Validity and reliability of the Spanish versions of the Bech-Rafaelsen’s mania and melancholia scales for bipolar disorders


Objective: To assess classical psychometric properties of the Spanish versions of the Bech-Rafaelsen’s mania (MAS) and melancholia (MES) scales.

Method: Observational, prospective, and multicentric study in bipolar out-patients. Convergent validity was assessed against the Young Mania Rating Scale and the Montgomery-Åsberg Depression Rating Scale. Discriminant validity, reliability, and sensitivity to change, were also assessed.

Results: One hundred and thirteen bipolar patients with a manic episode and 102 bipolar patients with a depressive episode were included. Both the MAS and the MES showed appropriate convergent validity ($r > 0.90$), discriminant validity ($P < 0.0001$), internal consistency (Cronbach’s alpha > 0.80), test–retest reliability [intraclass correlation coefficient (ICC) = 0.69 for the MAS and 0.94 for the MES], inter-rater reliability (ICC > 0.80), and sensitivity to change at 4 weeks since inception ($P < 0.0001$; within-group effect size ≥1.8).

Conclusion: The Spanish versions of both scales present appropriate psychometric estimates in bipolar patients treated in ambulatory care.

Significant outcomes

- The Spanish versions of the MAS and the MES show adequate psychometric properties regarding their reliability, validity, and sensitivity to clinical change in bipolar out-patients.
- The classical psychometric properties of the MAS and the MES are similar to those of competing severity scales as the YMRS and the MADRS in bipolar out-patients.

Limitations

- Contrary to other studies, this one was not embedded within an efficacy trial, and thus its reported estimates might be conservatively biased.
- This study was performed with bipolar out-patients on treatment as usual, and thus its results do not necessarily map to those obtained in other clinical settings.
- The MAS and the MES were translated to Castilian Spanish. As both scales must be administered by trained raters, their wording should be appropriately adapted to other local variations of Spanish if needed.
Introduction

There are just a few canonical scales which are customarily used to assess the severity – and its change over time – of symptomatic symptoms of mania and depression in bipolar patients. Among them, the Young Mania Rating Scale (YMRS) (1), the Hamilton Depression Rating Scale (2, 3), and the Montgomery-Asberg Depression Rating Scale (MADRS) (4). Others like the Bech-Rafaelsen Mania Scale (MAS) (5) or the Bech-Rafaelsen Melancholia Scale (MES) (6) have not attained a widespread use despite presenting in head-to-head comparisons to canonical scales comparable psychometric properties (7–10).

So far, the above scales have been recommended as main outcome measures in the analysis of efficacy from clinical trials with patients presenting bipolar disorders (11), and thus they have exerted a profound influence on the accumulation of knowledge in that area. As a consequence of their influence, most scales have been translated to languages other than the original (mainly English) for their use either in clinical or in experimental research. Unfortunately such translations (and further validations) have been commonly embedded with their use as primary outcome measures in pivotal clinical trials, or in postmarketing studies, in which the psychometric protocol was not the primary aim but an ancillary one. In the last years, our group undertook the task of validating the most used severity scales for the assessment of affective disorders by using proper psychometric designs and working with out-patients on treatment as usual (12–16). Spanish is currently spoken by over 352 million people worldwide; it is the fourth most spoken language after Chinese, English, and Hindi. Whereas there are dialects and local variations in different countries and regions, Castilian Spanish is a common language which is easily understood by all Spanish speakers in spite of the diversity local construction and vocabulary differences. The availability of valid and reliable versions of psychometric scales in Spanish is therefore crucial for research development and clinical assessment not only in Spanish speaking countries but also in countries speaking other languages but where immigration from Spanish speaking countries is an issue. As, to the best of our knowledge, there were not previous studies focusing upon the validation of both the MAS and the MES in their Spanish translation, apart from a former study devoted to the analysis of the construct validity of the MES (17), we decided to assess classical psychometric properties of the Spanish versions of both scales to ascertain their use as clinical and research instruments with bipolar patients.

