

## ORIGINAL ARTICLE

# SIGMA-VB: Validity and reliability of the Brazilian Portuguese version of the Montgomery-Åsberg Depression Rating Scale using the Structured Interview Guide for the MADRS

Fernando Fernandes,<sup>1</sup>  Adriana Carneiro,<sup>1</sup> Rodolfo N. Campos,<sup>2</sup> Marcio G. Soeiro-de-Souza,<sup>1</sup> Vivian B. Barros,<sup>3</sup> Ricardo A. Moreno<sup>1</sup>

<sup>1</sup>Grupo de Estudos de Doenças Afetivas (GRUDA), Departamento de Psiquiatria, Faculdade de Medicina, Universidade de São Paulo (USP), São Paulo, SP, Brazil. <sup>2</sup>Departamento de Saúde Mental e Medicina Legal, Faculdade de Medicina, Universidade Federal de Goiás (UFG), Goiânia, GO, Brazil. <sup>3</sup>Faculdade de Saúde Pública, USP, São Paulo, SP, Brazil.

**Objective:** The Montgomery-Åsberg Depression Rating Scale (MADRS) is widely used to assess depression severity. The Structured Interview Guide for the MADRS (SIGMA) was created to standardize MADRS assessment. The objective of this study was to translate and validate the original SIGMA into a Brazilian Portuguese version (SIGMA-VB).

**Methods:** We translated and cross-culturally validated the original SIGMA into the SIGMA-VB, and assessed its psychometric properties using data from 93 adult outpatients enrolled in the Integral Assessment in Unipolar Depression (AIUNI) trial. Participants were assessed by two raters on five visits over 8 weeks. We calculated multiple interrater reliability indexes for the SIGMA-VB and used the Hamilton Depression Rating Scale (HAM-D) for validation purposes.

**Results:** According to the SIGMA-VB, participants had moderate depression at baseline followed by mild depression at 8 weeks. We found over 90% of correlation between scores attributed by different raters using the SIGMA-VB. Correlations between the SIGMA-VB and the HAM-D were above 66%.

**Conclusion:** Our findings confirm that the SIGMA-VB is a valid and reliable instrument to assess depression severity in clinical research and practice. Its interrater reliability was similar to that of a previously published Japanese version of the SIGMA.

**Keywords:** Psychiatric status rating scales; depression; depressive disorder; validation studies as topic

## Introduction

The Montgomery-Åsberg Depression Rating Scale (MADRS) is a 10-item questionnaire widely used to assess depression severity in clinical trials and practice.<sup>1-3</sup> This instrument addresses several limitations of the Hamilton Depression Rating Scale (HAM-D), which was considered the gold standard for decades.<sup>4,5</sup> Although the HAM-D is a reliable and valid measure of depression severity, many problems with the scale have been described, including overrepresentation of vegetative symptoms and underrepresentation of atypical symptoms.<sup>6</sup> Furthermore, some items of the HAM-D have low interrater and retest coefficients.<sup>4</sup> Compared to the HAM-D, the MADRS has more precision in estimating depression and greater capacity to differentiate between responders and non-responders to antidepressants.<sup>7-9</sup>

Many studies in different settings have shown the MADRS to be valid, reliable, and sensitive to change.<sup>10-12</sup> Its sensitivity has been confirmed particularly in clinical noninferiority trials, where the difference in response among treatment arms was small.<sup>13</sup> Given its advantages, the MADRS has been translated from English into several languages, including Brazilian Portuguese, German, Spanish, Persian, and Bangla.<sup>14-20</sup> All validation studies reported high correlation indexes between the MADRS and the HAM-D (89% in the Brazilian Portuguese study, for instance).<sup>14</sup>

The original MADRS was published without suggestions on how to gather information to rate its items. To standardize MADRS assessment, Williams & Kobak<sup>8</sup> developed the Structured Interview Guide for the MADRS (SIGMA), based on feedback from raters. The SIGMA is an adaptation of the MADRS with detailed instructions on how

Correspondence: Fernando Fernandes, Departamento de Psiquiatria, Faculdade de Medicina, Universidade de São Paulo, Rua Dr. Ovidio Pires de Campos, 785, CEP 05403-010, São Paulo, SP, Brazil.  
E-mail: fernandes2000@gmail.com

Submitted Mar 22 2018, accepted Aug 11 2018, Epub Feb 18 2019.

