

BRIEF COMMUNICATION

The internal and external validity of the Major Depression Inventory in measuring severity of depressive states

L. R. OLSEN,¹ D. V. JENSEN, V. NOERHOLM, K. MARTINY AND P. BECH

From the Psychiatric Research Unit, Frederiksborg General Hospital, Hillerød; and Department of Rheumatology, Hoersholm General Hospital, Hoersholm, Denmark

ABSTRACT

Background. We have developed the Major Depression Inventory (MDI), consisting of 10 items, covering the DSM-IV as well as the ICD-10 symptoms of depressive illness. We aimed to evaluate this as a scale measuring severity of depressive states with reference to both internal and external validity.

Method. Patients representing the score range from no depression to marked depression on the Hamilton Depression Scale (HAM-D) completed the MDI. Both classical and modern psychometric methods were applied for the evaluation of validity, including the Rasch analysis.

Results. In total, 91 patients were included. The results showed that the MDI had an adequate internal validity in being a unidimensional scale (the total score an appropriate or sufficient statistic). The external validity of the MDI was also confirmed as the total score of the MDI correlated significantly with the HAM-D (Pearson's coefficient 0.86, $P \leq 0.01$, Spearman 0.80, $P \leq 0.01$).

Conclusion. When used in a sample of patients with different states of depression the MDI has an adequate internal and external validity.

INTRODUCTION

The most frequently used self-rating scales for depression are the Beck Depression Inventory (BDI) (Beck *et al.* 1961), the Zung Self-Rating Depression Scale (SDS) (Zung, 1965) and the Center for Epidemiological Studies Depression Scale (CES-D) (Radloff, 1977). These questionnaires have been psychometrically evaluated as scales to measure the severity of depressive states and also as screening instruments for the diagnosis of clinical depression.

With the introduction of DSM-III (APA, 1980) with symptom-based diagnostic criteria for mental disorders, the diagnosis of major depression is reached using an algorithm covering only nine symptoms. The three depression

questionnaires (BDI, SDS and CES-D) all contain around 20 symptoms which, however, have a limited coverage of the nine DSM-III symptoms of major depression. On this background, we developed the Major Depression Inventory (MDI) (Bech, 1998; Bech *et al.* 2001), which covers the whole spectrum of symptoms in both the DSM-III/DSM-IV (APA, 1994) 'major depression' and the ICD-10 (WHO, 1993) 'moderate to severe depression'.

On the basis of the algorithms for diagnosing depression in accordance with DSM-IV or ICD-10 the MDI showed a high sensitivity and specificity in a previous study (Bech *et al.* 2001).

In the present study we have investigated the MDI as a scale for measuring severity of depressive states. The analysis of the MDI has focused on both the internal validity (i.e. tests for unidimensionality) and the external validity (i.e. the correspondence with a clinician rated scale).

¹ Address for correspondence: Dr Lis Raabaek Olsen, Psychiatric Research Unit, Frederiksborg General Hospital, DK-3400 Hillerød, Denmark.

Objectives

The objectives of the study with the MDI have been to evaluate the internal validity of the scale (the total score being an appropriate or sufficient statistic) as well as the external validity of the scale (the correlation of the MDI with the Hamilton Depression Scale (HAM-D₁₇) (Hamilton, 1967; Bech *et al.* 1986)), which includes a standardization of the MDI with cut-off scores in terms of the HAM-D definitions of mild and moderate degrees of depression.

METHOD

Study population

The patients were selected from ongoing studies within the following range of depressive states.

(1) *Very mild, doubtful or no depression*

These were out-patients from our Department of Rheumatology; we consecutively included patients who had suffered from low back pain for more than 3 months without psychiatric co-morbidity.

(2) *Mild to moderate depression*

These out-patients were from a private psychiatric practice in Copenhagen and they had been screened for inclusion in a study on social adaptation.

(3) *Mild to moderate depression*

These out-patients from our Psychiatric Research Unit were participating in an ongoing study on light therapy in major depression without SAD (seasonal affective disorder).

(4) *Mild to marked depression*

These in-patients from our Psychiatric Hospital Department were participating in a study on the sensitivity and specificity of the MDI, using the Present State Examination (PSE) as the index of diagnostic validity (Bech *et al.* 2001).

