

Predicting response to treatment and discriminating bipolar and depression symptoms using Hamilton Depression Rating Scale

Preditor de resposta ao tratamento e discriminação de pacientes com depressão maior e transtorno bipolar pela Escala Hamilton de Depressão

Adriana Munhoz Carneiro¹, André Cavalcanti¹, Lucas de Francisco Carvalho², Ricardo Alberto Moreno¹

ABSTRACT

Objective: The present study aims to compare the diagnostic ability of the HAMD 17 items with shorter versions of 7 and 6 items. **Methods:** A total of 133 patients from a 6 month clinical trial diagnosed with mood disorders (60.2% with Major Depressive Disorder and 39.8% with bipolar type I disorder) were recruited. **Results:** The 17 items HAMD scale showed similar results as compared with shorter versions. Furthermore, almost all patients' diagnosed with Major Depressive Disorder scored more compared to Bipolar Disorder, but the difference was not significant. **Conclusion:** This study allows that the use of a shorter version of HAMD might be an adequate possibility, and also that depressive symptoms were similar among groups.

Keywords

Depressive disorder, psychiatric status rating scales, psychometrics, follow-up studies.

RESUMO

Objetivo: O presente estudo tem por objetivo investigar a habilidade diagnóstica da HAMD 17 comparada a versões breves de 7 e 6 itens. **Métodos:** Foram selecionados 133 participantes recrutados de um ensaio clínico (60,2% com Depressão Maior e 39,8% com Transtorno Bipolar tipo I). **Resultados:** A versão de 17 itens da HAMD mostrou resultados similares quando comparada às versões breves. Ademais, quase todos os pacientes diagnosticados com Depressão Maior apresentaram maiores pontuações, contudo, essa diferença não foi estatisticamente significativa. **Conclusão:** Este estudo demonstra que o uso de uma versão breve da HAMD pode ser uma possibilidade adequada e que os sintomas depressivos se mostraram similares entre os grupos de bipolares e depressivos.

Palavras-chave

Depressão, escalas de graduação psiquiátrica, psicometria, estudos de seguimento.

Received on
1/10/2017
Approved on
7/22/2017

DOI: 10.1590/0047-2085000000161

1 University of Sao Paulo, School of Medicine, Institute and Department of Psychiatry, Mood Disorders Unit (GRUDA).
2 Universidade Sao Francisco, Laboratory of quantitative measurement.

Address for correspondence: Adriana Munhoz Carneiro
Psychiatry Institute, Mood Disorders Unit
Dr. Ovidio Pires de Campos St., 785
3° floor North, Ceapesq, room 12
05403-010 – Sao Paulo, SP, Brazil
Telephone: 55-11-26616648
Fax: 55-11-26617894
Email: adriancarneiro01@gmail.com

INTRODUCTION

Mood disorders rating scales are essential for research as for clinical practice¹. In clinical trials, there is a need of efficacious measures procedures to evaluate drug efficacy compared to placebo or a standard gold drug. In clinical setting, rating scales are required as an effort to study psychometric properties of patients that can help clinician to evaluate the patient in a more objectively and standardized way². Therefore, the need of tools to assess depressive symptoms was needed to establish a universal language of what could be understood by depression. Then, Max Hamilton developed a scale consisting of 17 descriptors initially³, which was subsequently validated in different countries⁴⁻⁸, among them Brazil which confirm its ability to auxiliary in the diagnosis depressive disorder and severity detection⁹⁻¹¹.

Regarding to psychometric properties, there is a non-consensus about reliability coefficients^{1,12,13}. Bagby *et al.*¹² found that HAMD-17 is psychometrically and conceptually inconsistent. Using MEDLINE database of 70 selected studies published between 1979 – 2003 to evaluate psychometric properties they found adequate internal reliability, but poor interrater and retest reliability; and validity results indicated poor indices on content validity but an adequate result on convergent discriminant validity. These findings suggest that psychometric results are inconclusive to determine HAMD-17 consistency to evaluate depression and that more studies are necessary. In contrast, Trajković *et al.*¹ performed a meta-analysis (data obtained from a review of 409 articles available in MEDLINE and PsycINFO between the years 1960-2008) to detect the reliability properties of the HAMD-17 indicating good properties in patients with depression as a primary diagnosis and also with comorbidities and additionally, a good internal consistency index were found.

