Journal of Psychiatric Research 108 (2019) 84–89

Contents lists available at ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/jpsychires

Assessment of the psychometric properties of the 17- and 6-item Hamilton Depression Rating Scales in major depressive disorder, bipolar depression and bipolar depression with mixed features

Alexandre Kieslich da Silva[a,b,*], Mateus Reche[a,b], Ana Flávia da Silva Lima[c], Marcelo Pio de Almeida Fleck[d], Edison Capp[a], Flávio Milman Shansis[a,b,e,f]

[a] Gynecology and Obstetrics Postgraduate Program, Medical School, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil
[b] Mood Disorders Research and Educational Program (PROPESTH), Hospital Psiquiátrico São Pedro (HPSP), Porto Alegre, RS, Brazil
[c] Instituto de Avaliação de Tecnologia em Saúde (AITS), Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, RS, Brazil
[d] Psychiatry Postgraduate Program, Medical School, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil
[e] Collective Health Postgraduate Program, Medical School, Vale dos Sinos University (UNISINOS), São Leopoldo, RS, Brazil
[f] Medical School, Vale do Taquari University (UNIVATES), Lajeado, RS, Brazil

ARTICLE INFO

Keywords:
6-Item Hamilton Depression Rating Scale
Depression scale
Psychometric properties
Depression
Mixed features

ABSTRACT

Assessing therapeutic response in depression requires scales that adequately measure the core symptoms of depressive symptomatology. The main goal of this study was to assess the psychometric properties of the 17-item Hamilton Depression Rating Scale (HAM-D17) and the 6-item Hamilton Depression Rating Scale (HAM-D6) in patients with Major Depressive Disorder (MDD), bipolar depression and bipolar depression with mixed features. We conducted a reanalysis of a pragmatic clinical trial in an outpatient clinic for mood disorders that included eight weeks of follow-up. A Mokken analysis was performed to evaluate the unidimensionality of the HAM-D17 and HAM-D6, and the Spearman correlation was used to assess concurrent validity between the HAM-D17, the HAM-D6 and quality of life scale (SF-36 and WHOQOL-BREF) scores. A total of 237 patients with a mean age of 40.2 years (± 11.7) were included. According to the DSM-IV criteria, 58 (24.5%) were diagnosed with MDD and 73 (30.8%) were diagnosed with bipolar depression. Bipolar depression with mixed features was diagnosed in 106 (44.7%) patients according to the DSM-IV and supplemented by the Cincinnati criteria. Only the HAM-D6 scale proved to be unidimensional, showing strong homogeneity for evaluating MDD, moderate homogeneity for bipolar depression and weak homogeneity for bipolar depression with mixed features. Both the HAM-D17 and the HAM-D6 had inverse, significant correlations at baseline with SF-36 and WHOQOL-BREF scores. This is the first study to include bipolar depression patients with mixed features in an assessment of HAM-D6 unidimensionality.

1. Introduction

Although the most widely used scale in clinical trials to assess depressive symptoms is the 17-item Hamilton Depression Rating Scale (HAM-D17) (Hamilton, 1960; Khan et al., 2010), a number of studies have recently called its superiority into question (Bagby et al., 2004; Licht et al., 2005; Zimmerman et al., 2005). Their main criticisms is related to the HAM-D17's construct validity, since it may represent a multidimensional measure due to its inclusion of symptoms beyond the core symptoms of depression (Lecrubier and Bech, 2007). There are also questions about the HAM-D17's clinical validity, since its general score does not reflect depression severity as perceived by experienced psychiatrists (Bech et al., 1975). Moreover, the HAM-D17 has shown low sensitivity to clinical change in depressive patients (Ballesteros et al., 2007; Santor and Coyne, 2001).

One alternative that could compensate for the construct validity, clinical validity, and sensitivity problems of the HAM-D17 would be the 6-item Hamilton Depression Rating Scale (HAM-D6), which consists of the following HAM-D17 items: Item 1, depressed mood; Item 2, feelings of guilt; Item 7, work and interests; Item 8, psychomotor retardation;

* Corresponding author. Gynecology and Obstetrics Postgraduate Program, Hospital de Clínicas de Porto Alegre, Rua Ramiro Barcelos, 2.350, 11º andar, sala 1.125, 90035-903, Porto Alegre, RS, Brazil.
E-mail addresses: akieslich@gmail.com (A. Kieslich da Silva), matrech@msn.com (M. Reche), afbslima@gmail.com (A.F.d.S. Lima), mpafleck@gmail.com (M.P.d.A. Fleck), edcapp@ufrgs.br (E. Capp), fmshansis@gmail.com (F.M. Shansis).

https://doi.org/10.1016/j.jpsychires.2018.07.009
Received 4 April 2018; Received in revised form 16 July 2018; Accepted 18 July 2018

Item 10, psychic anxiety, and Item 13, general somatic symptoms (Bech et al., 1981). The HAM-D6 has been used in many clinical trials on depression (Bech et al., 2010; Fabre et al., 1995; Higuchi et al., 2016; Martiny et al., 2013). Moreover, its clinimetric properties have been evaluated in a number of studies, most of which investigated its unidimensionality (Bech et al., 2011; Timmerby et al., 2017), although others have described its psychometric properties, such as internal validity and concurrent validity (Bachner, 2016; Bobes et al., 2016). Some studies have also evaluated effect size to assess the HAM-D6's degree of superiority regarding responsiveness (Bech et al., 2010; Østergaard et al., 2016; Østergaard et al., 2014a; b).