Aims of the study

i) To assess the convergent validity of the MAS and the MES scales when compared with the YMRS and the MADRS respectively; ii) to assess the discriminant validity of both scales by comparing the distribution of their scores with the Clinical Global Impression (CGI); iii) to assess their reliability (internal consistency; inter-rater and test–retest reliability); and iv) to assess their sensitivity to the clinical change of bipolar out-patients on treatment as usual over an adequate follow-up time.

Material and methods

Design

The study was designed as a psychometric study on its own. It was not embedded into another observational or experimental design in which psychometrics would be included as an ancillary aim. It is an observational, short-term prospective (follow-up period of 1 or 4 weeks), and multicentric study (18 psychiatric centers across Spain were included), in out-patients diagnosed of Bipolar I or II Disorder (DSM-IV-TR or ICD-10 criteria) who were on treatment as usual.

All consecutive bipolar patients attending out-patient psychiatric services of the clinical centers involved in the study were approached to ascertain their willingness to participate in the study. Those patients who agreed and fulfilled the inclusion criteria were recruited into the study. The research protocol was approved by the corresponding Ethics and Research Board. Inclusion criteria were i) to get a diagnosis of Bipolar I or II Disorder according to DSM-IV-TR or ICD-10 criteria; ii) to present an adequate severity level at baseline in unstable patients to allow for a sensible analysis of the sensitivity to change of the MAS and MES scales (a minimum score of 18 and 22 points for the YMRS and MADRS respectively was recommended for inclusion but the final decision was left to the researchers on the basis of their clinical judgement); iii) an age ≥18 years; iv) to provide a signed informed consent to participate in the study; and v) not to present with any relevant cognitive problem which could interfere with the correct understanding of the clinical questions or with the completion of the clinical scales.

According to the methods reported in previous psychometric studies (13, 14), at study inception
patients were judged to be on stable or unstable clinical conditions by their psychiatrists on charge. Patients considered to be in stable condition were those who were not expected to change significantly in their clinical severity in 1-week time. On the contrary, unstable patients were those others in whom a clinical significant change in severity was expected (either by a change of treatment or because of presenting new or recurrent manic or depressive episodes). Stable patients underwent a test–retest reliability design at 1 week since inception. Unstable patients underwent a sensitivity to change design at 4 weeks since inception as part of the validity analysis. Both groups were analyzed at baseline to ascertain the internal consistency, and the convergent and discriminant validity of the MAS and the MES scales. Additionally a subgroup of patients presenting with manic or depressive episodes were assessed at baseline at three centers by two independent psychiatrists to evaluate the inter-rater reliability. The study protocol was approved by the corresponding clinical centers’ ethical and research committees. All patients gave their written informed consent before being included in the study. Figure 1 shows the flowchart of the study.

Sample size

Minimum estimates for sample size were calculated according to the psychometric characteristics to evaluate: validity, reliability, and sensitivity to change. All calculations were performed assuming 5% significance (two-tailed), and a 90% power. For the test–retest and inter-rater reliability, the minimum estimated sample size was of 20 pairs assuming a correlation among times or raters of 0.70. The minimum sample size to estimate the sensitivity to change, assuming a decrease of 10 points in the scales (SD = 11) and a moderate correlation (0.40) among baseline and final scores (4 weeks), was also of 20 pairs. For convergent validity, the estimated sample size was of 40 patients, assuming a correlation of 0.70 among validating and reference scales (correlation coefficient under the null hypothesis 0.30). Final minimum sample size estimates were increased assuming a 10% drop-out rate throughout the

---

**Fig. 1.** Flowchart of the validation study.
study. Figure 1 presents the sample size attained at inception and the sample size analyzed at the corresponding follow-up points.