**How to cite this article:** Fernandes F, Carneiro A, Campos RN, Soeiro-de-Souza MG, Barros VB, Moreno RA. SIGMA-VB: SIGMA-VB: Validity and reliability of the Brazilian Portuguese version of the Montgomery-Åsberg Depression Rating Scale using the Structured Interview Guide for the MADRS. Braz J Psychiatry. 2019;41:297-302. <http://dx.doi.org/10.1590/1516-4446-2018-0105>

to elicit each of its items. It standardizes clinician application of the MADRS, resulting in high reliability.<sup>5</sup> The SIGMA also facilitates training of new raters.

A Brazilian Portuguese version of the SIGMA is not yet available.<sup>14,15,21</sup> To the best of our knowledge, the original SIGMA has only been translated to Japanese.<sup>18,19,22,23</sup> Given this gap in the literature, we translated and validated the original version of the SIGMA into Brazilian Portuguese, using data from participants enrolled in an ongoing clinical trial. The present paper reports the results of this cross-cultural adaptation and validation process.

## Methods

### *Study design*

This observational study was designed to develop the Brazilian Portuguese version of the SIGMA (SIGMA-VB). We translated and cross-culturally validated the original SIGMA into the SIGMA-VB, and formally validated the SIGMA-VB using data from 93 adult outpatients enrolled in the Integral Assessment in Unipolar Depression (AIUNI) trial (ClinicalTrials.gov identifier: NCT02268487). The study was designed and reported per the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement.<sup>24</sup>

### *Setting*

The AIUNI is an open-label controlled clinical trial designed to assess early improvement in major depressive disorder after treatment with sertraline (n=55) or venlafaxine (n=38) at doses of 50-200 mg and 75-300 mg, respectively.

Dosage adjustments in the AIUNI trial were performed according to scores on the HAM-D and Udvalg for Kliniske Undersøgelser (UKU) Side Effect Rating Scale.<sup>25</sup> The latter is a 48-item clinician-rated structured interview developed by the Scandinavian Society for Psychopharmacology that assesses possible side effects of psychotropic medications.

Data collection occurred between 2013 and 2015. Our study was performed prior to a modification in the AIUNI trial protocol specifying sertraline as the only intervention for new participants.

### *Ethics*

Our study and the AIUNI trial were approved by the institutional review board of Instituto de Psiquiatria, Faculdade de Medicina, Universidade de São Paulo (protocols 58942 and 11848, respectively). All participants signed an informed consent form prior to any study procedure. The original SIGMA author and its copyright holders approved development of the SIGMA-VB.

### *Participants*

Participants enrolled in the AIUNI trial were a convenience sample of individuals aged 18-65 years with a diagnosis

of major depressive disorder according to a psychiatric interview and the Structured Clinical Interview for DSM-IV Axis I disorders (SCID-I/P). This initial assessment was performed by psychiatrists specialized in mood disorders. Recruitment for the AIUNI trial occurred via the Grupo de Estudos de Doenças Afetivas (GRUDA) website at [www.progruda.com](http://www.progruda.com). The exclusion criteria for the AIUNI trial were psychiatric comorbidities in the past year and history of bipolar disorder, schizophrenia, mental retardation, or Axis II disorders. Participants with unstable or untreated clinical comorbidities were also excluded. Participants undergoing treatment with psychotropic drugs underwent a washout period prior to the AIUNI trial (7 days for antidepressants, mood stabilizers, and antipsychotics except clozapine; 28 days for clozapine and fluoxetine).