However, as far as we know, no study has simultaneously investigated the unidimensionality and psychometric properties of the HAM-D6 in a heterogeneous population of patients with depressive symptomatology, including bipolar depression with mixed features.

1.1. Aims of the study

The current study aimed to evaluate the unidimensionality and psychometric properties of the HAM-D17 and HAM-D6 in a sample of patients with major depressive disorder, bipolar depression and bipolar depression with mixed features. We aimed to verify the concurrent validity of the HAM-D17 and HAM-D6 using quality of life scales (WHOQOL-BREF and SF-36). We also evaluated the HAM-D6's sensitivity to change through effect size and compared it with that of the HAM-D17.

2. Material and methods

Data from a pragmatic clinical trial on the effectiveness of algorithms for outpatient mood disorder treatment, which was conducted between October 2010 and October 2014 in the Brazilian public health system, were reanalyzed.

The patients were divided into groups according to three treatment algorithms originally developed for unipolar depressive episodes, bipolar depressive episodes and mixed episodes. The unipolar depression algorithm included a sequence of sertraline, potentiation with lithium, and substitution or association with nortriptyline. For bipolar depressive episodes, the sequence was lithium associated with sertraline, followed by the substitution or association of sertraline with nortriptyline. In both of these groups, a single-group, pre-test-post-test experimental approach was used. For mixed-episode patients, a pragmatic, cross-sectional, non-blinded multi-arm trial was performed. To maintain proportionality (1:1:1) in the three groups (lithium, valproic acid, and carbamazepine), simple randomization was performed (random computer-generated numbers). The sequence of this algorithm involved the substitution or association of one the other mood stabilizers, followed by association with risperidone. The registry numbers of the clinical trials are NCT02901249, NCT02870283 and NCT02918097 (Lima et al., 2017). All patients completed an informed consent form. This study was approved by the São Pedro Psychiatric Hospital Research Ethics Committee (case 700005332).

2.1. Study procedures and measures

The patients evaluated in Lima et al. (2017) were diagnosed as having a mood disorder according to DSM-IV criteria through the Mini International Neuropsychiatric Interview (MINI) version 5.0.0 (Lecrubier et al., 1997), which was confirmed in a clinical interview. In the clinical interview, the diagnosis of bipolar depression with mixed features was not restricted to the DSM-IV criteria for a mixed episode; it was also based on the Cincinnati criteria, defined as dysphoric mania due to the presence of three or more symptoms of major depression during a full manic or hypomanic episode according to DSM-III-R criteria (McElroy et al., 1992). The Cincinnati criteria are more inclusive than those for mixed features in the DSM-IV because they do not require all of the symptoms of the opposing pole (American Psychiatric Association [APA], 1994). In this respect, the Cincinnati criteria are more similar to the DSM-5, as shown in Table 1.

Table 1
Sociodemographic characteristics of the sample.

Sex	*n* (*n*%)
Female	188 (79.32)
Ethnicity	
White	167 (70.46)
Black	52 (21.94)
Other	18 (7.59)
Marital Status	
Single or without a partner	120 (50.63)
Years of Study	
from 1 to 4	29 (12.37)
from 5 to 8	71 (29.96)
from 9 to 12	88 (37.13)
more than 13 years	43 (18.14)
Work	
Unemployed	132 (55.70)
Brazilian Criterion- LSE-2012	
A and B	59 (24.89)
C	145 (61.18)
D and E	29 (12.24)
Patients with psychotic symptoms at baseline	20 (8.44%)

Key: *n:* absolute frequency; *n%*: relative frequency.

The present study assessed demographic data, the HAM-D17, the HAM-D6, and the Clinical Global Impression Severity Scale (CGI-S) (Guy, 1976), as well as the self-administered WHOQOL-BREF (Whoqol Group, 1998) and Short Form-36 (SF-36) (Ware and Sherbourne, 1992) quality of life scales, using scores from baseline and the eighth week of treatment.

2.2. Demographic data

Demographic data included age, gender, years of education, employment status, and socioeconomic status as defined by the Brazilian Economic Classification Criteria (ABEP, 2009), which categorizes social class as A, B, C, D, or E.

2.3. MINI 5.0.0

The MINI is a standardized diagnostic interview compatible with the DSM-III-R/IV and ICD-10 criteria used in psychiatric research and clinical practice (Lecrubier et al., 1997).

2.4. CGI-S

The CGI-S consists of a Likert-type scale that assesses the severity of depression, with scores ranging from 1 (normal) to 7 (extremely ill) (Guy, 1976).

