

RESEARCH

Open Access



Accuracy of the Arabic HCL - 32 and MDQ in detecting patients with bipolar disorder

Uta Ouali^{1,2,3}, Yosra Zgueb^{1,2,3*}, Lamia Jouini⁴, Amina Aissa^{1,2,3}, Rabaa Jomli^{1,2,3}, Abdelhafidh Ouertani^{1,3}, Adel Omrani⁵, Fethi Nacef^{1,2,3^}, Mauro G. Carta⁶ and Antonio Preti⁷

Abstract

Background Studies about the two most used and validated instruments for the early detection of Bipolar Disorder (BD), the 32 - item Hypomania Checklist (HCL - 32) and the Mood Disorder Questionnaire (MDQ), are scarce in non-Western countries. This study aimed to explore the reliability, factor structure, and criterion validity of their Arabic versions in a sample of Tunisian patients diagnosed with mood disorders.

Methods The sample included 59 patients with BD, 86 with unipolar Major Depressive Disorder (MDD) and 281 controls. Confirmatory factor analysis was applied to show that a single global score was an appropriate summary measure of the screeners in the sample. Receiver Operating Characteristic analysis was used to assess the capacity of the translated screeners to distinguish patients with BD from those with MDD and controls.

Results Reliability was good for both tools in all samples. The bifactor implementation of the most reported two-factor model had the best fit for both screeners. Both were able to distinguish patients diagnosed with BD from putatively healthy controls, and equally able to distinguish patients diagnosed with BD from patients with MDD.

Conclusion Both screeners work best in excluding the presence of BD in patients with MDD, which is an advantage in deciding whether or not to prescribe an antidepressant.

Keywords Bipolar disorder, Depression, Screening, Hypomania check list (HCL - 32), Mood disorder questionnaire (MDQ), Hypomania

Background

Bipolar disorder (BD) is a severe mental disorder with a chronic-recurring course. Since the first episode of a BD is often of a depressive kind [1], BD is often misdiagnosed as major depressive disorder (MDD) [2, 3] and treated as such. This may lead to adverse consequences, such as increased suicide risk [4, 5], greater probability of hospitalization [5], poorer response to antidepressants, and antidepressant-induced switch to mania [6]. Because of its course characterized by recurring episodes separated by periods of euthymia with no or scant symptoms of hypomania, BD may persist undiagnosed for a long time unless a frank episode of mania erupts [7, 8]. Minor hypomanic episodes are often overlooked, and, indeed, the differentiation of clinically elated and irritable

[^]Fethi Nacef is deceased.

*Correspondence:

Yosra Zgueb

zguebyosra@gmail.com

¹ Department Psychiatry A, Razi Hospital, Rue des Orangers, 2010 La Manouba, Tunisia

² Faculty of Medicine of Tunis, University of Tunis El Manar, Tunis, Tunisia

³ Research Laboratory LR18SP03, Tunis, Tunisia

⁴ Centre de Compétences en Psychiatrie et Psychothérapie, Pôle de Psychiatrie et Psychothérapie, Hôpital du Valais (HVS)- Centre Hospitalier du Valais Romand, Sion, Switzerland

⁵ Tunisian Bipolar Forum, Erable Médical Cabinet 324, Tunis, Tunisia

⁶ Department of Medical Sciences and Public Health, University

of Cagliari, Cagliari, Italy

⁷ Department of Neuroscience, University of Turin, Turin, Italy



mood or increased activity from “normal” variation in the population is often challenging. On average, the duration of undiagnosed, hence untreated, BD may last up to 10 years, and there is some evidence that up to one-third of patients with BD are misdiagnosed at least once during their lifetime [9, 10].

Early identification of BD is essential for appropriate treatment [11]. Several self-report tools have been developed to identify people with possible or probable BD [12]. Self-report screening tools are brief and cost-effective and can be preferred in the busy clinical setting to standardized interviews, which are more accurate but are time-consuming and require appropriate training for the administration and scoring. Nevertheless, caution should be applied in deriving epidemiologic estimates from case-finding based on screening tools [13]. Two of the most used and validated instruments for the early detection of BD are the 32-item Hypomania Checklist (HCL - 32) [11], and the Mood Disorder Questionnaire (MDQ) [14]. These two instruments have been validated in many countries [11, 15–24]. There is evidence that both the HCL - 32 and the MDQ have acceptable psychometric properties and appear to be useful screening tools for BD [25].

Most studies on the HCL - 32 and the MDQ have been carried out in Western countries. The epidemiology of BD shows minor variations by country and ethnicity [26, 27], and the disorder has a likely genetic basis rooted in evolutionary mechanisms [28, 29]. However, cultural factors may influence how symptoms leading to the diagnosis of BD are evaluated [26, 30]. For example, geographical variations in the prevalence of BD might be in part a reflection of the relevance given to the occurrence of psychotic features in BD. The diagnosis of schizophrenia is given priority when the possibility that psychotic features may also occur in the course of BD is overlooked [31]. More subtle influences are related to cultural variations in the patients’ attitudes towards their symptoms. For example, there is evidence from factor analysis that greater involvement in sexual activity, an oft-observed correlate of hypomania, is perceived as a favorable trait by Latin-Mediterranean patients while Asian patients attribute a negative value to hypersexuality, which they tend to associate with other risky behaviors, such as excessive spending or getting in troubles [32, 33]. Studies about the MDQ and the HCL - 32 in non-Western countries are scarce, and most of them are from Asian countries. So far, two studies had explored the reliability and the factor structure of the HCL - 32 [34] and the MDQ [35], respectively, in Arabic-speaking countries. This study aimed at further exploring the reliability, factor structure, and criterion validity of the Arabic version of the HCL - 32 and MDQ in a sample of Tunisian patients

diagnosed with a mood disorder (either MDD or BD) by comparison with a sample of putatively healthy people drawn from the general population of Tunisia.

Material and methods

The study complies with the guidelines of the 1995 Declaration of Helsinki and its revisions [36]. Approval to the study protocol has been granted by the Institutional Review Board (IRB) of Razi Hospital, Tunis, with the authorization signed on 8 Oct 2014.

Participants

The study was conducted between February 2015 and September 2019 and included a patient group and a control group.