### *Data collection*

Participants in the AIUNI trial were assessed over five visits: at baseline, and at weeks 1, 2, 4, and 8 of treatment. These assessments included a psychiatric evaluation, the SIGMA-VB, and the HAM-D.<sup>1,26,27</sup> The HAM-D was administered by a trained rater. The SIGMA-VB was administered by two raters, a psychiatrist and a clinical psychologist, specifically trained in both scales and in this research protocol. Interviews were conducted independently during the same day, and raters had no access to previous or current scores.

### *Data sources*

SIGMA and SIGMA-VB: structured interviews for MADRS

The SIGMA and the MADRS have the same sequence of questions. However, SIGMA questions are open-ended to encourage participants to report their experience, instead of providing responses constrained by alternatives. For instance, rather than "Are you sad or happy?," the SIGMA question is: "How have you been feeling since last [day of week]?" Each question of the SIGMA should be inquired literally. If further clarification is necessary, the SIGMA provides optional follow-up questions. Additionally, raters can ask other questions to improve accuracy. Furthermore, the SIGMA question about "apparent sadness" depends on the rater's clinical impression, and can rely on information from third parties (such as family members and nursing reports).

We translated the SIGMA from English into Brazilian Portuguese using the Brazilian Portuguese version of the MADRS and following the World Health Organization guidelines for World Mental Health Surveys.<sup>14,28</sup> Our first translation was reviewed by a team of psychiatrists who were fluent in English, specialists in mood disorders, and trained in application of the MADRS. These psychiatrists evaluated the whole scale and graded the understandability of each item as very low, reasonable, good, or excellent. Following their suggestions, we created the final version of the SIGMA-VB. This new version was then back-translated into English by a native speaker who had no prior knowledge of psychopathology or of the MADRS.

## Other instruments

We used the SCID-I/P to verify trial eligibility.<sup>29</sup> The SCID-I/P is a clinician-rated instrument designed to comprehensively assess major psychiatric diagnoses. It includes open-ended questions following DSM-IV-TR criteria, prompts for elaboration and examples, guidelines for rating, and an algorithm to reach final diagnosis and severity. The reliability and validity of the SCID-I/P have been reported in several studies.<sup>30</sup>

For validation purposes, we used the HAM-D, a 17-item questionnaire with multiple choice questions on depressive symptoms. According to the American Psychiatric Association, scores are classified as follows: remission (1-7); mild depression (8-13); moderate depression (14-18); severe depression (19-22); and very severe depression (23 and above).<sup>31</sup> The HAM-D has demonstrated psychometric robustness (validity, reliability and change sensitivity) in several studies.<sup>32</sup>

## Statistical analyses

All analyses were performed in the R<sup>33</sup> and the AgreeStat 2011.1 softwares. First, we evaluated central tendency, dispersion, and distribution of all continuous and categorical variables. Then, data were analyzed with nonparametric statistics, considering dropouts in the AIUNI trial. We used AgreeStat to calculate several interrater reliability indexes: Cohen's kappa, Scott's pi, the Brennan-Prediger coefficient, Gwet's AC2, and Krippendorff's alpha. We also used Gwet's and Krippendorff's coefficients as complements: the first was calculated considering all participants, while the later was calculated excluding participants with missing data. Finally, we used Spearman's correlation coefficients to measure correlation between raters' scores using the SIGMA-VB and between SIGMA-VB and HAM-D scores. For all analyses, we assumed a significance level of 0.05.

## Results

Our sample comprised 93 individuals with an average age of 42.5 years (standard deviation [SD] 11.2, range 19-65 years). Most participants were female (75%). A total of 122 participants were assessed for eligibility, but 29 were excluded (26 did not meet our inclusion criteria and 3 declined to participate). Twenty-one participants dropped out due to nonadherence to medication (n=16), intolerable adverse effects or clinical decline (n=4), and withdrawal (n=1). The number of participants in each week of the study ranged from 93 at baseline to 72 at 8 weeks.