2.5. WHOQOL-BREF

The WHOQOL-BREF consists of 26 questions, two of which address general quality of life and another 24 that are divided into four domains: physical, psychological, social relationships, and environment. The questions are answered on a Likert scale (from 1 to 5): the higher the score, the better the quality of life (Whoqol Group, 1998).

2.6. SF-36

The SF-36 is a multidimensional instrument for assessing quality of life that consists of 36 items which are grouped into eight domains: functional capacity, physical aspects, pain, general health status, vitality, mental health, and social and emotional aspects. Its scores can

range from zero to 100, which correspond to the worst and best health status, respectively (Ware and Sherbourne, 1992).

2.7. Statistical analysis

The Mokken scale analysis, which is a non-parametric model of Item Response Theory (IRT) (Mokken, 1971), was used to evaluate whether the HAM-D17 and HAM-D6 scales were unidimensional, that is, if each item in the scale conveys unique information about the severity of the latent syndrome. In contrast to parametric theory models, the Mokken scale analysis requires no assumption about the distribution of the latent characteristic in the population or the item parameters. The criteria indicating unidimensionality are described by Loevinger's scalability coefficient. An H coefficient ≥ 0.5 indicates strong homogeneity, 0.40 to 0.49 moderate homogeneity, and 0.30 to 0.39 weak homogeneity, while < 0.3 indicates that a scale is not homogeneous (Sijtsma and Molenaar, 2002). The Mokken analysis was based on the homogeneity coefficient at eight weeks follow-up to ensure a rating score distribution with sufficient dispersion to evaluate the test using correlation coefficients.

The Mokken analyses were performed using R version 3.4.1 with the Mokken package (van der Ark, 2012). The other statistical analyses were performed using SPSS version 19. The internal validity of the HAM-D17 and HAM-D6 were evaluated with Cronbach's alpha reliability coefficient, in which an alpha coefficient between 0.70 and 0.90 is considered suitable (Streiner, 2003). Concurrent validity assessments were performed using Spearman's correlation coefficient between the HAM-D17, the HAM-D6 and the SF-36 and WHOQOL-BREF scores. The ROC curve was used to obtain the cut-off point for HAM-D6 in comparison with the CGI-S scale, and the highest values of sensitivity and specificity were chosen.

Cohen's d effect size, with a 95% confidence interval, was calculated to compare the change arising from treatment with the HAM-D17 and HAM-D6 scales (Sawilowsky, 2009). According to Hopkins et al. the following parameters should be used to interpret the magnitude of Cohen's d effect size: between 0 and 0.19 is trivial, between 0.20 and 0.59 is small, between 0.60 and 1.19 is moderate, between 1.2 and 1.99 is large, and between 2.00 and 4.00 is very large (Hopkins et al., 2009). A significance level of $p < 0.05$ was set for all statistical analyses.

The sample calculation was estimated using WinPepi for Windows version 11.65, considering the differences between the means and the standard deviations of the HAM-17 and HAM-D6 scales after 4 weeks of study (Licht et al., 2005). Considering a sample power of 90%, with 10% loss and a 5% significance level, the final sample size needed to detect a difference of 4 units of means was 55 patients in each group (major depressive disorder [MDD], bipolar depression, and bipolar depression with mixed features).

3. Results

The sample consisted of 237 patients with a mean age of 40.2 years (± 11.7) and was predominantly female. Only 161 patients continued follow-up until the eighth week of treatment. According to the DSM-IV criteria, 58 (24.47%) patients had MDD, 73 (30.80%) had bipolar depression and 106 (44.73%) had bipolar depression with mixed features. Of these 106 patients, 63 (59.4%) met the DSM-IV criteria for mixed episode according to the MINI results, whereas 43 (40,6%) met the Cincinnati criteria for mixed symptoms through clinical assessment (McElroy et al., 1992). The sociodemographic data and the percentage of patients with psychotic symptoms at baseline are described in Table 1. Table 2 shows the medications that were being taken at eighth week of treatment.

Table 3 shows the Loevinger coefficient for the HAM-D17 and the HAM-D6 at the eighth week of treatment, which indicates that only HAM-D6 was a unidimensional scale. Regarding the diagnosis, according to the HAM-D17, only the MDD group obtained homogeneity, whereas according to the HAM-D6 all three patient groups obtained homogeneity, except that the homogeneity level was slightly below moderate in the bipolar depression with mixed features group.

Table 2
Medication use at the eighth week of treatment.

Medication	Major Depressive Disorder	Bipolar Depression	Bipolar Depression with mixed features
Sertraline *n* (%)	32 (72.72)	20 (45.45)	0 (0.00)
Nortriptyline	14 (31.81)	0 (0.00)	0 (0.00)
Valproic acid	0 (0.00)	4 (9.09)	27 (36.99)
Lithium	4 (9.09)	40 (90.09)	38 (52.05)
Carbamazepine	0 (0.00)	1 (2.27)	20 (27.40)
Risperidone	8 (18.18)	18 (40.90)	20 (27.40)
Total patients	44 (100.00)	44 (100.00)	73 (100.00)

Key: *n:* absolute frequency; *n*%: relative frequency.

