

SHORT REPORT

Open Access

# Hamilton depression rating scale and montgomery–asberg depression rating scale in depressed and bipolar I patients: psychometric properties in a Brazilian sample

Adriana Munhoz Carneiro<sup>\*</sup>, Fernando Fernandes and Ricardo Alberto Moreno

## Abstract

**Background:** The Hamilton Depression Rating Scale (HAM-D) and the Montgomery–Asberg Depression Scale (MADRS) are used worldwide and considered standard scales for evaluating depressive symptoms. This paper aims to investigate the psychometric proprieties (reliability and validity) of these scales in a Brazilian sample, and to compare responses in bipolar and unipolar patients.

**Methods:** The sample comprised 91 patients with either bipolar I or major depressive disorder from a psychiatric institute at São Paulo, Brazil. Participants were recruited and treated by clinicians through the Structured Interview for DSM-IV criteria, and had previously been interviewed by a trained, blind tester.

**Results:** Both scales indicated good reliability properties; however, the MADRS reliability statistics were higher than those of the HAM-D for detecting initial symptoms of unipolar depression. Correlation between the tests was moderate. Despite demonstrating adequate validity, neither test achieved the levels of sensitivity and specificity required for identification of a cutoff score to differentiate bipolar I and unipolar patients.

**Conclusions:** Both scales demonstrate adequate reliability and validity for assessing depressive symptoms in the Brazilian sample, and are good options to complement psychiatric diagnosis, but are not appropriate for distinguishing between the two affective disorder types.

**Keywords:** Validity, Reliability, Rating scales, Bipolar disorder, Depression

## Background

The general term depression can be applied to a wide range of states, and is defined by symptoms that can be present in a number of different clinical or psychiatric conditions, associated with the use of psychoactive drugs, or even manifest under normal conditions (such as grief or sadness) [1]. By contrast, patients with major depressive disorder present with a group of consistent symptoms that occur throughout most of the day, almost every day, for at least 2 weeks [2].

As is the case for certain other psychiatric disorders, specific biological markers are absent for mood disorders, which means that diagnosis tends to be based on the presence of a syndrome, i.e., a group of relatively stable or recurrent signs and symptoms. Furthermore, some symptoms can be confounded with physical disorders, such as sleep disturbances and changes in weight or libido; these complicating factors facilitate underdiagnosis and confusion with bipolar disorder [3-7].

Consequently, the use of clinical rating scales is required to improve diagnosis quality, to reduce bias caused by physical symptoms, to assess prognosis during treatment, and to evaluate outcomes [8,9]. A large number of scales for evaluating depressive conditions are available, but the Hamilton Depression Rating Scale (HAM-D) [8] and the Montgomery–Asberg Depression

\* Correspondence: [adrianacarneiro01@gmail.com](mailto:adrianacarneiro01@gmail.com)

Mood Disorders Unit (Grupo de Distúrbios Afetivos- GRUDA), Department and Institute of Psychiatry, School of Medicine, Universidade de São Paulo (USP), Dr. Ovídio Pires de Campos St., 785 – 3rd floor -Ala Norte, Cerqueira César, São Paulo, SP Post code: 05403-010, Brazil

Rating Scale (MADRS) [10] are the most common rating scales in use, and are commonly applied to establish clinical criteria for distinguishing levels of severity and for measuring evolution of and recovery from a depressive episode [1].

Despite the differences in content, form of scaling and number of items, most studies that compare HAM-D and MADRS have indicated that the scales are correlated. Some studies reveal correlation coefficients between 0.65 and 0.94 [9,11-13], which suggests that the two instruments may actually measure slightly different aspects of depression. For this reason, we will describe the instruments in sequence.

The HAM-D was developed for administration with patients previously diagnosed with affective disorder, as a measure to quantify the results obtained during the clinical interview, but it has largely been used to assess the efficacy of antidepressive treatment [8,14]. Considered the "gold standard" of depression measures, the HAM-D serves as a point of reference for more recently developed scales. The HAM-D is a hetero-rated scale: it requires a trained rater with sufficient knowledge of the instrument and the symptoms of the depressive syndrome. In order to ensure fidelity, a standardized interview guide was created as part of the training [11,15].

Since first published by Hamilton [8], the scale has been largely applied, and many studies of it have been performed. Bagby et al. [11] used the MEDLINE database to access and review studies that used the 17-item version and that were published between the instrument's original release in 1979 and 2003. Results indicated good internal, interrater, and retest reliability estimates for the overall scale, but weak interrater and retest coefficients at the item level. As expected, evidence for good convergent, discriminant, and predictive validity was also found. Other studies have found coefficients ranging from 0.83 to 0.94. Interrater reliability of the scale proved consistent, exceeding 0.85 [16]. Nonetheless, the authors highlight some deficits of the scale, mainly due to Hamilton's conception of depression, which was created more than 40 years ago.