Our SIGMA-VB reliability analysis included 93 validity assessments resulting in 288 interactions. Table 1 and Figure 1 report SIGMA-VB and HAM-D scores over the study visits. According to the SIGMA-VB mean scores, participants had moderate depression at baseline followed by mild depression at 8 weeks; according to the HAM-D mean scores, they had severe depression at baseline followed by remission at 8 weeks. Mean depression scores declined consistently during the course of the

**Table 1** Participants' SIGMA-VB and HAM-D scores

Instrument/week	Mean (SD)	Minimum	Maximum
<b>SIGMA-VB-A</b>			
0	28.6 (7.6)	10	47
1	24.1 (8.9)	6	44
2	21.2 (9.5)	1	42
4	18.3 (9.0)	0	37
8	14.9 (1.4)	0	40
<b>SIGMA-VB-B</b>			
0	29.7 (8.2)	13	48
1	24.2 (8.5)	4	44
2	21.0 (9.7)	1	44
4	19.2 (9.6)	0	40
8	15.1 (1.5)	0	41
<b>HAM-D</b>			
0	19.4 (5.7)	14	40
1	16.0 (6.2)	6	31
2	13.0 (6.5)	0	33
4	12.1 (5.7)	0	24
8	1.8 (6.6)	0	27

HAM-D = Hamilton Depression Rating Scale; SD = standard deviation; SIGMA-VB-A and B = Brazilian Portuguese version of the SIGMA applied by raters A and B, respectively.

visits. However, SD scores remained high until the fourth week of treatment, while maximum scores remained high throughout the study period.

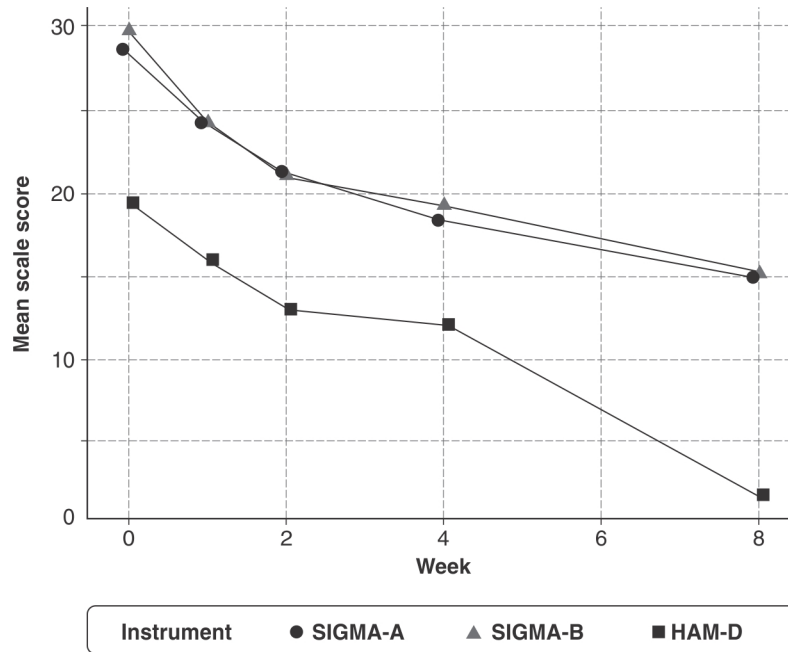
We found good interrater reliability indexes using the SIGMA-VB (Table 2). Raters were able to classify the same severity of depression as the clinicians over the five SIGMA-VB assessments. Furthermore, we found over 90% of correlation between scores attributed by different raters using the SIGMA-VB (Table 3).

Correlations between the HAM-D and SIGMA-VB were strong ( $> 0.66$ ) (Table 4). Correlation coefficients increased during the course of visits: we found lower correlation levels at baseline (when most participants had severe or moderate depression) and higher correlation levels at 8 weeks (when participants had mild depression or remission).

