ORIGINAL ARTICLES
(CC BY-SA)
UDC: 159.98::616.89-079.4
https://doi.org/10.2298/VSP171017176V

Montgomery–Asberg depression rating scale in clinical practice: Psychometric properties on Serbian patients

Montgomeri–Ašbergova skala za procenu depresivnosti u kliničkoj praksi: psihometrijska svojstva na bolesnicima u Srbiji

Petar Vojvodić*, Ana Andonov†, Dejan Stevanović‡, Ivana Peruničić Mladenović§, Goran Mihajlović‖, Jovana Vojvodić*

***Clinic for Psychiatric Disorders "Dr Laza Lazarević", Belgrade, Serbia; Faculty of Media and Communications, †Department of Clinical Psychology, Belgrade, Serbia; ‡Clinic for Neurology and Psychiatry for Children and Youth, Belgrade, Serbia; §Institute for Mental Health, Belgrade, Serbia; Clinical Center of Kragujevac, ‖Clinic of Psychiatry, Kragujevac, Serbia**

Abstract

Background/Aim. Various rating scales for depression are avalable, but the Montgomery-Asberg Depression Rating Scale (MADRS) is one of the most frequently used scales. The aim of this study was to analyze the measurement properties of the MADRS Serbian version for quantifying depression severity in the clinical setting. **Methods**. Two studies have been conducted in order to validate the MADRS. The first study included sixty-four adult patients with major depressive disorder (MDD), with test-retest situation, and the second one included 19 participants (also with MDD), who had six test-retest situations. Psychometric evaluation included descriptive analysis, internal consistency and test-retest reliability, and concurrent validity (correlations with the Hamilton Depression Rating Scale 17 – HAMD-17). **Results.** The internal consistency for test-retest reliability was 0.93 in total for the MADRS, and for six test-retest situations was 0.95. The MADRS had one factor structure, with explained variance of 66.26% for the first testing, and 61.29% for the retest. There were statistical significant correlations between the MADRS and HAMD-17 ($r = 0.96$ for test and $r = 0.94$ for retest). Also, it was shown a great correlation between all items on the MADRS, and for the instrument in total ($r = 0.89$). **Conclusion**. The MADRS was shown good statistical results, and it could be used in everyday clinical practice for discriminating MDD.

Key words:
depression; sensitivity and specificity; severity of illness index; surveys and questionnaires.

Apstrakt

Uvod/Cilj. Razne skale za procenu depresije su dostupne, ali je Montgomeri-Ašbergova skala za procenu depresivnosti (MADRS) jedna od najviše korišćenih. Cilj istraživanja bio je da se analiziraju merne karakteristike srpske verzije MADRS za procenu ozbiljnosti depresije u kliničkim uslovima. **Metode.** Sprovedena su dva istraživanja kako bi se validirala MADRS. Prva studija obuhvatila je 64 odraslih bolesnika sa velikim depresivnim poremećajem (MDD), sa test-retest situacijom, a druga je obuhvatila 19 učesnika (takođe sa MDD), koji su imali šest test-retest situacija. Psihometrijska procena se bazirala na deskriptivnoj analizi, unutrašnjoj konzistenciji i test-retest pouzdanosti, kao i konkurentnoj validnosti (korelacije sa Hamiltonovom skalom za procenu depresivnosti 17 – HAMD-17). **Rezultati.** Interna konzistentnost za test-retest pouzdanost iznosila je 0,93 za ceo MADRS instrument, dok je za šest test-retest situacija iznosila 0,95. MADRS je pokazala jednofaktorsku strukturu, koja objašnjava 66,26% varijanse za prvo testiranje, odnosno 61,29% za retest. Utvrđena je statistički značajna korelacija između MADRS i HAMD-17 ($r = 0,96$ za test i $r = 0,94$ za retest). Takođe, utvrđena je značajna korelacija između svih stavki na MADRS pojedinačno, ali i za ceo instrument ($r = 0,89$). **Zaključak**. Skala MADRS pokazala je dobre statističke rezultate i mogla bi se koristiti u svakodnevnoj kliničkoj praksi za diskriminaciju MDD.

Ključne reči:
depresija; osetiljvost i specifičnost; bolest, indeks težine; ankete i upitnici.