In Brazil, there are few studies about the psychometric proprieties of the HAM-D. The first study with the HAM-D-17 was conducted in the 1980s. In 1987, Dratcu, Ribeiro and Calil [17] assessed severity of patients' depressive levels with the HAM-D and compared these results with those of the MADRS. Results indicated a moderate correlation between the HAM-D and the MADRS, and low sensitivity of the HAM-D-17 for assessing severity of symptoms. Carvalho [18] designed a translation study, using the HAM-D with 63 bilingual undergraduate students; and Fleck et al. [19] examined the factorial structure of the HAM-D in 130 depressed

inpatients from France and Brazil. The authors found differences in the distribution of the factor solution, such that the Brazilian solution indicated four factors, one more than in the French version, as well as differences in the way anxiety items were distributed. No existing reliability studies with Brazilian samples were found for the HAM-D or the MADRS.

The MADRS items are designated to remedy the psychometric limitations of the HAM-D and to measure change during treatment. The scale's main advantage is its focus on core symptoms to the exclusion of somatic and psychomotor items. The MADRS has been translated into more than 24 languages. Reliability analyses have confirmed the ability to discriminate changes during treatment; interrater reliability coefficients have ranged from 0.89 when with a psychiatrist and a general practitioner and applied on a Swedish sample that lives on North America. Results indicated a reliability of 0.95 when applied by American researchers, 0.97 Swedish researchers, 0.97 with a Swedish and an English applicator, and 0.93 with an American psychiatrist with a nurse [10,20]. Zimmerman et al. [21] conducted a literature review of the MADRS to verify the cutoff for defining remission, using MEDLINE data from 1966 to 2002, resulting in ten studies. According to them,  $\geq 4$  was the best cutoff score to delineate significant symptoms of depression; whereas the majority of the literature defines an optimal cut off  $< 9$ . Zimmerman et al. [21] recommend caution in interpreting their data, however, because not all studies used the same inclusion criteria, and not all patients were being assessed for major depressive disorder (MDD). The previously mentioned work by Dratcu, Ribeiro and Calil [17] is the only Brazilian study on the MADRS.

Both scales have ample psychometric evidence that indicates they are valid measures of symptom outcome in major depression. Validity studies are primary conditions to a test, considering that ensures the degree to which an empirical evidence is supported [22,23].

Validity is often treated as a unitary concept, but the process of validation should be an ongoing process of accumulating various kinds of evidence [23]. Where translation of an instrument is concerned, the first form of evidence relates to content. A literal translation does not ensure that the test measures the same constructs as the original instrument did. Results obtained by Fleck et al. [19] comparing French and Brazilian samples demonstrated the importance of using appropriate statistical analysis, such as factor analysis, to examine and avoid translation threats. Variations in depression scores with the HAM-D have been reported in other studies, indicating the importance of considering cultural differences when validating scales [24,25]. Other important and common evidence is related to other variables, that

address the degree that measure are consistent with the construct, as in the study by Dractu et al. [17]; however, this study was performed more than 20 years ago, underscoring the dearth of Brazilian studies despite the scales' being the most commonly used worldwide.

As was mentioned earlier, no reliability study with a Brazilian sample was found. Reliability is an important component of verifying whether a test is able to measure the proposed construct; i.e., the scale should yield similar answers about the same or similar constructs [26]. In the case of the HAM-D and the MADRS, such reliability is particularly important for ensuring efficacy of clinical drug trials. Within this context, the present study aims to explore the psychometric properties (validity and reliability) of the HAM-D-17 and the MADRS. Secondly, it aims to compare responses of bipolar I and unipolar patients.

## Methods

### Participants

The sample comprised 91 adult patients experiencing a major depressive episode: 39 (43%) diagnosed according to APA criteria [27] with MDD and 52 (57%) with bipolar disorder type I. All participants were 18 to 59 years old ( $M = 33.74$ ;  $SD = 12.10$ ); 63.4% ( $n = 59$ ) were female. Participants were selected using a version of the *Structured Clinical Interview for DSM* (SCID) [28], which was administered by trained psychiatrists. Exclusion criteria were: presence of psychotic symptoms; comorbidities with other psychiatric disorder; and uncontrolled clinical diseases, such as hypothyroidism, diabetes, or hypertension; and who were receiving other psychiatric treatment.

### Instruments

The Structured Clinical Interview for DSM-IV Axis I Disorders – Clinical Version (SCID-CV) [28] was developed to standardize the psychiatric diagnostic procedures based on interview; the version that had been translated into Portuguese was utilized in the present study. The reliability study was performed in psychiatric patients from a hospital situated in São Paulo's interior. Test-retest methodology was employed using a two-day interval between interviews. Forty-five (45) patients, predominantly female (60%), with a mean age of 34.9 years ( $SD = 11.8$ ) participated. The degree of concordance for the diagnosis (Kappa) was 0.83, demonstrating that the instrument had good reliability.

