Pergamon

J. Child Psychol. Psychiat. Vol. 36. No. 2, pp. 327–334, 1995
Elsevier Science Ltd

0021-9630/95 $9.50 + 0.00

0021-9630(94)00121-9

Properties of the Mood and Feelings Questionnaire in Adolescent Psychiatric Outpatients: A Research Note

Alison Wood, Leo Kroll, Ann Moore and Richard Harrington

Abstract—This study examined the psychometric properties of the Mood and Feelings Questionnaire (MFQ) in 104 adolescent outpatients attending a psychiatric clinic. The clinic offers a special assessment and treatment service for young people with depressive disorders. In this sample the self-report version of the MFQ had acceptable reliability and was a satisfactory screen for major depressive disorder diagnosed by a standardised interview with the child. It was also a useful measure of clinical remission.

Keywords: Depression, child, adolescent, screening

Abbreviations: AUC: Area Under the Curve, c.i.: Confidence Interval, GAS: Global Assessment Scale, ICC: Intraclass Correlation Coefficient, K-SADS: Schedule for Affective Disorders and Schizophrenia—Child Version, MDD: Major Depressive Disorder, MFQ-C: Mood and Feelings Questionnaire—Child report, MFQ-P: Mood and Feelings Questionnaire—Parental report, ROC: Receiver Operating Characteristic.

Introduction

There is a growing consensus that depressive disorders are quite common in adolescent psychiatric patients (Kolvin *et al.*, 1991) and a variety of self-rating instruments have been devised to screen for these conditions in clinical samples (Harrington & Shariff, 1992). Unfortunately, the psychometric properties of many of these scales have not been satisfactory. For instance, in the Newcastle Child Depression Project (Kolvin *et al.*, 1991), self-report questionnaire measures of depression were found to have "disappointing" screening properties for the diagnosis of major depressive disorder at interview (Fundudis *et al.*, 1991).

The *Mood and Feelings Questionnaire* (MFQ) (Angold, Costello, Pickles & Winder, 1987) was specifically developed to select children and adolescents for an epidemiological study of depression. Its contents cover the symptoms of DSM-III-R major depressive disorder (MDD), and it was designed to be an effective

Accepted manuscript received 5 *September* 1994

Requests for reprints to: Professor Harrington, University Department of Child and Adolescent Psychiatry, Royal Manchester Children's Hospital, Hospital Road, Pendlebury, Manchester M27 1HA, U.K.

screening instrument for MDD (Costello & Angold, 1988). However, there are few data on its psychometric properties in clinical samples.

This paper describes the following properties of the MFQ in adolescent psychiatric outpatients: (1) reliability; (2) parent–child agreement; (3) screening efficacy for DSM-III–R major depression; (4) validity as a measure of remission from MDD.

Methods

Subjects

The sample consisted of a consecutive series of 104 adolescents, 43 boys and 61 girls, aged between 10 and 19 years (mean age 13.7 years, *SD* 1.9), who had been referrred to an outpatient psychiatric clinic in a University Department. The clinic provides a specialized assessment and treatment service for depressed or suicidal adolescents as well as a general district service. Sixty-one referrals were from general practitioners, 29 from child psychiatrists and the remainder from social services and paediatricians. None of the cases was an inpatient at the time of the assessment, though about 15% were subsequently admitted for inpatient treatment.

Forty-three depressed patients entered a treatment trial comparing cognitive–behaviour therapy and relaxation training. All of these cases were followed up on average 8 weeks after the first assessment, when the questionnaire and interview measures were repeated.

Measures

The Mood and Feelings questionnaire is a 32-item depression scale that is rated by the child (MFQ-C). There is also a parallel version for the parent to complete about the child (MFQ-P). Symptoms are rated for the past 2 weeks "true" ("2"), "sometimes true" ("1") or "not true" ("0"), yielding a maximum score of 64.

The "gold standard" against which the MFQ was judged was a diagnosis of DSM-III-R major depressive disorder (American Psychiatric Asssociation, 1987), which was made using the *Schedule for Affective Disorders and Schizophrenia—Child Version* (K-SADS) (Puig-Antich & Chambers, 1978). Since recent research suggests that children provide a more accurate account of their mental state than their parents (Barrett *et al.*, 1991; Harrington, 1993), the main source of data for the diagnosis was direct interview with the subject. However, in all cases this interview was corroborated by a brief interview with the primary caretaker/s, and the results were combined using a simple best estimate rule that gave priority to the presence of symptoms. All of the posttreatment assessments and most of the baseline assessments were conducted without knowledge of MFQ scores. However, in nine baseline assessments the interviewer was aware that the child had a high score on the MFQ. Removal of these cases from the analyses made no material difference to the results.