Table 3
Loevinger's H coefficient for the HAM-D17 and HAM-D6 at the eighth week of treatment stratified by clinical diagnosis.

	n	HAM-D17	HAM-D6
MDD	44	0.37	0.55
Bipolar Depression	44	0.30	0.41
Bipolar Depression with mixed features	73	0.26	0.36
All	161	0.29	0.42

MDD = major depressive disorder; $H < 0.3$ = without homogeneity; ≥ 0.3 and < 0.4 = weak homogeneity; ≥ 0.4 and < 0.5 = moderate homogeneity; $0.5 \geq$ strong homogeneity.

Both the HAM-D17 and the HAM-D6 showed an inverse and statistically significant correlation at baseline with WHOQOL-BREF and SF-36 scores. At the eighth week of treatment, there was no correlation between the depression scales and SF-36 score; however, regarding the WHOQOL-BREF score, there was a strong correlation with HAM-D17 and a moderate correlation with HAM-D6 (Table 4).

The HAM-D17 and HAM-D6 moderated the effect size of the response to pharmacological treatment in the three groups (Table 5). At baseline, the internal validities of the HAM-D17 and HAM-D6 according to Cronbach's alpha coefficient were 0.71 and 0.54, respectively. At the eighth week of treatment, the HAM-D17 and the HAM-D6 had Cronbach's alphas of 0.82 and 0.74, respectively. The HAM-D6 was highly correlated with the HAM-D17 at baseline ($r = 0.81$; $p < 0.001$) and at the eighth week ($r = 0.917$; $p < 0.001$).

4. Discussion

The HAM-D6 Loevinger coefficient found at the eighth week of treatment is consistent with other studies, which confirms that the HAM-D6 is a unidimensional depression scale (Bachner, 2016; Lee, 2017). Likewise, the internal validity results with Cronbach's alpha

Table 4
Concurrent validity of the HAM-D17 and HAM-D6 with the quality-of-life scales at baseline and the eighth week of treatment.

HAM-D17	WHOQOL-BREF			SF-36		
	n	r	p value	N	r	p-value
Baseline	219	0.324	< 0.001	218	0.281	< 0.001
8th week	89	0.546	< 0.001	90	0.138	= 0.195
HAM-D6						
Baseline	219	0.317	< 0.001	218	0.255	< 0.001
8th week	89	0.433	< 0.001	90	0.123	= 0.248

Table 5
Effect size of the response to pharmacological treatment by groups of patients assessed using the HAM-D17 and HAM-D6.

	Baseline		8th week		Effect size (95% confidence interval)
	n	Mean (SD)	n	Mean (SD)	
HAM-D17					
MDD	58	20.73 (6.27)	44	12.14 (8.54)	1.17 (0.74–1.59)
Bipolar Depression	73	19.18 (5.97)	44	12.93 (7.47)	0.95 (0.55–1.34)
Bipolar with mixed features	106	20.14 (6.95)	73	12.41 (7.25)	1.09 (0.77–1.41)
Total	237	20.04 (6.50)	161	12.48 (7.64)	1.08 (0.87–1.29)
HAM-D6					
MDD	58	10.70 (3.17)	44	6.34 (4.97)	1.08 (0.65–1.49)
Bipolar Depression	73	9.70 (2.75)	44	6.77 (3.91)	0.91 (0.51–1.29)
Bipolar with mixed features	106	9.41 (3.53)	73	6.42 (4.00)	0.80 (0.49–1.11)
Total	237	9.84 (3.26)	161	6.50 (4.24)	0.91 (0.70–1.12)

MDD = major depressive disorder.

coefficient were lower for the HAM-D6 than for the HAM-D17 at baseline and at the eighth week of treatment, which was expected since the HAM-D6 has fewer items. Nevertheless, at the eighth week of treatment, the HAM-D6 had a Cronbach's alpha of 0.74, which is considered acceptable (Dunn et al., 2013).

In a recent systematic review, Timmerby et al. evaluated the clinimetric properties of the HAM-D6 in comparison with the HAM-D17 and the Montgomery-Åsberg Depression Rating Scale (Timmerby et al., 2017). Of the 51 articles they evaluated, none included bipolar depression patients with mixed features and only two included patients with bipolar depression (Holmskov et al., 2017; K. Martiny et al., 2013). When these two studies involving bipolar patients were analyzed separately, it became evident that the presence of bipolar symptomatology did not affect the homogeneity of the HAM-D6 coefficient (Timmerby et al., 2017). In our study, the HAM-D6 presented a strong homogeneity coefficient for MDD patients and moderate one for bipolar depression patients, which agrees with Timmerby et al.'s systematic review.