Scales to validate
The MAS and the MES were developed to assess the severity of manic and depressive symptoms (5). Both contain 11 items which are scored according to a five-category Likert scale (0 = not present, 1 = very mild or doubtful, 2 = mild, 3 = moderate, 4 = severe). Their total score range from 0 to 44 points. In this study, the MAS and the MES underwent first an English–Spanish translation from their original formats as sent to us from one of the original researchers (18), and then a backward translation, by using four independent translators; two clinicians with English as second language and two English translators working at a Medical School with Spanish as second language. The final Spanish versions of both the MAS and the MES appear as annexes S1 and S2 to the manuscript.

Reference scales
The YMRS, and the MADRS were included as reference scales to assess convergent validity (1, 4). The YMRS includes 11 items which are scored on a five-category Likert scale from 0 to 4, except items 5, 6, 8, and 9 that score twice that range. Its total score ranges from 0 to 60 points. The MADRS includes ten items which are scored on a six-category Likert scale from 0 to 6; with each item presenting anchor definitions at even scores. The total score for the MADRS ranges from 0 to 60 points. Both the YMRS and the MADRS Spanish versions were previously validated (14, 19).

A modified CGI for use with bipolar patients, and also validated in Spanish (20), was used as a categorical information source of the symptomatic severity to ascertain discriminant validity and sensitivity to change. The CGI is defined on the customary seven-category Likert scale (1 = not ill, 2 = doubtfully ill, 3 = mildly ill, 4 = moderately ill, 5 = severe, 6 = severely ill, 7 = among the most severe patients).

Statistical analyses
Figure 1 displays, along with the flowchart of the study and the sample sizes attained, the main analyses conducted to assess the classical psychometric properties of the MAS and the MES scales. Estimates for their reliability were obtained by analysing, i) their internal homogeneity by the Cronbach’s alpha statistic, ii) their inter-rater reliability (two independent raters at three centers) by the intraclass correlation coefficient (ICC), and iii) their test–retest reliability by using also the ICC (1 week after recruitment into the study, and within the subgroup of patients considered to be clinically stable). It is important to note that Cronbach’s alpha estimate for internal reliability strongly depends on the scale’s number of items. In our case, its comparison among scales was not affected by such difference as all of them include a similar number of items. Estimates for their convergent validity at baseline were obtained by the Pearson’s correlation among the MAS and the YMRS total scores, and the MES and the MADRS total scores. Discriminant validity was assessed also at baseline by, i) one-way ANOVAS contrasting the mean differences of the MAS and the MES scales across the ordered categories of the CGI, and ii) by evaluating from ordinal regression models the significance of the linear trend in their scores according to the CGI categories. Sensitivity to change was assessed at week 4 since inception, for the subgroup of patients considered to be clinically unstable at baseline, with an appropriate within-group effect size (21). Additionally, the discriminant power of the MAS and the MES change scores since baseline was assessed against the final CGI categories [dichotomized as normal or borderline severity (1/2 scores) vs. mild to extreme severity (3/6 scores)]. For these analyses, the area under the receiving operating characteristic curve (ROC) was used. The non-independent ROC areas so obtained allowed us to compare the discriminant performance regarding the sensitivity to change of the validating and reference scales (MAS vs. YMRS, and MES vs. MADRS). All the data management and analyses were performed with Stata v9.2 (StataCorp, 2005; College Station, TX, USA).

Results
Sample description
Overall, 215 bipolar I and II patients from 18 clinical centers were recruited; 113 presenting with mania [mean age (SD) = 43.1 years (13.2); 60.7% women] and 102 presenting with depression [mean age = 47.3 years (12.7); 67.6% women]. Table 1 shows their main characteristics at baseline according to the current affective episode and clinical stability as predicted by the psychiatrists. As expected by the study inclusion criteria, the psychopathological severity at baseline was significantly different between stable and unstable
patients ($P < 0.0001$ for both the YMRS and the MADRS), even if the scores range presented some overlapping (YMRS: 5–41 and 8–46 for stable and unstable patients respectively; MADRS: 3–46 and 19–53 for stable and unstable patients respectively). However, the shift in the mean distribution of the scores for both groups was clearly apparent with 43 unstable patients over 56 (76.8%) scoring $\geq 18$ in the YMRS [24 over 57 (42.1%) for stable patients], and 43 unstable patients over 49 (87.7%) scoring $\geq 22$ in the MADRS [19 over 53 (35.8%) for stable patients].

