BMC Psychiatry

RESEARCH

Open Access

Patient-rated scales improve the classification accuracy for patients with depression and anxiety disorder: a linear discriminant analysis

Shanling Ji[1†], Jing Zhang[2†], Cong Zhou[1], Min Chen[1,2*] and Hao Yu[1*]

Abstract

Background The current study aimed to investigate the performances of clinical scales rated by clinicians and patients as well as cognitive function tests in distinguishing patients with affective and anxiety disorders from healthy controls (HCs).

Methods We recruited a total of 122 subjects, comprising 24 patients with bipolar disorder (BD), 34 patients with major depressive disorder (MDD), 29 patients with anxiety disorder (AD), and 35 matched HCs. Three clinician-rated scales and five patient-rated scales were used to quantify clinical symptoms, while four cognitive tests were employed to measure cognitive functions in all subjects. Fisher's discriminant analysis (FDA) was employed to distinguish patients from HCs, as well as to discriminate patient sub-groups from each other. In the FDA model, the prior probability of each group was set as 0.5 in the two-group classification and 0.25 in the four-group classification.

Results The results showed that patient-rated scales achieved higher classification accuracies than clinician-rated scales in identifying MDD and AD from HCs. In contrast, cognitive tests exhibited the lowest accuracy.

Conclusions These findings suggest that patient-rated scales might improve the classification accuracy for patients with MDD and AD.

Keywords Clinician-rated scales, Patient-rated scales, Cognition, Cross-validation, Classification, Discrimination

†Shanling Ji and Jing Zhang contributed equally to this work.

*Correspondence:
Min Chen
cm7697@163.com
Hao Yu
yuhao@mail.jnmc.edu.cn
[1]Institute of Mental Health, Jining Medical University, Jining, Shandong 272056, China
[2]Department of Psychiatry, Shandong Daizhuang Hospital, Shandong, China

Background

Bipolar disorder (BD), major depressive disorder (MDD), and anxiety disorder (AD) are the most prevalent mental illnesses worldwide, especially in China [1–3]. In clinical diagnosis and treatment, psychiatric clinicians quantify patients' symptoms using clinician-rated scales, such as the Hamilton Depression Rating Scale (HAMD) [4], Hamilton Anxiety Rating Scale (HAMA) [5], and Young Mania Rating Scale (YMRS) [6]. Patients quantify their subjective experience of symptoms using patient-rated scales, such as the Patient Health Questionnaire-9 (PHQ-9) [7] and Generalized Anxiety Disorder-7 (GAD-7) scale [8].

Among these evaluation tools, whether some scales or items are more sensitive in classifying patients and quantifying symptoms than others needs to be fully explored. The identification of some evaluation tools as more effective would reduce unnecessary expenditure on medical resources. It also helps clinicians choose the right kind of scales to be more accurate and efficient in clinical assessment. Moreover, with the use of smartphones and the development of artificial intelligence, patient-rated scales can help clinicians gather important information online before meeting patients for the first time. However, only a few studies have investigated the classification performance of clinician- or patient-rated scales in patients with diseases, such as BD [9, 10], MDD [11–16], cancer [17], stroke [18–21], and epilepsy [22]. PHQ-9 and its abbreviated eight-item (PHQ-8) and two-item (PHQ-2) versions were often used to screen individuals with depression [12, 14–16, 23], while PHQ-15 was used to distinguish BD type II from MDD [9, 23]. GAD-7 and its abbreviated two-item (GAD-2) version were often used to screen for generalized anxiety, panic, social anxiety, and post-traumatic stress disorder [8, 23].

Fisher's discriminant analysis (FDA) is a statistical technique used to classify one category from two or more categories by minimizing the differences within groups and maximizing the differences between groups [24]. It has no restrictions on data distribution and variance; thus, it is widely used in many scientific fields, such as medicine and biology [25–27].

The current study aimed to investigate the performances of clinician-rated scales, patient-rated scales, and cognitive tests in distinguishing patients with BD, MDD, and AD from healthy controls (HCs). We hypothesized that clinician-rated scales could distinguish patients from HCs well, but the accuracy would increase by combining patient-rated scales and cognitive tests. Additionally, we expected that the classification accuracy achieved by patient-rated scales would be equal to that of clinician-rated scales, as previous studies have found good performance of patient-rated scales in screening patients [9, 23]. To test these hypotheses, we recruited patients with BD, MDD, and AD as well as HCs. Three clinician-rated scales (HAMD, HAMA, and YMRS) were used to quantify the observed symptoms, five patient-rated scales were employed to measure the experiential symptoms, and four cognitive tests were conducted to assess the cognitive functions in all subjects. FDA was employed to distinguish patients from HCs, as well as discriminating patient sub-groups from each other.

Methods

Subjects

This study initially enrolled a total of 130 subjects, comprising 27 BD patients, 37 MDD patients, 31 AD patients, and 35 HCs through outpatient services and online advertisements from August 2018 to December 2019. All subjects were interviewed using Chinese versions of the Structured Clinical Interview for DSM-IV and the Mini International Neuropsychiatric Interview (MINI). At last, 8 patients did not complete the assessments and were excluded from the study, thereby leaving 122 subjects, comprising 24 BD patients, 34 MDD patients, 29 AD patients, and 35 HCs. Figure 1 shows the flowchart of this study.

Inclusion criteria for patients were: (1) 18–55 years old; (2) primary school education or above; (3) providing written informed consent; (4) currently meeting one of the following diagnoses in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV): AD (panic disorder, agoraphobia, social phobia, generalized anxiety disorder), MDD, or BD (types I and II); and (5) first episode of psychiatric illness or discontinuation of psychotropic drugs at least five half-lives prior to the assessment of clinical scales and cognitive tests [28]. Exclusion criteria were: (1) history of epilepsy or brain trauma; (2) severe physical illness; (3) high risk of suicide; (4) electroconvulsive therapy or transcranial magnetic stimulation within the past 6 months; and (5) pregnancy or breastfeeding.

In addition, 35 age, sex, and educational level-matched HCs were enrolled from the local community through online advertisement, ensuring that they did not have any history of psychiatric disorders, suicidality, or any family history of psychiatric illnesses.

Written informed consent was provided by all subjects, and the protocol was approved by the Institutional Review Board of Jining Medical University.