To our knowledge, ours is the first study to have evaluated mixed symptomatology with the HAM-D6. However, these patients were followed up using the HAM-D17 and the Young Mania Rating Scale (Lima et al., 2017), since the combined use of depression and mania scales is the most common strategy for evaluating response to psychopharmacological treatment of this disorder (Shansis et al., 2016). Nevertheless, this strategy is conceptually questionable, since mixed symptoms are more than the sum of depressive and manic symptoms, as the DSM-IV suggests (Zimmerman, 2017). Despite such concerns, this strategy has become widely used due to the scarcity of more specific instruments for assessing mixed symptomatology (Shansis et al., 2016). In our study, however, in addition to being diagnosed as mixed episode according to the DSM-IV, the patients were diagnosed with bipolar depression with mixed features according to the Cincinnati criteria (McElroy et al., 1992), which are more inclusive and similar to the current bipolar depression with mixed features criteria of the DSM-5 (Shansis et al., 2016), as shown in Fig. 1.

With the new categorical-dimensional approach of the DSM-5 (APA, 2013), the former mixed episode category in the DSM-IV was replaced with a mixed-features specifier, which could not only be present in manic episodes but in hypomanic episodes and in MDD as well, given that they are characterized by at least three symptoms of opposite polarity (Tortorella et al., 2015). This is still a controversial issue, but in general, according to McElroy and Keck, the mixed state may not be a single condition, but rather a group of heterogeneous clinical entities that vary according to the criteria selected (Mcelroy and Keck, 2017).

There are few psychometric scales that suitably measure mixed symptomatology (Berk et al., 2007). According to Zimmerman's recent review, only the Schedule for Affective Disorders and Schizophrenia (Spitzer et al., 1978) and the Bipolar Inventory of Symptoms Scale (Bowden et al., 2007) assess both depressive and mixed symptoms according to DSM-5 criteria. Others measures assess only depression, mania/hypomania or some of their symptoms (Zimmerman, 2017). It has been proposed, for example, that a score ≥10 should be used as a specific cut-off of in the HAM-D17 to define hypomanic mixed episodes and a score ≥8 on the Schedule for Affective Disorders and Schizophrenia should be used to define a depressive mixed episode (Swann et al., 2013). Moreover, González-Pinto et al. stated that five of the HAM-D21 items (depressive symptoms, suicidal ideation, guilt, psychic anxiety, and obsessional symptoms), which make up the 5-item Hamilton Depression Scale (HAM-D5), would better to assess and follow-up of depressive symptoms in mixed bipolar patients or a manic episode, since the HAM-D5 has been shown to be more specific and have greater discriminant validity than the HAM-D21 (González-Pinto et al., 2009). It should also be pointed out that the HAM-D6's content validity is predominantly in the depressive domain of bipolar depression with mixed features. Shansis et al. showed that manic-like symptoms in patients with mixed symptomatology are better evaluated by the Young Mania Rating Scale than the Bech-Rafaelsen Mania Scale or the Clinician-Administered Rating Scale for Mania in association with the HAM-D21 (Shansis et al., 2016).

It should also be mentioned that Østergaard et al. investigated the unidimensionality of the HAM-D17 and the HAM-D6 in relation to psychotic depression, finding that only the HAM-D6 was a valid measure of psychotic depression, which is a subtype of severe depression according to the ICD-10 and the DSM-5 (Østergaard et al., 2014a; b; APA, 2013; Lecrubier et al., 1997).

In the present study, the HAM-D17 and the HAM-D6 were inversely and significantly correlated at baseline and eighth week of treatment with WHOQOL-BREF scores. These data are agree Bachner's cross-sectional study on caregivers in a cancer hospital, which found an inverse correlation between the HAM-D6 and a quality-of-life scale (the World Health Organization Well-Being Index), an emotional exhaustion scale (Emotional Exhaustion Subscale of the Maslach Burnout Inventory) and a caregiver stress scale (Zarit Burden Inventory) (Bachner, 2016). Likewise, Østergaard et al. showed that changes in total scores on the HAM-D17 and HAM-D6 scales were highly correlated with scores on the Quality of Life Enjoyment and Satisfaction Questionnaire Scale throughout treatment (Østergaard et al., 2014a; b). In our study, the baseline SF-36 score was inversely correlated with the depression scales. This indicates that, prior to treatment, the greater the severity of depressive symptoms found in the HAM-D17 and HAM-D6, the lower the quality of life according to the SF-36, which was not observed by the eighth week of treatment.

Timmerby et al.'s review found that the effect sizes of HAM-D6 for assessing antidepressants were higher than those of the HAM-D17 in several randomized controlled trials (Timmerby et al., 2017). In the present study, there were no significant effect size differences between the HAM-D6 and the HAM-D17, which may have been due to the small sample size. Thus, the HAM-D6 can be used as an alternative to the HAM-D17 for assessing treatment efficacy, with the advantage that the HAM-D6 is more quickly administered (Helmreich et al., 2015). In our study, moreover, the effect size in the bipolar depression with mixed features group was slightly lower than in the other groups, probably due to their lower response to treatment, as has been described previously (Fountoulakis et al., 2012; Pae et al., 2012).