Feasibility

The mean time (SD) for the administration of the MAS was 10.9 min (7), and 9.2 min (5.6) for the MES.

Reliability

The mean elapsed time for the test–retest study was 5.8 days (SD = 1.5) for the MAS, and 6 days (1.2) for the MES. Table 2 presents the reliability estimates obtained for both scales and their validation counterparts (YMRS and MADRS). All estimates for the validating scales, but the test–retest correlation for the MAS (ICC = 0.69), were above 0.80, with similar estimates to those obtained by the reference scales.

Convergent validity

Figure 2 displays the linear correlations obtained among the MAS and the YMRS, and the MES and the MADRS. Both of them were well above 0.80.

<table>
<thead>
<tr>
<th>Variables</th>
<th>YMRS</th>
<th>MAS</th>
<th>MADRS</th>
<th>MES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency, Cronbach’s $\alpha$</td>
<td>[n = 113]</td>
<td>[n = 113]</td>
<td>[n = 102]</td>
<td>[n = 102]</td>
</tr>
<tr>
<td>Test–retest reliability at 1 week</td>
<td>0.84</td>
<td>0.88</td>
<td>0.90</td>
<td>0.91</td>
</tr>
<tr>
<td>Mean score test (SD)</td>
<td>16.3 (8.3)</td>
<td>12.9 (6.4)</td>
<td>20.2 (10.2)</td>
<td>14.1 (7.6)</td>
</tr>
<tr>
<td>Mean score retest (SD)</td>
<td>13.3 (7.5)</td>
<td>10.5 (5.8)</td>
<td>18.4 (10.5)</td>
<td>12.9 (8.0)</td>
</tr>
<tr>
<td>Test–retest correlation</td>
<td>0.88</td>
<td>0.76</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>ICC [95% CI]</td>
<td>0.85 (0.78–0.92)</td>
<td>0.69 (0.56–0.83)</td>
<td>0.93 (0.89–0.97)</td>
<td>0.94 (0.90–0.97)</td>
</tr>
<tr>
<td>Inter-rater reliability</td>
<td>[n = 48]</td>
<td></td>
<td>[n = 44]</td>
<td></td>
</tr>
<tr>
<td>Mean score rater 1 (SD)</td>
<td>NA</td>
<td>13.4 (5.0)</td>
<td>NA</td>
<td>15.5 (6.3)</td>
</tr>
<tr>
<td>Mean score rater 2 (SD)</td>
<td>NA</td>
<td>12.9 (5.2)</td>
<td>NA</td>
<td>15.3 (6.4)</td>
</tr>
<tr>
<td>ICC [95% CI]</td>
<td>NA</td>
<td>0.89 (0.80–0.97)</td>
<td>NA</td>
<td>0.98 (0.95–0.99)</td>
</tr>
</tbody>
</table>

YMRS, Young Mania Rating Scale; MAS, Bech-Rafaelsen Mania Scale; MADRS, Montgomery-Åsberg Depression Rating Scale; MES, Bech-Rafaelsen Melancholia Scale; ICC, intraclass correlation coefficient.
Discriminant validity

Both the MAS and the MES showed adequate discriminant validity when compared with the CGI severity scores at baseline (Table 3). The apparent linear trend observed in the table for the MAS and the MES scores were confirmed by further ordinal regression analyses with the CGI categories as dependent variables (P-value for linear trend ≤0.001 in both cases). The results obtained for the validating scales run in parallel with those obtained for the YMRS and the MADRS.