Measures

Clinician-rated scales

In the current study, all subjects underwent three commonly used clinical assessments, which were conducted by two experienced psychiatrists. Both psychiatrists receive thorough training on the assessment tools, diagnostic criteria, and procedures to ensure they have a shared understanding of how to evaluate cases. Both

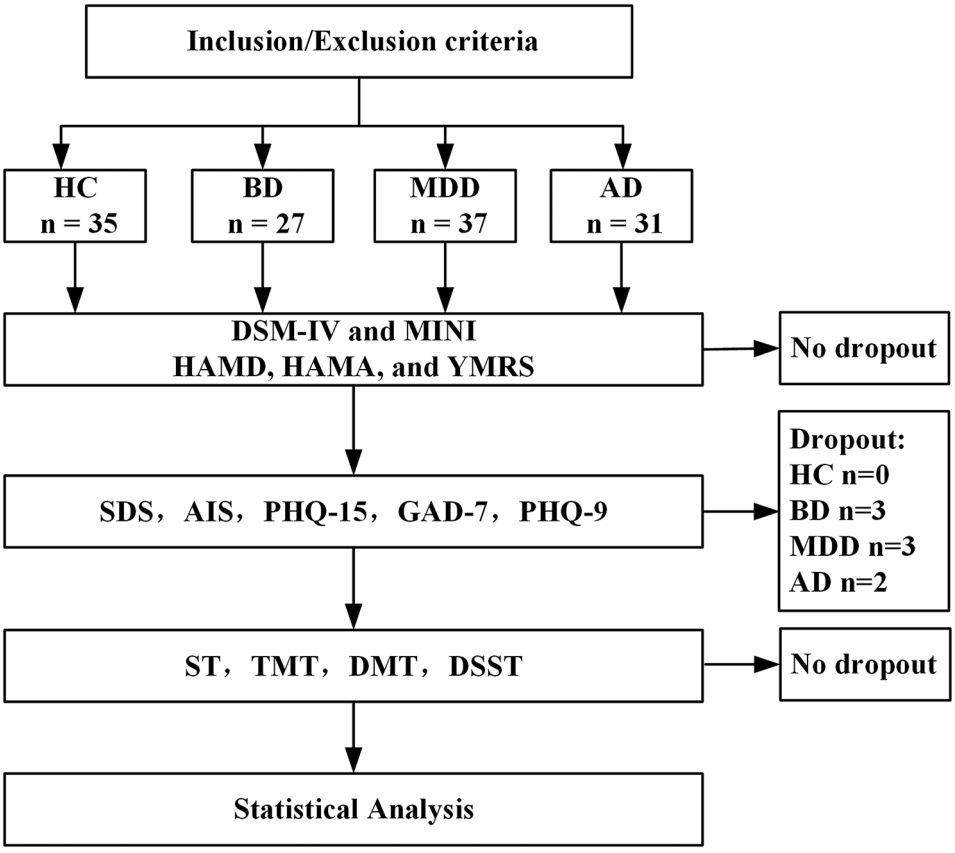

Fig. 1 Flowchart of this study. *Note Abbreviations* AD = anxiety disorder; AIS = Athens Insomnia Scale; BD = bipolar disorder; DMT = Digit Memory Test; DSST = Digit Symbol Substitution Test; GAD-7 = Generalized Anxiety Disorder-7; HAMA = Hamilton Anxiety Scale; HAMD = Hamilton Depression Scale; HC = healthy control; MDD = major depressive disorder; PHQ-9 = Patient Health Questionnaire-9; PHQ-15 = Patient Health Questionnaire; SDS = Sheehan Disability Scale; ST = Stroop Test; TMT = Trail Marking Test; and YMRS = Young Mania Rating Scale

psychiatrists perform evaluations independently and without prior discussion, ensuring that their judgments are unbiased. The inter-rater reliability was measured using Cohen's Kappa to quantify the agreement between the two psychiatrists. The values of Cohen's Kappa on HAMD, HAMA, and YMRS were 0.58, 0.50, and 0.47, indicating moderate agreement.

Hamilton depression rating scale-17 item (HAMD-17) HAMD-17 is a widely used clinician-rated measure of depressive symptoms. Each item is scored on a scale from 0 to 4 points or from 0 to 2 points, and the total score ranges from 0 to 52 points [4]. In this study, the HAMD-17 was used to quantify the depressive symptoms in patients with BD, MDD, and AD, as the depressive mood is one of the most commonly shared symptoms among patients with these disorders.

Hamilton anxiety rating scale-14 item (HAMA-14) HAMA-14 is a commonly used clinician-rated measure of anxiety in clinical and research settings [5]. Each item is scored on a scale from 1 to 4 points. The total score ranges from 0 to 56 points. In this study, the HAMA-14 was used to assess somatic anxiety and psychic anxiety in all patients, as anxious mood is one of the most commonly shared symptoms among patients with BD, MDD, and AD.

Young mania rating scale (YMRS) The YMRS is used to quantify the severity of manic symptoms in BD patients [6]. It includes 11 items, which assess the severity of manic symptomatology. Each item is scored on a scale from 0 to 4 points (seven items) or from 0 to 8 points (four items). The total score ranges from 0 to 60 points. In this study, the YMRS was mainly used to quantify the manic symptoms in BD patients.

Patient-rated scales

All subjects were asked to complete five patient-rated scales, which are widely used in the Chinese population. A clinical psychologist was responsible for explaining the requirements of rating these scales.

Patient health questionnaire-9 (PHQ-9) The PHQ-9 assesses the severity of depressive symptoms during the last 2 weeks. Each item is scored on a scale from 0 (not at all) to 3 points (almost daily), with the total score ranging from 0 to 27 points [7]. Given that depressive mood is one of the major complaints in patients with affective and anxiety disorders, this study used the PHQ-9 to measure the depressive symptoms in patients with BD, MDD, and AD.

Generalized anxiety disorder-7 (GAD-7) The GAD-7 quantifies anxiety symptoms during the last 2 weeks. Responses for each item can range from 0 (not at all) to 3 points (almost daily), with the total score ranging from 0 to 21 points [8]. In this study, the GAD-7 was employed to measure the anxious symptoms in all patients, as BD and MDD might be accompanied by anxiety symptoms.

Patient health questionnaire-15 (PHQ-15) The PHQ-15 measures somatic symptoms during the last month. Each item is scored on a scale from 0 (none) to 2 points (very many), with the total score ranging from 0 to 30 points [29]. Given that MDD and AD are often accompanied by somatic symptoms, the present study used the PHQ-15 to evaluate the somatic symptoms in patients with BD, MDD, and AD.

Sheehan disability scale (SDS) The SDS consists of five items that assess the degree of disability in daily life, the workplace/school, and social life in the last week. The first three items, which are directly related to disability, were used in this study. Each item is scored on a scale from 0 (not at all) to 10 points (extremely), with the total score ranging from 0 to 30 points [30]. In this study, the SDS was used to evaluate the degree of disability in patients with BD, MDD, and AD.

Athens Insomnia scale (AIS) The AIS consists of eight items, which are based on the International Classification of Diseases, Tenth Revision, diagnostic criteria for insomnia [31]. The first five items of this scale assess sleep in the nighttime, and the last three items assess the function in the daytime during the last month. The score of each item ranges from 0 (none) to 3 points (many), with the total score ranging from 0 to 24 points. Given that sleep quality influences affection and cognition, the current study used the AIS to measure the quality of sleep in patients with BD, MDD, and AD.

The PHQ-9, GAD-7, PHQ-15, SDS, and AIS are all Chinese versions and have been validated well in Chinese.

Cognitive function

To determine the differences in cognition functions between HC and patients as well as to verify the ability of cognition functions to identify patients, four cognitive function tests were performed on all subjects by a clinical psychologist.