Correspondence to: Petar Vojvodić, Clinic for Psychiatric Disorders "Dr Laza Lazarević", Višegradska 26, 11 000 Belgrade, Serbia. E-mail: petar.vojvodic@gmail.com

Introduction

The diagnostic code for major depressive disorder (MDD) is based on episodic course, current severity, presence of psychotic features, and remission status [1]. Quantifying MDD severity and defining remission in research and clinical settings is mainly based on symptom rating scales, which are self-ratings or administered by clinicians. Various rating scales for depression are available [2], but the Montgomery-Asberg Depression Rating Scale (MADRS) is one of the most frequently used scales to quantify severity in clinical trials and everyday clinical practice [3].

Accumulated evidence from studies with different groups of people with depressive disorders indicates that the MADRS has sound psychometric properties in terms of good internal consistency, test-retest stability, and convergent validity [4–7]. It was also shown that the MADRS total score has sound construct validity for an unidimensional measure targeting core depressive symptoms [4, 5], and it provides the most accurate reflection of depression severity in overall [7]. Some studies reported that the construct of the MADRS might be represented by two to three factors underlying different depressive symptoms, such as dysphoria, retardation, and vegetative symptoms [8, 9], which should be considered in evaluating depression treatment. Good reliability and validity were also reported for the MARDS in different language versions, such as Bangla [10], Brazilian [11], Chinese [12], French [13], Korean [14], Malay [15], Persian [16], Spanish [17, 18], and Thai [19].

Research on the compatibility of the scale between the original version of the MADRS showed that there is a moderate to high association between patient and physician results [13, 20]. Also, it was examined whether the results of the MADRS were better when it was done with or without a structured interview, and the results showed that the scale had satisfactory reliability, regardless of whether the structured interview was used or not [3]. Analyzing each item individually, the MADRS has all responsive responses and the end result is more sensitive to changes in treatment [21].

The MADRS shows greater sensitivity in distinguishing between moderate and severe depression compared to the Hamilton Depression Rating Scale (HAMD) (sensitivity 93.5%, specificity 83.3%) [22]. Also, in comparison with the HAMD, significantly higher results are obtained, and it is considered to be a calibration of the scope of both instruments, that is, that the results would be equated if the cut-off score for the MARDS depression was 12, instead of the original 6 [23]. Possible shortened versions for the HAMD and MADRS were also examined without items related to somatic symptoms (e.g. sleep, appetite, etc.) [24]. In case that only a rough screen is needed, shorts version of the instruments can be used, but if the scales are used for diagnostic purposes, then it is recommended to have a full version of both scales.

The translation into Serbian for the MADRS instrument was previonsly done twice in 2008 and in 2012.

The MADRS in Serbian language has not yet been standardized. The aim of this study was to analyze psychometric properties of the MADRS Serbian version in the clinical settings.

Methods

Study 1

Questionnaires

The first scale we used was the MADRS [3]. The MADRS is the clinician-rated 10-item scale with the following 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 to all items are given on the 7-point Likert scale ranging from 0 = not at all to 6 = definitively, with higher scores reflecting more severe depression symptoms. The total score is the sum of all answered items.

The second scale that was used was the HAMD, version with 17 items (HAMD-17 [25, 26]). The HAMD is the clinician-rated scale with the following items and response options: 1) depressed mood 0–4; 2) feelings of guilt 0–4; 3) suicide 0–4; 4) early insomnia 0–2; 5) middle insomnia 0–2; 6) late insomnia 0–2; (7) work and activities 0–4; 8) retardation 0–4; 9) agitation 0–4; 10) psychic anxiety 0–4; 11) somatic anxiety 0–4; 12) gastrointestinal somatic symptoms/appetite 0–2; 13) general somatic symptoms 0–2; 14) genital symptoms 0–2; 15) hypochondriasis 0–4; 16) loss of weight 0–2; and 17) insight 0–2. These symptoms are rated to cover the 1-week period prior to the interview. The total score is the sum of all answered items, with higher scores reflecting more severe depression. The HAMD had the internal consistency reliability of 0.90 in the present study.

Participants

All adults aged 18 year and above, admitted to daily hospital between June and September 2017, were eligible. The main inclusion criterion was the diagnosis of a unipolar MDD episode. Exclusion criteria were the presence of any other psychiatric and/or neurological disorder or a major somatic problem (e.g. chronic illness, impairment). All patients were diagnosed according to the International Classification of Diseases, 10th revision (ICD-10) [27] and to all was initiated some kind of treatment; antidepressant medications, social therapy, and/or psychotherapy.