## Discussion

To our knowledge, this is the first study to translate a structured interview guide for the MADRS into Brazilian Portuguese. Our participants were assessed at five time points during the acute phase of depression treatment (8 weeks), and thus at different depression severities. This aspect represents a strength of our study. Comparatively, the Japanese SIGMA study evaluated seven participants in a total of 18 interviews (one to four interviews per participant).

In all five of our SIGMA-VB assessments, correlation coefficients between raters exceeded 90%. These results are similar to the findings of the Japanese SIGMA study, in which interrater reliability ranged from 0.91 to 1.00.<sup>15</sup> We obtained high interrater reliability scores despite having raters with different backgrounds (a psychiatrist and a clinical psychologist). This suggests that the SIGMA-VB is appropriate for assessing depression severity even with some degree of interrater heterogeneity.



**Figure 1** Progression of mean SIGMA-VB and HAM-D over the study period. HAM-D = Hamilton Depression Rating Scale; SIGMA-VB-A and B = Brazilian Portuguese version of the SIGMA applied by raters A and B, respectively.

**Table 2** SIGMA-VB: two-rater chance-corrected agreement coefficients (quadratic-weighted)

Instrument/week	Correlation coefficient	Standard error of the mean	95%CI	p-value
<b>Cohen's kappa</b>				
0	0.90	0.02	0.86-0.94	< 0.001
1	0.95	0.01	0.93-0.97	< 0.001
2	0.95	0.01	0.93-0.97	< 0.001
4	0.93	0.02	0.88-0.98	< 0.001
8	0.98	0.00	0.97-0.99	< 0.001
<b>Scott's pi</b>				
0	0.90	0.02	0.86-0.94	< 0.001
1	0.95	0.01	0.93-0.97	< 0.001
2	0.95	0.01	0.93-0.97	< 0.001
4	0.93	0.02	0.88-0.98	< 0.001
8	0.98	0.00	0.97-0.99	< 0.001
<b>Brennan-Prediger</b>				
0	0.94	0.01	0.93-0.96	< 0.001
1	0.97	0.01	0.96-0.98	< 0.001
2	0.97	0.00	0.96-0.98	< 0.001
4	0.95	0.02	0.91-0.98	< 0.001
8	0.99	0.00	0.98-0.99	< 0.001
<b>Gwet's AC2</b>				
0	0.95	0.01	0.93-0.96	< 0.001
1	0.97	0.00	0.96-0.98	< 0.001
2	0.97	0.00	0.96-0.98	< 0.001
4	0.95	0.02	0.92-0.98	< 0.001
8	0.99	0.00	0.98-0.99	< 0.001
<b>Krippendorff's alpha</b>				
0	0.90	0.02	0.86-0.94	< 0.001
1	0.96	0.01	0.94-0.97	< 0.001
2	0.95	0.01	0.93-0.97	< 0.001
4	0.93	0.02	0.88-0.98	< 0.001
8	0.98	0.00	0.97-0.99	< 0.001

95%CI = 95% confidence interval.

**Table 3** SIGMA-VB: Spearman's correlation coefficients between raters

	Rater B				
	Week 0	Week 1	Week 2	Week 4	Week 8
<b>Rater A</b>					
Week 0	0.91*	0.66*	0.48*	0.50*	0.37 <sup>†</sup>
Week 1		0.95*	0.39*	0.69*	0.29 <sup>†</sup>
Week 2			0.95*	0.61*	0.54*
Week 4				0.92*	0.43*
Week 8					0.98*

SIGMA-VB = Brazilian Portuguese version of the SIGMA.