Patient group

All consecutive individuals who consulted for the first time at the Department of Psychiatry A of Razi Hospital La Manouba, Tunisia, for signs and symptoms of depression were invited to take part in the study. Individuals were included when the clinician formulated a diagnosis of a current major depressive episode. Thereafter, the Mood Disorder Section of the Tunisian Arabic adapted version of the Structured Clinical Interview for DSM-IV-TR (SCID) was administered by one single researcher (UO) to confirm the diagnosis of Major Depressive Episode and to ascribe the episode to a unipolar or bipolar mood disorder. Resulting diagnosis was (BD) I or II in case a past manic or hypomanic episode were identified, and Major Depressive Disorder (MDD) in case no past manic or hypomanic episode were identified. Additional inclusion criteria were: aged between 18 and 65 years old; having the capacity of providing informed consent. Exclusion criteria were: illiteracy or other cause of inability to read; documented history of mental retardation; and cognitive decline.

Healthy control group

Healthy control subjects were included from the general population upon completion of patient recruitment. Control subjects were gender- and age-matched. Inclusion criteria were the absence of a personal history of any psychiatric disorder or consultation in psychiatry, and the absence of a family history of psychiatric disorder in a first-degree relative. In addition, subjects had to answer “no” to both “A” criteria questions for a lifetime major depressive episode of the SCID.

After inclusion and the administration of the SCID in the patient group, subjects of both groups filled out the MDQ and the HCL - 32. All included subjects provided written informed consent.

Measures

The Arabic version of the MDQ has been used [35]. The MDQ is a self-report tool aimed at screening for potential lifetime indicators of a manic or hypomanic syndrome. It consists of 13 yes/no items evaluating manic symptoms according to DSM-IV criteria [14]. A cut-off of 7 out of 13 items is optimal, in terms of sensitivity and specificity, for identifying BD against healthy people or patients diagnosed with MDD [25].

A Tunisian Arabic version of the HCL - 32 has been used, which was prepared according to standard procedures [37]. At the time of the planning of the study, there was no Arabic version of the HCL - 32. An Arabic version of the HCL - 32 has been published thereafter only [34]. Moreover, each Arabic country has its own dialect, and although there is a standard Arabic language, many people grasp the concepts better in their local Arabic language. As the HCL - 32 has quite a few items which are culturally sensitive (and could therefore be interpreted differently if not understood at 100% - and for detecting hypomania, nuances can sometimes be very important), we preferred to develop a Tunisian Arabic version. Thus, the HCL - 32 was translated into the Tunisian Arabic language by a bilingual native editor, then back-translated into English by another bilingual native editor. A third, independent researcher, with a deep knowledge of the tool, contributed to harmonize the translation and back-translation of the HCL - 32. Potential issues in reading or unclear items were addressed in a pilot study with eight patients, whose help served to complete the translation of the HCL - 32 in its final form.

The HCL - 32 is a self-report questionnaire comprising a list of 32 possible hypomanic symptoms, to be rated as present or absent in a yes/no format. Additional questions concern the duration of the hypomanic experience and the impact on the family, social, and work life. A total score is yielded by the sum of all “yes” replies. A cut-off of 14 out of 32 items is optimal, in terms of sensitivity and specificity, for identifying BD [25].

Statistics

Data were imputed in Excel, then they were coded and analyzed using the Statistical Package for Social Sciences (SPSS) version 27. Specific analyses were done with dedicated packages running [38] in R. All tests were two-tailed, with alpha set at $p < 0.05$.

Descriptive statistics were reported as means with standard deviation, or as counts and percentages. Non-parametric tests were used to assess differences between groups or correlations among variables, except for age.

To assess the usability of the scale in the target population, we calculated floor and ceiling effects [39]. They

occur when more than 15% of respondents score at the minimum (in this case, zero) or the maximum scores (either 13 for the MDQ or 32 for the HCL-32). The occurrence of floor or ceiling effects indicates that extreme items are missing in the lower or upper end of the scale, indicating limited content validity.

Reliability was measured as internal coherence using Cronbach's alpha. The Bayesian reliability analysis, as implemented in JASP 0.14.1 version [40], has been used to calculate the Cronbach's alpha. According to a shared rule-of-thumb, Cronbach's alpha is considered “moderate” when it is > 0.6 and “good” when it is > 0.7 [41].

Before testing the criterion validity of the MDQ and the HCL-32, confirmatory factor analysis (CFA) was applied to the items of both questionnaires to make sure that a single global score was an appropriate summary measure of the screeners in the total sample. Preliminary analysis with the Mardia's test [42] revealed a violation of multivariate normality in the data for both the MDQ and the HCL-32 (skew's $p < 0.0001$ in both analyses). Therefore, the Diagonally Weighted Least Squares (DWLS) estimator was used in CFA. To assess goodness of fit estimation, we used the following parameters: the chi-square, the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). In the presence of a chi-square with $p < 0.001$, as expected with large samples ($n > 300$), RMSEA values of 0.08 or lower, SRMR values of 0.09 or lower, and CFI values of 0.90 or higher were considered an indication of acceptable fit according to conventional rules of thumb [43]. The following model were tested: an unidimensional model, which assumes all core items of the MDQ or the HCL-32 tap into a single dimension of propensity to the manic/hypomanic syndrome; a two-factor model of elated and irritable dimensions, as in Ouali et al., 2020 for the MDQ [35] and in Meyer et al., 2007 for the HCL-32 [15]; and these two-factor models' bifactor implementation [44], which assumes that most variance in the scores is attributable to a general factor resulting from the loading of all items on a single dimension of propensity to the manic/hypomanic syndrome, with an additional but residual variance purportedly explained by the loading of the items on the “elated” and the “irritable” dimensions, as defined above. To check for reasonable unidimensionality of the general factor extracted from the bifactor model, the explained common variance (ECV), the percentage of uncontaminated correlations (PUC), and the Omega Hierarchical (ω_H) were calculated [45]. We also calculated the construct replicability H index of Hancock and Mueller (2001) [46]. H values of .80 or higher indicate a well-defined latent variable, which is more likely to be stable across studies. The presence of multidimensionality might be discarded

when ECV is higher than .60 and $\omega_H > .70$ or $PUC > .70$ [45]. CFA models were tested with the “lavaan” package running in R [47]. The calculation of the bifactor indices was done with the “Bifactor Indices Calculator” package running in R [48].

