General Hospital Psychiatry 61 (2019) 69–75

Contents lists available at ScienceDirect

General Hospital Psychiatry

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

Psychometric properties of the Patient Health Questionnaire nine-item version (PHQ-9) for use among hospitalized non-psychiatric medical patients

F.M. Daray[a,b,*], F. Hunter[c], A. Anastasia[d], M. Fornaro[e]

[a] Institute of Pharmacology, School of Medicine, University of Buenos Aires, Argentina
[b] National Scientific and Technical Research Council (CONICET), Argentina
[c] José Ramos Mejia Hospital, Buenos Aires, Argentina
[d] National Institue for Social Security, Latina, Italy
[e] Federico II University, Section of Psychiatry, Department of Neuroscience, Reproductive Sciences and Dentistry, Naples, Italy

1. Introduction

Depression is a highly prevalent and disabling condition among the general population [1,2], and its prevalence is even higher among patients hospitalized due to medical illnesses [3–6]. It has been estimated that one out of four patients hospitalized for medical illness has a comorbid major depressive episode (MDE) [3].

The presence of depression in a patient hospitalized due to a medical illness worsens the medical prognosis [7], increases symptom burden [8–10], complicates self-care and treatment-adherence [11–13], extends the hospitalization stay [14], and inflates the costs of care and mortality [11,15–17]. Taking into account that patients with depression can be effectively treated with medication and psychotherapy, early diagnosis and treatment can significantly reduce this burden [18,19]. Despite this, almost 50% of patients who suffer comorbid depression when hospitalized for medical conditions do not have the condition detected by clinicians [20]. This is explained, at least in part, by the lack of quick and sensitive diagnostic tools, especially for use in specific languages and cultures. Structured clinical interviews such as the composite international diagnostic interview (CIDI) [21] or the mini international neuropsychiatric interview (MINI) [22] are the gold standard for diagnosing depression. However, the use of these instruments require administration by trained mental health professionals, are time-consuming and difficult to systematically apply in routine clinical practice. This highlights the need for reliable, valid, and easily applicable self-assessment screening tools to detect depression in patients hospitalized due to a non-psychiatric medical illness. Several self-screening tools are available for screening, tracking and documenting depressive symptoms with minimal clinician time, such as the Patient Health Questionnaire 9 (PHQ-9) [23,24], the Beck Depression Inventory-II (BDI-II) [25] and the Hospital Anxiety and Depression Scale (HADS) [26].

To our knowledge, no studies have compared, at the same time and with the same sample, the psychometric properties of these three self-rated questionaries' with a semi-structured clinical interview in patients hospitalized for non-psychiatric medical illness. To cover this gap, we designed the present study.

Our aims are to: (1) compare the operating characteristics of PHQ-9, BDI-II and HADS according to an independent criterion standard for MDE in patients hospitalized due to non-psychiatric medical illness; (2) determine the optimum cut-off scores of these self-rating scales for discriminating between patients with and without MDE in this specific group of patients. The results of this study will provide general practitioners with highly valuable information to choose an assessment tool for depression in this subgroup of patients.

2. Methods

2.1. Study design

The present multicenter, cross-sectional study was conducted across three facilities in Buenos Aires, Argentina: Bernardo Houssay Hospital in the Vicente Lopez district, Eva Perón Hospital in the San Martín district, and Bernardino Rivadavia Hospital in the Autonomous City district. The study was conducted between August 2013 and May 2014. These general hospitals mostly serve low-income patients without health insurance and are located in the metropolitan area of Buenos Aires, which has a population of 12.8 million. The study protocol was approved by the Institutional Review Boards of the participating hospitals, and written informed consent was obtained from all subjects.

2.2. Participants

In order to have heterogeneity of medical conditions as in the real

* Corresponding author at: Institute of Pharmacology, School of Medicine, University of Buenos Aires, Argentina, Paraguay 2155, piso 9, C1121ABG Ciudad de Buenos Aires, Argentina.
E-mail address: fdaray@hotmail.com (F.M. Daray).

https://doi.org/10.1016/j.genhosppsych.2019.10.004
Received 8 July 2019; Received in revised form 8 October 2019; Accepted 14 October 2019

world, we sampled all consecutive patients hospitalized in these General Hospitals for any medical illness over six months. Participants were adult inpatients with at least one primary medical diagnosis according to the ICD-10 [27] classification: infectious (22.0%), neurological (17.0%), metabolic (12.9%), pulmonary (11.1%), cardiovascular (10.4%), neoplastic (9.13%), urological (9.13%), gastrointestinal (7.46%) and hematological (4.56%).