\*p > 0.001; <sup>†</sup> p > 0.05.**Table 4** SIGMA-VB and HAM-D: Spearman's correlation indexes

	HAM-D				
	Week 0	Week 1	Week 2	Week 4	Week 8
<b>SIGMA-VB</b>					
<b>Rater A</b>					
Week 0	<b>0.66*</b>	0.56*	0.50*	0.45*	0.45*
Week 1	0.49*	<b>0.70*</b>	0.48*	0.58*	0.39*
Week 2	0.29 <sup>†</sup>	0.20 <sup>†</sup>	<b>0.74*</b>	0.50*	0.47*
Week 4	0.36*	0.27 <sup>†</sup>	0.39*	<b>0.87*</b>	0.47*
Week 8	0.1	0.20 <sup>†</sup>	0.32 <sup>†</sup>	0.51*	<b>0.88*</b>
<b>Rater B</b>					
Week 0	<b>0.66*</b>	0.60*	0.47*	0.49*	0.51*
Week 1	0.54*	<b>0.71*</b>	0.50*	0.57*	0.49*
Week 2	0.38*	0.29*	<b>0.78*</b>	0.46*	0.46*
Week 4	0.41*	0.39*	0.48*	<b>0.85*</b>	0.54*
Week 8	0.07	0.22 <sup>†</sup>	0.32 <sup>†</sup>	0.46*	<b>0.85*</b>

HAM-D = Hamilton Depression Rating Scale; SIGMA-VB = Brazilian Portuguese version of the SIGMA.

\*p > 0.001; <sup>†</sup> p > 0.05.

Bold type denotes correlation &gt; 0.60.

We found large correlations between the SIGMA-VB and the HAM-D, confirming the validity of the SIGMA-VB. The concurrent validity between the MADRS and the HAM-D was 0.85 and 0.95 in the German and Persian versions of the MADRS respectively<sup>16,18</sup>; no validity analysis was reported in the Japanese SIGMA study.

Correlation between the SIGMA-VB and the MADRS varied according to depression severity, with larger correlation coefficients among milder cases (0.64-0.88). In agreement with previous studies, the SIGMA-VB was more consistent in detecting severe depression and monitoring progress during treatment.<sup>1,13</sup> However, the HAM-D was more sensitive to non-cardinal symptoms of depression (such as anxiety and somatic symptoms) in participants with mild depression or euthymia.

Despite the relevance of translating the SIGMA into Brazilian Portuguese, our study does have limitations. First, we used a convenience sampling method, which reduces the generalizability of our findings. Also, we did not compare the reliability of the MADRS scale with and without the SIGMA-VB interview, as that would have increased the complexity of our data collection process. However, previous literature supports the SIGMA as a more reliable instrument.<sup>5</sup> In addition, we did not perform an analysis stratified for age, as that would have required a larger sample.

Our findings confirm that the Brazilian Portuguese version of the SIGMA is a reliable instrument to assess depression severity. The SIGMA-VB should be considered for use in research and practice, especially when dealing with severely depressed patients undergoing antidepressant treatment.

## Disclosure

The authors report no conflicts of interest.

## References

- 1 Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134:382-9.
- 2 Leucht S, Fennema H, Engel RR, Kaspers-Janssen M, Lepping P, Szegedi A. What does the MADRS mean? Equipercntile linking with the CGI using a company database of mirtazapine studies. *J Affect Disord*. 2017;210:287-93.
- 3 Quilty LC, Robinson JJ, Rolland JP, Fruyt FD, Rouillon F, Bagby RM. The structure of the Montgomery-Åsberg depression rating scale over the course of treatment for depression. *Int J Methods Psychiatr Res*. 2013;22:175-84.
- 4 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.
- 5 Williams JB, Kobak KA. Development and reliability of a structured interview guide for the Montgomery Asberg depression rating scale (SIGMA). *Br J Psychiatry*. 2008;192:52-8.