Some limitations should be considered in our study. The HAM-D6 was not designed to assess bipolar depression or bipolar depression

	Cincinnati criteria[a]	DSM-IV[b]	DSM-5[c]
A full syndrome of the actual episode	■	■	■
A full syndrome of the opposite pole		■	
Simultaneous presence of at least three of symptoms of the opposite pole	■		■
Symptoms occurring nearly every day for at least one week		■	
Manic and depressive symptoms occurring simultaneously or alternating rapidly, i.e., within minutes	■		
Manic and depressive symptoms are simultaneously present for at least 24 h	■		
Symptoms are present during the majority of days of the current or most recent episode			■

[a] (McElroy et al., 1992) [b] (American Psychiatric Association, 1994) [c](American Psychiatric Association, 2013)

Fig. 1. Mixed symptoms according to the Cincinnati criteria, DSM-IV and DSM-5.
[a] (McElroy et al., 1992) [b] (American Psychiatric Association, 1994) [c] (American Psychiatric Association, 2013).

with mixed features, since these patients may present other symptoms besides the core MDD symptoms. Nevertheless, our results show that the HAM-D6 can be used to assess patients with bipolar depression (Holmskov et al., 2017; K. Martiny et al., 2013) and bipolar depression with mixed features. Another limitation of our study is the use of quality of life scales and depression rating scales to assess concurrent validity, since they rate different constructs.

In the present study, the HAM-D6 was highly correlated with the HAM-D17, inversely correlated with WHOQOL-BREF and SF-36 scores, and did not differ significantly in effect size with the HAM-D17. The HAM-D6 was shown to be a unidimensional scale for assessing MDD, bipolar depression and bipolar depression with mixed features. Thus, we can conclude that the HAM-D6 can be used to assess depression severity in heterogeneous depression samples, with the advantage that the HAM-D6 can be applied more quickly than the HAM-D17. Although it is an abbreviated version, the HAM-D6 has been shown to be more informative than the HAM-D17. Therefore, the use of HAM-D6 could contribute not only as a randomized controlled trial outcome measure, but in treatment evaluation, especially in public health settings. Finally, we reiterate that, to the best of our knowledge, this is the first study to have included bipolar depression patients with mixed features in an assessment of HAM-D6 unidimensionality.

Author disclosure

The authors declare that they have no conflicts of interest.

References

American Psychiatric Association [APA], 2013. Diagnostic and Statistical Manual of Mental Disorders - DSM-5, 5 ed. (Arlington, VA).

American Psychiatric Association [APA], 1994. Diagnostic and Statistical Manual of Mental Disorders - DSM IV, IV. ed. (Washington, DC).

Associação Brasilieira de Empresas de Pesquisa [ABEP], 2009. Critério de Classificacão Econômica Brasil. http://www.abep.org/Servicos/Download.aspx?id=04, Accessed date: 13 January 2017.

Bachner, Y.G., 2016. Psychometric properties of responses to an Arabic version of the Hamilton depression rating scale (HAM-D6). J. Am. Psychiatr. Nurses Assoc. 22, 27–30. https://doi.org/10.1177/1078390316629959.

Bagby, R.M., Ryder, A.G., Schuller, D.R., Marshall, M.B., 2004. The Hamilton depression rating scale: has the gold standard become a lead weight? Am. J. Psychiatr. 161, 2163–2177.

Ballesteros, J., Bobes, J., Bulbena, A., Luque, A., Dal-Ré, R., Ibarra, N., Güemes, I., 2007. Sensitivity to change, discriminative performance, and cutoff criteria to define remission for embedded short scales of the Hamilton depression rating scale (HAMD). J. Affect. Disord. 102, 93–99. https://doi.org/10.1016/j.jad.2006.12.015.

Bech, P., Allerup, P., Gram, L., Rosenberg, R., Jacobsen, O., Nagy, A., 1981. The Hamilton Depression Scale: evaluation of objectivity using logistic models. Acta Psychiatr. Scand. 63, 290–299.

Bech, P., Boyer, P., Germain, J., Padmanabhan, K., Haudiquet, V., Pitrosky, B., Tourian, K., 2010. HAM-D17 and HAM-D6 sensitivity to change in relation to desvenlafaxine dose and baseline depression severity in major depressive disorder. Pharmacopsychiatry 43, 271–276.

Bech, P., Fava, M., Trivedi, M.H., Wisniewski, S.R., Rush, a J., 2011. Factor structure and dimensionality of the two depression scales in STAR*D using level 1 datasets. J. Affect. Disord. 132, 396–400. https://doi.org/10.1016/j.jad.2011.03.011.

Bech, P., Gram, L.F., Dein, E., Jacobsen, O., Vitger, J., Bolwing, T.G., 1975. Quantitative rating of depressive states. Acta Psychiatr. Scand. 51, 161–170.

Berk, M., Malhi, G.S., Cahill, C., Carman, A., Hadzi-Pavlovic, D., Hawkins, M.T., Tohen, M., Mitchell, P.B., 2007. The bipolar depression rating scale (BDRS): its development, validation and utility. Bipolar Disord. 9, 571–579.