Included subjects were: [a] native Spanish speakers, [b] aged ≥18 years old and [c] able to provide written informed consent to participate. Subjects were excluded if they were: [a] unable to participate in clinical assessments or to complete symptom-ratings because of illness, medication, sensory or speech impairment, or lack of language fluency, or [b] scored < 25 on a preliminary Mini-Mental State Examination (MMSE) [28] suggesting dementia or delirium.

2.3. Procedure measures

Participants were interviewed at days 2–6 after admission, and assessed by a psychiatrist or psychiatric resident using the MINI (MINI, Spanish version 5.0) to look for the presence of an MDE. Then each participant went through the administration of the following rating questionaries based on a random sequence in order to limit re-test artifacts: the PHQ-9, the BDI-II, and the HADS. A psychiatrist or psychiatric resident trained for this purpose, held blind to study findings, examined all study participants, collecting demographic and descriptive data, and administering the MINI and the three self-rating scales.

2.4. Assessment tools

The PHQ-9 was developed in the mid-1990s for screening depression in a primary care setting and for the detection of this condition in large epidemiological studies [29–31]. It is a nine-item, self-reporting measure designed to detect and rate the presence and severity of depression with continuous scoring (0–3/item, a total of 0–27). It has been extensively validated in many countries and is one of the most commonly used tools for diagnosis and severity assessment of depression [32]. This instrument is based on DSM- IV-TR [33] criteria for MDE. Its psychometric properties have been extensively studied in the general population and clinical settings [32,34]. In Argentina, scores ≥8 are considered consistent with clinical MDE [35].

The BDI-II has been extensively used since 1996 in clinical and research settings; it is a self-reporting questionnaire rating 21 items for presence and severity (0–3/item) for the previous two weeks and the current day, with possible scores of 0–63. In Argentina, a cut-off score of ≥14 is considered positive for depression [36]. The BDI-II differs from the PHQ-9 in how it asks about the directionality of sleep, appetite, and weight changes; also, the BDI-II asks about additional items such as punishment feelings, crying, indecisiveness, irritability, and libido.

The HADS differ from the other two questionnaires in important ways such as the exclusion of somatic symptoms; given that diagnostic interviews for depression include somatic items, comparison of a diagnostic referent standard with the HADS is essential. The HADS contains two subscales measuring symptoms of depression (HADS-D) and anxiety (HADS-A) during the previous week. It includes seven statements on each disorder, and each response consists of a four-point rating scale (0 to 3); a higher score indicates a worse condition. For each subscale, the total score is at most 21. A score of ≥11 is considered a clinically significant disorder, whereas a score between 8 and 10 suggests a mild disorder [26,37]. This questionnaire has been validated in Argentina [38].

2.5. Statistical analysis

Data distribution was ascertained using the Kolmogorov-Smirnov test. Concerning descriptive statistics, the χ^2 test and the independent-samples *t*-test were adopted to assess frequency distributions and to compare continuous outcomes. Data transformations were performed as necessary in order to avoid non-parametric tests (which may have limited power compared to the parametric equivalents). Also, Bayes' factors were reported beyond the sole null-hypothesis statistics. Bayes factors instead of *p*-values (only) since they are independent of intentions and sample size, aiding in the critical interpretation of negative p-values in particular. The ability to consider model uncertainty within a single framework, although currently underused, is a major advantage of Bayesian method. Since it is not void of limitations, the canonical "frequentistic statistics" ("*p*-values") were likewise reported as standard part of the results [39].

Bayesian analyses included the receiver operating characteristics (ROC) curve, which was used to discriminate between groups and to ascertain the sensitivity (probability of positive test result [PHQ-9$^+$]) given that the individual tested actually has the disorder [MDE], and specificity (the probability of a negative screening at the PHQ-9 [PHQ-9$^-$]) given that the individual tested does not have the disorder - or "true negative" rate, analyzed as 1-specificity or "false positive" rate) at various cut-off scores. The same ROC analyses were applied to the BDI-II and the HADS in order to compare the psychometric properties of the PHQ-9 among general medicine Argentinean inpatients.