The receiver operating characteristics (ROC) curve was used to test for the criterion validity of the tools. Criterion validity was intended the degree to which the scores of the instrument were an adequate reflection of a “gold standard” [49]. For the purposes of this study, we used the diagnosis assigned after the SCID interview as a “gold standard” for reference. Thus, the ROC curve analysis was used to distinguish between diagnostic groups for both the MDQ and the HCL-32. Sensitivity was defined as the probability of a true positive case, i.e. the probability of identifying a patient with BD. Specificity was the probability of a true negative case, i.e. the probability of identifying a patient without BD. We also derived the positive predictive value (PPV), i.e., the probability that a person is a case of BD when a positive test result is observed; the negative predictive value (NPV), i.e., the probability that a person is not a case of BD when a negative test result is observed; and the positive diagnostic likelihood ratio, which is the odds ratio that a positive test will be observed in a population of people with BD compared to the odds that the same result will be observed among a population of people without BD. The accuracy in the prediction was estimated from the area under the curve (AUC; with 95% confidence interval). Agreed threshold for the AUC were: $\leq .70$, poor; between .70 and .80, fair; between .80 and .90, good; above .90, excellent [50].

We used the “pROC” package running in R to perform the ROC analysis [51], while the best cut-off point for the MDQ and the HCL-32 was established according to the Youden (1950) method with the “Optimal Cut points” package [52]. The comparison of the two paired ROC curves for MDQ and HCL-32 in the same sample was done with a bootstrap test according to Hanley and McNeil (1983). The test was performed with the “pROC” package.

Sample size estimation and power analysis

CFA and ROC analysis impose some requirements for sample size. As for the CFA, with DWLS applied to binary or ordinal data, a sample size between 200 and 500 subjects is enough for model convergence and parameters’ estimation, according to Monte Carlo simulation studies (Bandalos, 2014). Thus, the global sample size in this study was sufficient to conduct CFA.

As for the ROC analysis, with alpha set at 0.05 and power at 80% (beta = 0.20), with 59 cases of BD and 281 controls, we could detect an AUC as low as 0.612, which is even lower than the minimum fair AUC (0.700). With

the same parameters and 59 cases of BD and 86 cases of MDD, we could test the diagnostic ability of the screeners in discriminating the two diagnoses detecting an AUC as low as 0.632. This power analysis was performed with the “pROC” package running in R [53].

Results

The sample included 86 patients diagnosed with MDD, 22 patients diagnosed with BD-I and 37 patients diagnosed with BD-II. There were also 281 putatively healthy controls (Table 1).

There were no differences by gender or maximum education level among participants; controls were marginally younger than the patients (partial eta-squared = 0.020).

Clinical data were available for patients only. There was no relevant difference in the age of onset of the psychopathology among groups. A family history of depression was observed more often in patients diagnosed with BD-II, while a family history of BD was observed in just 5% of patients diagnosed with MDD and in about 25% of those diagnosed with BD (Table 1 for details).

Patients diagnosed with BD-I were more likely to have attempted suicide and have been more often admitted to a psychiatric service than patients with MDD or BD-II. A prescription of an antidepressant was received by most patients, with no differences by diagnosis. A second-generation antipsychotic was prescribed in about 10% of cases, again with no difference by diagnosis. Lithium was rarely prescribed and only in patients diagnosed with BD-I.

Overall, 86 patients with MDD, 58 patients with BD (either BD-I or BD-II), and 265 controls completed the MDQ; while the HCL-32 was completed by 64 patients with MDD, 32 with BD, and 225 controls.

Floor or ceiling effects

There were no floor effects for the MDQ: 25 controls (8.9%) and just 1 with MDD (1%) scored zero on the MDQ ($\chi^2 = 11.85$; $df = 2$; $p = 0.003$). However, a modest ceiling effect was observed for the MDQ: 4 controls (1.4%) and 11 patients with BD (17.7%) scored 13 on the MDQ ($\chi^2 = 44.38$; $df = 2$; $p < 0.0001$).

There were no floor and ceiling effects for the HCL-32. Overall, in the sample 7 patients scored zero on the HCL-32: 5 controls, 2 with MDD, none with BD ($\chi^2 = 1.28$; $df = 2$; $p = 0.52$). No participants scored 32 on the HCL-32.

Reliability of the questionnaires

Cronbach’s alpha for MDQ was 0.79 (95%CI: 0.76–0.83) in controls; 0.78 (0.75–0.82) in patients with MDD; and 0.71 (0.60–0.81) in patients diagnosed with BD. Cronbach’s alpha for HCL-32 was, respectively, 0.85

Table 1 General characteristics of the participants included in the study

	MDD N = 86	BD-I N = 22	BD-II N = 37	Healthy Controls N = 281	
Sex					$\chi^2 = 4.52; df = 3; p = 0.21$
Men	30 (35%)	13 (59%)	16 (43%)	110 (39%)	
Women	56 (65%)	9 (41%)	21 (57%)	171 (61%)	
Age	42 (11)	42 (9)	41 (9)	38 (13)	$F[3;421] = 2.85; p = 0.037$
Education					$\chi^2 = 12.03; df = 9; p = 0.21$
College or University	39 (45%)	8 (36%)	17 (46%)	102 (55%)	
Age of onset of psychopathology	35 (12)	35 (10)	31 (10)	–	$F[2;142] = 1.34; p = 0.26$
Family history of depression	26 (31%)	6 (27%)	21 (57%)	–	$\chi^2 = 8.40; df = 2; p = 0.015$
Family history of bipolar disorder	4 (5%)	6 (27%)	9 (24%)	–	$\chi^2 = 12.9; df = 2; p = 0.002$
Suicide attempt	14 (16%)	12 (54%)	11 (30%)	–	$\chi^2 = 14.03; df = 2; p = 0.001$
Admission to psychiatric services	16 (18%)	11 (50%)	10 (27%)	–	$\chi^2 = 9.14; df = 2; p = 0.010$
Received a prescription of antidepressants	70 (84%)	19 (86%)	33 (89%)	–	$\chi^2 = 0.73; df = 2; p = 0.69$
Received a prescription of a second-generation antipsychotic	9 (10%)	3 (13%)	4 (11%)	–	Freeman-Halton extension of Fisher's exact test: $p = 0.93$
Received a prescription of lithium	0 (0%)	2 (9%)	0 (0%)	–	Freeman-Halton extension of Fisher's exact test: $p = 0.026$

Data were expressed as counts and percentages within brackets or as mean and standard deviation within brackets

(0.82–0.87) in controls, 0.80 (0.74–0.85) in MDD, and 0.76 (0.68–0.85) in BD.

Confirmatory factor analysis of the factor structure of the MDQ and the HCL-32

For both the MDQ and the HCL-32, the bifactor implementation of the two-factor model had the best fit according to the predefined parameters (Table 2).