Stroop task (ST) In the ST, we used three separate conditions of 100 stimuli each in which subjects were instructed to (i) read color words (color), (ii) name color bars (word), and (iii) name the color of color words while inhibiting the prepotent response of reading the word (conflict). The dependent measures collected were the number of errors and the mean time to complete 100 trials for each condition. This test measured the disinhibition of subjects.

Trail marking test (TMT) In the TMT, there are 26 English letters and 10 Arabic numerals. Subjects were asked to connect the English letters and Arabic numerals in ascending order. The test was used to measure the spatial perception, coordination, and cognitive flexibility of subjects.

Digit memory test (DMT) In the DMT, subjects were required to recite numbers one minute after the psychologist read them once. This test measured the short-term memory of subjects.

Digit symbol substitution test (DSST) The DSST consists of numbers and corresponding symbols. Subjects were required to fill the correct symbols into the spaces under numbers within 90 s, and the number of correct answers was the final score. This test measured the learning ability and cognitive flexibility of subjects.

Statistical analyses

Statistical software Statistical Product and Service Solutions version 20 (SPSS 20) was used for preliminary analysis. Age, clinician- and patient-rated scales, and cognitive tests underwent a one-way multivariate analysis of variance to determine whether their distributions were different among the four groups. The chi-squared (χ^2) test was used to identify the differences in gender among groups. The Bonferroni method was used for post hoc analysis. Patients with BD, MDD, and AD were distinguished from HCs using Fisher's linear discriminant analysis. The prior probability of each group was set as $p=0.5$. Leave-one-out cross-validation was used for validating models. Patient groups were also distinguished from each other using the same method (BD vs. MDD, MDD vs. AD, BD vs. AD). Finally, we performed a four-group classification with a prior probability of $p=0.25$ for each group. The classification performances were measured by accuracy, specificity, sensitivity, and area under the curve (AUC). AUC is the area under the Receiver Operating Characteristic curve, which is an important index to measure the classification ability of a model. The higher the AUC value, the better the ability

of the model to distinguish between positive and negative samples. AUC is independent of the class distribution of samples, and can effectively evaluate model performance even if the ratio of positive and negative samples is unbalanced. To obtain the best classification model, we used two kinds of combined data: The first was using the total score of each scale and test, and the second was adding the scores of all items in each scale. We report the results following the STROBE guidelines [32] which are shown in the Supplementary Materials Table S1.

Results

Demographics and clinical characteristics of subjects

Analysis revealed statistically significant differences in scales and cognitive tests among the four groups [$F(3, 118)=10.87$, $p<0.001$, Wilks' lambda=0.06; partial $\eta^2=0.62$]. No significant differences were found in gender, age, and education years among the four groups. Significant differences were observed in all clinician- and patient-rated scales and all cognitive tests except DMT (Table 1). Supplementary Materials Tables S2 and S3 show the post hoc results.

Two-group classification performances

The results of all two-group classifications showed that the total scores of scales and cognitive tests exhibited better performances than their items.

Figure 2 shows the classification performances (accuracy, specificity, sensitivity, and AUC) for BD vs. HC, MDD vs. HC, AD vs. HC, BD vs. MDD, BD vs. AD, and MDD vs. AD of clinician-rated scales, patient-rated scales, and cognitive tests used separately as well as in different combinations.

In distinguishing BD patients from HCs, the clinician-rated scales showed the highest accuracy of 97.43%, with 96% sensitivity, 100% specificity, and 100% AUC, which did not increase when patient-rated scales and cognitive tests were added. The key variables in this optimal discrimination were YMRS ($\beta=1.03$) and HAMD ($\beta=0.97$). The cognitive tests achieved the lowest classification accuracy of 66.70%. Additionally, we used a single clinician-rated scale (HAMD, HAMA, YMRS) to distinguish BD patients from HCs and found that the HAMD and YMRS achieved the same accuracy of 90% and that HAMA achieved an accuracy of 85% (Supplementary Materials Table S5).

In distinguishing MDD patients from HCs, patient-rated scales exhibited a classification accuracy of 95.92%, which was higher than that of clinician-rated scales (93.94%). The combination of clinician- and patient-rated scales achieved the highest classification accuracy of 98.00%, with 97% sensitivity, 100% specificity, and 99% AUC, which resulted in three key variables: HAMD ($\beta=0.32$), HAMA ($\beta=0.66$), and SDS ($\beta=0.79$). The cognitive tests achieved the lowest classification accuracy of 55.14%.

In distinguishing AD patients from HCs, patient-rated scales exhibited a classification accuracy of 95.51%, which was higher than that of clinician-rated scales (93.22%). The combination of clinician- and patient-rated scales and cognitive tests achieved the highest classification accuracy of 98.00%, with 90% sensitivity, 100% specificity, and 99% AUC, which resulted in seven key variables,

Table 1 Demographics and clinical characteristics of subjects ($n=122$, mean ± *SD*)

Measurements	HC ($n=35$)	BD ($n=24$)	MDD ($n=34$)	AD ($n=29$)	F/χ^2	*p* (two-tailed)	Partial η^2
Gender (M/F)	11/24	9/15	11/23	14/15	3.33 [a]	0.34	0.18
Age	28.53 ± 9.14	28.17 ± 7.41	29.00 ± 8.12	32.45 ± 8.60	1.53	0.21	0.04
Education(year)	5.76 ± 0.82	6.13 ± 0.35	5.83 ± 0.76	5.71 ± 0.97	4.01	0.07	0.14
HAMD	0.20 ± 0.56	13.79 ± 8.46	18.82 ± 6.69	13.48 ± 6.16	28.02	0.000	0.46
HAMA	0.53 ± 1.13	14.88 ± 8.51	18.32 ± 7.51	22.83 ± 7.75	32.19	0.000	0.50
YMRS	0.20 ± 0.77	12.71 ± 6.52	1.68 ± 2.68	1.10 ± 2.04	60.05	0.000	0.65
SDS	0.27 ± 1.03	15.88 ± 10.25	22.61 ± 9.14	22.21 ± 10.15	24.86	0.000	0.43
AIS	2.00 ± 1.89	9.33 ± 4.78	9.85 ± 4.41	10.41 ± 4.97	13.92	0.000	0.30
PHQ-15	2.27 ± 2.71	11.33 ± 5.96	12.76 ± 5.88	14.03 ± 5.14	17.65	0.000	0.35
GAD-7	0.47 ± 1.06	10.42 ± 6.67	12.50 ± 5.38	11.79 ± 5.31	19.64	0.000	0.38
PHQ-9	1.93 ± 2.76	13.04 ± 7.85	16.91 ± 5.58	14.10 ± 7.05	19.82	0.000	0.38
ST	47.20 ± 12.54	35.75 ± 11.57	41.21 ± 13.71	33.29 ± 11.97	5.13	0.002	0.14
TMT	56.13 ± 23.61	68.92 ± 22.86	76.33 ± 27.33	89.50 ± 48.82	5.25	0.002	0.14
DMT	6.86 ± 1.61	7.00 ± 5.24	6.13 ± 1.88	5.55 ± 1.64	1.35	0.263	0.04
DSST	73.64 ± 13.42	63.29 ± 13.40	62.19 ± 18.83	52.00 ± 17.84	6.12	0.001	0.16