The multidimensional configuration of HAMD is a recurring point of discussion on the literature. Some argues that depression is a result of multiple causes that makes unfair the definition of depression as a one-dimensional condition. This way, a scale with a good coverage is appropriated when assumes that depression is a result of a set of clinical features, however, the scale must also generate one-dimensional subscales for each clinical feature, in order to evaluate clinical results of pharmacological studies; on the other hand, some researchers criticizes that subscale scores could implies in outperformed results, due to the unspecific items to asses depression symptoms, for example, hypochondriasis and anxiety, that affects directly on depression severity estimates and perceived changes during depressive episode¹³⁻¹⁸. Therefore, a short version of HAMD to improve psychometric characteristics and consequently to reduce outperformed results has been developed, to focus in core symptoms and exclude items that are related with medication effects or comorbidities.

Comparing unidimensional HAMD-17 and MADRS six items core symptoms subscales, Maier *et al.*¹⁹ found that internal consistency and sensitivity to change were similar and recommended the use of smaller versions for clinical practice. Faries *et al.*²⁰ studying the utility of unidimensional core symptoms of HAMD versus full HAMD concluded that inadequate forms to measure effectiveness can lead to the use of larger samples and to increase the time of recruitment, financial costs, complexity in the execution of the study and less validity. A study comparing the 17 and 6 item version of HAMD indicated that the six-item scale has a strongly relationship with 17 items in baseline and at endpoint of 143 MDD patients (double and melancholic depression) in four antidepressant drug treatment trial. Results lead authors concluded that the six-item scale is apparently sensitive to changes over time as the 17 items scale²¹. Bech *et al.*¹⁵ found that a reduced six-item HAMD scale, results on a unidimensional measure of severity of depressive, states episodes and is more sensitive to change than the full scale.

Isacsson and Adler²² reanalyzes data from six randomized clinical trials to investigate whether HAMD six-item subscale might explain findings of low/absent efficacy of antidepressants drugs in 597 patients with MDD with HAMD score less than eight points. Comparing the HAMD 17 and the 6 item scale, they conclude that the six-item has better perform to explain a large variance and that the HAMD 17 item provides unreliable data, such as, low effect sizes and sensitivity to changes. They suggest that HAMD-17 could be inappropriate to assess depression severity.

Helmreich *et al.*²³ investigates the predictive ability for treatment outcome of HAMD-17 item and a HAMD core depressive symptoms version, which includes depressed mood, physic anxiety, low self-esteem, feeling of guilt and work and activities; other items as, suicide, agitation, retardation, somatic anxiety, general somatic symptoms and libido were included separately. The study was based on a severe 210 MDD outpatients evaluated at inclusion, baseline and after two weeks of antidepressant treatment and the improvement criteria was a reduction of equal to or more than 20% after baseline evaluation. An analysis of sensibility was performed to indicate if the predictive capacity of Toronto 7 item scale and Evans 6 item scale was similar, nonetheless, all subscales showed good sensitivity (80-96%) and moderated specificity (36-54%). Additionally, the best discriminative item between stable remitter and non remitter was *work and activities* and *depressed mood*. The authors concluded that core versions could be considered appropriated to predict outcomes as good as 17 item version. The use of 17 item version was recommended only on baseline and week 2 to predict response or treatment failure in the early phase of treatment, and Toronto and Evans scale in the subsequent weeks. Finally, they suggest the need of more prospective studies comparing HAMD-17 and six-

time subscales to investigate its ability to predict treatment outcome. Therefore, studies about psychometric properties and clinical implications of shorter versions of HAMD on trials indicated that those scales might be good options for clinical trials, however, clinical data are insufficient. Our study aims to investigate psychometric properties (validity and reliability) of HAMD-17 compared with shorter versions to identify the better form to detect changes during treatment, and secondarily, to investigate differences between MDD and bipolar depressions.