All the interviewers (AW, AM, LK, RH) were trained in the use of the K-SADS. Inter-rater agreement on the K-SADS depression scale was assessed by the independent rating of audiotaped interviews. The intraclass correlation coefficient (ICC) was .93 ($N = 10$, $F = 27.5$, $p < .001$).

Statistical methods

The sensitivity and specificity of the MFQ against the diagnosis of MDD and against response to treatment were evaluated at various MFQ cutpoints with Receiver Operating Characteristic (ROC) analyses (Swets, 1988). Fombonne (1991) gives a helpful account of the use of these methods in child psychiatry. Briefly ROC analysis involves the plotting of true positive rate (sensitivity) against false positive rate (1-specificity). The area under this curve (AUC) is then estimated and used as an index of diagnostic accuracy. The AUC ranges from 0.5 to 1.0, with values of 0.5–0.7 usually being interpreted as indicating "low" test accuracy, 0.7–0.9 "moderate" accuracy, and > 0.9 "high" accuracy (Henderson, 1993). Calculation of the AUC allows direct comparisons of the performances of different diagnostic techniques (Hsiao, Bartko & Potter, 1989).

Results

Characteristics of the sample

Fifty five of the 104 cases had MDD, of whom 15 were male and 40 female. Their mean age was 14.1 years (SD = 1.7 years) and the mean duration of the present episode of depression was 59.7 weeks (SD = 62.6 weeks). They were severely impaired, as judged by the finding that virtually all of them (51/55) had a Global Assessment Scale (GAS) score (Shaffer *et al.*, 1983) of "50" or less. Scores on the GAS can range from "100" ("no symptoms") to "0" ("needs constant supervision"), with a score of 50 being defined as "any serious symptomatology or impairment in functioning that most clinicians would think obviously requires treatment". The mean GAS score of MDD cases was 41.0 (SD = 8.7, range 21–61). In line with previous studies the depressed cases had many comorbid disorders (which were diagnosed without hierarchy), including overanxious disorder (N = 35), dysthymic disorder (N = 25), separation anxiety (N = 7), oppositional disorder (N = 23) and conduct disorder (N = 11).

The 49 cases without MDD comprised 28 males and 21 females, mean age 13.2 years (SD = 2.1 years). Seven of them did not have a mental disorder. Among the remainder the most common diagnoses, which were diagnosed without hierarchy, were oppositional disorders (N = 21), conduct disorders (N = 12), minor depressive disorder (N = 6), overanxious disorder (N = 3) and separation anxiety (N = 2). Their mean duration of episode was 106.2 weeks (SD = 110.1 weeks) and the mean GAS score (excluding those without a psychiatric disorder) was 52.0 (SD = 17.7, range 10–90).

Reliability

Both the child (MFQ-C) and parent (MFQ-P) versions of the MFQ had high internal consistency (alpha = .94 and .92, respectively).

The MFQ-C was also examined for stability in a random sample of 15 cases with MDD, in whom the MFQ was repeated on average 18 days after the first assessment, but before the treatment started. The mean MFQ-C score tended to decline between the two assessments, mean scores of 40.6 (SD = 11.1) and 37.5 (SD = 12.5) respectively, though the difference was not significant (t = 1.6, df = 14, p = .14) and there was a substantial correlation between the two MFQ-C scores (ICC = .78).

Parent–child agreement

In line with many other studies of questionnaire ratings of depression, children reported more symptoms than their parents did about them, with mean MFQ scores of 29.2 (SD = 15.1) and 23.2 (SD = 13.0) respectively (t = 4.1, df = 93, $p < .001$) (MPQ-P data were missing or incomplete on 10 cases).

The correlation between child and parent total MFQ scores was 0.51. The average kappa coefficient of agreement between child and parent on individual symptoms ("true" and "sometimes" versus "not true") was .31.

Screening characteristics

The mean MFQ-C and MFQ-P scores of the cases with major depression were 36.9 (*SD* = 12.4) and 26.6 (*SD* = 13.0) respectively, compared with MFQ-C and MFQ-P scores in nondepressed cases of 20.5 (*SD* = 13.2) and 18.2 (*SD* = 12.4) respectively.

