ORIGINAL RESEARCH

Validity and test–retest reliability of the Persian version of the Montgomery–Asberg Depression Rating Scale

Mohammad Ahmadpanah[1]
Meisam Sheikhbabaei[1]
Mohammad Haghighi[1]
Fatemeh Roham[1]
Leila Jahangard[1]
Amineh Akhondi[2]
Dena Sadeghi Bahmani[3]
Hafez Bajoghli[4]
Edith Holsboer-Trachsler[3]
Serge Brand[3,5]

[1]Behavioral Disorders and Substances Abuse Research Center, Hamadan University of Medical Sciences, Hamadan, Iran; [2]Hamadan Educational Organization, Ministry of Education, Hamadan, Iran; [3]Center for Affective, Stress, and Sleep Disorders, Psychiatric Clinics of the University of Basel, Basel, Switzerland; [4]Iranian National Center for Addiction Studies (INCAS), Tehran University of Medical Sciences, Tehran, Iran; [5]Department of Sport, Exercise and Health Science, Sport Science Section, University of Basel, Basel, Switzerland

Correspondence: Serge Brand
Center for Affective, Stress, and Sleep Disorders, Psychiatric Clinics of the University of Basel, Wilhelm Klein-Strasse 27, 4012 Basel, Switzerland
Tel +41 61 325 5114
Fax +41 61 325 5513
Email serge.brand@upkbs.ch

Background and aims: The Montgomery–Asberg Depression Rating Scale (MADRS) is an expert's rating tool to assess the severity and symptoms of depression. The aim of the present two studies was to validate the Persian version of the MADRS and determine its test–retest reliability in patients diagnosed with major depressive disorders (MDD).

Methods: In study 1, the translated MADRS and the Hamilton Depression Rating Scale (HDRS) were applied to 210 patients diagnosed with MDD and 100 healthy adults. In study 2, 200 patients diagnosed with MDD were assessed with the MADRS in face-to-face interviews. Thereafter, 100 patients were assessed 3–14 days later, again via face-to-face-interviews, while the other 100 patients were assessed 3–14 days later via a telephone interview.

Results: Study 1: The MADRS and HDRS scores between patients with MDD and healthy controls differed significantly. Agreement between scoring of the MADRS and HDRS was high ($r=0.95$). Study 2: The intraclass correlation coefficient (test–retest reliability) was $r=0.944$ for the face-to-face interviews, and $r=0.959$ for the telephone interviews.

Conclusion: The present data suggest that the Persian MADRS has high validity and excellent test–retest reliability over a time interval of 3–14 days, irrespective of whether the second assessment was carried out face-to-face or via a telephone interview.

Keywords: major depressive disorders, Montgomery–Asberg Depression Rating Scale, validation, reliability

Introduction

Murray and Lopez[1] estimated that major depressive disorders (MDD) will be the third leading cause of health burden worldwide by 2020, suggesting therefore that MDD are among the most prevalent lifetime psychiatric disorders. Furthermore, Lockwood et al[2] reported that MDD are associated with chronic lifelong risk for recurrent relapse, and high morbidity, comorbidity, and mortality. In Iran, prevalence rates for MDD vary between 4.29%,[3] and 12.7%,[4] indicating therefore that, as in Western countries, MDD are a major health concern.

Although diagnoses are strictly based on internationally accepted classifications such as the International Classification of Diseases-10 (ICD-10) or the *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edition, to assess symptoms of MDD and illness severity, the scientific community essentially relies on two tools, the Hamilton Depression Rating Scale (HDRS)[5] consisting of 21 items, and the Montgomery–Asberg Depression Rating Scale (MADRS).[6] The ten items forming the latter scale assess the following symptoms: 1) apparent sadness; 2) reported sadness; 3) inner tension; 4) reduced sleep; 5) reduced appetite; 6) concentration difficulties; 7) lassitude;

8) inability to feel; 9) pessimistic thoughts; and 10) suicidal thoughts. Answers are given on a 7-point Likert scale ranging from 0 (= not at all) to 6 (= definitely), with higher scores reflecting higher symptoms of depression.

The aim of the present study was to evaluate the concurrent validity and reliability of the Persian/Farsi version of the MADRS. To do so, we conducted two separate studies comprising of 410 patients diagnosed with MDD and 100 healthy controls.