Although the HAM-D [8] is regarded as the “gold standard” for assessing severity of depressive episodes in patients with mood disorders, it was not originally designed to be a diagnostic instrument for depression. Carvalho [18] translated the questionnaire version for use in Brazil by administering the back-translated

version to 63 bilingual university graduates. The 17-item version was employed in the present study. In this version, each item is scored from 0 to 2 or from 0 to 4; total scores can range from 0 to 52. The HAM-D scale was not originally designed with cutoff points to designate levels of severity of the depressive condition; therefore, we used Blacker's research [29] to define cutoff points and severity levels as follows:  $> 23 =$  very severe;  $19–22 =$  severe;  $14–18 =$  moderate;  $8–13 =$  mild; and  $< 7 =$  remission. The scale predominantly assesses cognitive and vegetative symptoms, with relatively few items related to social, motor, anxiety, and mood factors. Results are categorized as mild, moderate, or severe depression. Reliability of the instrument is based on non Brazilian studies, and lies in the 0.83–0.94 range [16].

The MADRS [10] is a 10-item semi-structured scale specifically designed to indicate the severity of the depressive condition. The validation study for use in Brazil was carried out by Dractu et al. [17], who compared the scale against the HAM-D and the Visual Analog Mood Scale in 40 depressed patients. The MADRS assesses mood symptoms exhibited over the preceding 2 weeks, scoring items from 0 to 6, to give a maximum total score of 60 points. Müller [30] proposed the following cutoff points for severity:  $0–8 =$  remission;  $9–17 =$  mild;  $18–34 =$  moderate; and  $>35 =$  severe. The items assess somatic, cognitive, vegetative and anxious symptoms. The inter-item reliability of the instrument in an international study was 0.86 [31].

### Procedures

Patients were recruited from the Institute of Psychiatry, located at the Clinical Hospital of the University of São Paulo, Brazil, School of Medicine, after approval of ethic committee CAPPesq (CAAE: 02222012.8.0000.0068). They were diagnosed according to the DSM-IV TR [27]. They were included in the study only after they had read, understood, and signed the Informed Consent Form. Participants were allocated randomly to one of two blocks via a sequence generated by a biostatistician, and were assessed individually by a blind tester trained to administer the scales. The scales took an average of 1 hour to administer. The patients were assessed with the scales prior to treatment (V0), and at four (V4) and eight (V8) weeks after start of treatment, according to the Lithium and Carbamazepine compared to lithium and Valproic acid in the treatment of young bipolar patients (LICAVAL) study [32].

### Statistical analysis

The reliability of the HAM-D-17 was determined on the following basis: (1) interrater reliability, calculated by the intraclass correlation coefficient; (2) internal consistency, calculated using Cronbach's alpha. Validity was assessed

as follows: (1) sensitivity and specificity among groups, using the Receiver Operation Characteristic (ROC) curve; (2) correlation between the HAM-D and the MADRS, using Spearman’s rank correlation; and (3) mean differences in response to MADRS and HAM-D items at V0, calculated by the Mann–Whitney test. All significance tests used  $p < .05$  as the minimal criterion.

**Results**

The internal consistency results for the HAM-D-17 total, according to Cronbach’s Alpha ( $\alpha$ ) were  $\alpha = 0.83$  (V0),  $\alpha = 0.71$  (V4) and  $\alpha = 0.85$  (V8), whereas the MADRS statistics were  $\alpha = 0.84$ ;  $\alpha = 0.70$  and  $\alpha = 0.80$ , respectively, showing excellent reliability of both instruments at first and last assessment and good reliability at V4 [26]. Results for the intraclass correlation coefficient are given in Table 1.

Table 1 shows that the intraclass coefficient of the scales ranged from 0.70 to 0.85, and that reliability declined at V4 in both groups for both scales. In the bipolar I group, a reduction in reliability on the HAM-D-17 with increased weeks of treatment was observed, while on the MADRS, this was evident only in comparison of V0 with V8. According to the reliability parameters adopted, scores over 0.60 indicate adequate reliability. Therefore, all scores qualified as good or excellent [26]. The ROC curve was plotted to investigate the sensitivity and specificity among groups and determine a possible cutoff point (Figure 1).

Results obtained with the ROC curve indicated that the area under the curve for all the variables ranged between 0.34 and 0.64, i.e., below the minimum criterion stipulated for differentiation (0.70). In addition, the curve indicated no definable cutoff score for discriminating symptoms of bipolar I patients from those of unipolar patients.

Group differences in responses to items at time of entry for treatment (V0) were assessed using the Mann–Whitney U test. Results are presented in Table 2.

Of the 17 items in the HAM-D-17, four (**Insomnia late, Somatic symptoms–general, Hypochondriasis, and Insight**) revealed significant group differences in mean scores [33,34]. According to mean scores, depressive patients reported more sleep-related difficulties and

somatic symptoms and some inappropriate worry about their health in their initial interview than did bipolar I patients, who reported more knowledge about their depressive state at time of entry. Unipolar subjects had higher overall scores than did bipolar I patients, although this difference was not statistically significant. The same analysis was performed on the results from the MADRS; results are described in Table 3.