ROC analyses showed that the MFQ-C had moderate diagnostic accuracy (AUC = .82, *SE* = .04, 95% c.i. .74–.91) for major depression at interview, whereas the MFQ-P had only low accuracy (AUC = .69, *SE* = .05, 95% c.i. .58–.79). The difference between these AUCs was significant, as assessed by the test described by Hanley and McNeil (1983) for comparing the areas under ROC curves derived from the same cases ($z = 2.4$, $p < .05$).

To see whether the screening efficacy of the MFQ could be improved, various combinations of data from the MFQ-C and MFQ-P were computed but none of the combinations produced an AUC that was higher than that obtained from the MFQ-C. For instance, data from the MFQ-C and MFQ-P were combined by recording a symptom positive only if both questionnaire responses were "true" or "sometimes". The AUC for this composite scale was .73 (*SE* = .06).

AUCs were also calculated for males (*N* = 43, MFQ-C AUC = .82, *SE* = .07) and females (*N* = 61, MFQ-C AUC = .77, *SE* = .07), and for younger subjects (aged 13 or less, *N* = 44, MFQ-C AUC = .85, *SE* = .06) and older subjects (age 14 or more, *N* = 60, MFQ-C AUC = .76, *SE* = .07). Neither the difference in AUCs between males and females nor the difference in AUCs between younger and older subjects reached statistical significance at the 5% level ($z = 0.55$, and $z = 0.97$, respectively). Similar analyses were conducted according to GAS score, duration of episode and comorbidity with anxiety and with conduct disorders, but in none of these analyses did the differences between AUCs reach significance at the 5% level.

Figures 1 and 2 plot sensitivity and specificity against various cut-off points on the MFQ-C and MFQ-P, respectively. The misclassification rate is also shown (i.e. the number of false negatives and false positives as a proportion of the total number of subjects). This graphical presentation is the same as that in the Newcastle Childhood Depression Study (Fundudis *et al.*, 1991), allowing direct comparisons to be made.

On the MFQ-C (Fig. 1) the cut-off score with the best diagnostic confidence, as determined by the intersect point of sensitivity and specificity, is 27. This cutpoint gives a sensitivity of .78 (95% c.i. .67–.89), a specificity of .78 (95% c.i. .66–.89) and a misclassification rate of .22 (95% c.i. .14–.30).

Figure 2 shows that the point of intersection on the MFQ-P was lower than on the MFQ-C, at just 21 (rounded down from 21.3), with sensitivity .63 (95% c.i. .50–.77), specificity .61 (95% c.i. .47–.76) and misclassification rate .38 (95% c.i. .28–.47).

Measurement of clinical remission

The usefulness of the MFQ for measuring remission from MDD was assessed by comparing MFQ scores posttreatment with the K-SADS rating of remission made independently by the interviewer. The interview method was the same; K-SADS

rating of child interview with parental corroboration. Remission was defined using predetermined criteria (Puig-Antich *et al.*, 1987), which require that K-SADS scores in both depressed mood and anhedonia are "2" ("slight, of questionable clinical significance") or less. With these criteria, 16 of the 43 cases who completed treatment were judged to have remitted.

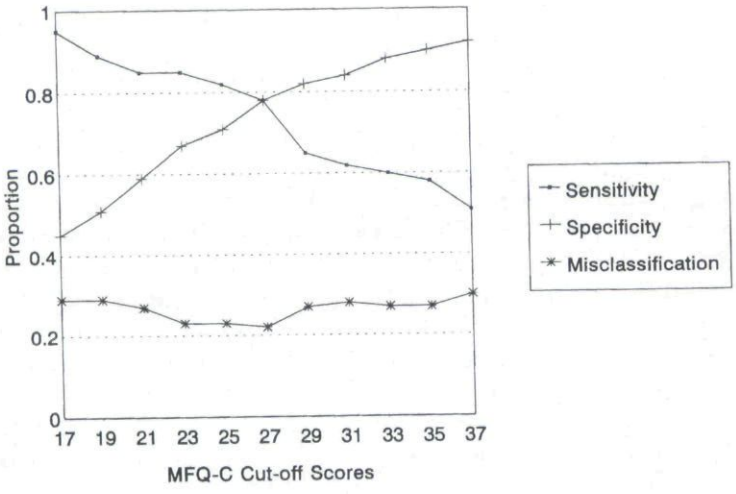

Fig. 1. Sensitivity (■), specificity (+) and misclassification (*) proportions at different cut-offs on the MFQ-Child.