METHODS

This is a follow-up study reporting data obtained from the LICAVAL protocol²⁴, collected during a six month clinical trial conducted in a Psychiatric Institute in Brazil. A total of 133 patients from a Mood Disorders Unity (67.7%, $n = 90$ were females; mean age = 34.50 ± 12.45 ; and single 36.1% $n = 48$) were included. Patients were previously diagnosed with Bipolar Disorder Type I –BD I ($n = 53$) and Major Depressive Disorder – MDD ($n = 80$) according to DSM IV TR²⁵ criteria and Structured Clinical Interview – SCID I²⁶, applied by trained psychiatrists. Exclusion criteria were: neurological disorders, previous head trauma, any illness requiring medical intervention, had undergone electroconvulsive therapy in the preceding six months, suicidality, comorbidities such as anxiety disorders or substance abuse and psychotic symptoms.

Assessment: the Hamilton Depression Rating Scale – HAMD were used to evaluate depressive symptomatology assessed by trained clinicians prior to treatment (V0) and at weeks four (V4) and eight (V8) after treatment. In order to compare psychometric properties, the items of HAMD 17 were transformed into the following subscales: Bech melancholia scale^{14,15}, Gibbons global depression severity¹³ and Toronto Scale²⁷. Additionally, patients were assessed with the Montgomery-Åsberg Depression Rating Scale¹⁸ (MADRS), to serve as a criteria for compare response.

Procedures: patients were recruited from a tertiary outpatient Mood Disorders Unit. Bipolar and MDD patients were diagnosed according to the DSM-IV TR²⁵ and included in the study after they had read, understood, and signed the Informed Consent Form, according to the Lithium and Carbamazepine compared to lithium and VALproic acid in the treatment of young bipolar patients (LICAVAL) protocol study²⁴. Patients were assessed with the scales which took an average of 30 minutes to apply at entering phase (V0), two weeks (V2) and four weeks after treatment (V4).

Analysis: statistical analyses were conducted with SPSS 20.0. Reliability coefficients were calculated using Cronbach's Alpha indices, Spearman-Brown prophecy²⁸ for sensibility and specificity; validity was assessed by relationship with internal variables (correlation between items and by diagnosis) and

predictive validity (regression analysis). The significance level was set at 0.05 and all tests were two tailed.

RESULTS

Internal consistency reliability coefficients (α) were calculated considering the four different versions at V0, V2 and V4 stages (data are presented in Table 1).

Table 1. Internal consistency reliability coefficients

	V0	V2	V4
	α	α	α
HAMD 17	.78	.82	.72
HAMD 7	.68 (.67)	.75 (.74)	.71 (.68)
HAMD6a	.61 (.58)	.69 (.63)	.59 (.64)
HAMD 6b	.44 (.51)	.68 (.68)	.58 (.53)
MADRS	.79	.75	.84

The correction by spearman-brown prophecy is presented inside the parentheses; HAMD: Hamilton Depression Scale; MADRS: Montgomery-Asberg Depression Rating Scale.
V0: entering on treatment; V2: after two weeks of treatment; V4: after four weeks of treatment.

The more reliable stage was at V2 and the least was at V0, and MADRS showed the best internal consistency coefficient at V0, HAMD-17 at V2 and MADRS again at V4. The spearman-brown prophecy did not increase the reliability coefficients. Figure 1 shows area under the curve (AUC) for all HAMD versions and MADRS and Table 2 their respective values.

Visually, Figure 1 shows that MADRS draw the best AUC in the three stages, which is confirmed in Table 2. In general, the HAMD version with better AUC was 7-items, but always less than MADRS; for V0 the 17-items showed the worst performance. Table 2 shows the statistic results for HAMD versions and MADRS.

The results indicated that the sensitivity and specificity for distinguishing MDD and BD were low in both versions of HAMD and in MADRS scale in all applications, even though, the MADRS showed better results. Furthermore, there was a tendency to sensitivity exceed specificity.

Correlations between the versions of the HAMD with MADRS were high, with few occurrences presenting magnitude lower than 0.80. It should also be considered that the correlation magnitudes seem to show a tendency to increase during the applications. In addition, the logistic regression was calculated, in order to determine logistica Table 3 the t test and Cohen's d are presented, verifying mean differences between groups.

Bipolar group showed expressive differences ($d \geq .20$) in 7-items and 6 a-items Hamilton versions and in Montgomery, in almost all cases, the MDD means were higher than Bipolar. The differences in the stages were similar, but the difference between MADRS at V0 and V4 was statistically significant to discriminate bipolar and Major Depressive Disorders.