Results of the Mann–Whitney U test showed statistically significant group differences on seven of the 10 items as well as on the total score of the MADRS, with the unipolar group consistently scoring higher. Item-specific comparisons showed that unipolar patients had greater sadness and apparent sadness, tension, psychomotor inhibition, reduced appetite, concentration difficulties, and inability to feel and total score. To conclude, correlation analyses between the scales were performed, in order to determine the degree of similarity of HAM-D and MADRS at the three administration time points. Results revealed correlations of  $r_s = 0.78$  ( $p = 0.01$ ) at V0,  $r_s = 0.84$  ( $p = 0.001$ ) at V4 and  $r_s = 0.76$  ( $p = 0.001$ ) at V8, i.e., excellent levels of correlation [33].

**Discussion**

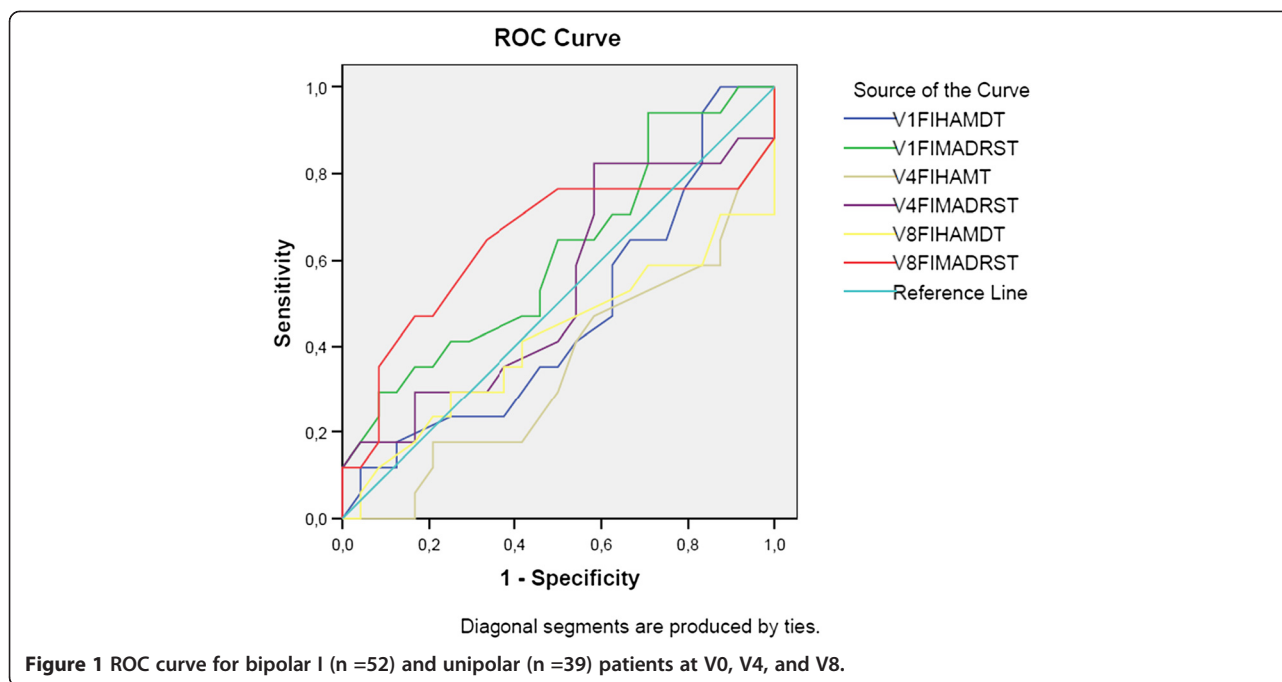
The aim of the present study was to investigate the psychometric properties of MADRS and HAM-D-17 in a Brazilian sample from São Paulo. The results showed good levels of reliability and validity for both scales, indicating that they were able to measure gravity of depressive symptoms. Reliability of the HAM-D ranged from good to excellent at the three administration time points, although few items served to distinguish the more specific symptoms of unipolar depressed subjects. By contrast, 70% of the MADRS items were able to differentiate unipolar depression versus bipolar I depression, and the reliability scores ranged from good to excellent for the three experimental time points.