Additional analyses included calculation of the area-under-the-curve (AUC) and of positive predictive value (PPV) and the negative predictive value (NPV), defined as "the probability of having the disorder when given a positive test result" and "the probability of not having the disorder when given a negative test result" respectively. The following values were considered for the AUC: "0.9 = high accuracy, 0.7–0.9 = moderate accuracy, 0.5–0.7 = low accuracy, and 0.5 a chance result." [40]. A principal component analysis using oblique (promax) rotation was used for the PHQ-9 [41]. The Kaiser-Meyer-Olkin (KMO) measure verified how suited the data were for the PCA. The test measures sampling adequacy for each variable in the model and for the complete model. The KMO statistic is a measure of the proportion of variance among variables that might be common variance; the Bartlett's test of sphericity assessed the degree of inter-correlation between variables, also selecting the "anti-image" option and the Anderson-Rubin method. Eigenvalues > 1 were initially retained (Kaiser's criterion). The Scree-plot output and clinical considerations decided the final number of factors to retain. Items having an absolute loading score substantially overlapping across two or more factors were then disregarded in a second-passage PCA. The internal consistency of the PHQ-9 was determined using Cronbach's alpha ("split-half" method). The level of significance was set at 0.05 (two-tailed) using 95% confidence intervals (CI) across the analyses. Missing data were excluded pairwise. Finally, data were analyzed using the Microsoft Windows® versions v.25 of both IBM SPSS Statistics®.

3. Results

3.1. General characteristics of the sample

A total of 754 participants were screened consecutively, and 365 were classified as potentially eligible. From these, 15 were excluded as they were younger than 18 years, 7 were excluded as they were unable to communicate, 33 refused to consent, and 4 were discharged from the hospital before personnel could invite them. Forty-nine additional participants were excluded due to incomplete outcome data, leaving a total of 257 participants for analysis. Each participating center recruited about the same number of participants.

Subjects included 152 (59.1%) men and 105 (40.9%) women, with an average age of 54.15 ± 16.57 years. Essential demographic and clinical data of the sample have been outlined in Table 1. From the sample, 69 participants (26.85%) were considered to have current MDE by expert clinical assessment using MINI. The observed prevalence of MDE across different medical illnesses is described in Table 1 and

Table 1
Demographic and clinical features of the included inpatients.

Total sample, n = 257	(Re)Current MDE (n = 69, of whom single episode n = 33, recurrent n = 36)	NO MDE (n = 188)	χ^2 *or t, (df)*	*p*	BF_{01}
Mean age, in years	52(19.5)	57(25)	6.0405(0.793)*	ns	8.61
Sex (M:F) (n,%)	n = 29(42):n = 40(58)	n = 123 (65.4):n = 65(34.6)	55.101(1)	0.001	0.026
Marital status					
Unmarried, *n*(%)	21(24.4)	65(75.6)	72.344(3)	< 0.001	4.57
Married, *n*(%)	33(30.6)	75(69.4)			
Divorced, *n*(%)	7(25.9)	20(74.1)			
Widow, *n*(%)	8(22.9)	27(77.1)			
Education level, in years, mean(ds)	8(4)	9(5)	1.712(249)	ns	2.17
Major medical condition					
Infective	16(30.8)	36(69.2)	94.541(10)	< 0.001	3.11
Oncological	9(39.1)	14(60.9)			
Hematological	2(16.7)	10(83.3)			
Metabolic	8(25.8)	23(74.2)			
Neurological	8(19.5)	33(80.5)			
Cardiovascular	7(25.9)	20(74.1)			
Pulmonary	2(10)	18(90)			
Gastrointestinal	5(25)	15(75)			
Dermatological	2(40)	3(60)			
Immunological	1(100)	0(0)			
Nephro-urological	8(34.8)	15(65.2)			
Diabetes mellitus	18(26.1)	47(25.0)	61.275(1)	< 0.001	5.46
Hypertension	22(31.9)	69(36.7)	21.887(1)	< 0.001	3.99
Essential substance use conducts					
Substance use disorder, n (%)	1(1.4)	10(5.3)	2.655(3)	ns	8.05
Smoking, n (%)	25(36.2)	43(22.9)	19.530(12)	ns	13.7
Essential rating of depression					
PHQ-9 total score	14.16 (5.25)	5.70 (5.42)	11.182(255)	< 0.001	0
HADS total score	19.35(± 6.133)	9.71 (6.42)	−10.785(254)	< 0.001	0
HADS-D total score	8.61 (3.59)	4.20 (3.63)	−8.504(254)	< 0.001	0
BDI-II total score	23.52 (10.47)	10.70 (8.82)	−9.806(255)	< 0.001	0