Rating scales

MDI

The items of the scale cover the ten ICD-10 symptoms of depression. These symptoms are identical with the DSM-IV major depression symptoms apart from one symptom, low self-esteem, which in DSM-IV is incorporated in the

item of guilt. Thus, the MDI contains 10 items, however, items 8 and 10 are divided into two sub-items, a and b (Appendix 1). Only the highest scores of items 8 and 10 (either a or b) are included in the statistical analysis. On a 6-point Likert scale, the individual items measure how much of the time the symptoms have been present during the past 14 days. The scale goes from 0 (the symptom has not been present at all) to 5 (the symptom has been present all of the time). The various steps refer to the frequency of the symptoms during the last 2 weeks and are defined by adverbs or adjectives (Appendix 1) with only indirect definitions. In a previous study (Bent-Hansen *et al.* 1995) it had been found that depressed patients prefer such indirect stipulations to a direct or definite item manual for the individual items.

The MDI is intended to be used both as a diagnostic instrument with the algorithms leading to the DSM-IV or ICD-10 categories 'major' or 'moderate to severe' depression (Bech *et al.* 2001), and as a measuring instrument in which the total score is a sufficient statistic. When used as a measuring instrument, the 10 items are added up, with a theoretical score range from 0 to 50.

Hamilton Depression Scale

We used the 17-item (HAM-D₁₇) version endorsed by Max Hamilton and published by Bech *et al.* (1986). This version has been used in the studies performed by the Danish University Antidepressant Group (e.g. DUAG, 1990). The HAM-D raters who participated in the present study had been trained as investigators in the DUAG trials. The intraclass coefficients of reliability in the DUAG trials are 0.75 or higher (Stage *et al.* 2001).

Statistical analysis

Internal validity

Classical psychometric approach

A factor analysis in terms of a principal component analysis was performed (Nunnally & Bernstein, 1994). A scree plot was used to determine the numbers of factors to be taken into consideration. A 'general factor' was defined as a factor explaining at least 50% of the variance.

Cronbach's coefficient alpha was used to evaluate internal consistency. A coefficient of 0.80 or

higher was considered adequate (Nunnally & Bernstein, 1994).

Modern psychometric approach

The Rasch analysis (Rasch, 1960; Bech *et al.* 1981; Allerup, 1997) was used to test for unidimensionality of the scale. The test of fit of the Rasch model for the total scale score being a sufficient statistic was performed by use of the one parameter logistic programme in which the criteria of males *versus* females and patients with low scores *versus* patients with high scores were tested (Verhelst & Glass, 1995). The non-parametric evaluation of the data structure in accordance with the Rasch model was performed using the Mokken analysis (Mokken, 1971; De Jong & Molenaar, 1987; Molenaar *et al.* 1994). The Mokken analysis of homogeneity or unidimensionality is a measure of the extent to which an extra item fits into the structure provided by the other items of the scale. The test of fit of the individual items analogously to the Mokken analysis was within the Rasch analysis performed as described by Allerup (1997). Each item was first dichotomized by rescaling grades 0, 1 and 2 as 0, and grades 3, 4 and 5 as 1. The level of rejection of unidimensionality in the Rasch analysis was $P \leq 0.01$. As external criterion the level of acceptance according to the Mokken analysis was a coefficient of homogeneity of ≥ 0.40 , while a coefficient of 0.30 to 0.39 was considered only to be just acceptable (Mokken, 1971).

External validity

The Hamilton Depression Scale (HAM-D) was used as the index of external validity in the 17-item version (HAM-D₁₇). As to the measure of correlation, the Pearson coefficient is reported in some studies while the Spearman coefficient is reported in other studies. Therefore, the strength of the association between the MDI and the HAM-D₁₇ was expressed in terms of the Pearson coefficient (Altman, 1991) as well as the Spearman coefficient (Siegel, 1956). The association between the MDI score and the HAM-D score to standardize the MDI was estimated by linear regression analysis in which the MDI score was considered the dependent variable. The standardization included prediction intervals of the estimated values.

Estimation of sample size

Traditionally, the number of patients needed psychometrically when using principal component analysis is approximately 10 times the number of items in the scale under examination (Aiken, 1995). As the MDI contains ten items, the number of patients should be approximately 100.

RESULTS

In total, 91 patients (24 males, 67 females; mean age 45.5 years, s.d. 15.2) were included in the psychometric analysis of the MDI. Of those, 18 patients were recruited from the Department of Rheumatology (5 males, 13 females, mean age 43.0 years, s.d. 15.2, HAM-D₁₇ mean score 6.1, s.d. 5.9), 11 patients were recruited from our Out-patient Research Unit (4 males, 7 females; mean age 48.5 years, s.d. 11.1, HAM-D₁₇ mean score 20.6, s.d. 4.6), 40 patients were recruited from the private psychiatric practice (13 males, 27 females; mean age 40.6 years, s.d. 15.5, HAM-D₁₇ mean score 18.9, s.d. 7.5), and 22 in-patients were recruited from our Psychiatric Department (2 males, 20 females; mean age 55.0 years, s.d. 15.1). In the latter sample of in-patients, all patients had a mood disorder, 15 patients had a current diagnosis of major depression (HAM-D₁₇ mean score 21.5, s.d. 5.5) and the remaining 7 patients had major depression in remission (HAM-D₁₇ mean score 11.1, s.d. 6.3). In the different groups, the percentage of patients with a HAM-D₁₇ score of ≥ 18 ranged from 5.6% to 66.7%.