The reliability of HAM-D-17, as measured by Cronbach’s alpha, was taken from previous data in the literature [8,11]. The mean scores in both groups proved similar to those reported by Rehm, Michael and O’Hara [12], who found statistically significant mean differences on items 10 (Anxiety, psychic), 15 (Hypochondriasis) and 17 (Insights), versus differences found in the present

**Table 1 Intraclass correlation coefficient of HAM-D-17 and MADRS for total sample (n = 91), bipolar I (n =52), and unipolar (n =39) patients**

Table	HAM-D-17			MADRS		
	V0	V4	V8	V0	V4	V8
TOTAL	0.80 (0.73–0.86)	0.71 (0.60–0.80)	0.85 (0.79–0.90)	0.84 (0.78–0.89)	0.70 (0.58–0.80)	0.80 (0.71–0.87)
Bipolar	0.78 (0.65–0.87)	0.73 (0.58–0.85)	0.63 (0.41–0.80)	0.75 (0.62–0.86)	0.79 (0.68–0.89)	0.73 (0.55–0.85)
Unipolar	0.83 (0.74–0.90)	0.70 (0.51–0.83)	0.85 (0.76–0.92)	0.86 (0.76–0.92)	0.85 (0.78–0.92)	0.83 (0.72–0.91)

CI =95%.



study on items 6 (Insomnia late), 13 (Somatic symptoms), 15 (Hypochondriasis), and 17 (Insight).

Concerning the MADRS, reliability was also as expected, showing scores in agreement with those found in other studies [9,10,21]. As already discussed, most items were able to differentiate unipolar from bipolar I subjects. Despite its extensive use, MADRS lacks published results in Brazil, precluding comparisons against the findings of the present study.

In addition, besides differences among items, it is important to note that the ROC curve failed to indicate a cutoff point for differentiating unipolar from bipolar I depression. It is known that bipolar I patients are often initially diagnosed as unipolar because they tend to seek medical assistance during the depressive phase [4]. For this reason, scales that are sensitive enough to discriminate the two conditions may be promising for assisting diagnosis by health professionals, and should be considered in further studies.

A secondary aim was to compare responses between the unipolar and bipolar I group. Regarding HAM-D score differences, the scale has determined that unipolar depressive patients have a greater tendency to weep and to exhibit more nonverbal expressions of sadness and hopelessness, according to other studies [3,6]. Other results relevant to MDD diagnosis are related to items 7 (work activities) and 8 (psychomotor retardation). Unipolar patients scored higher on these items, although the difference was statistically nonsignificant.

It is also pertinent to highlight some of the observed qualitative differences in depressive and bipolar I

symptomology, even those that were not statistically significant. According to the HAM-D results, depressed and bipolar I patients scored similarly on Guilt, Suicide and Psychic anxiety, and Loss of weight. This indicates that both patients frequently entertain ideas of guilt or rumination, and that sometimes these thoughts include the notion that life is not worth living and ideas about the possibility of death, but also excessive fear and concerns with minor matters. It is important to consider that the differences in this kind of ideation were not significant between groups, which is unexpected, considering that greater suicidal ideation is expected for depressed patients [3,7,16]. Additionally, bipolar I patients had a higher (but also nonsignificant) mean score on psychomotor agitation, which is a rating based on observation and indicates motor restlessness. This could be an important nonverbal sign to consider during interviews, and is supported by results from other studies [4-6]. Finally, regarding the statistical difference on item 17 (Insight), it is known that people with bipolar disorder are more likely to seek help when they are depressed than when they are experiencing mania or hypomania [5], so this is an expected result.

Concerning the MADRS, unipolar patients had greater Reported and Apparent sadness, Inner tension, Reduced appetite, Psychomotor inhibition and Inability to feel. It is notable that, of the symptoms listed in Table 3, pessimistic thoughts, Suicide thoughts, and Reduced sleep had higher (but not statistically significant) mean scores among unipolar than bipolar I patients.

**Table 2 Differences in mean scores on the HAM-D-17 between bipolar I (n = 52) and unipolar (n = 39) groups at time V0**