T = t-student test; χ^2 = chi-square. Note: df = degree of freedom. Transformations applied to non-normal distributions.
BF_{01} = Bayes factor, which is the ratio of the probability of the data given the alternative hypothesis to that for the null hypothesis. Bayes factors > 1 suggest that the observed data are more likely given the alternative hypothesis that given the null. Values < 1 suggest the opposite. Values between 1 and 3 reflect evidence for the alternative hypothesis that is "barely worth mentioning", values between 1 and 3is evidence that "has substance", and values between 3 and 10 are "strong" evidence [64].
Legend: HADS = Hospital Anxiety and Depression Scale; HADS-D = Hospital Anxiety and Depression − Depression Sub-Scale; BD-II = Beck Depression Inventory-II; PHQ-9 = Patient Health Questionnaire-9.
Note: numbers to adding up to 100% of the total indicates the presence of missing value(s): pairwise exclusion. Please note that the percentages reported in the present table are aimed at highlighting the between- rather the within group comparisons.

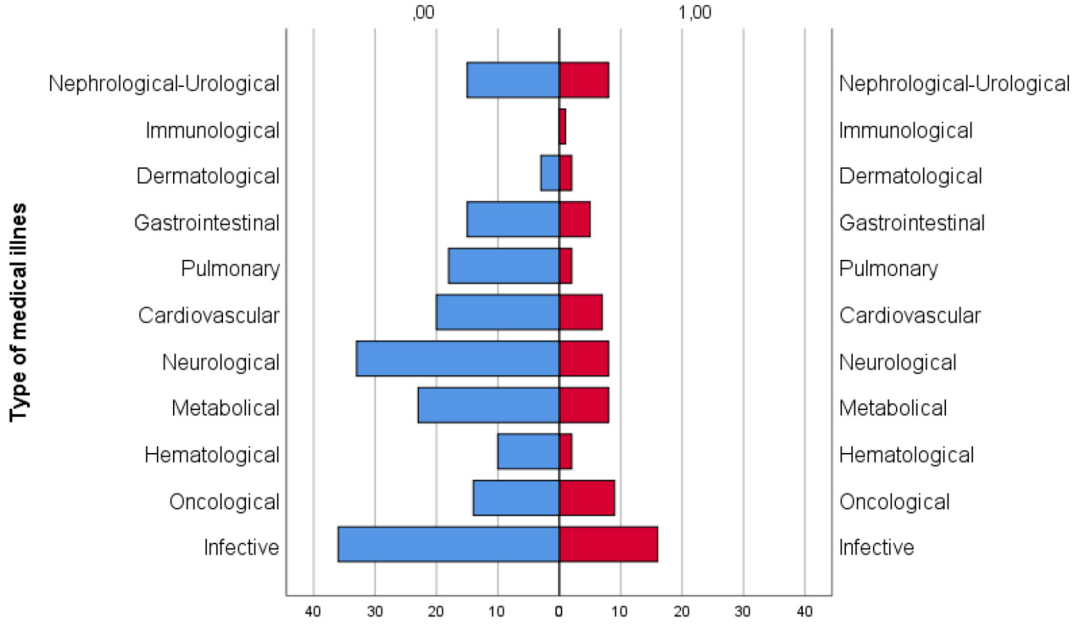

Fig. 1. Comparison of medical condition profile in non-depressed vs. depressed inpatients. Depressed cases on the right hand of the plot.

Table 2
Comparison of the operating characteristics of the optimal cutoff points (the greater value of the Youden' index) of the BDI, HADS; HADS (depression subscale) and PHQ-9 against structured clinical interview (MINI) among medically hospitalized patients.