A total of 64 patients, from which 36 (56.3%) were females and 28 (43.8%) males, participated in the research. Age of subjects varied from 24 to 68 years with mean of 46.11 [standard deviation (SD) = 10.85] years. The subject who were included in the study were only those who provided all the data, and only they were considered in each shown analysis.

Assessment

The MADRS was administered to all subjects independently by the first author. The same rater administered the HAMD-17. The MADRS and HAMD-17 were administered again to all subjects by the same rater two weeks later (test-retest assessment). Only subjects who appeared on the scheduled assessment after four weeks were assessed with the MARDS.

Psychometric analysis

The reliability assessment of the MADRS included internal consistency tested by the intraclass correlation coefficient (ICC), the two-way random method of absolute agreement [28]. Concurrent validity was assessed using the Pearson's correlation coefficient, and paired sample *t*-test for comparing between item results.

Study 2

Only the MADRS instrument, which characteristics were described previously, was used in the study 2. The administration of the instrument was done by the fist author, and unlike the study 1, where only one test and retest had been done, in this study six tests were done.

Participants

A total of 19 subjects participated from which 9 (47.4%) were females and 10 (52.6%) males. There was one dropout from the study, because the patient (female) was not shown to control after fourth administration. The age of the participants varied from 28 to 63 (mean = 47.32, SD = 11.06) years. Participants in this study had been also included in the study 1, but in that study were only subjected to the first two tests.

Psychometric analysis

The similar assessments were done like in the study 1: test-retest reliability by the interclass correlation coefficient, the Pearson's correlation for concurrent validity, and *t*-test for six testing situations.

Results

Study 1

Differences in impacts between items on the HAMD-17, and those on the MADRS were estimated with paired sample *t*-test. It was shown that there was statistically significant difference between several items (Tables 1 and 2).

Statistically significant results for both test and retest situations were observed in items listed in Table 2. According to the Cohen's *d*, we found that the HAMD-17 items 3, 6, 7, 10, 11, and 13 had small impact, item 9 had no significant impact, and only item 8 had moderate impact. In case of the MADRS, the Cohen's *d* showed that items 1, 6, 7, 8, and 10 had significant, but small impact, and the item 3 had no impact. For both instruments sums were statistically significant, and *d* had small effect size ($d = 0.31$ for the MADRS, and $d = 0.32$ for the HAMD-17).

The ICC for test-retest reliability was 0.93 in total [95% confidence interval (CI) 0.88–0.96; $p < 0.001$] for the

Table 1

The HAMD-17 test and retest results using *t*-test, ICC and reliability for each item

Number of item	Test	Retest	*t*	Cohen's d	*r*	ICC	95% CI	α	
	mean ± SD	mean ± SD						T1	T2
1	1.36 ± 1.44	1.11 ± 1.24	1.98	0.19	0.09	0.84*	0.73-0.90	0.90	0.89
2	0.86 ± 0.94	0.70 ± 0.85	1.86	0.18	0.09	0.84*	0.73-0.90	0.90	0.90
3	0.33 ± 0.69	0.16 ± 0.44	3.01*	0.29	0.15	0.82*	0.70-0.89	0.91	0.90
4	0.58 ± 0.75	0.48 ± 0.59	1.76	0.15	0.07	0.89*	0.82-0.93	0.90	0.89
5	0.50 ± 0.69	0.44 ± 1.31	0.36	0.06	0.03	0.21	-0.30.-0.52	0.91	0.90
6	0.55 ± 0.75	0.38 ± 0.58	3.01*	0.25	0.13	0.87*	0.79-0.92	0.91	0.89
7	1.65 ± 1.27	1.25 ± 1.19	3.40*	0.33	0.16	0.84*	0.74-0.90	0.90	0.89
8	0.98 ± 0.77	0.61 ± 0.68	5.20*	0.51	0.25	0.81*	0.69-0.89	0.91	0.89
9	0.86 ± 1.17	0.67 ± 0.99	2.55**	0.18	0.09	0.92*	0.87-0.95	0.91	0.89
10	1.05 ± 1.02	0.73 ± 0.84	4.07*	0.34	0.17	0.88*	0.80-0.93	0.91	0.89
11	0.80 ± 1.04	0.53 ± 0.85	3.28*	0.28	0.14	0.87*	0.78-0.92	0.91	0.89
12	0.30 ± 0.53	0.28 ± 0.52	0.30	0.04	0.02	0.81*	0.69-0.88	0.91	0.90
13	0.77 ± 0.81	0.61 ± 0.77	2.01**	0.20	0.10	0.82*	0.70-0.89	0.92	0.90
14	0.48 ± 0.64	0.41 ± 0.58	1.69	0.11	0.06	0.90*	0.84-0.94	0.91	0.90
15	0.31 ± 0.66	0.28 ± 0.65	0.47	0.05	0.02	0.80*	0.68-0.88	0.91	0.90
16	0.19 ± 0.47	0.16 ± 0.48	0.62	0.06	0.03	0.79*	0.65-0.87	0.91	0.90
17	0.09 ± 0.39	0.03 ± 0.18	1.43	0.20	0.10	0.48**	0.15-0.69	0.91	0.90
HAMD-17 total	11.71 ± 9.66	8.83 ± 8.50	4.75*	0.32	0.16	0.92*	0.88-0.95	0.94	0.95