For the bifactor model of the MDQ, $H = 0.79$, $ECV = 0.54$, $PUC = 0.60$, and $\omega H = 0.64$.

For the bifactor model of the HCL-32, $H = 0.80$, $ECV = 0.33$, $PUC = 0.48$, and $\omega H = 0.37$.

Thus, for both the MDQ and the HCL-32 there is some indication in favor of a single, reproducible latent

component. However, the multidimensionality in the data might influence the results that can be derived from a global summary score.

Discriminant capacity of the MDQ and the HCL-32

Patients diagnosed with BD scored higher than patients diagnosed with MDD and controls on both the MDQ and the HCL-32 (Table 3).

According to the epsilon-squared effect size (Tomczak and Tomczak, 2014), about 20% of the variance in the sample was attributable to the differences in MDQ by groups, and 10% was attributable to the differences in HCL-32 by groups.

Table 2 Confirmatory factor analysis of the MDQ and the HCL-32. Goodness-of-fit indices of the tested models

Model	χ^2	df	p	CFI	RMSEA (90%CI)	SRMR
<i>MDQ</i>						
Unidimensional	329.2	65	<0.0001	0.894	0.100 (0.089–0.111)	0.089
Two-factor	213.8	53	<0.0001	0.924	0.086 (0.074–0.098)	0.045
Bifactor	77.3	53	0.016	0.990	0.034 (0.015–0.049)	0.045
<i>HCL-32</i>						
Unidimensional	1864.6	464	<0.0001	0.798	0.097 (0.093–0.102)	0.109
Two-factor	1562.0	463	<0.0001	0.842	0.086 (0.081–0.091)	0.098
Bifactor	715.8	432	<0.0001	0.959	0.045 (0.039–0.051)	0.067
Threshold for fit			$p > 0.05$	0.90	≥ 0.08	≥ 0.09

ROC analysis

The MDQ and the HCL-32 were able to distinguish patients diagnosed with BD from putatively healthy controls, with better AUC in MDQ (82.7; 95%CI: 75.3–90.2) than in HCL-32 (73.4; 63.9–83.0) (Fig. 1).

The MDQ (AUC: 88.9; 81.4–96.3) and the HCL-32 (AUC: 83.3; 74.5–92.1) were equally able to distinguish patients diagnosed with BD from patients with MDD (Fig. 2).

Table 3 Scores of the HCL-32 and the MDQ by subgroup of participants

	MDD	BD	Healthy Controls	Kruskal-Wallis	
	N = 86	N = 58	N = 265		
MDQ	5.3 (2.3)	10.0 (2.5)	5.7 (3.4)	H = 77.9; df = 2; p < 0.0001	$\Sigma^2 = 0.191$
	N = 64	n = 32	N = 225		
HCL-32	11.9 (5.8)	19.5 (5.3)	14.5 (6.1)	H = 31.4; df = 2; p < 0.0001	$\Sigma^2 = 0.098$

Data were expressed as mean and standard deviation within brackets

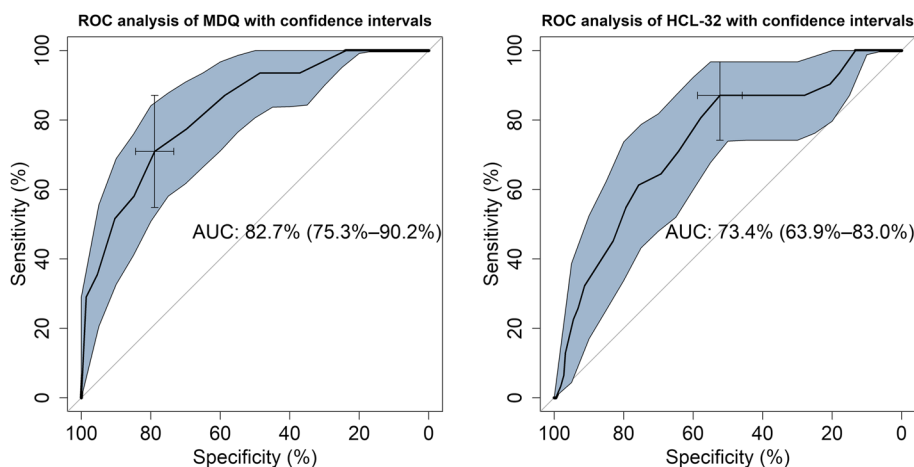


Fig. 1 Receiver operator characteristic (ROC) curve of the predictive capacity of the Tunisian MDQ (on the left) and the Tunisian arabic HCL-32 (on the right) in differentiating patients with BD from healthy controls. Sensitivity and specificity are reported as percentages, with a cross indicating on the curve the best compromise between them (corresponding to the cut-off). The area under the ROC curve (AUC) is reported alongside its 95% confidence interval

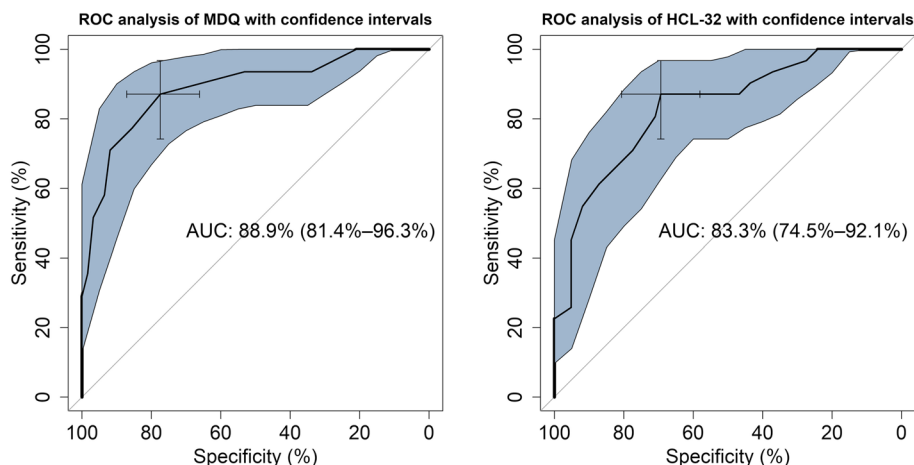


Fig. 2 Receiver operator characteristic (ROC) curve of the predictive capacity of the Tunisian MDQ (on the left) and the Tunisian arabic HCL-32 (on the right) in differentiating patients with BD from patients with MDD. Sensitivity and specificity are reported as percentages, with a cross indicating on the curve the best compromise between them (corresponding to the cut-off). The area under the ROC curve (AUC) is reported alongside its 95% confidence interval

When compared with the Hanley and McNeil’s test, the MDQ was confirmed better than the HCL-32 in distinguishing patients with BD from putatively healthy controls, while no difference was found between the two screeners in the differentiation of patients with BD from those with MDD (Fig. 3).