	DIAGNOSIS	N	Mean rank	Sum of ranks	U	z	p*
1. Depressed mood	Bipolar	39	33,95	1324,00	544	-1.917	0.055
	Unipolar	37	43,30	1602,00			
2. Guilt	Bipolar	39	37,10	1447,00	667	-0.613	0.540
	Unipolar	37	39,97	1479,00			
3. Suicide	Bipolar	39	37,19	1450,50	671	-0.619	0.536
	Unipolar	37	39,88	1475,50			
4. Insomnia early	Bipolar	39	36,96	1441,50	662	-0.690	0.490
	Unipolar	37	40,12	1484,50			
5. Insomnia middle	Bipolar	39	36,27	1414,50	635	-1.036	0.300
	Unipolar	37	40,85	1511,50			
<b>6. Insomnia late</b>	<b>Bipolar</b>	<b>39</b>	<b>34,04</b>	<b>1327,50</b>	<b>548</b>	<b>-2.210</b>	<b>0.027</b>
	<b>Unipolar</b>	<b>37</b>	<b>43,20</b>	<b>1598,50</b>			
7. Work and activities	Bipolar	39	35,13	1370,00	590	-1.397	0.162
	Unipolar	37	42,05	1556,00			
8. Psychomotor retardation	Bipolar	39	35,26	1375,00	595	-1.458	0.145
	Unipolar	37	41,92	1551,00			
9. Psychomotor agitation	Bipolar	39	41,42	1615,50	608	-1.334	0.182
	Unipolar	37	35,42	1310,50			
10. Anxiety, psychic	Bipolar	39	38,73	1510,50	713	-0.101	0.919
	Unipolar	37	38,26	1415,50			
11. Anxiety, somatic	Bipolar	39	35,96	1402,50	623	-1.082	0.279
	Unipolar	37	41,18	1523,50			
12. Loss of appetite (Somatic symptoms, Gastrointestinal)	Bipolar	38	34,18	1299,00	558	-1.741	0.082
	Unipolar	37	41,92	1551,00			
<b>13. Somatic symptoms, general</b>	<b>Bipolar</b>	<b>39</b>	<b>32,90</b>	<b>1283,00</b>	<b>503</b>	<b>-2.438</b>	<b>0.015</b>
	<b>Unipolar</b>	<b>37</b>	<b>44,41</b>	<b>1643,00</b>			
14. Sexual interest (Genital symptoms)	Bipolar	39	34,05	1328,00	548	-1.955	0.051
	Unipolar	37	43,19	1598,00			
<b>15. Hypochondriasis</b>	<b>Bipolar</b>	<b>39</b>	<b>34,95</b>	<b>1363,00</b>	<b>583</b>	<b>-1.976</b>	<b>0.048</b>
	<b>Unipolar</b>	<b>37</b>	<b>42,24</b>	<b>1563,00</b>			
16. Loss of weight	Bipolar	39	38,85	1515,00	708	-0.202	0.840
	Unipolar	37	38,14	1411,00			
<b>17. Insight</b>	<b>Bipolar</b>	<b>39</b>	<b>43,28</b>	<b>1688,00</b>	<b>535</b>	<b>-3.057</b>	<b>0.002</b>
	<b>Unipolar</b>	<b>37</b>	<b>33,46</b>	<b>1238,00</b>			
TOTAL SCORE	Bipolar	38	34,34	1305,00	564	-1.475	0.140
	Unipolar	37	41,76	1545,00			

\*significant considering 2-tailed curve; CI: 95%.

Considering these results, the MADRS items appear to be more appropriate for assessing unipolar depression, in view that discriminated more significant differences between bipolar and unipolar patients. Finally, a high correlation was found between the two instruments, and was similar to those reported by Dractu, Ribeiro and Calil [17]. Importantly, the sample used

by those authors included only unipolar subjects. Müller et al. [14] used a similar sample size (n =40) and found a correlation of ( $r_s = 0.66$ ;  $p > 0.005$ ). Together with previous data, our findings suggest that the HAM-D and the MADRS are adequate scales for measuring depressive symptoms before and during treatment.

**Table 3 Differences in mean scores on the MADRS between bipolar I (n = 52) and unipolar (n = 39) groups at time V0**

	DIAGNOSIS	N	Mean rank	Sum of ranks	U	z	p*
Reported sadness	Bipolar	39	29,12	1135,50	356	-3.880	0.000
	Unipolar	37	48,39	1790,50			
Apparent sadness	Bipolar	39	30,74	1199,00	419	-3.212	0.001
	Unipolar	37	46,68	1727,00			
Inner tension	Bipolar	39	29,76	1160,50	381	-3.669	0.000
	Unipolar	37	47,72	1765,50			
Reduced sleep	Bipolar	39	37,94	1479,50	700	-0.241	0.810
	Unipolar	37	39,09	1446,50			
Reduced appetite	Bipolar	39	31,73	1237,50	458	-3.021	0.003
	Unipolar	37	45,64	1688,50			
Concentration difficulties	Bipolar	40	34,40	1376,00	556	-1.978	0.048
	Unipolar	37	43,97	1627,00			
Psychomotor inhibition	Bipolar	39	28,36	1106,00	326	-4.266	0.000
	Unipolar	37	49,19	1820,00			
Inability to feel	Bipolar	40	30,80	1232,00	412	-3.441	0.001
	Unipolar	37	47,86	1771,00			
Pessimistic thoughts	Bipolar	40	36,15	1446,00	626	-1.227	0.220
	Unipolar	37	42,08	1557,00			
Suicidal thoughts	Bipolar	40	35,92	1437,00	617	-1.385	0.166
	Unipolar	37	42,32	1566,00			
Total score	Bipolar	40	26,79	1071,50	252	-3.172	0.002
	Unipolar	24	42,02	1008,50			

\*Sig (2-tailed); CI= 95%.