	Cut-off	Sensitivity	Specificity	PPV	NPV	AUC	Cronbach's α
PHQ-9	10	81%	79%	59%	92%	0.87 ± 0.0.02	0.86
BDI-II	16	77%	78%	56%	90%	0.83 ± 0.0.03	0.90
HADS	15	83%	75%	55%	92%	0.86 ± 0.0.02	0.84
HADS-D	7	67%	73%	56%	86%	0.80 ± 0.03	0.74

PHQ-9: Patient Health Questionnaire-9, BDI-II: Beck Depression Inventory-II, HADS: Hospital Anxiety and Depression Scale, HADS-D: Hospital Anxiety and Depression Scale – Depression Subscale. PPV: positive predictive values, NPV: negative predictive values.

Fig. 1.

3.2. ROC curve analysis and predictive values

The Receiver Operating Characteristic (ROC) curve of the performance of PHQ-9 in discriminating between "MDE (PHQ-9$^+$)" vs. non-depressed (PHQ-9$^-$)" across inpatients is shown in Supplemental Fig. 1. The AUC was 0.87 ± 0.02, $p \leq .001$; 95% C.I. (lower bound = 0.83; upper bound = 0.91) with an optimal cut-off 10 (sensitivity = 81%; specificity = 79%; PPV = 59% and NPV = 92%).

When analyzing the operating characteristic of the BDI-II, the AUC = 0.83 ± 0.03, $p \leq .001$; 95% C.I. (lower bound = 0.77; upper bound = 0.88), Supplemental Fig. 3) with an optimal cut-off of 16 (sensitivity = 77%; specificity = 78%; PPV = 56% and NPV = 90%). For the HADS, the AUC = 0.86 ± 0.02, p ≤ .001; 95% C.I. (lower bound = 0.81; upper bound = 0.90) with an optimal cut-off of 15 (sensitivity = 83%; specificity = 75%; PPV = 55% and NPV = 92%). In order to ease the interpretation of the change in sensitivity and specificity at various cut-offs for these questionnaires, a scatter dot plot of cut-off scores against sensitivity and specificity was also provided (Supplemental Figs. 1–4).

When comparing the AUC between the varying scales, the PHQ-9 performed slightly better than the BDI-II and the HADS (Table 2).

3.3. Reliability analysis

The internal consistency of the PHQ-9 was good [42], with a Cronbach's α value of 0.86. Comparably, the BD-II and the depression subscale of the HADS, and the total HADS (anxiety and depression modules altogether) had 0.90, 0.74, and 0.84 as Cronbach's α values, respectively.

3.4. Factor analysis

A principal component analysis (PCA) was initially conducted on all nine items of the PHQ-9 using an oblique (promax) rotation method. Four components had eigenvalues over Kaiser's criterion of 1 and, overall, explained 71.23% of the variance. The scree plot could have allotted for a two- (most likely) or multiple-components solutions [Fig. 2]. However, despite the chance of a "researcher-controlled bias," owing to clinical interpretability issues, and consistency issues with previous reports, a second-passage PCA using promax rotation was conducted forcing a two-factor solution, owing to interpretability issues (Table 3).

4. Discussion

The present study is the first to compare the psychometric properties of three widely used screening tools for depression (PHQ-9, BDI-II and HADS) against each other at the same time and with the same sample, using an independent criterion for the diagnosis of an MDE among patients hospitalized for non-psychiatric medical conditions. We observed that the PHQ-9 with an optimal cut-off score equal to ten performs better than the BDI-II and HADS in detecting MDE in this specific group of patients. Moreover, the Argentine version of the PHQ-9 questionnaire showed very good validity and reliability for the screening of MDEs in these patients.

Depression among patients hospitalized for non-psychiatric medical conditions is highly relevant not only for its frequency but also because of its clinical and economic burden on the health system [4,5,11,15]. However, depression and anxiety often co-occur even in non-psychiatric medical outpatients [43].

Validated self-assessment questionnaires are needed for the early detection of comorbid depressive symptoms in patients hospitalized for clinical illness. Several self-rating questionnaires such as the PHQ-9, BDI-II and HADS are available for depression screening; however, for most of these the psychometric properties and cut-off scores were estimated in the general population or outpatient psychiatric settings, so data regarding their properties in this specific group of patients are scarce [44]. In patients hospitalized for non-psychiatric medical conditions, symptoms of their clinical illness may overlap with the somatic symptoms of depression. This means that extrapolating cut-off scores calculated in the general population or outpatient clinical settings may give a high number of false positives.