The best threshold for the differentiation of patients with BD from patients with MDD was 7 for the MDQ (Fig. A1) and 15 for the HCL-32 (Fig. A2).

Sensitivity and specificity at the best threshold were 87 and 77%, respectively, for the MDQ, and 87 and 69% for the HCL-32. Both screeners had a better NPV (92.3 and 91.4%, respectively) than PPV (65.8 and 58.7%). The positive diagnostic likelihood ratio was modestly higher for the MDQ (3.86) than for the HCL-32 (2.84).

In the investigated samples, 109 controls (41.1%), 21 patients with MDD (24.4%), and 52 patients with BD (89.7%) scored at or above the cut-off on the MDQ ($\chi^2 = 63.14$; $df=2$; $p < 0.0001$). The corresponding figures for the HCL-32 were 108 (48%) among controls, 21 (32.8%) among patients with MDD, and 28 (87.5%) among patients with BD ($\chi^2 = 25.78$; $df=2$; $p < 0.0001$).

Discussion

In this study, both the MDQ and the HCL-32 were able to distinguish patients diagnosed with BD from patients diagnosed with MDD, with a good accuracy (when measured with AUC) and an informative positive diagnostic likelihood ratio (above 2). Both screeners were more able to exclude the presence of a BD than to confirm it, on the basis of their PPV and NPV. Reliability was good for both the MDQ and the HCL-32. In controls, too, the reliability of the two screeners was good to excellent.

The controls were probably likely to admit socially acceptable hyperthymic traits, such as being more

sociable than their peers or being exuberant in social circumstances. This might explain the higher fraction of controls than of MDD patients scoring at or above the cut-off for screening a BD. However, the reporting of hypomanic-like symptoms by controls does not necessarily correspond to real, true episodes of hypomania. Moreover, the higher reporting of hyperthymic traits and hypomanic-like symptoms by controls was not corroborated by an independent source.

This is the first study to have tested a bifactor structure of the MDQ and the HCL-32. In past investigations, a two-factor structure was repeatedly reported to explain the distribution of the scores of the two screeners, with some items reflecting a propensity to elated behaviors, and another set of items being a reflection of an impulsive/irritable mood [24, 32, 54, 55]. In this study, this two-factor solution did not show a good fit according to the predefined parameters. The bifactor implementation of this two-factor model, instead, showed a good fit to the data. The excessive reliance on the exploratory factor analysis over the confirmatory factor analysis of past studies might in part explain the difference between this and previous investigations of the topic. It should be noted that both the MDQ and the HCL-32 are usually applied as a single factor screener, thus a bifactor model of a multidimensional structure of the screeners is the best approximation to the expected factor structure of the tools and to its current use. It should be noted that in this study, the indicators of the appropriateness of the general factor of the bifactor model were below the accepted threshold for full acceptance of the general factor as a single summary score of the tools. This may depend on the application of the model to a sample that included both patients and putatively healthy controls. This might have inflated the impact of the

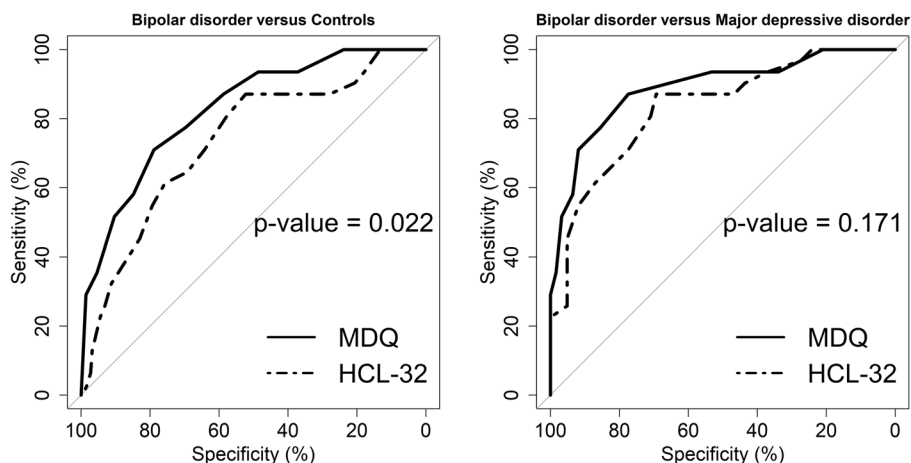


Fig. 3 Comparison with the Hanley and McNeil’s test between the Tunisian arabic MDQ and the Tunisian arabic HCL-32 in distinguishing patients with BD from putatively healthy controls (on the left), or from patients with MDD (on the right)

multidimensionality of both tools, since the elated and impulsive/irritable experience of the patients might be qualitatively different from the corresponding experience in people without a mood disorder.

In this sample, the best cut-off for the HCL-32 was close to the one reported in past studies that were carried out in the Western samples, usually about 14 or 15. However, in some non-Western samples, such as in the Arabian study of Fornaro et al. (2015) [34] or the Brazilian sample of patients of Soares et al. (2010), higher cut-offs were reported, around 17/18. Fornaro et al. (2015) included inpatients, while Soares et al. (2010) [18] enrolled outpatients. Probably both severity and cultural differences in admitting some hypomanic symptoms might have had a role in explaining the higher cut-offs in those studies. In this study, the sensitivity and specificity of the HCL-32 in discriminating patients with BD from those with MDD were, respectively, .87 and .69, somehow higher than the corresponding figures in the Soares et al. study (.75 and .58), and close to the values observed by Perugi et al. (2012) [56] in their large Italian study (.85 and .78). Fornaro et al. [34] found similar values of sensitivity (.82) and specificity (.77) of their version of the HCL-32 in the discrimination of Arabic patients with MDD from those with BD. Both the Perugi et al. (2012) study and Fornaro et al. [34] study found a higher specificity of the applied version of the HCL-32, suggesting that sample composition might affect the detection of hypomanic symptoms. Indeed, in the present study, we enrolled a larger fraction of patients with BD-II than with BD-I, while the Fornaro et al. [34] study had a ratio of BD-I to BD-II = 4.7. This might be considered a limitation of the present study, but in community samples, the lifetime prevalence of BD-II tends to be higher (1.57%; 95%CI: 1.15–1.99) than the lifetime prevalence of BD-I (1.06%; 0.81–1.31) [26]. Moreover, in past studies, patients had already received a diagnosis of BD, thus might have been more prone to admit hypomanic symptoms.