Bobes, J., Bulbena, A., Luque, A., Dal-ré, R., Ballesteros, J., Ibarra, N., Psicométricas, D.E., 2016. Evaluación psicométrica comparativa de las versiones en español de 6, 17 y 21 ítems de la Escala de valoración de Hamilton para la evaluación de la depresión, vol. 120. pp. 693–700.

Bowden, C.L., Singh, V., Thompson, P., et al., 2007. Development of the bipolar inventory of symptoms scale. Acta Psychiatr. Scand. 116, 189–194.

Dunn, T.J., Baguley, T., Brunsden, V., 2013. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br. J. Psychol. 105, 399–412.

Fabre, L., Abuzzahab, F., Ami, M., Claghorn, J., Mendels, J., Petrie, W., Dubé, S., Small, J., 1995. Sertraline safety and efficacy in major depression: a double-blind fixed-dose comparison with placebo. Biol. Psychiatr. 38, 592–602.

Fountoulakis, K.N., Kontis, D., Gonda, X., et al., 2012. Treatment of mixed bipolar states. Int. J. Neuropsychopharmacol. 15, 1015–1026.

González-Pinto, A., Mosquera, F., Reed, C., Novick, D., Barbeito, S., Vega, P., Bertsch, J., Alberich, S., Haro, J.M., 2009. Validity and reliability of the Hamilton depression rating scale (5 items) for manic and mixed bipolar disorders. J. Nerv. Ment. Dis. 197, 682–686. https://doi.org/10.1097/NMD.0b013e3181b3b3a0.

Guy, W., 1976. ECDEU Assessment Manual for Psychopharmacology. Department of Health, Education, and Welfare, Rockville,MD, U.S 1976.

Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23, 56–62.

Helmreich, I., Wagner, S., König, J., Kohnen, R., Szegedi, A., Hiemke, C., Tadić, A., 2015. Hamilton depression rating subscales to predict antidepressant treatment outcome in the early course of treatment. J. Affect. Disord. 175, 199–208. https://doi.org/10.1016/j.jad.2014.12.043.

Higuchi, T., Kamijima, K., Nakagome, K., Itamura, R., Asami, Y., Kuribayashi, K., Imaeda, T., 2016. A randomized, double-blinded, placebo-controlled study to evaluate the efficacy and safety of venlafaxine extended release and a long-term extension study for patients with major depressive disorder in Japan. Int. Clin. Psychopharmacol. 31, 8–19. https://doi.org/10.1097/YIC.0000000000000105.

Holmskov, J., Licht, R., Andersen, K., Stage, T., Nilsson, F.M., Stage, K., Valentin, J., Bech, P., 2017. Diagnostic conversion to bipolar disorder in unipolar depressed patients participating in trials of antidepressants. Eur. Psychiatr. 40, 76–81.

Hopkins, W.G., Marshal, l S.W., Batterham, A.M., Hanin, J., 2009. Progressive statistics for studies in sports medicine and exercise science. Med. Sci. Sports Exerc. 41, 3–13.

Khan, A., Bhat, A., Kolts, R., Thase, M., Brown, W., 2010. Why has the antidepressant-placebo difference in antidepressant clinical trials diminished over the past three decades? CNS Neurosci. Ther. 16, 217–226.

Lecrubier, Y., Bech, P., 2007. The Ham D(6) is more homogenous and as sensitive as the Ham D(17). Eur. Psychiatr. 22, 252–255. https://doi.org/10.1016/j.eurpsy.2007.01.1218.

Lecrubier, Y., Sheehan, D., Weiller, E., Amorim, P., Bonora, I., Sheehan, K., Janavs, J., Dunbar, G., 1997. The Mini International Neuropsychiatric Interview (M.I.N.I.), a short diagnostic interview: reliability and validity according to the CIDI. Eur. Psychiatr. 12, 232–241.

Lee, C.P., 2017. Psychometric evaluation of a 6 item Chinese version of the Hamilton Depression Rating Scale : Mokken scaling and item analysis. Asia Pac. Psychiatr. 95, 5–7. https://doi.org/10.1111/appy.12287.

Licht, R.W., Qvitzau, S., Allerup, P., et al., 2005. Validation of the Bech-Rafaelsen Melnacholia Scale and the Hamilton Depression Scale in patients with major depression: is the total score a valid measure of illness severity? Acta Psychiatr. Scand. 111, 144–149.

Lima, A.F., Miguel, S.R., Cohen, M., Zimmermann, J.J., Shansis, F.M., Cruz, L.N., 2017.