In the present study, we compared data obtained at the same time and from the same sample, evaluating the psychometric properties of three widely used self-rating questionnaires with semi-structural clinical interviews in patients hospitalized for medical conditions. When comparing the AUC, the PHQ-9 performed better with an AUC value of 0.87 ± 0.02, which suggests a high diagnostic accuracy. The sensitivity at the cut-off score of ten was 82%, and the specificity was 79%. This cut-off score is higher than the score of eight previously reported in Argentina in a psychiatric ambulatory setting (35), however, it is within the range reported in a meta-analysis; the adequate cut-off score for diagnosing MDE with the PHQ-9 ranged from eight to eleven [46,47].

Previous studies using the PHQ-9 in different settings reported a Cronbach's α that ranged from 0.67 to 0.89 [48–54]. It has been suggested that a Cronbach's α of 0.70 or higher should be regarded as acceptable for a self-reported instrument [55]. The present study showed that the PHQ-9 has a Cronbach's α of 0.86, which is within the highest level of the range. Although the PHQ-9 attempts to measure one construct, the depressive syndrome, previous results of factorial analysis of this instrument revealed different factorial solutions of 1 or 2 factorial structure [41,56–59]. In the present study, the factor analysis of PHQ-9 identified a two-factorial structure, which was consistent with previous studies [41,56,57,59,60]. Specifically, component 1 comprised the following items: 1, 3, 4, 5, 8; component 2 comprised 2, 6, 7, 9. We interpreted component 1 as: "somatic/neuro-vegetative" core of depression (which we hypothesized to be underpinned by the primary medical illness), and we interpreted component 2 as: "psychological/cognitive" component of depression. Confirmatory factor analysis likewise suggested the satisfactory goodness-of-fit of PHQ-9.

Consistent with previous studies carried out on patients in the primary care setting, the present study demonstrated that PHQ-9 is a valid and reliable screening tool for depression in medically hospitalized patients, and the optimal sensibility and specificity for the best detection of MDE are observed with a cut-off score equal to ten. This is of great importance since the PHQ-9 is well-known and used widely by