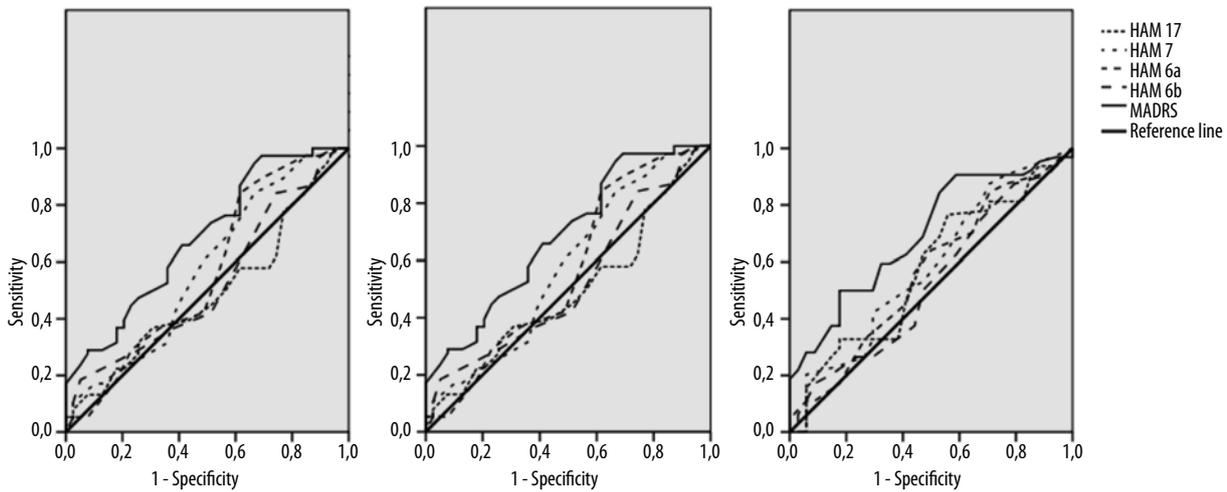


Figure 1. Area under the Curve (AUC) for HAMD versions and MADRS. HAMD: Hamilton Depression Scale; MADRS: Montgomery-Asberg Depression Rating Scale.

Table 2. AUC values for Hamilton (HAMD) and Montgomery Depression Rating Scale (MADRS)

Tests	AUC (95% CI); p	Raw	Ss (95% CI)	Sp (95% CI)	PPV	NPV	
V0	HAM 17	.49 (.38-.60); .89	13	84.2 (74.0-91.6)	15.4 (5.71-29.8)	65.3 (55.0-76.6)	33.3 (13.3-59.0)
	HAM 7	.57 (.46-.69); .19	10	84.2 (72.8-90.7)	33.3 (17.0-47.6)	70.3 (59.8-79.4)	48 (27.8-68.7)
	HAM 6a	.56 (.44-.68); .27	9	84.2 (74.0-91.6)	38.5 (22.7-54.2)	71.9 (61.4-80.9)	55.5 (35.3-74.5)
	HAM 6b	.52 (.41-.64); .65	8	84.2 (74.0-91.6)	25.6 (12.7-41.2)	68.1 (57.7-77.3)	45.4 (24.4-67.8)
	MADRS	.68 (.58-.79); .01	20	86.8 (77.3-93.5)	38.5 (23.3-55.4)	73.3 (63-82.1)	60 (38.7-78.9)
V2	HAM 17	.42 (.31-.53); .20	13	37.1 (25.9-49.5)	43.7 (27.2-62.1)	57.8 (42.1-72.3)	25.4 (15-38.4)
	HAM 7	.42 (.31-.54); .21	10	31.4 (20.8-43.6)	59.4 (39.3-73.4)	59.4 (42.1-75.2)	29.4 (19-41.7)
	HAM 6a	.42 (.29-.54); .19	9	32.9 (22.1-45.1)	53.1 (36.6-71.2)	59 (42.1-74.4)	28.8 (18.3-41.2)
	HAM 6b	.43 (.31-.55); .26	8	34.3 (23.3-46.6)	53.1 (36.6-71.2)	60 (43.3-75.1)	29.2 (18.6-41.8)
	MADRS	.55 (.43-.67); .37	20	44.3 (32.4-56.6)	59.4 (42.1-77.1)	70.4 (54.8-83.2)	33.9 (22.1-47.4)
V4	HAM 17	.57 (.45-.69); .24	13	32.8 (21.6-45.7)	61.8 (44.9-78.5)	61.7 (43.5-77.8)	33.8 (22.6-46.6)
	HAM 7	.58 (.46-.70); .19	10	32.8 (21.6-45.7)	70.6 (53.7-85.3)	67.7 (48.6-83.3)	36.7 (25.4-49.3)
	HAM 6a	.56 (.44-.68); .31	9	32.8 (21.6-45.7)	73.5(53.7-85.3)	67.7 (48.6-83.3)	36.7 (25.4-49.3)
	HAM 6b	.52 (.40-.65); .67	8	26.6 (16.3-39.1)	70.6 (50.7-83.1)	60.7 (40.6-78.5)	33.8 (23-46)
	MADRS	.69 (.59-.80); .01	20	50 (36-61.1)	82.4 (65.5-93.2)	84.2 (68.7-94)	45.1 (32.5-58.3)