Effectiveness evaluation of mood disorder treatment algorithms in Brazilian public healthcare patients. Rev. Bras. Psiquiatr 0, 000–000.
Martiny, K., Refsgaard, E., Lund, V., Lunde, M., Sørensen, L., Thougaard, B., Lindberg, L., Bech, P., 2013. The day-to-day acute effect of wake therapy in patients with major depression using the HAM-D6 as primary outcome measure: results from a randomised controlled trial. PLoS One 8, e67264. https://doi.org/10.1371/journal.pone.0067264.
McElroy, S., Keck Jr., P., Pope Jr., H., Hudson, J., Faedda, G., Swann, A., Keck, P.J., Pope, Hgj, Hudson, J., Faedda, A., Swann, G., 1992. Clinical and research implications of the diagnosis of dysphoric or mixed mania or hypomania. Am. J. Psychiatr. 149, 1633–1644.
Mcelroy, S.L., Keck, P., 2017. Dysphoric mania, mixed states, and mania with mixed features specifier : are we mixing things up . CNS Spectr. 22, 170–176. https://doi.org/10.1017/S1092852916000717.
Mokken, R.J., 1971. Atheory and Procedure of Scale Analysis. (Paris).
Østergaard, S.D., Bech, P., Miskowiak, K.W., 2016. Fewer study participants needed to demonstrate superior antidepressant efficacy when using the Hamilton melancholia subscale (HAM-D_6) as outcome measure. J. Affect. Disord. 190, 842–845. https://doi.org/10.1016/j.jad.2014.10.047.
Østergaard, S.D., Bech, P., Trivedi, M.H., Wisniewski, S.R., Rush, A.J., Fava, M., 2014a. Brief, unidimensional melancholia rating scales are highly sensitive to the effect of citalopram and may have biological validity: implications for the Research Domain Criteria (RDoC). J. Affect. Disord. 163, 18–24. https://doi.org/10.1016/j.jad.2014.03.049.
Østergaard, S.D., Meyers, B.S., Flint, A.J., Mulsant, B.H., Whyte, E.M., Ulbricht, C.M., Bech, P., Rothschild, A.J., 2014b. Measuring psychotic depression. Acta Psychiatr. Scand. 129, 211–220. https://doi.org/10.1111/acps.12165.
Pae, C.U., V€ohringer, P.A., Holtzman, N.S., et al., 2012. Mixed depression: a study of its phenomenology and relation to treatment response. J. Affect. Disord. 136, 1059–1061.
Santor, D.A., Coyne, J.C., 2001. Examining symptoms expression as a function of symptom severity: item performance on the Hamilton Rating Scale for depression. Psychol. Assess. 13, 127–139.
Sawilowsky, S., 2009. New effect size rules of thumb. J. Mod. Appl. Stat. Meth. 8, 467–474.
Shansis, F.M., Reche, M., Capp, E., 2016. Evaluating response to mood stabilizers in patients with mixed depression: a study of agreement between three different mania rating scales and a depression rating scale. J. Affect. Disord. 197, 1–7. https://doi.org/10.1016/j.jad.2016.02.064.
Sijtsma, K., Molenaar, I., 2002. Introduction to Nonparametric Item Response Theory. Sage Publications, Thousand Oaks, CA.
Spitzer, R.L., Endicott, J., Robins, E., 1978. Research diagnostic criteria: rationale and reliability. Arch. Gen. Psychiatr. 35, 773–782.
Streiner, D.L., 2003. Being inconsistent about consistency: when coefficient alpha does and doesn't matter. J. Pers. Assess. 80, 217–222.
Swann, A.C., Lafer, B., Perugi, G., Frye, M.A., Bauer, M., Bahk, W.M., Scott, J., Ha, K., Suppes, T., 2013. Bipolar mixed states: an international society for bipolar disorders task force report of symptom structure, course of illness, and diagnosis. Am. J. Psychiatr. 170, 31–42. https://doi.org/10.1176/appi.ajp.2012.12030301.
Timmerby, N., Andersen, J.H., Søndergaard, S., Østergaard, S.D., Bech, P., 2017. A systematic review of the clinimetric properties of the 6-item version of the Hamilton depression rating scale. Psychother. Psychosom. 86, 141–149. https://doi.org/10.1159/000457131.
Tortorella, A., Albert, U., Nivoli, A.M.A., Erfurth, A., Colom, F., Maina, G., 2015. Mixed states: still a modern psychopathological syndrome? J. Psychopathol. 21, 332–340.
van der Ark, L.A., 2012. New developments in Mokken scale analysis in R. J. Stat. Software 48, 1–27.
Ware, J.J., Sherbourne, C.D., 1992. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med. Care 30, 473–483.
Whoqol Group, 1998. Development of the World health organization WHOQOL-BREF quality of life assessment. The WHOQOL group. Psychol. Med. 28, 551–558. https://doi.org/10.5.12.
Zimmerman, M., 2017. Measures of the DSM – 5 mixed-features specifier of major depressive disorder. CNS Spectr. 196–202. Apr;22. https://doi.org/10.1017/S1092852916000857.
Zimmerman, M., Posternak, M.A., Marshall, M.B., et al., 2005. Is it time to replace the Hamilton Depression Rating Scale as the primary outcome measure in treatment studies of depression? J. Clin. Psychopharmacol. 25, 105–110.