Abbreviations AD=anxiety disorder; AIS=Athens Insomnia Scale; BD=bipolar disorder; DMT=Digit Memory Test; DSST=Digit Symbol Substitution Test; GAD-7=Generalized Anxiety Disorder-7; HAMA=Hamilton Anxiety Scale; HAMD=Hamilton Depression Scale; HC=healthy control; MDD=major depressive disorder; PHQ-9=Patient Health Questionnaire-9; PHQ-15=Patient Health Questionnaire; SDS=Sheehan Disability Scale; ST=Stroop Test; TMT=Trail Marking Test; and YMRS=Young Mania Rating Scale. [a] Chi-square (χ^2) test

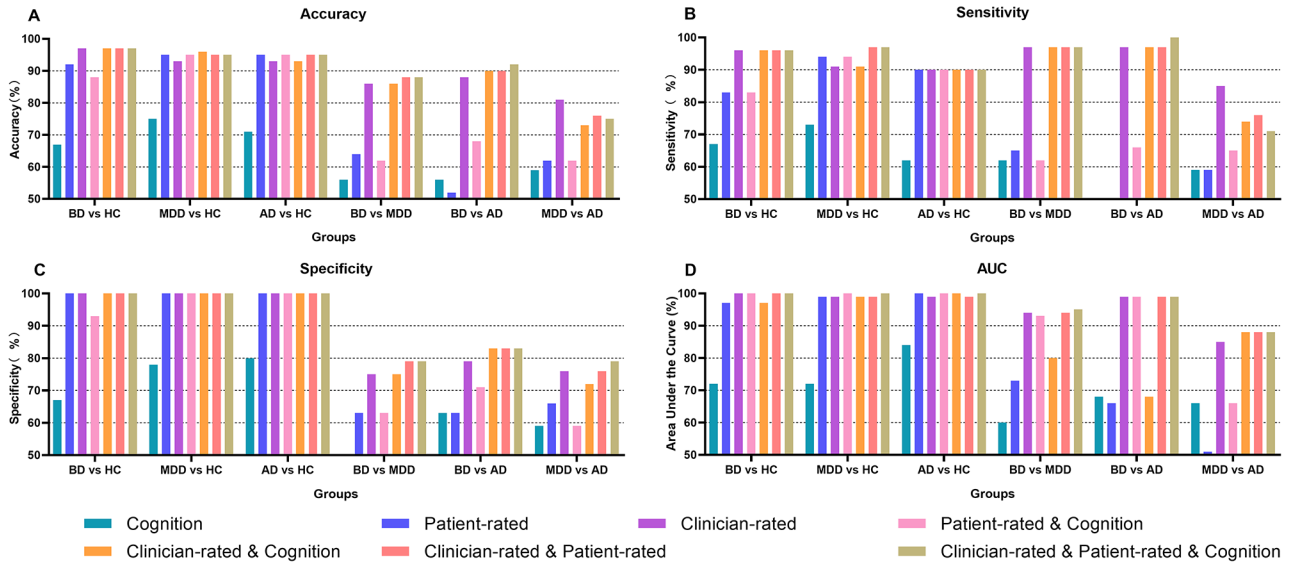

Fig. 2 Two-group classification performances. *Note* Accuracy, specificity, sensitivity, and AUC were obtained using leave-one-out cross-validation. Clinician-rated scales included the Hamilton Anxiety Scale (HAMA), Hamilton Depression Scale (HAMD), and Young Mania Rating Scale (YMRS). Patient-rated scales included the Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder-7 (GAD-7), Patient Health Questionnaire-15 (PHQ-15), Sheehan Disability Scale (SDS), and Athens Insomnia Scale (AIS). Cognition tests included the Stroop Test (ST), Trail Marking Test (TMT), Digit Memory Test (DMT), and Digit Symbol Substitution Test (DSST). *Abbreviations* AUC = area under the curve; AD = anxiety disorder; BD = bipolar disorder; HC = healthy control; and MDD = major depressive disorder

including HAMA (β=0.93), HAMD (β=0.40), PHQ-15 (β=0.31), and AIS (β = −1.05). The cognitive tests achieved the lowest classification accuracy of 65.9%.

In distinguishing BD patients from MDD patients, the combination of clinician- and patient-rated scales showed the highest accuracy of 88.30%, with 97% sensitivity, 79% specificity, and 94% AUC, and this accuracy did not increase when cognitive tests were added. The key variable was YMRS (β=0.22).

In distinguishing BD patients from AD patients, the combination of clinician- and patient-rated scales and cognitive tests showed the highest accuracy of 92.33%, with 100% sensitivity, 83% specificity, and 99% AUC. The key variables were HAMD (β=0.68), HAMA (β = −0.55), and YMRS (β=0.96).

In distinguishing MDD patients from AD patients, the combination of clinician- and patient-rated scales and cognitive tests showed the highest accuracy of 75.13%, with 71% sensitivity, 79% specificity, and 88% AUC. The key variables were HAMD (β=0.73) and HAMA (β = −0.87).

Four-group classification performances

In optimal discrimination for four groups, the combination of clinician- and patient-rated scales and cognitive tests achieved the highest total accuracy of 82.4%, and the cognitive tests alone achieved the lowest classification accuracy of 34.32%. Figure 3 shows the accuracy, specificity, sensitivity, and AUC for each group. Three Fisher discriminant functions were formulated using non-standardized coefficients (Supplementary Materials). The summary and validity test of discriminant functions are shown in Supplementary Materials Tables S6 and S7. The standardized coefficients of discriminant functions show the relative contribution of the predictor variables to the composition of the discriminant function (Table 2). The larger the absolute value of the coefficient, the more important for the discriminant function. YMRS is the most important for the first function, the HAMA and SDS are the most important for the second function, and HAMD (item 1) is the most important for the third function.

Figure 4 presents the distribution of four groups in the space generated by discriminant functions 1 and 2. The four group means evaluated using Fisher discriminant functions with non-standardized coefficients are shown in Supplementary Materials Table S8.