Methods

The two studies to evaluate the validity and reliability of the Persian/Farsi MADRS are described in more detail in the sections below. In both studies, the participants were fully informed about the study aims and the voluntary nature of participation. Furthermore, the participants were informed that all data would be gathered anonymously, and written informed consent was obtained. Both studies took place between spring and summer 2015 at the Frashchian Hospital, Hamadan University of Medical Sciences (Hamadan, Iran). Both studies were approved by the Review Board of the Hamadan University of Medical Sciences (Iran), and were performed in accordance with the ethical standards laid down in the Declaration of Helsinki.

Study 1
Samples
Patients

A total of 210 inpatients diagnosed with MDD (mean age: =34.44 years, standard deviation [SD] =11.65; 60.5% females) took part in the study. Psychiatrists and clinical psychologists not involved in the data analysis conducted the clinical interviews based on the Mini-International Neuropsychiatric Interview (MINI)[7] to ensure that only patients with MDD were enrolled in the study. The inclusion criteria for the patients were: 1) aged 18–65 years; 2) diagnosed with MDD by trained psychiatrists or clinical psychologists; 3) willing and able to participate in the study; and 4) gave written and informed consent. The exclusion criteria were: 1) not meeting the inclusion criteria; 2) unwilling or unable to participate in clinical interviews; and 3) psychiatric comorbidities such as substance abuse, bipolar disorders, anxiety disorders, personality disorders, or schizophrenia.

Healthy controls

A total of 100 healthy controls (mean age: =26.07 years, SD =6.13; 62% females) took part in the study. They were recruited via advertisements in the hospital and at the University of Hamadan (Iran). Again, psychiatrists and clinical psychologists not involved in the data analysis performed a clinical interview based on the MINI[7] to ensure that only psychopathologically healthy participants were enrolled. The inclusion criteria were: 1) aged 18–65 years; 2) no diagnosis of any psychiatric disorders, as assessed and confirmed by trained psychiatrists or clinical psychologists; 3) willing and able to participate in the study; and 4) gave written and informed consent. The exclusion criteria were: 1) not meeting the inclusion criteria; and 2) unwilling or unable to participate in clinical interviews.

Assessment tools
Montgomery–Asberg Depression Rating Scale

First, the MADRS[6] was translated from English into Farsi; we rigorously followed the procedure proposed by Brislin;[8] that is to say, the English items were translated into Farsi, and then back-translated into English by an independent translator. Consensus was reached on a final version that was subjected to the translation–retranslation process.[8] Thereafter, clinical psychologists and psychiatrists not otherwise involved in the data analysis conducted the clinical interview based on this final version of the MADRS. As described in the "Introduction" section, the MADRS consists of the following ten items: 1) apparent sadness; 2) reported sadness; 3) inner tension; 4) reduced sleep; 5) reduced appetite; 6) concentration difficulties; 7) lassitude; 8) inability to feel; 9) pessimistic thoughts; and 10) suicidal thoughts. Answers are given on a 7-point Likert scale ranging from 0 (= not at all) to 6 (= definitely), with higher scores reflecting more severe symptoms of depression.

Hamilton Depression Rating Scale

Psychologists and psychiatrists not involved in the data analysis assessed the depression severity of the patients with the HDRS (version with 21 items; Persian version, including psychometric indices).[5,9] The rating scale consisted of 21 items asking about symptoms related to depression, including low mood, suicidality, irritability, tension, loss of appetite, loss of interests, and somatic symptoms. Answers were given on different rating scales ranging from 3-, 4- or 5-point ratings: (eg, "insomnia early": 0= no difficulty falling asleep; 1= complains of occasional difficulty falling asleep – ie, more than 0.5 hours; 2= complains of nightly difficulty falling asleep), with higher scores reflecting more marked depressive symptoms. Scores were additionally categorized as follows:

0–7 points: no depressive symptom/remission; 8–17 points: mild depressive disorder; 18–24 points: moderate depressive disorder; 25 and more points: severe depressive disorder (Cronbach's alpha =0.88).

Statistical analysis

A series of Pearson's correlations was performed to explore the association between the MADRS and HDRS, both for the entire group, and separately for patients and healthy controls.

Next, two Student's t-tests were performed to calculate the differences in MADRS and HRDS scores between patients and healthy controls.

A binary logistic regression was performed to calculate sensitivity, that is, the number of participants correctly identified as patients, and specificity, that is, the number of participants correctly identified as healthy controls. The variable "patients vs controls" was the dependent variable, and MADRS score was the independent variable.

The level of significance was set at alpha <0.05. All statistical computations were performed with SPSS® 22.0 (IBM Corporation, Armonk, NY, USA) for Apple Mac®.