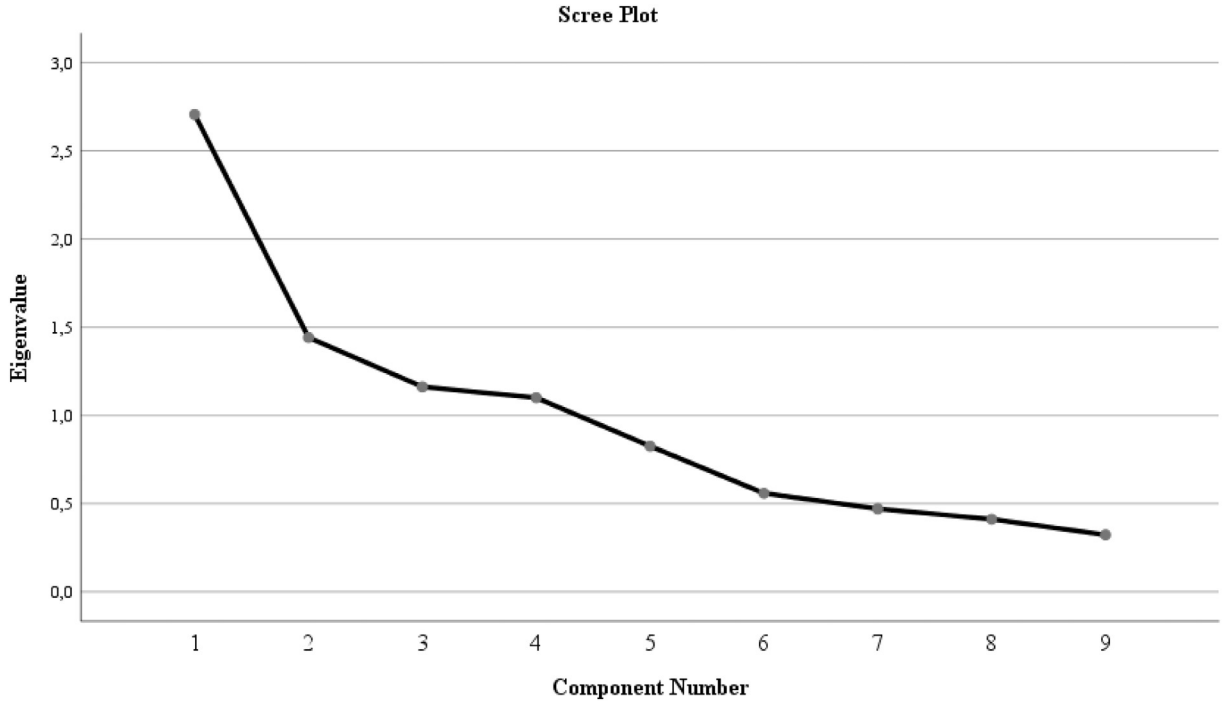

Fig. 2. Scree plot obtained by performing an exploratory factor analysis using promax rotation applied to the nine items of the PHQ-9. Owing to interpretability considerations, we retained a two-component solution. Eigenvalues: component 1 = 2.706; component 2 = 1.442.

clinical practitioners, allows quick evaluations (it requires about 1 min to complete) and can be used free of charge.

The HADS is a questionary designed to measure depression and anxiety in patients with physical illness since this instrument excludes somatic symptoms that may overlap with psychological/cognitive symptoms. The HADS contains two subscales measuring symptoms of depression (HADS-D) and anxiety (HADS-A). When comparing the operating characteristic of the complete scale (HADS) vs. depression subscale (HADS-D) we observed that the full scale performed better in terms of sensitivity and specificity than the sole depression module of the HADS-D. A possible explanation for this finding is that when considering anxiety items, numerous patients with anxious depression cross the detection threshold and are included as positive for depression. This suggests that depression in the medical setting is characterized by prominent anxiety features. Similar findings have been previously reported in cardiac patients [61] and in hemodialysis patients [62]. Although "anxious depression" is more frequently observed among the bipolar spectrum, in the present study we do not have the information of which cases of MDE correspond to unipolar and bipolar patients to answer this question, but future studies could explore it. Future studies should also include instruments specifically developed for the detection of either sub- or full-threshold bipolarity, with the aim of better instructing therapeutic choices, even among patients primarily hospitalized for a non-psychiatric medical condition. Furthermore, these results suggest that taking the MINI as the gold standard for MDE in patients with non-psychiatric medical conditions allows the inclusion of more subjects with depression and prominent anxiety symptoms. The results of the present study could also explain why MINI rated a higher

Table 3
PCA of the PHQ-9 using promax with Kaiser normalization, oblique rotation, a two-factor solution was forced upon inspection of the scree plot (please see below).

	Component	
	1	2
Item 1: Little interest or pleasure in doing things/*Poco interés o placer en hacer cosas*	**0.449**	0.298
Item 2: Feeling down, depressed, or hopeless/*Sentirse decaído/a, deprimido/a o desesperanzado/a*	0.348	**0.689**
Item 3: Trouble falling or staying asleep, or sleeping too much/*Dificultad para dormirse o para mantener el sueño, o dormir demasiado*	**0.682**	0.135
Item 4: Feeling tired or having little energy/*Sentirse cansado/a o con poca energía*	**0.643**	0.234
Item 5: Poor appetite or overeating/*Con poco apetito o comer excesivamente*	**0.715**	0.126
Item 6: Feeling bad about yourself — or that you are a failure or have let yourself or your family down/*Sentirse mal con usted mismo/a o sentir que usted es un fracaso o que le ha fallado a su familia o a sí mismo/a*	0.150	**0.864**
Item 7: Trouble concentrating on things, such as reading the newspaper or watching television/*Dificultad para concentrarse en cosas, tales como leer el diario o ver televisión*	0.287	**0.452**
Item 8: Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual/*Estar más lento/a que lo habitual para moverse o hablar, o, por el contrario, estar más inquieto/a e intranquilo/a, moviéndose más que lo habitual*	**0.734**	0.214
Item 9: Thoughts that you would be better off dead or of hurting yourself in some way/*Pensamientos de que usted estaría mejor muerto/a, o de hacerse daño a sí mismo/a de alguna manera*	0.091	**0.687**

Bold Eigenvalues loaded on the corresponding component in the columns.
Specifically, component 1 comprised the following items: 1, 3, 4, 5, 8; component 2 comprised the following: 2, 6, 7, 9.
Components were assigned either to C1 or C2 depending on a value ≥ 0.4.
We interpreted C1 as: "somatic/neurovegetative" core of depression (which we hypothesized to be underpinned by the primary medical illness), and we interpreted C2 as: "psychological/cognitive" component of depression.

proportion of people as depressed in comparison with other fully structured diagnostic interviews [63].