***Correlation is significant at p < 0.01; **Correlation is significant at p < 0.05.**
HAMD – Hamilton Depression Rating Scale; ICC – intraclass correlation coefficient;
SD – standard deviaton; CI – confident interval.

Table 2

The MADRS test and retest results using *t*-test and ICC for each item

Number of item	Test	Retest	*t*	Cohen's d	r	ICC	95%CI	α	
	mean ± SD	mean ± SD						T1	T2
1	2.63 ± 1.83	1.94 ± 1.73	3.83*	0.39	0.19	0.81*	0.68-0.88	0.94	0.92
2	2.05 ± 1.89	1.69 ± 1.82	1.81	0.19	0.10	0.78*	0.63-0.86	0.93	0.91
3	1.80 ± 1.57	1.55 ± 1.44	2.29**	0.17	0.08	0.91*	0.85-0.94	0.94	0.92
4	1.61 ± 1.89	1.41 ± 1.67	1.85	0.11	0.06	0.94*	0.89-0.96	0.93	0.91
5	1.11 ± 1.72	0.88 ± 1.32	1.49	0.15	0.08	0.80*	0.67-0.88	0.94	0.92
6	1.98 ± 1.84	1.45 ± 1.60	3.82*	0.31	0.15	0.88*	0.81-0.93	0.93	0.91
7	2.27 ± 1.71	1.69 ± 1.58	3.84*	0.35	0.17	0.85*	0.75-0.91	0.93	0.91
8	1.88 ± 1.78	1.45 ± 1.55	2.68*	0.26	0.13	0.84*	0.73-0.90	0.93	0.91
9	1.39 ± 1.39	1.11 ± 1.20	1.99	0.22	0.11	0.76*	0.61-0.86	0.94	0.92
10	0.70 ± 1.14	0.30 ± 0.63	4.46*	0.43	0.21	0.81*	0.69-0.89	0.94	0.92
MADRS-tot	17.41 ± 13.67	13.45 ± 11.44	4.82*	0.31	0.16	0.93*	0.88-0.96	0.95	0.94

***Correlation is significant at p < 0.01; **Correlation is significant at p < 0.05. MADRS – Montgomery-Asberg Depression Rating Scale; ICC – intraclass correlation coefficient; CI – confident interval.**

MADRS, and 0.92 for the HAMD-17 in total (95% CI 0.88-0.95; $p < 0.001$). As for each item, all items on the the MADRS had significant and large impact (ICC = 0.76–0.94), and the HAMD-17 had similar results. Exception was the item 5, about transitory insomnia, where was no statistical significance. All other items had ICC values that were high and significant (ICC = 0.81–0.92). These results showed that both instruments were stable through time, and they could show changes in a patient's reaction in treatment of depression.

All items on the HAMD-17 showed significant reliability, with α = 0.89 or higher, and α = 0.91 or higher for the MADRS. According to George and Mallery [29], all α values above 0.7 are acceptable, 0.8 are good, and 0.9 are excellent. Following that rule, in this research it was shown that the MADRS had better reliability coefficients for each item than the HAMD-17, but the total scores showed similar reliability that was considered excellent (the HAMD-17: α = 0.94 for test, α = 0.95 for retest; the MADRS: α = 0.95 for test, and α = 0.94 for retest).