Results

MADRS and HDRS scores between patients and healthy controls

MADRS scores differed significantly between patients (mean score [M] =30.14; SD =11.54; Cronbach's alpha =0.90) and healthy controls (M =8.34, SD =5.25; Cronbach's alpha =0.92; t[308] =18.01, P<0.0001).

HDRS scores differed significantly between patients (M =37.58; SD =12.50; Cronbach's alpha =0.89) and healthy controls (M =6.23, SD =1.05; Cronbach's alpha =0.93; t[308] =21.04, P<0.0001).

Correlations between the MADRS and HDRS scores

The correlation coefficient between the MADRS and HDRS was r=0.92 for the entire sample, r=0.96 for patients, and r=0.88 for healthy controls.

Identifying patients and healthy controls based on the MADRS scores

Results from the binary logistic regression analysis (variable "patients vs healthy controls" as dependent variable and MADRS scores as independent variable) showed a sensitivity of 96% and a specificity of 97%, corresponding to an overall precision of 96.5%.

Discussion

The key finding of study 1 was that the scores of the MADRS, translated into Farsi to assess the symptoms of depression among patients diagnosed with MDD, very closely matched the scores of depressive symptoms as derived from a validated and established tool, the HDRS.[5,9] Furthermore, the MADRS differentiated with high sensitivity and specificity between patients and healthy controls.

Study 2

With study 1, the results were consistent with the evidence that the version showed acceptable levels of concurrent validity. The aim of study 2 was to measure the test–retest reliability using two different approaches, first a face-to-face interview, and second a telephone interview.

Sample

A total of 200 patients diagnosed with MDD (M =36.13 years, SD =12.24; 30% females) took part in the study. As in study 1, psychiatrists and clinical psychologists not involved in the data analysis performed the clinical interview based on the MINI[7] to ensure that only patients with MDD were enrolled in the study. The inclusion criteria for the patients were identical to study 1.

Procedure

After the diagnosis of MDD as described in study 1, psychiatrists and clinical psychologists not involved in the data analysis rated the patients' symptoms and symptom severity with the MADRS. Next, the patients were randomly assigned either to the face-to-face condition or to the telephone interview condition. In the first condition, psychiatrists and clinical psychologists not involved in the data analysis conducted a face-to-face interview with patients to assess the symptoms and symptom severity. In the second condition, psychiatrists and clinical psychologists not involved in the data analysis interviewed the patients via a phone call to make these assessments. The second interview took place 3–14 days after the first interview.

Statistical analysis

Two Student's t-tests were performed to compare MADRS scores at the beginning and at retest between the two study conditions (face-to-face vs telephone interviews). To assess the reliability of the test–retest, three separate intraclass coefficients (ICCs) were computed: one for the whole sample, one for the face-to-face interview, and one for the telephone

interview; in each case, scores of the first assessment were compared with scores of the second. Furthermore, three correlations were computed between the second MADRS scores and the time lapse (days) between the two assessments, again for the whole sample, and separately for the two study conditions.

Results

MADRS scores at the beginning; comparison between the two study conditions

Mean scores at the beginning of the study differed significantly between the two groups (t[198] =4.35, $P<0.001$; effect size: d=0.64= medium effect), with patients in the telephone interview having higher MADRS scores (M =37.97, SD =5.97; Cronbach's alpha =0.93) than patients in the face-to-face interview condition (M =32.37; SD =11.35; Cronbach's alpha =0.91).

MADRS scores at retest; comparison between the two study conditions

Mean scores at the retest of the study differed significantly between the two groups (t[198] =5.09, $P<0.001$; effect size: d=0.74= medium effect), with patients in the telephone interview showing higher MADRS scores (M =38.61, SD =6.11; Cronbach's alpha =0.90) compared with patients in the face-to-face interview condition (M =32.11; SD =11.22; Cronbach's alpha =0.90).

Test–retest reliability

The ICC for both groups was ICC =0.956; for the face-to-face-interview condition, ICC =0.944, and for the telephone condition ICC =0.959.

Time lapse of days between test and retest

The time lapse in days between test and retest was M =5.45, SD =0.98; correlation coefficients between the time lapse and the retest scores were 0.00, 0.01, and 0.01 (over the whole sample; face-to-face interview; telephone interview).

Discussion

The key finding of study 2 was that test–retest reliability was very high, irrespective of the method of retest (face-to-face interview vs telephone interview). Furthermore, the length of the interval between the test and retest was not associated with retest scores.