Discussion

In the present study, we investigated the performances of clinician-rated scales, patient-rated scales, and cognitive tests in classifying patients with BD, MDD, and AD. The results indicated that clinician-rated scales demonstrated the highest accuracy in BD patients and HCs in the absence of self-rated scales for mania or hypomania. Meanwhile, patient-rated scales were more accurate than clinician-rated scales in distinguishing MDD and AD patients from HCs, indicating that subjective experience might be more important for MDD and AD [33, 34]. Furthermore, when discriminating patient subgroups

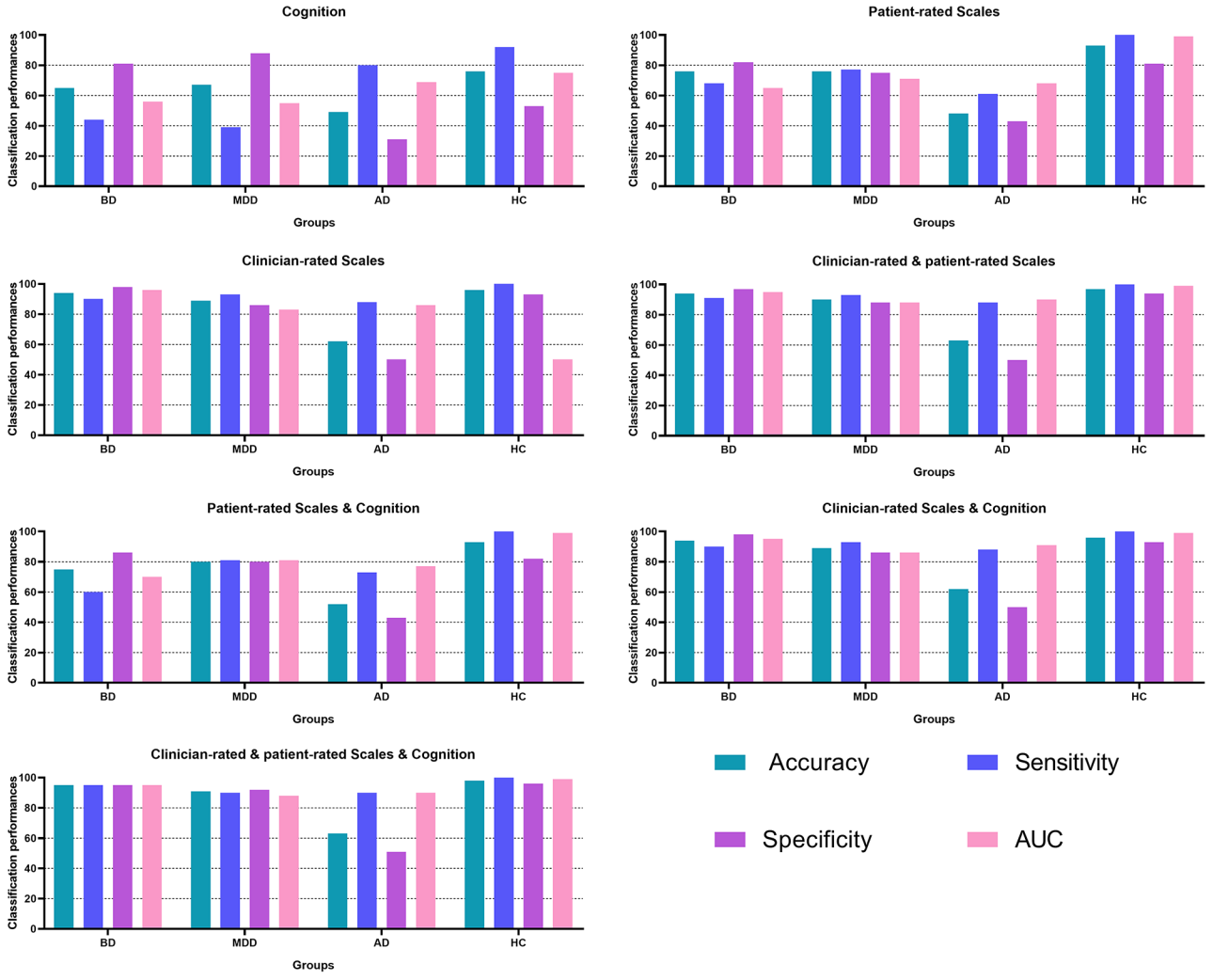

Fig. 3 Four-group classification performances. *Note* Accuracy, specificity, sensitivity, and AUC were obtained using leave-one-out cross-validation. Clinician-rated scales included the Hamilton Anxiety Scale (HAMA), Hamilton Depression Scale (HAMD), and Young Mania Rating Scale (YMRS). Patient-rated scales included the Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder-7 (GAD-7), Patient Health Questionnaire-15 (PHQ-15), Sheehan Disability Scale (SDS), and Athens Insomnia Scale (AIS). Cognition tests included the Stroop Test (ST), Trail Marking Test (TMT), Digit Memory Test (DMT), and Digit Symbol Substitution Test (DSST). *Abbreviations* AUC = area under the curve; AD = anxiety disorder; BD = bipolar disorder; HC = healthy control; and MDD = major depressive disorder

from each other, the accuracy of distinguishing BD from AD was the highest, while the accuracy of classifying MDD from AD was the lowest. This result suggests that BD patients might have higher heterogeneity with AD patients, while MDD patients might share more common symptomatic features with AD patients [35]. These outcomes highlight the importance of clinical scales and cognitive tests in distinguishing patients with BD, MDD, and AD in clinical practice and relevant research.

In the two-group classifications, we observed that the SDS showed the maximum weight in distinguishing MDD patients from HCs, which implied that MDD patients might experience more severe impairment in daily life, such as learning, working, housework, and childrearing [30, 36, 37]. Meanwhile, the nocturnal symptoms (AIS) showed more sensitivity in distinguishing AD patients from HCs, which suggested that AD patients might be more prone to poor sleep quality, as was reported in previous studies [33, 38]. These results suggest that MDD and AD patients experience different mood symptoms [35, 39, 40].

In the four-group classification, clinician- and patient-rated scales and cognitive tests showed contributions to the accuracy of classification, which further demonstrated that subjective experience and cognitive functions should not be ignored in screening patients, as suggested by a previous study [41]. HAMA and SDS were equally important for function 2, which revealed that daily disability and anxiety symptoms might be the shared symptoms among patients with BD, MDD, and AD [42–44].

Table 2 Standardized coefficients of discriminant functions for four-group classification

Predictor variables	Standardized coefficients (β)		
	Function 1	Function 2	Function 3
YMRS	1.53	0.23	−0.54
HAMA	−0.23	0.86	−0.36
HAMD (item 1)	0.96	−0.35	0.58
TMT	−0.88	0.17	−0.34
DMT	−0.81	−0.17	0.07
PHQ-15 (item 13)	−0.52	−0.05	0.22
SDS (item 3)	0.83	−0.33	0.44
SDS (item 4)	−1.15	0.05	−0.44
SDS	0.02	0.86	0.16
Constant	0.89	-0.41	0.90

Abbreviations DMT=Digit Memory Test; HAMA=Hamilton Anxiety Scale; HAMD=Hamilton Depression Scale; PHQ-15=Patient Health Questionnaire; SDS=Sheehan Disability Scale; TMT=Trail Marking Test; and YMRS=Young Mania Rating Scale