Overall, the two screeners revealed ease of use, albeit requiring some degree of literacy. Time to fill in was in general minimal for patients with adequate reading skills, but sometimes it requires more time in older patients. Nevertheless, both the MDQ and the HCL-32 might represent valuable help in busy primary care settings, favoring the recognition of cases in need of closer evaluation.

Strengths and limitations

The major strength of the study is its design, which was as close as possible to clinical reality, as we included patients only complaining of depressive signs and symptoms, but did not have any precompiled diagnosis of unipolar or bipolar depression when they first presented. This is a major difference from most of the

other studies about MDQ and HCL-32, which often included patients that had already received a diagnosis of BD [1, 34], and might have received some clue about the symptoms they are expected to admit [57]. Several limitations have to be taken into account. Some of the questionnaires, either MDQ or HCL-32, were incomplete, especially among patients with BD. This depended mainly on patients leaving blank some items, such as item 6 (about wanting to travel) or 7 (about risky driving) of the HCL-32 because they do not habitually do the enquired action (they do not travel or drive a car), thus they didn't know how to reply to the question. As a consequence, we had to discard some of the cases and this resulted in a loss of power for the analysis. In particular, we had not enough cases with BD-II to test the discriminant capacity of the tools with respect to MDD, the main usage of a screening tool to identify BD. Indeed, while manic episodes are more likely to be recognized by clinicians and to be remembered by the patients, the hypomanic episodes are precisely those that complicate the diagnosis of BD in the clinical setting.

Conclusion

Despite its limitations, this study showed the good capacity of both the MDQ and the HCL-32 as screening tools to be used to differentiate patients with BD from patients with MDD. Both screeners work best in excluding the presence of BD in patients with MDD, which is an advantage in deciding whether or not to prescribe an antidepressant, which can have known negative effects in patients with BD [58]. When the screener is positive for the presence of BD, it may prompt a deeper investigation of past manic/hypomanic episodes that might have been overlooked at the first assessment.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12888-023-04529-x>.

Additional file 1. Operating characteristics of the Tunisian arabic MDQ for various threshold scores among patients diagnosed with a current episode of major depressive disorder either in the course of a unipolar or bipolar mood disorder as diagnosed with the SCID. Specificity and sensitivity are plotted per percentage of subjects and the number of items checked positive on the screener.

Additional file 2. Operating characteristics of the Tunisian arabic HCL-32 for various threshold scores among patients diagnosed with a current episode of major depressive disorder either in the course of a unipolar or bipolar mood disorder as diagnosed with the SCID. Specificity and sensitivity are plotted per percentage of subjects and the number of items checked positive on the screener.

Additional file 3.

Acknowledgements

None.

Authors' contributions

UO, AOm and FN designed the study. Data collection was conducted by UO, YZ, LJ, AA, RJ, AOu. Data were analyzed by AP and MC. Data was interpreted by AP, MC, UO, AOm, and FN. AP and UO have drafted the work, which was reviewed by all authors. All authors read and approved the final manuscript.

Funding

No funding was provided for this study.

Availability of data and materials

The dataset of this study is available from the corresponding author on reasonable request.

Declarations**Ethics approval and consent to participate**

The study was approved by the Institutional Review Board of Razi Hospital, with the authorization signed on 8 Oct 2014. All participants provided informed consent to participate in the study.

Consent for publication

The authors provide consent for publication.

Competing interests

The authors declare no competing interests.