## Conclusions

The Brazilian version of the HAM-D and MADRS scales showed good reliability (coefficients ranging from 0.71 to 0.85) and intraclass correlation, as well as evidence of discriminant and convergent validity (based on the relationship with other variables). The scales proved consistent for assessing depressive symptoms, but limitations in their ability to discriminate unipolar and bipolar I patients were found. This result was expected only for the HAM-D, but was also found for the MADRS. Nevertheless, and consistent with other findings, the MADRS has proven to be a more reliable tool than the HAM-D for assessing depressive symptoms. Further investigations should replicate the present study with larger samples and incorporate scales that may be more sensitive for detecting the different spectrums in order to reduce current limitations.

Moreover, is important to highlight that the MADRS and HAM-D have a number of particularities that should be considered by the clinician, including the number of constituent items and the distribution of symptoms. For instance, the HAM-D-17 contains somatic items that are not included in the MADRS, which

for its part, represents a good option in cases where assessment of somatic symptoms of depression is not required, and where comparisons of observed and reported symptoms is desired. Therefore, although the scales constitute useful tools that can help reduce health professionals' subjectivity, they should be considered as supplemental to rigorous clinical-psychological evaluation.

## Abbreviations

BD: Bipolar disorder; MDD: Major depressive disorder; HAM-D: Hamilton Depression Rating Scale; MADRS: Montgomery-Asberg Depression Rating Scale; ROC: Receiver operating characteristic.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

AMC designed the study, performed the statistical analyses and interpretation, and contributed to writing and reviewing the manuscript. RAM designed the study and managed the statistical analyses. FF interpreted the data and drafted the manuscript and contributed to reviewing it. All authors have read and approved the final manuscript.

## Acknowledgments

The authors would like to thank Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP for supporting this project. This study is part of a Master's degree, so I would like to thank the Grupo de Distúrbios Afetivos -GRUDA team for contributing to my research.