Sensitivity to clinical change

The mean elapsed time to assess the sensitivity to change was 29.5 days (SD = 8) for the MAS, and 28.5 days (8.1) for the MES. Table 4 shows the estimates obtained for the sensitivity to change as evaluated after ≈ 4 weeks on treatment as usual. The within-group effect size for the validating scales and their counterparts were large enough (>1.5) to support an appropriate sensitivity to clinical change, as were the results of the paired t-tests among the baseline and final scores. There were not significant differences for the within-group effect sizes when both validating scales were compared against their counterparts (MAS vs. YMRS difference: z = 1.1, P = 0.29; MES vs. MADRS difference: z = 0.48, P = 0.63). The
discriminative power for the estimates of the manic scales (MAS, YMRS), as evaluated by the ROC area, was somehow lower than the values obtained by the depression scales (MES, MADRS). The ROC area for the MAS was not different to the ROC area for the YMRS \(X^2 (1 \text{ df}) = 0.87, P = 0.35\), and the ROC area for the MES was not different to the ROC area for the MADRS \(X^2 (1 \text{ df}) = 0.06, P = 0.80\).

Discussion

The use of valid and reliable instruments is essential to psychiatry given the subjective nature of many symptoms and the lack of external validators (22, 23). In this study both the Spanish versions of the MAS and the MES have shown adequate psychometric properties regarding reliability, and validity, including their sensitivity to change in bipolar out-patients. Moreover, their psychometric estimates are closely comparable to those obtained for the YMRS and the MADRS as reference scales. Our results thus add to the findings reported by other authors regarding the psychometric equivalence of the MAS and the MES to other competing scales (7–10, 24).

As our study was not embedded within an efficacy trial but was designed to reflect, as closer as possible, the usual clinical practice with bipolar patients (broad inclusion criteria and patients in treatment as usual), its results could actually be underestimates of those reported previously within the framework of efficacy trials (see, for instance, the scores range for mania and depression at baseline, which cover a broader range of symptomatic severity – from mild to severe – than the criteria usually followed for the inclusion of bipolar patients in efficacy trials). However, this does not seem to be true, at least regarding the comparisons with the results reported by others for internal homogeneity, inter-rater reliability, discriminative validity, and sensitivity to change (7–9, 25, 26).

In summary, in this study, both the MAS and the MES have shown adequate and comparable psychometric results to those achieved by other canonical scales (YMRS, MADRS). Both scales could be then used in bipolar out-patients to assess their symptomatic profile, their severity, and their change over time; and both scales may be then added to the few severity scales adequately translated and validated into Spanish to assess bipolar patients for clinical or research purposes. Probably, the last step – if any – needed for definitely including both the MAS and the MES among the canonical scales usually used for the assessment of bipolar disorders, would be further research on their standardization against independent criteria for clinical response and remission, and on their comparative performance against patient research outcomes designed to tap those clinical constructs (27).