In bold the test with better sensibility and specificity. V0: entering on treatment; V2: after two weeks of treatment; V4: after four weeks of treatment.

DISCUSSION

Considering the importance of HAMD scale to assess depression, this study aims to compare the diagnostic capacity of HAMD versions compared with MADRS. The psychometric properties of HAMD scale has been questioned, and considered as an imprecise scale. This way, tentatives to refine this scale was proposed, using versions based on core symptoms^{8,12,13}.

In the present study, HAMD 17 scores were reliable and satisfactory, and this could be comparable to Trajković *et al.*¹ study, that related good properties in depression patients (primary diagnosis or with comorbidities) and with Bagby *et*

*al.*¹² revision, that found a Cronbach's Alpha ($\alpha > 0.70$) and the internal reliability Referring to short versions, satisfactory; however, the 7 item version was the most adequate. It is also important mentioning that MADRS as HAMD showed results that confirm their reliability and the idea that this scale are designed to be particularly sensitive to change in patients with antidepressant medication¹⁸, although, HAMD was the best option at V0 and MADRS during treatment (V2 and V4). This result might indicate that, during treatment, when core symptoms are more easily to be detected, MADRS is more indicated to detect change differences.

Applying the area under the curve (AUC) for HAMD versions and MADRS to analyze the sensitivity of both scales

Table 3. Groups comparison using Hamilton (HAMD) and Montgomery Depression Rating Scale (MADRS)

		Groups	N	M (SD)	t (df)	p	d
V0	HAMV17	Bipolar	40	20.10 (7.31)	.27 (114)	.79	.05
		MDD	76	19.71 (7.49)			
	HAMV7	Bipolar	40	11.90 (4.51)	-1.24 (114)	.21	.24
		MDD	76	12.89 (3.85)			
	HAMV6a	Bipolar	40	9.87 (3.65)	-1.36 (114)	.17	.27
		MDD	76	10.72 (2.92)			
	HAMV6b	Bipolar	40	9.47 (3.13)	-.69 (114)	.49	.14
		MDD	76	9.89 (3.09)			
	MADRS	Bipolar	40	21.17 (10.05)	-3.01 (108)	.003	.59
		MDD	76	27.30 (10.44)			
V2	HAMV17	Bipolar	40	13.44 (6.09)	0.945 (102)	.34	.20
		MDD	76	11.95 (8.10)			
	HAMV7	Bipolar	40	8.62 (4.35)	1.173 (103)	.24	.24
		MDD	76	7.50 (4.78)			
	HAMV6a	Bipolar	40	7.68 (3.92)	1.439 (103)	.15	.30
		MDD	76	6.55 (3.72)			
	HAMV6b	Bipolar	40	6.77 (3.63)	1.176 (103)	.24	.24
		MDD	76	5.84 (3.89)			
	MADRS	Bipolar	40	17.42 (10.18)	-.842 (101)	.40	.18
		MDD	76	19.28 (10.58)			
V4	HAMV17	Bipolar	40	10.31 (6.43)	-.999 (97)	.32	.21
		MDD	76	11.64 (6.25)			
	HAMV7	Bipolar	40	6.45 (4.85)	-1.212 (97)	.23	.26
		MDD	76	7.62 (4.42)			
	HAMV6a	Bipolar	40	5.74 (3.83)	-1.059 (97)	.29	.22
		MDD	76	6.57 (3.70)			
	HAMV6b	Bipolar	40	5.54 (3.73)	-.507 (97)	.61	.11
		MDD	76	5.95 (3.90)			
	MADRS	Bipolar	40	12.02 (8.12)	-3.078 (98)	.003	.65
		MDD	76	18.4 (10.60)			