Table 3

Pearson's correlation for the HAMD-17 and MADRS for test and retest situations

HAMD-17 items		MADRS items	
df = (N-2)	0.01	df = (N-2)	0.01
1	0.726*	1	0.676*
2	0.723*	2	0.634*
3	0.761*	3	0.834*
4	0.825*	4	0.886*
5	0.141	5	0.687*
6	0.613*	6	0.800*
7	0.720*	7	0.735*
8	0.688*	8	0.723*
9	0.864*	9	0.624*
10	0.797*	10	0.807*
11	0.784*		
12	0.680*		
13	0.691*		
14	0.822*		
15	0.672*		
16	0.648*		
17	0.424*		
Sum(17)	0.866*	Sum(10)	0.878*

***Correlation is significant at p < 0.001 level; HAMD – Hamilton Depression Rating Scale; MADRS – Montgimery-Asberg Depression Rating Scale.**

The correlation analysis showed that there were high correlations between items on test and retest (Table 3). The MADRS had significant correlations for each item on test and retest, and coefficinet of correlation (*r*) varied from 0.62 to 0.89 ($p < 0.001$). The sum results also showed high correlation ($r = 0.89$, $p < 0.001$). Similar correlations were found also for the HAMD-17, with correlations between items on test and retest demonstrating significant correlations for all items except one (the item 5 for test and retest showed nonsignificant correlations). The coefficients of correlations varied from 0.42 to 0.86 ($p < 0.001$); for the sum, correlations were also significant ($r = 0.87$, $p < 0.001$). There were statistically significant correlations between the MADRS and HAMD-17. For the first testing, coefficient of correlation was $r = 0.96$ ($p < 0.001$), and for the retest, it was $r = 0.94$ ($p < 0.001$).

Factor analysis showed that it could be extracted one factor for both test and retest items (Table 4). For the test situation, it was shown that one factor explains 66.26% ofthe variance, and for the retest, it was explained by 61.29% of the variance. These results were as it was hypothesized, because it was supposed to be extracted one factor for the MADRS, supposing that it was measuring one factor – depression.

Study 2

The ICC for test-retest reliability was 0.95 in total (95% CI 0.90–0.98; $p < 0.001$), as it was shown in Table 5. All the items for six test-retest situations showed significance at the level $p < 0.001$, and the ICC varied from 0.77 to 0.95. This showed that with six tests, the MADRS still had good stability throughout time, at least for a period of one and a half month of the treatment in clinical conditions.

As for the reliability analysis, all six test had α = 0.91 or higher for each item, as it was the case for the sum results implying an excellent reliability by each item and in total for the MADRS (Table 5).

Table 4

Factor loadings and communalities for the MADRS based on a principal components analysis for 10 items, for both test and retest situations

Number of item	Test		Retest	
	factor loading	variance explained	factor loading	variance explained
	1		1	
1	0.731	0.534	0.772	0.596
2	0.878	0.771	0.814	0.662
3	0.810	0.656	0.747	0.558
4	0.858	0.735	0.854	0.730
5	0.751	0.564	0.617	0.381
6	0.907	0.822	0.837	0.701
7	0.817	0.667	0.839	0.705
8	0.849	0.721	0.868	0.753
9	0.738	0.545	0.668	0.446
10	0.781	0.610	0.773	0.597

***Note*: Every loading greater than 0.30 is considered significant.**
MADRS – Montgomery-Asberg Depression Rating Scale.