Internal validity

Principal component analysis of the MDI identified only one factor when the scree plot was analysed. This factor explained 56% of the variance while the second factor explained 10%, the third factor 8% and the fourth factor 5% of the variance. Table 1 shows the factor loadings for the individual items according to the principal component analysis, indicating a higher loading in the top-listed items compared to the bottom-listed. Cronbach's coefficient alpha was 0.90. Table 1 shows also the results of the Mokken analysis with the Loevinger coefficient of homogeneity for the total scores and for the individual items. Although two of the items had Loevinger coefficients of < 0.40 (item 9 and 10)

Table 1. *The psychometric Mokken analysis of the 10 items of the MDI with the corresponding factor loadings from the principal component analysis. The items are listed in terms of inclusiveness (rank-ordered), i.e. highest mean score for 'lack of energy' and lowest mean score for 'suicidal thoughts'*

Item no. Content (mean score value)	Loevinger's coefficient of homogeneity	Factor loading
3 Lack of energy (3.03)	0.60	0.83
8 Restlessness (2.99)	0.61	0.85
1 Depressed mood (2.93)	0.63	0.86
4 Lack of self-confidence (2.90)	0.60	0.84
2 Lack of interests (2.60)	0.60	0.86
7 Poor concentration (2.60)	0.53	0.76
10 Reduced appetite (2.32)	0.39	0.56
5 Guilt feelings (2.24)	0.51	0.74
9 Sleep (2.02)	0.25	0.36
6 Suicidal thoughts (1.49)	0.47	0.66
Sum of all items	0.52	Eigen-value = 5.59

the coefficient for testing to what extent all the MDI items taken together concerning unidimensionality of the scale was acceptable (0.52). Table 1 shows the rank-order of the MDI items when using the mean score value for the individual items as index of inclusiveness. Thus, at the top is placed item 3 (lack of energy) and at the bottom item 6 (suicidal thoughts).

The Rasch analysis confirmed that the 10 items of MDI constitute one dimension. According to the Rasch analysis the same rank-order of the individual items was found both when males were compared with females and when patients with low total MDI scores were compared with patients with high total MDI scores. Where discrepancies emerged in rank-order between the Mokken analysis and the Rasch analysis the difference was only of the order of one rank. The item with the lowest coefficient in the Mokken analysis was item 9 (sleep), which also was the weakest in the Rasch analysis.

External validity

When the MDI scores were correlated to the HAM-D₁₇ scores ($N=91$) the Pearson coefficient was 0.86 ($P \leq 0.01$), (the corresponding non-parametric Spearman coefficient was 0.80 ($P \leq 0.01$)).

By linear regression in which the MDI score was considered as the dependent variable the following equation was estimated (confidence interval of 95%):

$$\text{MDI score} = 4.54 + 1.65 \times \text{HDS score} (\pm 1.6).$$

For this estimation the value of R^2 is 0.73, i.e. the proportion of the total variation of the dependent variable explained by this model is 73%. Table 2 shows the standardization of the MDI using the conventional cut-off scores on the HAM-D₁₇ as index of validity (Bech *et al.* 1975).

DISCUSSION

The range of scores obtained on the Hamilton Depression Scale (HAM-D₁₇) across the sample population in the present study had a distribution which was adequate for an analysis of a self-rating scale such as the MDI, i.e. a scale for patients with mild to marked degrees of depressive states. All patients in the present study were able to complete the MDI, indicating a high degree of applicability.

The 10 items of the MDI obviously have a high content validity when compared to the diagnostic systems (DSM-IV or ICD-10) as the scale is based on the universe of symptoms within these systems. Although symptoms with a high diagnostic validity do not necessarily have a high validity for measuring severity (Frances *et al.* 1990; Kessler & Mroczek, 1995), the present study showed that the MDI is a unidimensional scale. This was supported both with classical psychometric tests (e.g. principal component analysis and Cronbach's coefficient alpha) and with modern psychometric tests (e.g. the Mokken analysis and the Rasch analysis). The rank-order of inclusiveness showed almost the same pattern when applying the two different types of modern psychometric tests. The structure of inclusiveness shows that the core symptoms of depression according to DSM-IV and ICD-10 (depressed mood, lack of energy and lack of interests) are among the most inclusive items of the MDI (Table 1) indicating a 'ceiling effect', while the items of guilt feelings and suicidal thoughts were most exclusive indicating a 'floor effect'. The somatic items (sleep and appetite) showed suboptimal fitting in the Mokken analysis (Loevinger's coefficients < 0.40) as well as in the

