1959, Ullman, 1957). In many studies, self-report questionnaires whose items
often reflect inter-personal difficulties, are used as measures of adjustment (Northway, 1952). In both cases the adjustment measures are loaded with sociability items, and while this constitutes interesting validation for sociometry, it is a narrower definition of adjustment than that adopted in this research. It is perhaps not surprising that significant correlations with positive sociometric choices have commonly been found. Semler (1960), in his discussion of the concept of adjustment, recognizes that it is naive to think that this can be estimated on the basis of a single test score. In his study he looks at the agreement between self-report, sociometric and teacher ratings and finds moderately high correlations between the three. Rejection was not measured in these studies, so that comparison with our main findings is not possible.

Bower's (1969) report describes his sociometric instrument in detail. His approach is somewhat more indirect than our own in that children are invited to nominate classmates for a 'class play'. While this variation has the advantage of being less direct than the children's popularity and the acceptance and rejection measures. The correlations were again high, particularly between teacher ratings and acceptance, but it should be noted again that the teacher was being asked to make a rating of popularity, not of deviant behaviour and educational problems as in our study. In an in-depth study by Cox (reported in Roff et al., 1972), a composite index of acceptance minus rejection is used and is found to relate to many family and personal variables of the child. However, a comparison of the relationships of these with acceptance and rejection separately is not possible from this report. We have provided a more extensive critique of the means of analysis used by Roff which can be shown to lead to an inflation of the levels of correlation (Kolvin et al., 1977).

One way in which our findings could be distorted is through the nature of our data. It is established (Rutter, 1970) that teachers tend

<table>
<thead>
<tr>
<th>Table IX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group versus Criterion Group: Mean scores of behaviour data deriving from parents (Rutter, A scale)</td>
</tr>
<tr>
<td>Mean scores</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Antisocial score</td>
</tr>
<tr>
<td>Neurotic score</td>
</tr>
<tr>
<td>Total score</td>
</tr>
</tbody>
</table>

* Significant difference from controls at 5 per cent level.
** Significant difference from controls at 1 per cent level.
to over-report antisocial as opposed to neurotic behaviour. If, as seems likely, the rejection scale is related more to antisocial behaviour, a close relationship to teacher reports might be expected. The fact that rejection correlates more closely than isolation with parent reports, as shown in Table IX, weighs against this argument.

Various authors (Northway, 1952) have felt that the rejection criterion is artificial because people are not usually interested in those with whom they do not associate. However, we must agree with Harper (1968) that the claim that the rejection measure may evoke resistance, resentment and unwillingness to admit negative feelings has not in practice proved to be true provided the research worker uses a diplomatic phrasing of the questions. Furthermore, it appears to measure a highly meaningful feature of deviant and maladjusted children. We took steps, both in our instructions to the children, and in the phrasing of the questions, to ensure that the likelihood of distressing the subjects was minimized. In practice there were few objections from teachers, and only one from a parent and no apparent adverse effects amongst children. Furthermore, the rejection measure was the more reliable of the two, a fact that also has a bearing on the validity.

Finally, it should be added that the relationship between sociometric measures and deviant behaviour cannot be regarded as a fixed quantity. Hargreaves (1967), for example, showed that whereas well-behaved children were chosen by their peers in the upper streams of a secondary modern school, poorly behaved children were chosen in lower streams. Grygier (1976) has gone further in suggesting that the correlation of deviance scores, as rated by staff caring for children, with sociometric measures provides an index of the atmosphere of residential treatment settings. This illustrates the way in which the environment may in some instances affect behavioural measures of individuals.

To summarize: measures of social isolation have long been lauded, for example by the New York State Youth Commission (1952), as a means of identifying children with poor psychological adjustment. However, we have not found social isolation to be related impressively to neurotic or antisocial behaviour judged by teachers or parents. In these respects, peer rejection has proved to be the more important criterion. Apparently it is wrong to assume that lack of popularity is synonymous with poor adjustment: isolation may also be due to introversion and thus is not necessarily associated with neuroticism.