Table 5

The MADRS for six test-retest results using ICC and reliability for each item

Number of item	ICC	95% CI	α					
			T1	T2	T3	T4	T5	T6
1	0.93*	0.87-0.97	0.94	0.94	0.93	0.94	0.93	0.91
2	0.93*	0.87-0.97	0.94	0.94	0.93	0.94	0.93	0.92
3	0.77*	0.56-0.90	0.95	0.95	0.95	0.95	0.94	0.93
4	0.88*	0.77-0.95	0.95	0.95	0.94	0.95	0.94	0.94
5	0.92*	0.85-0.97	0.95	0.95	0.96	0.95	0.94	0.93
6	0.95*	0.90-0.98	0.95	0.94	0.94	0.94	0.93	0.92
7	0.95*	0.90-0.98	0.95	0.94	0.94	0.94	0.93	0.91
8	0.94*	0.88-0.97	0.94	0.94	0.93	0.94	0.93	0.92
9	0.93*	0.87-0.97	0.95	0.94	0.94	0.94	0.93	0.92
10	0.90*	0.81-0.96	0.96	0.95	0.95	0.95	0.94	0.93
MADRS-tot	0.95*	0.90-0.98	0.94	0.93	0.93	0.93	0.94	0.95

***Correlation is significant at p < 0.001.**
MADRS – Montgomery-Asberg Depression Rating Scale;
ICC – intraclass correlation coefficient; CI – confidence interval.

The correlation results showed that there were high correlations (Table 6). The MADRS had significant correlations for all six retests, and the coefficients of correlations varied from 0.51 to 0.98, with significance at $p < 0.01$ or $p < 0.05$. Higher correlations were shown for tests that had a closer time interval, unlike those that had more distant time interval. Also, higher correlations at the significance level $p < 0.01$ were shown in the first testing, and for the sixth retest showed smaller correlation at $p < 0.05$.

Discussion

The multivariable analysis showed that the MADRS possesses appropriate reliability and concurrent validity. The internal consistency reliability of the MADRS in Serbian language was high as well as corrected item-total correlations, what pictures high homogeneity among the items in measuring the intended concept and the consistency in rating the severity across the items even when considering individual assessments [28, 29]. The ICC for the study 1 was 0.93 in total (95% CI 0.88–0.96; $p < 0.001$) for the MADRS, and 0.95 in total for the study 2 (95% CI 0.90–0.98; $p < 0.001$). High internal consistency reliability for the MADRS total score, with Cronbach's alpha coefficient above 0.8, was previously observed across studies using the original and different language versions [5, 7, 16, 19]. In addition, the test-retest reliability of the MADRS in Serbian was excellent, for both studies 1 and 2, whereas in both α was 0.91 and higher, indicating satisfactory stability in repeated measurements.

The factor analysis showed that one factor explained most of the variance (66.26% of the variance for the first testing, and for the retest it was explained by 61.29% of the variance), as it was expected. Other studies have found more factors that could explain variance, that is, three [30], or two [31], depending on the study. This may be due to smaller sample size in our study, and these results should be confirmed in later research.

Finally, concurrent validity reported previously [11, 15, 16] was also evident for the MADRS total score for the Serbian version when tested against the HAMD-17 total score. The correlation of MADRS with HAMD was high and significant ($r = 0.96$; $p < 0.001$ for test, and $r = 0.94$; $p < 0.001$ for the retest). Other studies have shown smaller correlations, $r = 0.58$ [32]. Higher correlations in our research might be because of the smaller sample size, so the results might be different in the future research with bigger sample. There were also significant correlations between items on the MADRS (both for test and retest; $r = 0.62–0.89$, $p < 0.001$), and on the HAMD-17 (test and retest; $r = 0.42–0.86$, $p < 0.001$). Correlations between six tests in the study 2 also were significant ($r = 0.51–0.98$, mostly $p < 0.01$). Significant correlations were also between the MADRS and HAMD-17 ($r = 0.96$; $p < 0.001$ for the first test, and $r = 0.94$; $p < 0.001$ for the retest). These results have been confirmed in other studies, where the correlations exist between items, and between the MADRS and HAMD-17 [31, 33].

Table 6

The Pearson's correlation for the MADRS for six test-retest situations

Number of retest	1	2	3	4	5	6
1	-					
2	0.979*	-				
3	0.865*	0.889*	-			
4	0.703*	0.741*	0.899*	-		
5	0.581*	0.638*	0.775*	0.918*	-	
6	0.511**	0.553**	0.657*	0.812*	0.942*	-

***Correlation is significant at p < 0.01;**
****Correlation is significant at p < 0.05.**
MADRS – Montgomery-Asberg Depression Rating Scale.

Limitations

There are several limitations of the study. First, a small number of participants did not allow to study changes in mental health in those who deteriorated during the study period. Also, the samples in both studies were small and this limited the generalizability of the studies to other settings. Further research should include a bigger sample and also comparison to a general population, for the purpose of better validity testing. The bigger sample is referred to both studies, 1 and 2.