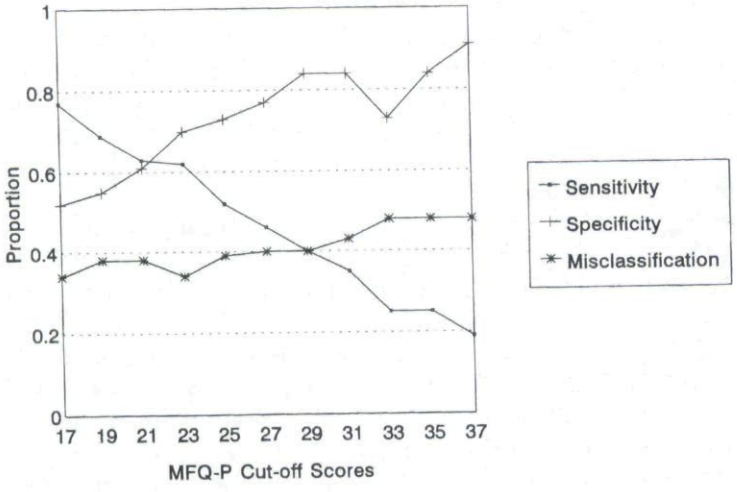

Fig. 2. Sensitivity (■), specificity (+) and misclassification (*) proportions at different cut-offs on the MFQ-Parent.

ROC analyses showed that the MFQ-C had high diagnostic accuracy (AUC = .91, *SE* = .04) for remission at interview. The intersect point of sensitivity and specificity was 16 (rounded down). This cutpoint gave a sensitivity of .75 (95% c.i. .54–.96), a specificity of .81 (95% c.i. .67–.96), and a misclassification rate of .21 (95% c.i. .09–.33).

By contrast, the MFQ-P had only low accuracy (AUC = .56, *SE* = .1) for remission. The difference between the MFQ-C AUC and the MFQ-P AUC was significant at the 1% level ($z = 3.0$). The correlation between MFQ-C and MFQ-P was low ($r = .16$, n.s.)

Discussion

In interpreting these results it must be borne in mind that this study was based on referrals to a clinic that offered a specialized assessment and treatment service for adolescents with depression. The sample therefore, contained a high proportion of cases with depressive disorder, and the results may not generalize to samples where the rate of MDD is lower. It should also be noted that our comparison cases without MDD had low rates of anxiety disorders and high rates of behavioural disorders. It may be that the MFQ would have different screening properties against a comparison group with higher rates of anxiety.

Moreover, it will be appreciated that the usefulness of a screen depends not only on its screening statistics (such as specificity), but also on the base rate of the disorder in the population being studied. This base rate problem is highlighted by considering examples where specificity and misclassification rate are held constant and the base rate changes. At a base rate of 50% MDD, and with a misclassification rate and specificity as in the present study (.23 and .78, respectively), then in a sample of 100 referrals, 78% (38/49) of cases who were screened positive would have MDD at interview. However, at a base rate of 25% (roughly the rate of MDD found in studies of general child psychiatric outpatient samples, Carlson & Cantwell, 1980; Kolvin *et al.*, 1991) the percentage of interview positive cases drops to 53%. At a base rate of 10%, a scale with the same psychometric properties would result in about three nondepressed cases being interviewed for every case who turned out to have the disorder.

Our results suggest that in samples containing a high proportion of depressed cases the MFQ-C may be a useful screening device. Nevertheless, its screening properties do not seem to be substantially better than those of other instruments. For example, the sensitivity (.78), specificity (.78) and misclassification rate (.22) of the MFQ-C at the intersect point are very similar to the values found in the Newcastle Study (Fundudis *et al.*, 1991) with the Children's Depression Inventory (Kovacs, 1981) (misclassification rate = .23; sensitivity and specificity = .75), and with the Depression Self-Rating Scale (Birleson, 1981) (misclassification rate = .25; sensitivity = .75; specificity = .73). Nor are they different from those obtained with the Beck Depression Inventory in adolescents referred to an outpatient clinic for depression (Ambrosini, Metz, Bianchi, Rabinovich & Undie, 1991). Similarly, the MFQ-C AUC found in the present study (.82) was identical to that reported by Rey

and Morris-Yates (1992) in a study of referred adolescents, in which a composite depression scale was derived from the Child Behaviour Checklist and used as a screen for MDD.