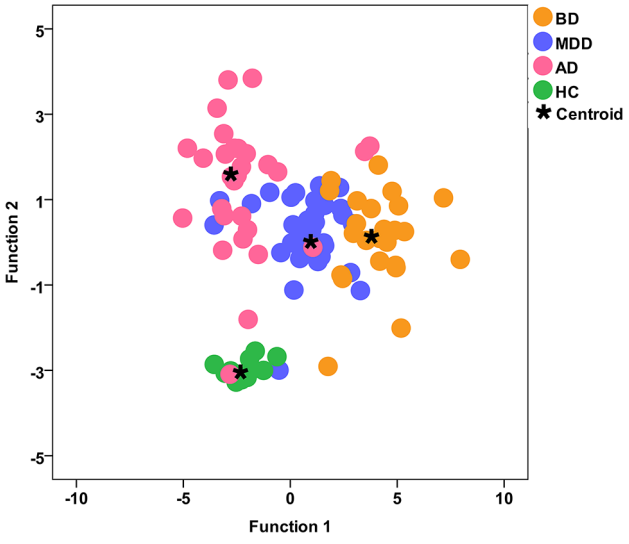

Fig. 4 Scatter plots for four groups. This coordinate system consists of functions 1 and 2. The centroid represents the central distribution of each group. *Abbreviations* AD=anxiety disorder; BD=bipolar disorder; HC=healthy control; and MDD=major depressive disorder

On the other hand, the high scores of depressive and anxious symptoms in these three disorders observed in the current study were consistent with previous studies [42, 45–48].

Moreover, we noticed that some scales, especially the GAD-7 and PHQ-9, showed little contribution to the classification of MDD or AD (Supplementary Materials Table S4). In previous studies, the GAD-7 is specific to anxiety symptoms [8, 49], and the PHQ-9 is more sensitive to MDD [15, 50, 51]. However, they might not be as comprehensive as HAMA and HAMD in the measurement of symptoms for patients, and they also might not be as sensitive as the PHQ-15 and SDS to somatic symptoms [29] and daily life disability [37, 52, 53]. On the other hand, the failure of GAD-7 to classify AD might be attributed to the heterogeneity of patients as a previous study, which has demonstrated poor specificity and a high false positive rate of GAD-7 for all anxiety disorders [54].

In clinical diagnosis, psychiatrists conducted a comprehensive assessment and diagnosis for patients according to the DSM-IV, ICD-10, and MINI, mainly taking the severity and duration of pathological emotion and impaired cognitive function into consideration [1, 55–57]. However, psychiatric symptoms might not be fully apparent in a relatively short time frame and might fluctuate over time [58–60]. Moreover, the phenotype of psychiatric disorders varies from patient to patient [34, 61–63]. Furthermore, clinician-rated scales, such as HAMD and HAMA, might not cover all symptoms of all patients [60, 64, 65]. Thus, more treatment benefits might be achieved by recognition of individual experience, as suggested by a recent study [35].

The findings of this study suggest that the observed symptoms assessed by psychiatrists (HAMD, HAMA, and YMRS), the subjective experience rated by patients themselves (AIS, PHQ-15, GAD-7, and PHQ-9), and the cognitive functionsthat patients experienced, assessed under professional guidance (ST, TMT, DMT, and DSST), should be considered comprehensively in clinical diagnosis [41]. In particular, the self-report scales were used to screen mood disorders in patients with psychiatric disorders [10, 16, 66], cancer [17], stroke [18–21], and epilepsy [22] in general hospitals, where there might be a shortage of psychiatric services.

The present study has several limitations that deserve mentioning. First, as the BD patients in this study were not in manic episodes, none of the patient-rated mania scales was used to rate subjective mania symptoms. However, the YMRS scores indicated that some BD patitents exhibited subthreshold mania or hypomanic symptoms. This omission led to a potentially overlooking clinician-rated scales in classifying BD patients. Future studies should take subjective mania symptoms into the analysis. Second, there are significant rates of comorbidity among MDD, BD, and AD according to previous studies [67–69]. This study did not include patients with comorbid MDD and AD or BD and AD because its primary focus was to assess the classification accuracy of scales and cognitive tests commonly used in clinical practice. Future research should prioritize examining the classification performance of these scales and cognitive tests in the context of comorbidity. Third, neither symptom validity testing nor performance validity testing was administered in the current study. Consequently, this omission could undermine the overall credibility of the findings and conclusions in this study. Future works should include symptom and performance validity testing to ensure that only valid data is included in analyses. Last, the small sample size in

this study may lead to an overestimation of accuracy, sensitivity, specificity, and AUC, resulting in inflated performance measures due to reduced variability and sampling bias. Therefore, conducting further research with a larger sample size would be valuable.

In summary, our findings indicate that subjective experience might be more effective in distinguishing patients with MDD and AD from HCs. Furthermore, daily life disability is the most sensitive to classify MDD patients, while nocturnal symptoms is the most important to AD patients. This work highlights the importance of the self-reported subjective experience in screening patients with BD, MDD, and AD.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12888-024-06237-6.

Supplementary Material 1

Acknowledgements
We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Author contributions
SJ: conceptualization, methodology, investigation, software, formal analysis, visualization, validation, writing – original draft preparation, writing – review and editing. JZ: investigation, software, validation, writing – review and editing. CZ: investigation, methodology, writing – review and editing. MC: conceptualization, methodology, writing – review and editing. HY: funding acquisition, project administration, supervision, writing – review and editing. SJ and JZ contributed equally to this work.

Funding
This work was supported in part by the National Natural Science Foundation of China (81901358), Natural Science Foundation of Shandong Province (ZR2019BH001 and ZR2021YQ55), Young Taishan Scholars of Shandong Province (tsqn201909146), and Key research and development program of Jining City (2021YXNS087).

Data availability
The datasets used and analysed during the current study available from the corresponding author on reasonable request.

Declarations

Ethical approval and consent to participate
Informed written consent was provided by all subjects, and the protocol was approved by institutional review board of Jining Medical University.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 23 April 2024 / Accepted: 30 October 2024
Published online: 11 November 2024