Table 2. Standardization of the Major Depression Inventory (MDI) using the HAM-D₁₇ as index of validity

Category	HAM-D ₁₇ score	MDI-score Predicted value (95% prediction interval)
No depression/full remission	7	13.3 (11.3, 15.3)
Probable major depression/mild depression	13	20.9 (19.4, 22.3)
Major depression/moderate depression	18	27.1 (25.8, 28.5)
Major depression/marked depression	25	35.9 (34.0, 37.8)

Rasch analysis. Furthermore, in the principal component analysis the two somatic items showed the lowest factor loadings. However, the somatic items had no impact on the overall validity of the MDI indicating that the total score is a sufficient statistic. The three self-rating scales developed before the release of the DSM-III (BDI, SDS and CES-D) have all previously been correlated with the HAM-D₁₇ and coefficients between 0.6 and 0.8 have been reported (e.g. Brown & Zung, 1972; Bech *et al.* 1975; Biggs *et al.* 1978; Radloff, 1977). The correlation coefficient of 0.86 found in the present study is, therefore, very acceptable.

The standardization of the MDI indicated that a cut-off score of 27 corresponds to a score of 18 on HAM-D₁₇ (or major depression), which is in agreement with our analysis of the MDI when compared to the diagnosis of major depression based on a psychiatric interview (Bech *et al.* 2001). As shown by Paykel (1990) a HAM-D₁₇ score of 18 equals major depression while a score of 13 equals probable major depression.

Because the MDI scale is a brief scale, consisting of only 10 items that are presented to the patient on a single page (Appendix 1), the MDI can easily be used in the setting of general practice or in somatic hospital departments both as a screening instrument for detecting depression (Bech & Wermuth, 1998; Bech *et al.* 2001) and to monitor the effect of antidepressive therapy analogous to the use of the Hamilton Depression Scale as outcome measure.

This study has some limitations. The diagnoses were not made by structured research interview, instead the Hamilton Depression Scale was used as reference, conducted by trained psychiatrists. Comparison to SCAN interview has previously been made (Bech *et al.* 2001). In the present study we used a sample covering the spectrum from no depression to severe

depression to fulfil the objective of the study. The scale might perform differently in a more homogenous sample of depressed people, studies to inspect this are now in progress. Additional items could have been added, e.g. an item about hypersomnia, which is included in the DSM-IV but not in the ICD-10. Nevertheless, the purpose with this scale was to make it as short as possible while still covering enough information to make diagnoses as well as to rate severity of depression.

In conclusion, this study has shown that the total score of the MDI is a sufficient statistic to measure severity of depressive states. Moreover, a linear correlation to the Hamilton Depression Scale has been found, resulting in a standardization of the MDI by using the HAM-D₁₇ as index of validity.

REFERENCES

- Aiken, L. R. (1995). *Personality Assessment. Methods and Practices (2nd revised edn.)*. Hogrefe and Huber: Toronto.
- Allerup, P. (1997). Statistical analyses of data from the IEA reading literacy study. In *Applications of Latent Trait and Latent Class Models in the Social Sciences* (ed. J. Rost and R. Langeheine), pp. 50–58. Waxman: Berlin.
- Altman, D. (1991). *Practical Statistics for Medical Research*. Chapman & Hall: London.
- American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders, 3rd edn. (DSM-III)*. APA: Washington, DC.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders, 4th edn. (DSM-IV)*. APA: Washington, DC.
- Bech, P. (1998). *Quality of Life in the Psychiatric Patient*. Mosby-Wolfe: London.
- Bech, P. & Wermuth, L. (1998). Applicability of the Major Depression Inventory in patients with Parkinson's disease. *Nordic Journal of Psychiatry* **52**, 305–309.
- Bech, P., Gram, L. F., Dein, E., Jacobsen, O., Vitger, J. & Bolwig, T. G. (1975). Quantitative rating of depressive states. *Acta Psychiatrica Scandinavica* **51**, 161–170.
- Bech, P., Allerup, A., Gram, L. F., Rejsby, N., Rosenberg, R., Jacobsen, O. & Nagy, A. (1981). The Hamilton Depression Scale. Evaluation of objectivity using logistic models. *Acta Psychiatrica Scandinavica* **63**, 290–299.
- Bech, P., Kastrup, M. & Rafaelsen, O. J. (1986). Mini-compendium of rating scales for states of anxiety, depression, mania, schizophrenia