Conclusion

In summary, this study of the MADRS in Serbian demonstrated that it is appropriate measure for routine, clinical assessments of individuals with MDD. It showed that the measure could produce reliable and valid assessments of MDD severity with possibility to distinguish a clinically important improvement from measurement error with a large amount of certainty. However, with awareness of the limitations of the present study, additional investigations will be needed with different samples in order to set the MADRS as a gold standard in routine psychiatric practice.

R E F E R E N C E S

1. *American Psychiatric Association.* Diagnostic and statistical manual of mental disorders. (DSM–5). 5th ed. Washington, DC: American Psychiatric Association Publishing; 2013.
2. *Lam RW, Michalaak EE, Swinson RP.* Assessment scales in depression, mania and anxiety. Assessment scales in depression, mania and anxiety. London: CRC Press; 2004.
3. *Montgomery SA, Asberg M.* A new depression scale designed to be sensitive to change, Br J Psychiatry 1979; 134: 382–9.
4. *Bernstein IH, Rush AJ, Stegman D, Macleod L., Witte B, Trivedi MH.* A Comparison of the QIDS-C16, QIDS-SR16, and the MADRS in an Adult Outpatient Clinical Sample. CNS Spectr 2010; 15(7): 458–68.
5. *Carmody TJ, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D,* et al. The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. Eur Neuropsychopharmacol 2006; 16(8): 601–11.
6. *Davidson J, Turnbull CD, Strickland R, Miller R, Graves K.* The Montgomery-Åsberg Depression Scale: reliability and validity. Acta Psychiatr Scand 1986; 73(5): 544–8.
7. *Uher R, Farmer A, Maier W, Rietschel M, Hauser J, Marusic A*, et al. Measuring depression: comparison and integration of three scales in the GENDEP study. Psychol Med 2008; 38(2): 289–300.
8. *Bech P, Allerup P, Larsen ER, Csillag C, Licht RW.* The Hamilton Depression Scale (HAM-D) and the Montgomery-Åsberg Depression Scale (MADRS). A psychometric re-analysis of the European genome-based therapeutic drugs for depression study using Rasch analysis. Psychiatry Res 2014; 217(3): 226–32.
9. *Suzuki A, Aoshima T, Fukasawa T, Yoshida K, Higuchi H, Shimizu T,* et al. A three-factor model of the MADRS in major depressive disorder. Depress Anxiety 2005; 21(2): 95–7.
10. *Soron TR.* Validation of Bangla Montgomery Asberg Depression Rating Scale (MADRSB). Asian J Psychiatr 2017; 28: 41–6.
11. *Carneiro AM, Fernandes F, Moreno RA.* Hamilton depression rating scale and montgomery-asberg depression rating scale in depressed and bipolar I patients: psychometric properties in a Brazilian sample. Health Qual Life Outcomes 2015; 13: 42.
12. *Liu J, Xiang YT, Lei H, Wang Q, Wang G, Ungvari GS,* et al. Guidance on the conversion of the Chinese versions of the Quick Inventory of Depressive Symptomatology-Self-Report