- 6 Zimmerman M, Posternak MA, Chelminski I. Is it time to replace the Hamilton depression rating scale as the primary outcome measure in treatment studies of depression? *J Clin Psychopharmacol.* 2005;25:105-10.
- 7 Huijbrechts IP, Haffmans PM, Jonker K, van Dijke A, Hoencamp E. A comparison of the 'Hamilton rating scale for depression' and the 'Montgomery-Asberg depression rating scale'. *Acta Neuropsychiatr.* 1999;11:34-7.
- 8 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;16:601-11.
- 9 Khan A, Brodhead AE, Kolts RL. Relative sensitivity of the Montgomery-Asberg depression rating scale, the Hamilton depression rating scale and the Clinical Global Impressions rating scale in anti-depressant clinical trials: a replication analysis. *Int Clin Psychopharmacol.* 2004;19:157-60.
- 10 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;121:268-72.
- 11 Kearns NP, Cruickshank CA, McGuigan KJ, Riley SA, Shaw SP, Snaith RP. A comparison of depression rating scales. *Br J Psychiatry.* 1982;141:45-9.
- 12 Davidson J, Turnbull CD, Strickland R, Miller R, Graves K. The Montgomery-Asberg depression scale: reliability and validity. *Acta Psychiatr Scand.* 1986;73:544-8.
- 13 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;43:1049-56.
- 14 Dratcu L, da Costa Ribeiro L, Calil HM. Depression assessment in Brazil. The first application of the Montgomery-Asberg depression rating scale. *Br J Psychiatry.* 1987;150:797-800.
- 15 Takahashi N, Tomita K, Higuchi T, Inada T. The inter-rater reliability of the Japanese version of the Montgomery-Asberg depression rating scale (MADRS) using a structured interview guide for MADRS (SIGMA). *Hum Psychopharmacol.* 2004;19:187-92.
- 16 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;139:51-65.
- 17 Lobo A, Chamorro L, Luque A, Dal-Ré R, Badia X, Baró E, et al. [Validation of the Spanish versions of the Montgomery-Asberg depression and Hamilton anxiety rating scales]. *Med Clin (Barc).* 2002;118:493-9.
- 18 Ahmadpanah M, Sheikhababaei 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.
- 19 Cano JF, Gomez Restrepo C, Rondón M. [Validation of the Montgomery-Asberg depression rating scale (MADRS) in Colombia]. *Rev Colomb Psiquiatr.* 2016;45:146-55.
- 20 Soron TR. Validation of Bangla Montgomery Asberg depression rating scale (MADRSB). *Asian J Psychiatr.* 2017;28:41-6.
- 21 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;13:42-42.
- 22 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.
- 23 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.
- 24 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The strengthening of reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg.* 2014;12:1495-9.
- 25 Lingjaerde O, Ahlfors UG, Bech P, Dencker SJ, Elgen K. The UKU side effect rating scale. A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. *Acta Psychiatr Scand Suppl.* 1987;334:1-100.
- 26 Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960;23:56-62.
- 27 Rohan KJ, Rough JN, Evans M, Ho SY, Meyerhoff J, Roberts LM, et al. A protocol for the Hamilton rating scale for depression: item scoring rules, rater training, and outcome accuracy with data on its application in a clinical trial. *J Affect Disord.* 2016;200:111-8.
- 28 World Health Organization (WHO). Process of translation and adaptation of instruments [Internet]. 2009 [cited 2017 Dec 01]. [www.who.int/substance\\_abuse/research\\_tools/translation/en/](http://www.who.int/substance_abuse/research_tools/translation/en/).
- 29 First MB, Spitzer RL, Gibbon M, Williams JBW. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition (SCID-I/P). New York: Biometrics Research; 2002.
- 30 Skre I, Onstad S, Torgersen S, Kringlen E. High interrater reliability for the structured clinical interview for DSM-III-R Axis I (SCID-I). *Acta Psychiatr Scand.* 1991;84:167-73.
- 31 Kriston L, von Wolff A. Not as golden as standards should be: interpretation of the Hamilton rating scale for depression. *J Affect Disord.* 2011;128:175-7.
- 32 Blacker D. Psychiatric rating scales. In: Sadock B, Sadock V, editors. Kaplan and Sadock's comprehensive textbook of psychiatry. 8th ed. Philadelphia: Lippincott Williams & Wilkins; 2005. p. 755-83.
- 33 R Core Team. R: A language and environment for statistical computing Vienna: R Foundation for Statistical Computing; 2013.