References

1. Charlson FJ, Baxter AJ, Cheng HG, Shidhaye R, Whiteford HA. The burden of mental, neurological, and substance use disorders in China and India: a systematic analysis of community representative epidemiological studies. Lancet. 2016;388(10042):376–89.
2. Chekroud AM, Loho H, Krystal JH. Mental illness and mental health. Lancet Psychiatry. 2017;4(4):276–7.
3. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, Yu Y, Kou C, Xu X, Lu J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. Lancet Psychiatry. 2019;6(3):211–24.
4. M.Hamilton. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23:56–62.
5. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol. 1959;23:50–9.
6. Young RCBJ, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry: J Mental Sci. 1978;133:429–35.
7. Williams KKLSBW. The PHQ-9 validity of a brief depression severity measure. J Gen Intern Med. 2001;16(7):606–13.
8. Delgadillo J, Payne S, Gilbody S, Godfrey C, Gore S, Jessop D, Dale V. Brief case finding tools for anxiety disorders: validation of GAD-7 and GAD-2 in addictions treatment. Drug Alcohol Depend. 2012;125(1–2):37–42.
9. Sun DL, LQH, Li MJ, Chen JH. Patient health questionnaire-15 (PHQ-15) to distinguish bipolar II disorder from major depressive disorder. Psychiatry Res. 2020;290:113026.
10. Zimmerman M, Galione JN, Chelminski I, Young D, Ruggero CJ. Performance of the Bipolar Spectrum Diagnostic Scale in psychiatric outpatients. Bipolar Disord. 2010;12(5):528–38.
11. Okamoto K, Harasawa Y. Prediction of symptomatic depression by discriminant analysis in Japanese community-dwelling elderly. Arch Gerontol Geriatr. 2011;52(2):177–80.
12. Harry ML, Waring SC. The measurement invariance of the Patient Health Questionnaire-9 for American Indian adults. J Affect Disord. 2019;254:59–68.
13. Lee D. The convergent, discriminant, and nomological validity of the Depression anxiety stress Scales-21 (DASS-21). J Affect Disord. 2019;259:136–42.
14. Dajpratham P, Pukrittayakamee P, Atsariyasing W, Wannarit K, Boonhong J, Pongpirul K. The validity and reliability of the PHQ-9 in screening for post-stroke depression. BMC Psychiatry. 2020;20(1):291.
15. Levis B, Sun Y, He C, Wu Y, Krishnan A, Bhandari PM, Neupane D, Imran M, Brehaut E, Negeri Z, et al. Accuracy of the PHQ-2 alone and in Combination with the PHQ-9 for screening to detect Major Depression: systematic review and Meta-analysis. JAMA. 2020;323(22):2290–300.
16. Hu C, Xiang Y-T, Wang G, Ungvari GS, Dickerson FB, Kilbourne AM, Lai KYC, Si T-M, Fang Y-R, Lu Z, et al. Screening for bipolar disorder with the Mood disorders Questionnaire in patients diagnosed as major depressive disorder — the experience in China. J Affect Disord. 2012;141(1):40–6.
17. Lafont C, Chah Wakilian A, Lemogne C, Gouraud C, Fossey-Diaz V, Orvoen G, Lhuillier N, Paillaud E, Bastuji-Garin S, Zebachi S, et al. Diagnostic performance of the 4-Item geriatric depression scale for Depression Screening in older patients with Cancer: the ELCAPA Cohort Study. Oncologist. 2021;26(6):e983–91.
18. Wang EY, Meyer C, Graham GD, Whooley MA. Evaluating screening tests for Depression in Post-stroke older adults. J Geriatr Psychiatr Neurol. 2018;31(3):129–35.
19. McCrory M, Murphy DF, Morris RC, Noad RF. Evaluating the GAD-2 to screen for post-stroke anxiety on an acute stroke unit. Neuropsychological Rehabilitation. 2022;33(3):480–96.
20. Myhre P, Radakovic R, Ford C. Validation of the self-rated dimensional apathy scale in community stroke survivors. J Neurol Sci. 2022;434:120103
21. Yue Y, Liu R, Chen J, Cao Y, Wu Y, Zhang S, Li H, Zhu J, Wu A, Yuan Y. The reliability and validity of Post Stroke Depression Scale in different type of Post Stroke Depression patients. J Stroke Cerebrovasc Dis. 2022;31(2).
22. Fiest KM, Patten SB, Jetté N. Screening for depression and anxiety in Epilepsy. Neurol Clin. 2016;34(2):351–61.
23. Kroenke K, Spitzer RL, Williams JB, Lowe B. The Patient Health Questionnaire somatic, anxiety, and depressive Symptom scales: a systematic review. Gen Hosp Psychiatry. 2010;32(4):345–59.
24. Belhumeur PN. Kriegman Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell. 1997;19:711–20.
25. Leo H, Chiang ELR, Richard D, Braatz. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. Chemometr Intell Lab Syst. 2000;50(2):243–352.
26. Bernardo LS, Quezada A, Munoz R, Maia FM, Pereira CR, Wu W, de Albuquerque VHC. Handwritten pattern recognition for early Parkinson's disease diagnosis. Pattern Recognit Lett. 2019;125:78–84.