(C-QIDS-SR) and the Montgomery-Asberg Scale (C-MADRS) in Chinese patients with major depression. J Affect Disord 2014; 152–154: 530–3.
13. *Bondolfi G, Jermann F, Rouget BW, Gex-Fabry M, McQuillan A, Dupont-Willemin A,* et al. Self- and clinician-rated Montgomery-Asberg Depression Rating Scale: evaluation in clinical practice. J Affect Disord 2010; 121(3): 268–72.
14. *Ahn YM, Lee KY, Yi JS, Kang MH, Kim DH, Kim JL*, et al. A validation study of the Korean-version of the Montgomery-Asberg depression rating scale. J Korean Neuropsychiatr Assoc 2005; 44(4): 466–76. (Korean)
15. *Yee A, Yassim AR, Loh HS, Ng CG, Tan KA.* Psychometric evaluation of the Malay version of the Montgomery- Asberg Depression Rating Scale (MADRS-BM). BMC Psychiatry 2015; 15: 200.
16. *Ahmadpanah M, Sheikhbabaei M, Haghighi M, Roham F, Jahangard L, Akhondi A,* et al. Validity and test-retest reliability of the Persian version of the Montgomery-Asberg Depression Rating Scale. Neuropsychiatr Dis Treat 2016; 12: 603–7.
17. *Cano JF, Gomez Restrepo C, Rondón M.* Validation of the Montgomery-Åsberg Depression Rating Scale (MADRS) in Colombia. Rev Colomb Psiquiatr 2016; 45(3): 146–55. (Spanish)
18. *Lobo A, Chamorro L, Luque A, Dal-Ré R, Badia X, Baró E.* Grupo de Validación en Español de Escalas Psicométricas (GVEEP). Validación de las versiones en español de la Montgomery-Asberg Depression Rating Scale y la Hamilton Anxiety Rating Scale para la evaluación de la depresión y de la ansiedad. Med Clín 2002; 118(13): 493–9.
19. *Satthapisit S, Posayaanuwat N, Sasaluksananont C, Kaewpornsawan T, Singhakun S.* The comparison of Montgomery and Asberg Depression Rating Scale (MADRS thai) to diagnostic and statistical manual of mental disorders (DSM) and to Hamilton Rating Scale for Depression (HRSD): validity and reliability. J Med Assoc Thai 2007; 90(3): 524–31
20. *Cunningham JL, Wernroth L, von Knorring L, Berglund L, Ekselius L.* Agreement between physicians' and patients' ratings on the Montgomery-Åsberg Depression Rating Scale. J Affect Disord 2011; 135(1–3): 148–53.
21. *Santen G, Danhof M, Della Pasqua O.* Sensitivity of the Montgomery Asberg Depression Rating Scale to response and its consequences for the assessment of efficacy. J Psychiatr Res 2009; 43(12): 1049–56.
22. *Müller MJ, Himmerich H, Kienzle B, Szegedi A.* Differentiating moderate and severe depression using the Montgomery-Asberg depression rating scale (MADRS). J Affect Disord 2003; 77(3): 255–60.
23. *Zimmerman M, Posternak MA, Chelminski I.* Derivation of a definition of remission on the Montgomery-Asberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. J Psychiatr Res 2004; 38(6): 577–82.
24. *Reijnders JS, Lousberg R, Leentjens AF.* Assessment of depression in Parkinson's disease: the contribution of somatic symptoms to the clinimetric performance of the Hamilton and Montgomery-Asberg rating scales. J Psychosom Res 2010; 68(6): 561–5.
25. *Hamilton M.* Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967; 6(4): 278–96.
26. *Williams JB.* A structured interview guide for the Hamilton Depression Rating Scale. Arch Gen Psychiatry 1988; 45(8): 742–7.
27. *World Health Organization.* The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. Geneva: World Health Organization; 2016.
28. *Streiner D, Norman G.* Health Measurement Scales: A Practical Guide to Their Development and Use. 4rd ed. Oxford: Oxford University Press; 2008.
29. *George D, Mallery P.* SPSS for Windows step by step: A simple guide and reference. 11.0 update. 4th ed. Boston: Allyn & Bacon; 2003.
30. *Benazzi F.* Factor analysis of the Montgomery Asberg Depression Rating Scale in 251 bipolar II and 306 unipolar depressed outpatients. Prog Neuropsychopharmacol Biol Psychiatry 2001; 25(7): 1369–76.
31. *Lobo A, Chamorro L, Luque A, Dal-Ré R, Badia X, Baró E* Grupo de Validación en Español de Escalas Psicométricas (GVEEP). Validation of the Spanish versions of the Montgomery-Asberg depression and Hamilton anxiety rating scales. Med Clin (Barc) 2002; 118(13): 493-9. (Spanish)
32. *Yi JS, Bae SO, Ahn YM, Park DB, Noh KS, Shin HK,* et al. Validity and Reliability of the Korean Version of the Hamilton Depression Rating Scale(K-HDRS). J Korean Neuropsychiatr Assoc 2005; 44(4): 456–65.
33. *Schmidtke A, Fleckenstein P, Moises W, Beckmann H.* Studies of the reliability and validity of the German version of the Montgomery-Asberg Depression Rating Scale (MADRS)]. Schweiz Arch Neurol Psychiatr (1985) 1988; 139(2): 51–65. (German)

Received on October 17, 2017.
Accepted on October 31, 2018.
Online First November, 2018.