Received: 3 August 2014 Accepted: 11 March 2015

Published online: 02 April 2015

## References

- Snaith RP. Present use of the Hamilton Depression Rating Scale: observation on method of assessment in research of depressive disorders. *Br J Psychiatry: J Mental Sci.* 1996;168(5):594–7. PubMed Epub 1996/05/01. eng.
- American Psychiatric Association[APA]. *Diagnostic and Statistical Manual of Mental Disorders- Fifth edition. DSM 5.* Washington DC: American Psychiatric Publishing; 2013.
- Judd LL. The clinical course of unipolar major depressive disorders. *Arch Gen Psychiatry.* 1997;54(11):989–91. doi:10.1001/archpsyc.1997.01830230015002.
- Bowden CL. A different depression: clinical distinctions between bipolar and unipolar depression. *J Affect Disorders.* 2005;84(2–3):117–25. doi:10.1016/S0165-0327(03)00194-0.
- Hirschfeld RM. *Psychiatric Management, from "Guideline Watch: Practice Guideline for the Treatment of Patients With Bipolar Disorder, 2nd Edition"*. [http://psychiatryonline.org/pb/assets/raw/sitewide/practice\\_guidelines/guidelines/bipolar-watch.pdf](http://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/bipolar-watch.pdf). Accessed on January 1, 2015.
- Mitchell PB, Frankland A, Hadzi-Pavlovic D, Roberts G, Corry J, Wright A, et al. Comparison of depressive episodes in bipolar disorder and in major depressive disorder within bipolar disorder pedigrees. *Br J Psychiatry.* 2011;199:303. doi:10.1192/bjp.bp.110.088823.
- Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases and decrements in health: results from the World Health Surveys. *Lancet.* 2007;370(9590):851–8. [http://dx.doi.org/10.1016/S0140-6736\(07\)61415-9](http://dx.doi.org/10.1016/S0140-6736(07)61415-9).
- Hamilton M. A rating scale for depression. *J Neurosurg Psychiatry.* 1960;23:56–62. doi:10.1136/jnnp.23.1.56.
- Cusin C, Yang H, Yeung A, Fava M. Rating scales for depression. In: Baer, MA, Blais (eds.), *Handbook of Clinical Rating Scales and Assessment in Psychiatry and Mental Health, Current Clinical Psychiatry.* 2009. doi:10.1007/978-1-59745-3.
- Montgomery AS, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry.* 1979;134:382–9. doi:10.1192/bjp.134.4.382.
- Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has true gold standard become a lead weight? *Am J Psychiatr.* 2004;161:2163–77. doi:10.1176/appi.ajp.161.12.2163.
- Rehm LP, O'Hara MW. Item characteristics of the Hamilton Rating Scale for Depression. *J Psychiatric Res.* 1983;19(1):31–41. Rehm LP, O'Hara MW. Item characteristics of the Hamilton Rating Scale for Depression. *Journal of Psychiatric Research.* 1983; 19(1): 31–41. [http://dx.doi.org/10.1016/0022-3956\(85\)90066-4](http://dx.doi.org/10.1016/0022-3956(85)90066-4).
- Iannuzzo RW, Jaeger J, Goldberg JF, Kafantaris V, Sublette ME. Development and reliability of the HAM-D –D/MADRS Interview: an integrated depression symptom rating scale. *Psychiatry Res.* 2006;145:21–37. doi:10.1016/j.psychres.2005.10.009.
- Müller MJ, Himmerich H, Szegedi A. Differentiating moderate and severe depression gradations for the Montgomery-Asberg Depression Rating Scale. *J Affect Disord.* 2003;77:137–40. [http://dx.doi.org/10.1016/S0165-0327\(02\)00120-9](http://dx.doi.org/10.1016/S0165-0327(02)00120-9).
- Williams JBW. Standardizing the Hamilton Depression Rating Scale: past, present and future. *Eur Arch Psychiatry Clin Neurosci.* 1988;2:116–12. doi:10.1007/BF03035120.
- López-Pina JA, Sánchez-Meca J, Rosa-Alcázar AI. The Hamilton Rating Scale for Depression: a meta analytic reliability generalization study. *Int J Clin Health Psychol.* 2009;9(1):143–59.
- Dractu L, da Ribeiro CL, Caili HM. Depression assessment in Brazil: the first application of the Montgomery-Asberg Depression Rating Scale. *Br J Psychiatry.* 1987;150:797–800. PubMed Epub 1987/06/01. eng. doi:10.1192/bjp.150.6.797.
- de Carvalho FR, Lima MG, Azevedo RCG, Soares RC, Caetano D. Tradução do inglês para o português do Questionário de Auto-avaliação da Escala de Hamilton para a depressão. *J Bras Psiquiatr.* 1993;42(5):255–60.
- de Fleck MPA, Chaves MLF, Poirier-Littré MF, Bourdel MC, Henri L, Guelfi JD. Depression in France and Brazil: factorial structure of the 17 item Hamilton Depression Scale in Inpatients. *J Nerv Mental Disord.* 2004;192(2):103–10.
- Stuart A, Montgomery N, Smeyatksy MR, Montgomery DB. Profiles of antidepressant activity with the Montgomery –Asberg depression rating scale. *Acta Psychiatr Scand.* 1985;72(320):38–42. doi:10.1111/j.1600-0447.1985.tb08073.x.
- Zimmerman M, Chelminski I, Posternak MC. Implications of using different cut-offs on symptom severity to define remission from depression. *Int Clin Psychopharmacol.* 2004;19:1–7. doi:10.1097/01.yic.0000130232.57629.46.
- Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull.* 1955;52:281–302.
- American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME) *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association. 1999.
- Cheng AT. Case definition and culture: are people all the same? *Br J Psychiatry.* 2001;179:1–3. doi:10.1192/bjp.179.1.1.
- Simon GE, VonKorff M, Piccinelli M, Fullerton C, Ormel J. An international study of the relation between somatic symptoms and depression. *N Engl J Med.* 1999;341:1329–35. doi:10.1056/NEJM199910283411801.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297–334.
- American Psychiatric Association[APA]. *Diagnostic and Statistical Manual of Mental Disorders, 4th edition, text revision (DSM-IV-TR).* Washington DC: American Psychiatric Publishing; 2000.
- Del-Ben CM, Vilela JAA, Crippa JAS, Hallak JEC, Labate CM, Zuardi AW. Test-retest reliability of the Structured Clinical Interview for DSM-IV - Clinical Version (SCID-CV) translated into portuguese. *Rev Bras Psiquiatr.* 2001;23:156–9. <http://dx.doi.org/10.1590/S1516-44462001000300008>.
- Blacker D. *Psychiatric rating scales.* In: Sadock B, Sadock V, editors. *Comprehensive textbook of psychiatry*, vol. 7th ed. Philadelphia: Lippincott Williams & Wilkins; 2000. p. 755–83.
- Müller-Thomsen T, Arlt S, Mann U, Maß R, Ganzer S. Detecting depression in Alzheimer's disease: evaluation of four different scales. *Arch Clin Neuropsychol.* 2005;20:271–6. doi:10.1016/j.acn.2004.03.010.
- Corruble E, Purper D, Payan C, Guelfi J. Inter rater reliability of two depression scales, MADRS and DRRS, based on videotape records of structured interviews. *Eur Psychiatry.* 1998;13(5):264–6. doi:10.1016/S0924-9338(98)80032-1.
- Campos RN, Costa LF, Bio DS, Soeiro de Souza MG, Demetrio FN, Moreno DH, Moreno RA LICAVAL: combination therapy in acute and maintenance treatment of bipolar disorder. *Trials.* 2010;23:11–72. doi:10.1186/1745-6215-11-72.
- Zar JH. Significance testing of the Spearman Rank Correlation Coefficient. *J Am Stat Assoc.* 1972;67(139):578–80. doi:10.1080/01621459.1972.10481251.
- Hamilton M. Comparative value of rating scales. *Br. Journal of Clinical Pharmacology.* 1976;Supp 58–60. doi:10.1111/j.1365-2125.1976.tb03714.x.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