Discussion

The key findings of the present two studies were that the Farsi version of the MADRS closely matched scores of the

established and previously validated HDRS, that this new version clearly differentiated between patients diagnosed with MDD and healthy controls (thus, validity was high; study 1), and that test–retest was very stable, irrespective of whether the retest was conducted face-to-face or via a telephone interview (study 2).

Compared with the HDRS, the advantage of the MADRS is its brevity with regard to items (ten items vs 21/17 items in the case of the HDRS) and time (a few minutes vs up to 10 minutes), while the MADRS, as the HDRS, allows to fully assessing core symptoms and symptom severity and intensity of a MDD. Further, compared with the MADRS, the HDRS focuses more on anxiety and physical symptoms of distress, along with strictly psychiatric symptoms such as depersonalization, paranoid feelings, obsessional feelings, feelings of guilt, and agitation (however, the presence of agitation seems to be predictive of [poor] treatment outcomes for patients suffering from MDD).[10,11] In our view, a further additional advantage of the MADRS is the fixed scaling of seven points (from 0 [= not at all] through 6 [= definitively]), while scoring on the HDRS ranges across a smaller number of anchor points (usually from 0 [= not at all] to 4 [= definitively]), and varies from item to item.

The strengths of the present studies are the large samples of patients diagnosed with MDD, the inclusion of healthy controls, and the results (high validity and high reliability). Nonetheless, the following limitations should be considered: first, no self-ratings were made, although both the validity and reliability of the MADRS might have been improved had self- and experts' ratings been compared. Second, for the patient sample, only those diagnosed with MDD were enrolled in the study, and it would have been interesting and important to explore the extent to which this version of the MADRS could have provided an assessment of patients with bipolar disorders, dysthymia, or cyclothymia. Third, the time lapse between the test–retest ranged from 3–14 days; future studies might assess the retest reliability over a longer interval. Last, we did not apply advanced psychometric methods such as item response theory to examine, for example, the reliability levels at different levels of symptom severity.[12–14]

Conclusion

The pattern of results of the two separate studies showed that the Farsi version of the MADRS had high concurrent validity and test–retest reliability.

Acknowledgment

Disclosure
The authors report no conflicts of interest in this work.

References
1. Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet*. 1997; 349(9063):1436–1442.
2. Lockwood LE, Su S, Youssef NA. The role of epigenetics in depression and suicide: A platform for gene-environment interactions. *Psychiatry Res*. 2015;228(3):235–242.
3. Mohammadi MR, Davidian H, Noorbala AA, et al. An epidemiological survey of psychiatric disorders in Iran. *Clin Pract Epidemiol Ment Health*. 2005;1:16.
4. Sharifi V, Amin-Esmaeili M, Hajebi A, et al. Twelve-month prevalence and correlates of psychiatric disorders in Iran: the Iranian Mental Health Survey, 2011. *Arch Iran Med*. 2015;18(2):76–84.
5. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23:56–62.
6. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134:382–389.
7. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*. 1998;59 Suppl 20:22–33.
8. Haghighi M, Bajoghli H, Angst J, Holsboer-Trachsler E, Brand S. The Farsi version of the Hypomania Check-List 32 (HCL-32): applicability and indication of a four-factorial solution. *BMC Psychiatry*. 2011;11:14.
9. Shabani A, Akbari M, Dadashi M. Reliability and validity of the Bipolar Depression Rating Scale on an Iranian sample. *Arch Iran Med*. 2010;13(3):217–222.
10. Wollmer MA, Kalak N, Jung S, et al. Agitation predicts response of depression to botulinum toxin treatment in a randomized controlled trial. *Front Psychiatry*. 2014;5:36.
11. Jung S, Wollmer MA, Kruger TH. The Hamburg-Hannover Agitation Scale (H2A): Development and validation of a self-assessment tool for symptoms of agitation. *J Psychiatr Res*. 2015;69:158–165.
12. Levine SZ, Rabinowitz J, Rizopoulos D. Recommendations to improve the positive and negative syndrome scale (PANSS) based on item response theory. *Psychiatry Res*. 2011;188(3):446–452.
13. Levine SZ, Leucht S. Psychometric analysis in support of shortening the Scale for the Assessment of Negative Symptoms. *Eur Neuropsychopharmacol*. 2013;23(9):1051–1056.
14. Wilson JE, Niu K, Nicolson SE, Levine SZ, Heckers S. The diagnostic criteria and structure of catatonia. *Schizophr Res*. 2015;164(1–3): 256–262.