Received: 3 July 2022 Accepted: 6 January 2023

Published online: 26 January 2023

References

- An D, Hong KS, Kim J-H. Exploratory factor analysis and confirmatory factor analysis of the Korean version of hypomania Checklist-32. *Psychiatry Investig.* 2011;8:334. <https://doi.org/10.4306/pi.2011.8.4.334>.
- Phillips ML, Kupfer DJ. Bipolar disorder diagnosis: challenges and future directions. *Lancet.* 2013;381:1663–71. [https://doi.org/10.1016/S0140-6736\(13\)60989-7](https://doi.org/10.1016/S0140-6736(13)60989-7).
- Culpepper L. Misdiagnosis of bipolar depression in primary care practices. *J Clin Psychiatry.* 2014;75:e05. <https://doi.org/10.4088/JCP.13019tx1c>.
- McCombs JS, Ahn J, Tencer T, Shi L. The impact of unrecognized bipolar disorders among patients treated for depression with antidepressants in the fee-for-services California Medicaid (Medi-Cal) program: a 6-year retrospective analysis. *J Affect Disord.* 2007;97:171–9. <https://doi.org/10.1016/j.jad.2006.06.018>.
- Buoli M, Cesana BM, Fagiolini A, Albert U, Maina G, de Bartolomeis A, et al. Which factors delay treatment in bipolar disorder? A nationwide study focussed on duration of untreated illness. *Early Interv Psychiatry.* 2021;15:1136–45. <https://doi.org/10.1111/eip.13051>.
- Smith D, Ghaemi S, Craddock N. The broad clinical spectrum of bipolar disorder: implications for research and practice. *J Psychopharmacol (Oxf).* 2008;22:397–400. <https://doi.org/10.1177/0269881108089585>.
- Dagani J, Signorini G, Nielssen O, Bani M, Pastore A, de Girolamo G, et al. Meta-analysis of the interval between the onset and Management of Bipolar Disorder. *Can J Psychiatr.* 2017;62:247–58. <https://doi.org/10.1177/0706743716656607>.
- Lublóy Á, Keresztúri JL, Németh A, Mihalicza P. Exploring factors of diagnostic delay for patients with bipolar disorder: a population-based cohort study. *BMC Psychiatry.* 2020;20:75. <https://doi.org/10.1186/s12888-020-2483-y>.
- Drancourt N, Etain B, Lajnef M, Henry C, Raust A, Cochet B, et al. Duration of untreated bipolar disorder: missed opportunities on the long road to optimal treatment: duration of untreated bipolar disorder. *Acta Psychiatr Scand.* 2013;127:136–44. <https://doi.org/10.1111/j.1600-0447.2012.01917.x>.
- Lish JD, Dime-Meenan S, Whybrow PC, Price RA, Hirschfeld RMA. The National Depressive and manic-depressive association (DMDA) survey of bipolar members. *J Affect Disord.* 1994;31:281–94. [https://doi.org/10.1016/0165-0327\(94\)90104-X](https://doi.org/10.1016/0165-0327(94)90104-X).
- Angst J, Adolfsson R, Benazzi F, Gamma A, Hantouche E, Meyer T, et al. The HCL-32: towards a self-assessment tool for hypomanic symptoms in outpatients. *J Affect Disord.* 2005;88:217–33. <https://doi.org/10.1016/j.jad.2005.05.011>.
- Meyer TD, Crist N, La Rosa N, Ye B, Soares JC, Bauer IE. Are existing self-ratings of acute manic symptoms in adults reliable and valid? A systematic review. *Bipolar Disord.* 2020;22:558–68. <https://doi.org/10.1111/bdi.12906>.
- Zimmerman M. Using screening scales for bipolar disorder in epidemiologic studies: lessons not yet learned. *J Affect Disord.* 2021;292:708–13. <https://doi.org/10.1016/j.jad.2021.06.009>.
- Hirschfeld RMA, Williams JBW, Spitzer RL, Calabrese JR, Flynn L, Keck PE, et al. Development and validation of a screening instrument for bipolar spectrum disorder: the mood disorder questionnaire. *Am J Psychiatry.* 2000;157:1873–5. <https://doi.org/10.1176/appi.ajp.157.11.1873>.
- Meyer TD, Hammelstein P, Nilsson L-G, Skeppar P, Adolfsson R, Angst J. The hypomania checklist (HCL-32): its factorial structure and association to indices of impairment in German and Swedish nonclinical samples. *Compr Psychiatry.* 2007;48:79–87. <https://doi.org/10.1016/j.comppsy.2006.07.001>.
- Twiss J, Jones S, Anderson I. Validation of the mood disorder questionnaire for screening for bipolar disorder in a UK sample. *J Affect Disord.* 2008;110:180–4. <https://doi.org/10.1016/j.jad.2007.12.235>.
- Wang YT, Yeh TL, Lee IH, Chen KC, Chen PS, Yang YK, et al. Screening for bipolar disorder in medicated patients treated for unipolar depression in a psychiatric outpatient clinic using the mood disorder questionnaire. *Int J Psychiatry Clin Pract.* 2009;13:117–21. <https://doi.org/10.1080/13651500802550008>.
- Soares OT, Moreno DH, de Moura EC, Angst J, Moreno RA. Reliability and validity of a Brazilian version of the hypomania checklist (HCL-32) compared to the mood disorder questionnaire (MDQ). *Rev Bras Psiquiatr.* 2010;32:416–23. <https://doi.org/10.1590/S1516-44462010000400015>.
- Bech P, Christensen EM, Vinberg M, Bech-Andersen G, Kessing LV. From items to syndromes in the hypomania checklist (HCL-32): psychometric validation and clinical validity analysis. *J Affect Disord.* 2011;132:48–54. <https://doi.org/10.1016/j.jad.2011.01.017>.
- Yang H-C, Xiang Y-T, Liu T-B, Han R, Wang G, Hu C, et al. Hypomanic symptoms assessed by the HCL-32 in patients with major depressive disorder: a multicenter trial across China. *J Affect Disord.* 2012;143:203–7. <https://doi.org/10.1016/j.jad.2012.06.002>.
- Chou CC, Lee IH, Yeh TL, Chen KC, Chen PS, Chen WT, et al. Comparison of the validity of the Chinese versions of the hypomania symptom Checklist-32 (HCL-32) and mood disorder questionnaire (MDQ) for the detection of bipolar disorder in medicated patients with major depressive disorder. *Int J Psychiatry Clin Pract.* 2012;16:132–7. <https://doi.org/10.3109/13651501.2011.644563>.
- Yang H, Yuan C, Liu T, Li L, Peng H, Liao C, et al. Validity of the 32-item hypomania checklist (HCL-32) in a clinical sample with mood disorders in China. *BMC Psychiatry.* 2011;11:84. <https://doi.org/10.1186/1471-244X-11-84>.
- Mosolov SN, Ushkalova AV, Kostukova EG, Shafarenko AA, Alfimov PV, Kostyukova AB, et al. Validation of the Russian version of the hypomania checklist (HCL-32) for the detection of bipolar II disorder in patients with a current diagnosis of recurrent depression. *J Affect Disord.* 2014;155:90–5. <https://doi.org/10.1016/j.jad.2013.10.029>.
- Gamma A, Angst J, Azorin J-M, Bowden CL, Perugi G, Vieta E, et al. Transcultural validity of the hypomania Checklist-32 (HCL-32) in patients with major depressive episodes. *Bipolar Disord.* 2013;15:701–12. <https://doi.org/10.1111/bdi.12101>.
- Wang Y-Y, Xu D-D, Liu R, Yang Y, Grover S, Ungvari GS, et al. Comparison of the screening ability between the 32-item hypomania checklist (HCL-32) and the mood disorder questionnaire (MDQ) for bipolar disorder: a meta-analysis and systematic review. *Psychiatry Res.* 2019;273:461–6. <https://doi.org/10.1016/j.psychres.2019.01.061>.
- Clemente AS, Diniz BS, Nicolato R, Kapczynski FP, Soares JC, Firmo JO, et al. Bipolar disorder prevalence: a systematic review and meta-analysis of the