However, the MFQ-C did seem to work well as a measure of change, as judged by its discrimination between those who "remitted" and those who did not. It may, therefore, be especially useful in clinical trials of treatment for MDD in adolescents.

The psychometric properties of the MFQ-P seemed to be much less satisfactory. It should be noted, however, that in our study the "gold standard" diagnosis of MDD was based largely on the interview with the child. Resources did not permit us to employ an interviewer to conduct an independent K-SADS with the parent. If the gold standard had been a parental interview, it is probable that the MFQ-P would have fared better.

In conclusion, our findings suggest that the child version of the MFQ has satisfactory psychometric properties in a sample containing a high proportion of depressed cases. Clinicians as well as researchers may find it useful as a measure of clinical remission.

Acknowledgements—This research was supported by grants from the Mental Health Foundation and the Medical Research Council. Thanks are due to Sally Hollis, statistician at North West Regional Health Authority, for her advice on data analysis.

References

Ambrosini, P. J., Metz, C., Bianchi, M. D., Rabinovich, H. & Undie, A. (1991). Concurrent validity and psychometric properties of the Beck Depression Inventory in outpatient adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry,* **30,** 51–57.

American Psychiatric Association (1987). *Diagnostic and Statistical Manual of Mental Disorders—DSM-III-R (third edition—revised).* Washington DC: American Psychiatric Association.

Angold, A., Costello, E. J., Pickles, A. & Winder, F. (1987). *The development of a questionnaire for use in epidemiological studies of depression in children and adolescents.* London: Medical Research Council Child Psychiatry Unit.

Barrett, M. L., Berney, T. P., Bhate, S., Famuyiwa, O. O., Fundudis, T., Kolvin, I. & Tyrer, S. (1991). Diagnosing childhood depression. Who should be interviewed—parent or child? The Newcastle Child Depression Project. *British Journal of Psychiatry,* **159,** 22–27.

Birleson, P. (1981). The validity of depressive disorder in childhood and the development of a self-rating scale: a research project. *Journal of Child Psychology and Psychiatry,* **22,** 73–88.

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

Costello, E. J. & Angold, A. (1988). Scales to assess child and adolescent depression: checklists, screens and nets. *Journal of the American Academy of Child Psychiatry,* **27,** 726–737.

Fombonne, E. (1991). The use of questionnaires in child psychiatry research: measuring their performance and choosing an optimal cut-off. *Journal of Child Psychology and Psychiatry,* **32,** 677–693.

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

Hanley, J. A. & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology,* **148,** 839–843.

Harrington, R. C. (1993). *Depressive disorder in childhood and adolescence.* Chichester: Wiley.

Harrington, R. C. & Shariff, A. (1992). Choosing an instrument to assess depression in young people. *Newsletter of the Association for Child Psychology and Psychiatry,* **14,** 279–282.

Henderson, A. R. (1993). Assessing test accuracy and its clinical consequences: a primer for receiver operating characteristic curve analysis. *Annals of Clinical Biochemistry,* **30,** 521–539.

Hsiao, J. K., Bartko, J. J. & Potter, W. Z. (1989). Diagnosing diagnoses. Receiver operating characteristic methods and psychiatry. *Archives of General Psychiatry,* **46,** 664–667.

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

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

Puig-Antich, J. & Chambers, W. (1978). *The Schedule for Affective Disorders and Schizophrenia for school-aged children.* New York State Psychiatric Institute: Unpublished interview schedule.

Puig-Antich, J., Perel, J. M., Lupatkin, W., Chambers, W. J., Tabrizi, M. A., King, J., Goetz, R., Davies, M. & Stiller, R. L. (1987). Imipramine in prepubertal major depressive disorders. *Archives of General Psychiatry,* **44,** 81–89.

Rey, J. M. & Morris-Yates, A. (1992). Diagnostic accuracy in adolescents of several depression rating scales extracted from a general purpose behavior checklist. *Journal of Affective Disorders,* **26,** 7–16.

Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H. & Aluwahlia, S. (1983). A children's Global Assessment Scale (C-GAS). *Archives of General Psychiatry,* **40,** 1228–1231.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science,* **240,** 1285–1293.