27. Lauren P, Qu G, Zhang F, Lendasse A. Discriminant document embeddings with an extreme learning machine for classifying clinical narratives. Neurocomputing. 2018;277:129–38.
28. Andrade C. Psychotropic drugs with long Half-Lives: implications for drug discontinuation, Occasional missed doses, dosing interval, and pregnancy planning. J Clin Psychiatry. 2022;83(4):22f14593.
29. Zijlema WL, Stolk RP, Lowe B, Rief W, BioShaRe, White PD, Rosmalen JG. How to assess common somatic symptoms in large-scale studies: a systematic review of questionnaires. J Psychosom Res. 2013;74(6):459–68.
30. Hodgins DC. Reliability and validity of the Sheehan disability scale modified for pathological gambling. David C Hodgins. 2013;13(177).
31. Soldatos CR, Dikeos DG, Paparrigopoulos TJ. The diagnostic validity of the Athens Insomnia Scale. J Psychosom Res. 2003;55(3):263–7.
32. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of Observational studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344–9.
33. Sun DL, Liu QH, Li MJ, Yang Y, Zhang R, Xiang SY, Chen JH. Patient health questionnaire-15 (PHQ-15) to distinguish bipolar II disorder from major depressive disorder. Psychiatry Res. 2020;290:113026.
34. Koyuncu A, Ertekin E, Ertekin BA, Binbay Z, Yuksel C, Deveci E, Tukel R. Relationship between atypical depression and social anxiety disorder. Psychiatry Res. 2015;225(1–2):79–84.
35. Allsopp K, Read J, Corcoran R, Kinderman P. Heterogeneity in psychiatric diagnostic classification. Psychiatry Res. 2019;279:15–22.
36. Leon AC, Olfson M, Portera L, Farber L, Sheehan DV. Assessing psychiatric impairment in primary care with the Sheehan disability scale. Int J Psychiatry Med. 1997;27(2):93–105.
37. Gonzalez-Blanch C, Fernando H-d-H, Munoz-Navarro R, Ruiz-Rodriguez P, Medrano LA, Moriana JA, Cano-Vindel A, Psic APRG. Domain-specific associations between disability and depression, anxiety, and somatization in primary care patients. Psychiatry Res. 2018;269:596–601.
38. Zhou L, Zhang H, Luo Z, Liu X, Yang L, Hu H, Wang Y, Wang C, Wang F. Abnormal night sleep duration and inappropriate sleep initiation time are associated with elevated anxiety symptoms in Chinese rural adults: the Henan Rural Cohort. Psychiatry Res. 2020; 291:113232.
39. Ziebold C, Goldberg DP, Reed GM, Minhas F, Razzaque B, Fortes S, Robles R, Lam TP, Bobes J, Iglesias C, et al. Dimensional analysis of depressive, anxious and somatic symptoms presented by primary care patients and their relationship with ICD-11 PHC proposed diagnoses. Psychol Med. 2019;49(5):764–71.
40. Williams LM. Defining biotypes for depression and anxiety based on large-scale circuit dysfunction: a theoretical review of the evidence and future directions for clinical translation. Depress Anxiety. 2017;34(1):9–24.
41. Walsh-Messinger J, Jiang H, Lee H, Rothman K, Ahn H, Malaspina D. Relative importance of symptoms, cognition, and other multilevel variables for psychiatric disease classifications by machine learning. Psychiatry Res. 2019;278:27–34.
42. Couillard Larocque M, Fortin-Vidah G, Angers M, Garceau L, Gros L, Fournel I, Provencher MD. Anxiety in bipolar disorder: a review of publication trends. J Affect Disord. 2023;320:340–7.
43. Seon Q, Hum S, Tuineag M, Pavlova B, Beaulieu S, Linnaranta O. Properties of common anxiety scales among patients with bipolar disorder. J Affect Disord. 2021;281:972–9.
44. Cullen C, Kappelmann N, Umer M, Abdolizadeh A, Husain MO, Bonato S, Sharma G, Xue S, Ortiz A, Kloiber SM, et al. Efficacy and acceptability of pharmacotherapy for comorbid anxiety symptoms in bipolar disorder: a systematic review and meta-analysis. Bipolar Disord. 2021;23(8):754–66.
45. Goldberg D, Fawcett J. The importance of anxiety in both Major Depression and Bipolar Disorder. Depress Anxiety. 2012;29(6):471–8.
46. Pizzagalli DA. Psychobiology of the intersection and divergence of depression and anxiety. Depress Anxiety. 2016;33(10):891–4.
47. Zhu JY, Plamondon A, Goldstein AL, Snorrason I, Katz J, Björgvinsson T. Dynamics of daily positive and negative affect and relations to anxiety and depression symptoms in a transdiagnostic clinical sample. Depress Anxiety. 2022;39(12):932–43.
48. Bauer EA, MacNamara A. Comorbid anxiety and depression: opposing effects on the electrocortical processing of negative imagery in a focal fear sample. Depress Anxiety. 2021;38(7):690–9.
49. Mossman SA, LMJ, Schroeder HK, et al. The generalized anxiety disorder 7-item scale in adolescents with generalized anxiety disorder: Signal detection and validation. Ann Clin Psychiatry. 2017;29(4):227–34.
50. Kroenke KSRL, Williams JBW. The PHQ-9 validity of a brief Depression Severity measure. J Gen Intern Med. 2001;16(9):606–13.
51. Monahan PO, Shacham E, Reece M, Kroenke K, Ong'or WO, Omollo O, Yebei VN, Ojwang C. Validity/Reliability of PHQ-9 and PHQ-2 Depression scales among adults living with HIV/AIDS in Western Kenya. J Gen Intern Med. 2008;24(2):189–97.
52. Leu SH, Chou JY, Lee Pc, Cheng HC, Shao WC, Hsien WL, Huang CL, Chen VCH. Validity and reliability of the Chinese version of the Sheehan disability scale (SDS-C). Asia-Pacific Psychiatry. 2015;7(2):215–22.
53. Gutiérrez-Rojas L, Jurado D, Gurpegui M. Factors associated with work, social life and family life disability in bipolar disorder patients. Psychiatry Res. 2011;186(2–3):254–60.
54. Beard C, Björgvinsson T. Beyond generalized anxiety disorder: psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. J Anxiety Disord. 2014;28(6):547–52.
55. Miquel Roca1 MV. 2 Emilio López-Navarro1,2 Javier García-Campayo2,3 Margalida Gili1,2: cognitive impairments and depression.a critical review. Actas Esp Psiquiatr. 2015;43(5):187–93.
56. Salagre E, Sole B, Tomioka Y, Fernandes BS, Hidalgo-Mazzei D, Garriga M, Jimenez E, Sanchez-Moreno J, Vieta E, Grande I. Treatment of neurocognitive symptoms in unipolar depression: a systematic review and future perspectives. J Affect Disord. 2017;221:205–21.
57. Xu G, Lin K, Rao D, Dang Y, Ouyang H, Guo Y, Ma J, Chen J. Neuropsychological performance in bipolar I, bipolar II and unipolar depression patients: a longitudinal, naturalistic study. J Affect Disord. 2012;136(3):328–39.
58. Liu RT, Walsh RFL, Sheehan AE. Prebiotics and probiotics for depression and anxiety: a systematic review and meta-analysis of controlled clinical trials. Neurosci Biobehav Rev. 2019;102:13–23.
59. Fang L, Marchetti I, Hoorelbeke K, Koster EHW. Do daily dynamics in rumination and affect predict depressive symptoms and trait rumination? An experience sampling study. J Behav Ther Exp Psychiatry. 2019;63:66–72.
60. Rodriguez-Seijas C, Thompson JS, Diehl JM, Zimmerman M. A comparison of the dimensionality of the Hamilton Rating Scale for anxiety and the DSM-5 anxious-distress specifier interview. Psychiatry Res. 2020;284:112788.
61. Saunders R, Buckman JEJ, Cape J, Fearon P, Leibowitz J, Pilling S. Trajectories of depression and anxiety symptom change during psychological therapy. J Affect Disord. 2019;249:327–35.
62. Xin LM, Chen L, Su YA, Yang FD, Wang G, Fang YR, Lu Z, Yang HC, Hu J, Chen ZY, et al. Prevalence and clinical features of atypical depression among patients with major depressive disorder in China. J Affect Disord. 2019;246:285–9.
63. Lojko D, Buzuk G, Owecki M, Ruchala M, Rybakowski JK. Atypical features in depression: Association with obesity and bipolar disorder. J Affect Disord. 2015;185:76–80.
64. Nutt D. The Hamilton Depression scale–accelerator or break on antidepressant drug discovery? J Neurol Neurosurg Psychiatry. 2014;85(2):119–20.
65. Zimmerman M, Thompson JS, Diehl JM, Balling C, Kiefer R. Is the DSM-5 anxious distress specifier interview a valid measure of anxiety in patients with generalized anxiety disorder: a comparison to the Hamilton anxiety scale. Psychiatry Res. 2020;286:112859.
66. Zimmerman M, Galione JN, Chelminski I, Young D, Dalrymple K. Psychiatric diagnoses in patients who screen positive on the Mood Disorder Questionnaire: implications for using the scale as a case-finding instrument for bipolar disorder. Psychiatry Res. 2011;185(3):444–9.
67. Adams GC, Wrath AJ, Mondal P, Asmundson GJG. Depression with or without comorbid social anxiety: is attachment the culprit? Psychiatry Res. 2018;269:86–92.
68. Paavonen V, Luoto K, Lassila A, Leinonen E, Kampman O. Temperament clusters associate with anxiety disorder comorbidity in depression. J Affect Disord. 2018;236:252–8.
69. Winter DM, Cipolletta C. Construing and body dissatisfaction in chronic depression: a study of body psychotherapy. Psychiatry Res. 2018;270:845–51.

Publisher's note