- literature. *Rev Bras Psiquiatr.* 2015;37:155–61. <https://doi.org/10.1590/1516-4446-2012-1693>.
27. Rowland TA, Marwaha S. Epidemiology and risk factors for bipolar disorder. *Ther Adv Psychopharmacol.* 2018;8:251–69. <https://doi.org/10.1177/2045125318769235>.
 28. Carta MG, Moro MF, Piras M, Ledda V, Prina E, Stocchino S, et al. Megacities, migration and an evolutionary approach to bipolar disorder: a study of Sardinian immigrants in Latin America. *Braz J Psychiatry.* 2020;42:63–7. <https://doi.org/10.1590/1516-4446-2018-0338>.
 29. de Zwart SMC, Brouwer RM, Agartz I, Alda M, Aleman A, Alpert KI, et al. The association between familial risk and brain abnormalities is disease specific: An ENIGMA-relatives study of schizophrenia and bipolar disorder. *Biol Psychiatry.* 2019;86:545–56. <https://doi.org/10.1016/j.biopsych.2019.03.985>.
 30. Johnson KR, Johnson SL. Cross-national prevalence and cultural correlates of bipolar I disorder. *Soc Psychiatry Psychiatr Epidemiol.* 2014;49:1111–7. <https://doi.org/10.1007/s00127-013-0797-5>.
 31. Jongsma HE, Turner C, Kirkbride JB, Jones PB. International incidence of psychotic disorders, 2002–17: a systematic review and meta-analysis. *Lancet Public Health.* 2019;4:e229–44. [https://doi.org/10.1016/S2468-2667\(19\)30056-8](https://doi.org/10.1016/S2468-2667(19)30056-8).
 32. Massidda D, Giovanni Carta M, Altoè G. Integrating different factorial solutions of a psychometric tool via social network analysis: the case of the mood disorder questionnaire. *Methodology.* 2016;12:97–106. <https://doi.org/10.1027/1614-2241/a000113>.
 33. Carta MG, Massidda D, Moro MF, Aguglia E, Balestrieri M, Caraci F, et al. Comparing factor structure of the Mood Disorder Questionnaire (MDQ): In Italy sexual behavior is euphoric but in Asia mysterious and forbidden. *J Affect Disord.* 2014;155:96–103. <https://doi.org/10.1016/j.jad.2013.10.030>.
 34. Fornaro M, Elassy M, Mounir M, Abd-Elmoneim N, Ashour H, Hamed R, et al. Factor structure and reliability of the Arabic adaptation of the hypomania check List-32, second revision (HCL-32-R2). *Compr Psychiatry.* 2015;59:141–50. <https://doi.org/10.1016/j.comppsy.2015.02.015>.
 35. Ouali U, Jouini L, Zgueb Y, Jomli R, Omrani A, Nacef F, et al. The factor structure of the mood disorder questionnaire in Tunisian patients. *Clin Pract Epidemiol Ment Health.* 2020;16:82–92. <https://doi.org/10.2174/1745017902016010082>.
 36. World Medical Association. Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2000;284:3043. <https://doi.org/10.1001/jama.284.23.3043>.
 37. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol.* 1993;46:1417–32. [https://doi.org/10.1016/0895-4356\(93\)90142-N](https://doi.org/10.1016/0895-4356(93)90142-N).
 38. R Core Team. A language and environment for statistical computing. Vienna: R Foundation for statistical computing; 2020. Austria. <https://www.R-project.org/>. n.d
 39. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>.
 40. JASP Team. JASP (Version 0.14.1) [Computer software]. New York: JASP team; 2020.
 41. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res Sci Educ.* 2018;48:1273–96. <https://doi.org/10.1007/s11165-016-9602-2>.
 42. Mardia KV. Measures of multivariate skewness and kurtosis with applications. *Biometrika.* 1970;57:519–30. <https://doi.org/10.1093/biomet/57.3.519>.
 43. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model Multidiscip J.* 1999;6:1–55. <https://doi.org/10.1080/10705519909540118>.
 44. Reise SP. The rediscovery of Bifactor measurement models. *Multivar Behav Res.* 2012;47:667–96. <https://doi.org/10.1080/00273171.2012.715555>.
 45. Rodríguez A, Reise SP, Haviland MG. Evaluating bifactor models: calculating and interpreting statistical indices. *Psychol Methods.* 2016;21:137–50. <https://doi.org/10.1037/met0000045>.
 46. Hancock GR, Mueller RO. Rethinking construct reliability within latent variable systems. In: Cudeck R, du Toit S, Sörbom D, editors. *Structural equation modeling: Present and future—A Festschrift in honor of Karl Jöreskog*. Lincolnwood: Scientific Software International; 2001. pp. 195–216.
 47. Rosseel Y. lavaan: An R package for structural equation modeling. *J Stat Softw.* 2012;48. <https://doi.org/10.18637/jss.v048.i02>.
 48. Dueber D. Bifactor Indices Calculator: Bifactor Indices Calculator: R package version 0.2.2; 2021. <https://github.com/ddueber/BifactorIndicesCalculator>
 49. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63:737–45. <https://doi.org/10.1016/j.jclinepi.2010.02.006>.
 50. Hosmer DW, Lemeshow S. *Applied Logistic Regression*, 2nd Ed. Chapter 5. New York: Wiley; 2000. pp. 160–164.
 51. Robin X, Turck N, Hainard N, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12:77. <https://doi.org/10.1186/1471-2105-12-77>.
 52. López-Ratón M, Rodríguez-Álvarez MX, Suárez CC, Sampedro FG. Optimal cut points: An R package for selecting optimal Cutpoints in diagnostic tests. *J Stat Softw.* 2014;61. <https://doi.org/10.18637/jss.v061.i08>.
 53. Bandalos DL. Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Struct Equ Model Multidiscip J.* 2014;21:102–16. <https://doi.org/10.1080/10705511.2014.859510>.
 54. Benazzi F, Akiskal HS. The dual factor structure of self-rated MDQ hypomania: energized-activity versus irritable-thought racing. *J Affect Disord.* 2003;73:59–64. [https://doi.org/10.1016/S0165-0327\(02\)00333-6](https://doi.org/10.1016/S0165-0327(02)00333-6).
 55. Perugi G, Fornaro M, Maremmani I, Canonico PL, Carbonatto P, Mencacci C, et al. Discriminative hypomania checklist-32 factors in unipolar and bipolar major depressive patients. *Psychopathology.* 2012;45(6):390–8. <https://doi.org/10.1159/000338047>.
 56. Angst J, Meyer TD, Adolfsson R, Skeppar P, Carta M, Benazzi F, et al. Hypomania: a transcultural perspective. *World Psychiatry.* 2010;9:41–9. <https://doi.org/10.1002/j.2051-5545.2010.tb00268.x>.
 57. Carta MG, Hardoy MC, Fryers T. Are structured interviews truly able to detect and diagnose bipolar II disorders in epidemiological studies? The king is still nude! *Clin Pract Epidemiol Ment Health CP EMH.* 2008;4:28. <https://doi.org/10.1186/1745-0179-4-28>.
 58. Young A, Seim D. Review: long term use of antidepressants for bipolar disorder reduces depressive episodes but increases risk of mania. *Evid Based Ment Health.* 2009;12:49. <https://doi.org/10.1136/ebmh.12.2.49>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