Acknowledgements

We are indebted to Professor Per Bech for giving us free permission to validate the 1996 versions of the MAS and the MES scales in Spanish. Eduard Vieta, Ana González-Pinto, Antonio Lobo, Anabel Martínez-Arán, María Reinares, José Manuel Goikolea and Antoni Benabarre (Hospital Clinic i Provincial, Barcelona); Julio Bobes, María Paz García-Portilla and Pilar Saiz (CSM II, Oviedo); José Cañete and Silvia Canizares (Hospital Mataró, Barcelona); José Vicente Baeza and Milagros Fuentes (Hospital General de Elche, Alicante); José Ramón Domènech and Antonio Córdoba (Hospital Dos de Maig, Barcelona); Antonio Bulbena, Purificación Salgado and Carles Masip (Hospital del Mar, Barcelona); Diego Palao, Blanca Navarro and Jesús Mendoza (Hospital General de Vic, Barcelona); Josep Gascón, María José Martin and Maria Luque (Hospital Mutua de Tarrasa, Barcelona); Antonio Higuera, Guillermo Pardo and José Eduardo Muñoz (Hospital Virgen de las Nieves, Granada); Ana González-Pinto, Eider Tapia, Ainara Jiménez, María Gracia Domínguez, Saioa Azpiazu, Patricia Vega and Sara Barbeito (Hospital Santiago Apóstol, Vitoria); José Ramón Gutiérrez, Javier Busto and Fernando Galán (Hospital Infantia Cristina, Badajoz); Eduardo García-Camba, Elena Ezquiaga and Alejandra Martín (Hospital de la Princesa, Madrid); Ignacio Tortajada, Julio Alonso, Jorge Pérez, Mario Páramo and Alicia Crespi (Hospital de Conxo, Santiago de Compostela); Antonio Lobo, Federico Dourdil and Raúl López (Hospital Clínico Universitario Lozano Blesa, Zaragoza); Jesús de la Gándara and Nuria Español (Hospital Divino Valles, Burgos); Pedro Pozo, María Angles de Haro and Juan Francisco Tello (Hospital General Universitario José M Morales Meseguer, Murcia); Manuel Serrano, Juan Carlos Díez, Domingo de Miguel and María José Avila (Complejo Hospitalario Universitario Juan Canalejo, La Coruña); Miguel Roca and María Jesús Serrano (Hospital Joan March, Palma de Mallorca), Sara Ruiz (Quintiles).
Vieta et al.

Manuel Goikolea, Antoni Benabarre and Javier Ballesteros thank the support of the Spanish Ministry of Health, Instituto de Salud Carlos III, RETICS RD06/0011 (REM-TAP Network). These authors plus Julio Bobes are currently included in the Spanish network for psychiatric research CIBER-SAM (Spanish Ministry of Health, Instituto de Salud Carlos III).

Declaration of interest

This study was funded by AstraZeneca Farmacéutica Spain SA.
Dr Eduard Vieta has received grant support, acted as consultant, or given presentations for the following pharmaceutical companies: Almirall, AstraZeneca, Bial, Bristol-Myers-Squibb, Eli Lilly, Glaxo-Smith-Kline, Janssen-Cilag, Merck Sharp & Dohme, Lundbeck, Novartis, Organon, Otsuka, Pfizer, Sanofi-Aventis, Servier, and UCB. Dr Julio Bobes has received grant support, acted as consultant, or given presentations for the following pharmaceutical companies: AstraZeneca, Bristol-Myers-Squibb, Eli Lilly, Glaxo-Smith-Kline, Janssen-Cilag, Otsuka, Pfizer, and Sanofi-Aventis. He is also in the Editorial Board of Acta Psychiatraca Scandinavica. Dr Javier Ballesteros has received grant support from Glaxo-Smith-Kline, Eli Lilly, and AstraZeneca. Dr Ana González-Pinto has received grant support, acted as consultant, or given presentations for the following pharmaceutical companies: Almirall, Astra-Zeneca, Bristol-Myers-Squibb, Eli Lilly, Glaxo-Smith-Kline, Janssen-Cilag, Sanofi-Aventis, Lundbeck, Novartis, and Pfizer. Mr Antonio Luque was a key figure both in the design and in the administrative follow-up of this study as Head of Therapeutical Area of Neuroscience in the Medical Department of AstraZeneca Pharmaceuticals Spain until his departure on October 30, 2006. Mrs Nora Ibarra does not present any conflict of interest.

Supplementary material

The following supplementary material is available for this article:

Annexe S1. Spanish version of the Bech-Rafaelsen’s Mania Scale (MAS)
Annexe S2. Spanish version of the Bech-Rafaelsen’s Melancholia Scale (MES)

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1600-0447.2007.01133.x (This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References


Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.