V0: entering on treatment; V2: after two weeks of treatment; V4: after four weeks of treatment.

to predict cases, results indicated that full version showed the worst performance, whereas HAMD 7 and MADRS, the best. Sensitivity reflects how much the scale is effective to identify correctly individuals who don't have depression (false-negative results), this way, the MADRS and HAMD 7 version are the most indicated rating scales to evaluate depressive individuals on this sample. Nonetheless, reliability results for short versions indicated that that scales has good sensitivity and specificity scores²³ and HAMD was not sensitive to specify depression symptoms as expected^{12,13,17}.

Additionally, in contrast to the other subscales the McIntyre *et al.*⁸ subscale perform the best reliability and

sensitivity scores, and was the best version to predict results as HAMD 17, and showed the best correlation indices. So far, MADRS was the unique scale able to differentiate depressive symptoms between groups, showing that at entering time (V0) and after one month after medication (V4) the scale was able to detect and differentiate groups, indicating that MDD group has more symptoms than BD.

CONCLUSION

In conclusion, some limitations should be punctuated. First, only cases with MDD moderate to severe was considered,

while BD patients could present results were considered and included with mild depression. This can contribute for interpreting depressive symptoms among groups. Second point is that there are limitations about the sample size. Considering that was applied classical methods of analyzes that could be inflated by the “n”, we try to reduce this bias with other methods, as r prophecy and *Cohen D*.

Despite this limitations, these findings indicates that core versions could be considered appropriated to assess depressive symptoms among ambulatory patients as good options considering time to administrate scales and for clinical trials, principally HAMD 7 version item. Concerning to 17 item version, results indicated a satisfactory internal reliability, but a unsatisfactory sensibility to discriminate false cases. Making an allowance for MADRS, it is important to consider that this scale was the one that presented results that are more reliable.

INDIVIDUAL CONTRIBUTIONS

Adriana Munhoz Carneiro, Lucas de Francisco Carvalho – Contributed to conception and design, analysis and interpretation of data.

Ricardo Alberto Moreno – Had substantially contributed to drafting the article or revising it critically for important intellectual content.

André Cavalcanti – Had given the final approval of the version to be published.

CONFLICTS OF INTEREST

The author declare that there is no conflict of interest related to this research.

Clinical trial registration: NCT00976794

ACKNOWLEDGEMENTS

This work was supported by grants from the *Fundação de Amparo à Pesquisa do Estado de São Paulo* (Fapesp). AMC and RAM received research grants from Fapesp.