(e) Reading

Many studies have demonstrated a relationship between under-achievement and behaviour. Malmquist (1958) revealed an association between backwardness in reading, and poor persistence, concentration and self-confidence, but maladjustment was not as closely linked with reading ability. In a recent review, Rutter et al (1970) point out that while some authors assert that maladjustment precedes poor reading ability (Blau, 1946; Pond, 1967) others claim that the reverse is true (Critchley, 1962). From their own study, Rutter et al report that antisocial disorders are associated with a greatly increased rate of reading retardation, but neurotic disorders are not. Other workers, such as Swift and Spivack (1969) have been able to demonstrate that educational under-achievement is related to much broader classroom functioning than just behaviour, such as pupil intelligence and pupil-teacher interactions.

Some clues as to the predictability of educational achievement by behaviour at an earlier age come from Swift and Spivack (1968) and from Feldhuizen et al (1970), who were able to relate previous behaviour with subsequent educational achievement, even in those studies in which IQ was controlled by statistical methods. Atwell et al (1967) similarly found that kindergarten behaviour ratings frequently predict later educational achievements. Our research confirms that there is a significant relationship between reading and conduct disorders but not between reading and neurotic disorders.

(f) Absenteeism

Bower (1969) reported that consistent unexcused school absences occurred with greater frequency in their maladjusted sample than in
their control group, but this was not sufficient to discriminate between the two groups. Both the Gluecks (1950) and the New York State Youth Commission (1952) reported that truancy was a good predictor of delinquency. However, in the Isle of Wight study (Rutter et al., 1970) it was reported that reading achievement was not generally associated with a high school absence rate, but was associated with emotional or behavioural disorder or physical disorder: nevertheless, in one of the tables (p 225) provided in the report from this research group there is a clear relationship between absenteeism from school for trivial reasons or for truanting and reading retardation.

In our study, absenteeism had the lowest yield and made the least independent contribution. Nevertheless, the absentee group have the highest mean neurotic scores on both the Rutter Teacher behaviour scale and Parent Scale. This suggests that absenteeism is an important symptom, even if not common enough to be useful for screening purposes.

(g) Behaviour (Rutter B Scale)

The use of this scale in epidemiological research has already been described and validated by the Maudsley group (Rutter et al., 1970, 1975); we will therefore merely confine ourselves to the comment that in our research the behaviour criterion had the highest yield and made the greatest independent contribution.

(h) High risk rate

Our results suggested that about a third of the children studied had a high risk of developing emotional and educational disorders. This gives a high-risk rate of 32.2 per cent. However, not all those identified would in fact develop disorders, and therefore it would be unwise to attempt to base an accurate quantitative estimate of services required, on such figures. All such a screen does is draw attention to those children and schools which may need a greater allocation of special resources. Such vulnerability in children can decrease, remain static, or increase; only study of the natural history of such vulnerability will indicate the eventual outcome. At present, therefore, attention should be directed only to those children who are markedly maladjusted or those whose condition deteriorates.

Acknowledgements

This study was supported by a grant from the Department of Education and Science. Some of the field-work was undertaken by Mrs L. Walker and Mr C. Hulbert. We would like to thank our planning committee for help in the early stages (Dr L. Mills and Dr K. Bailey of the Newcastle L.E.A., Mr A. Arnott and Miss M. McLaren of the University of Newcastle; Mr R. Stansfield of MIND; and Mr C. Norman of the D.E.S.). We are most grateful to the Local Education Authorities (Newcastle and Gateshead) for their cooperation.

References


SCREENING SCHOOLCHILDREN FOR HIGH RISK OF EMOTIONAL AND EDUCATIONAL DISORDER


I. Kolvin, B.A., M.D., F.R.C.Psych., Reader in Child Psychiatry,
R. F. Garside, B.Sc., Ph.D., F.B.Ps.S., Senior Lecturer,
R. Nicol, B.Sc., M.Phil., M.R.C.P., M.R.C.Psych., Consultant Psychiatrist,
I. Leitch, M.Sc., M.B.C.S., Senior Research Associate,
A. Macmillan, M.A., M.Sc., A.B.Ps.S. Senior Psychologist,

Human Development Unit, University of Newcastle upon Tyne; and Nuffield Psychology and Psychiatry Unit at the Fleming Hospital, Great North Road, Newcastle upon Tyne

(Received 28 June 1976; revised 23 February 1977)