REFERENCES

1. Trajković G, Starčević V, Latas M, Leštarević M, Ille T, Bukumirić Z, et al. Reliability of the Hamilton Rating Scale for Depression: a meta-analysis over a period of 49 years. *Psychiatry Res.* 2011;189:1-9.
2. Iannuzzo RW, Jaeger J, Goldberg JF, Kafantaris V, Sublette ME. Development and reliability of the HAM-D/MADRS Interview: an integrated depression symptom rating scale. *Psychiatry Res.* 2006;145:21-37.
3. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960;23:56-62.
4. Faries D, Herreraa J, Rayamajhi B, DeBrotta D, Demitracka M, Pottera WZ. The responsiveness of the Hamilton Depression Rating Scale. *J Psychiatr Res.* 2000;34:3-10.
5. Fleck MP de A, 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 Ment Dis.* 2004;192:103-10.
6. Kunigi H, Koga N, Hashikura M, Noda T, Shimizu Y, Kobayashi T, et al. Validation of computer-administered clinical rating scale: Hamilton Depression Rating Scale assessment with Interactive Voice Response technology--Japanese version *Psychiatry Clin Neurosci.* 67:253-8.
7. López-Pina JA, Sánhez-Meca J, Rosa-Alcázar AI. The Hamilton Rating Scale for Depression: a meta analytic reliability generalization study. *Int J Clin Health Psychol.* 2009;9:143-59.
8. McIntyre RS, Konarski JZ, Mancini DA, Fulton KA, Parikh SV, Grigoriadis S, et al. Measuring the severity of depression and remission in primary care: validation of the HAMD 7 scale. *CMAJ.* 2005;11:1327-34.
9. Dractu L, Ribeiro L da C, Calil HM. Depression assessment in Brazil: the first application of the Montgomery-Asberg Depression Rating Scale. *Br J Psychiatry.* 1987;150:797-800.
10. 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.
11. Freire MA, Figueiredo VLM de, Gomide A, Jansen K, Silva RA da, Magalhaes PVS, et al. Escala Hamilton: estudo das características psicométricas em uma amostra do sul do Brasil. *J Bras Psiquiatr.* 2014;63:281-9.
12. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry.* 2004;161:2163-77.
13. Gibbons RD, Clark DC, Kupfer DJ. Exactly what does the Hamilton Depression Rating Scale measure? *J Psychiatr Res.* 1993;27(3):259-73.
14. Bech P, Gram LF, Dein E, Jacobsen O, Vitger J, Bolwig TG. Quantitative rating of depressive states. *Acta Psychiatr Scand.* 1975;51:161-70.
15. Bech P, Allerup P, Gram LF, Reisby N, Rosenberg R, Jacobsen O, et al. The Hamilton Depression Scale: evaluation of objectivity using logistic models. *Acta Psychiatr Scand.* 1981;63:290-9.
16. Bech P, Allerup N, Reisby N, Gram LF. Assessment of symptom change from improvement curves on the Hamilton depression scale in trials with antidepressants. *Psychopharmacology.* 1984;84:276-81.
17. Isacsson G, Adler G. Randomized clinical trials underestimate the efficacy of antidepressants in less severe depression. *Acta Psychiatrica Scandinavica.* 2012;125:453-9.
18. Montgomery AS, Åsberg M. New Depression Scale Designed to be Sensitive to Change. *Br J Psychiatry.* 1979;134:382-9.
19. Maier W, Philipp M, Heuser I, Schlegel S, Buller R, Wetzel H. Improving depression severity assessment--I. Reliability, internal validity and sensitivity to change of three observer depression scales. *J Psychiatr Res.* 1988;22:3-12.
20. Faries D, Herrera J, Rayamajhi J, DeBrotta D, Demitrack M, Potter WZ. The responsiveness of the Hamilton Depression Rating Scale. *J Psychiatr Res.* 2000;34:3-10.
21. Hooper CL, Bakish D. An examination of the sensitivity of the six item Hamilton Rating Scale for Depression in a sample of patients suffering from major depressive disorder. *J Psychiatry Neurosci.* 2000;25:178-84.
22. Isacsson G, Adler M. The Hamilton Depression Rating Scale cannot reliably measure mild or moderate depression. *Eur Psychiatry.* 2012;1:1.
23. Helmreich I, Wagner S, König J, Kohnen R, Szegedi A, Hiemke C, et al. Hamilton depression rating subscales to predict antidepressant treatment outcome in the early course of treatment. *J Affect Disord.* 2015;175:199-208.
24. Campos RN, Costa LF, Bio DS, de Souza MG, Garcia CRL, Demétrio FN, et al. LICAVAL: combination therapy in acute and maintenance treatment of bipolar disorder. *Clinical Trials.* 2010;11:72. Registration NCT00976794.
25. American Psychiatric Association [APA]. Diagnostic and statistical manual of mental disorders: DSM-IV-TR, American Psychiatric Publishing, Inc. 2002.
26. First MB; Spitzer RL, Williams JB. Structured Clinical Interview for DSM-IV Axis I Disorders Scid-I. Washington, DC: American Psychiatric Press, Inc. 1996.
27. McIntyre RS, Konarski JZ, Mancini DA, Fulton KA, Parikh SV, Grigoriadis S, et al. Measuring the severity of depression and remission in primary care: validation of the HAMD-7 scale. *CMAJ.* 2005;173:1327-34.
28. Brown JD. Testing in language programs. Upper Saddle River, NJ: Prentice Hall Regents, 1996.