Finally, the BDI-II was the questionary with the lowest performance for the detection of MDE in patients with non-psychiatric medical conditions.

4.1. Strengths and limitations

One of the strengths of the study is that depression was assessed with two different methods, a self-rating scale and a structural clinical interview conducted face-to-face with the participant on the same day as screening with the questionnaires, and the researchers were blinded to the results of the questionnaires. Furthermore, the study was performed in three different metropolitan General Hospitals and with different medical conditions, which makes the sample more representative of daily practice.

A limitation of the study is that these are General Hospitals that serve low-income patients without health insurance, factors that might contribute to a higher risk of depression, so inferences should be made with caution. Moreover, we have no information about comorbid psychiatric diagnoses among participants with an MDE. Finally, medical diagnoses/categories were mutually exclusive, hindering the appreciation of medical comorbidities.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.genhosppsych.2019.10.004.

Funding

Supported by a research grant - “Ramón Carrillo-Arturo Oñativa for Multicentric Studies” (2015) from the commission “Salud Investiga” of the Ministry of Health and Social Action of Argentina (to FMD). Grant n° 1853.

Declaration of competing interest

No author or an immediate family member has financial relationships with commercial entities that might appear to represent a potential for conflicts of interest.

Acknowledgment

The authors want to gratefully acknowledge the study participants for their collaboration, and the field teams in each city.

References

[1] Vos T, Barber RM, Bell B, Bertozzi-Villa A, Biryukov S, Bolliger I, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet 2015;386:743–800.
[2] Marcus M, Yasamy M, van Ommeren M, Chisholm D, Saxena S. Depression: a global public health concern. WHO Department of Mental Health and Substance Abuse; 2012.
[3] Yanzon de la Torre A, Oliva N, Echevarrieta PL, Perez BG, Caporusso GB, Titaro AJ, et al. Major depression in hospitalized Argentine general medical patients: prevalence and risk factors. J Affect Disord 2016;197:36–42.
[4] Katon W, Ciechanowski P. Impact of major depression on chronic medical illness. J Psychosom Res 2002;53:859–63.
[5] Katon WJ. Epidemiology and treatment of depression in patients with chronic medical illness. Dialogues Clin Neurosci 2011;13:7–23.
[6] Olver JS, Hopwood MJ. Depression and physical illness. Med J Aust 2013;199:S9–12.
[7] Ludman EJ, Katon W, Russo J, Von Korff M, Simon G, Ciechanowski P, et al. Depression and diabetes symptom burden. Gen Hosp Psychiatry 2004;26:430–6.
[8] Katon W, Sullivan M, Walker E. Medical symptoms without identified pathology: relationship to psychiatric disorders, childhood and adult trauma, and personality traits. Ann Intern Med 2001;134:917–25.
[9] Katon W, Lin EH, Kroenke K. The association of depression and anxiety with medical symptom burden in patients with chronic medical illness. Gen Hosp Psychiatry 2007;29:147–55.
[10] Gureje O, Simon GE, Von Korff M. A cross-national study of the course of persistent pain in primary care. Pain 2001;92:195–200.
[11] Ciechanowski PS, Katon WJ, Russo JE. Depression and diabetes: impact of depressive symptoms on adherence, function, and costs. Arch Intern Med 2000;160:3278–85.
[12] DiMatteo MR, Lepper HS, Croghan TW. Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence. Arch Intern Med 2000;160:2101–7.
[13] Lin EH, Katon W, Von Korff M, Rutter C, Simon GE, Oliver M, et al. Relationship of depression and diabetes self-care, medication adherence, and preventive care. Diabetes Care 2004;27:2154–60.
[14] Saravay SM, Steinberg MD, Weinschel B, Pollack S, Alovis N. Psychological comorbidity and length of stay in the general hospital. Am J Psychiatry 1991;148:324–9.
[15] Katon WJ, Lin E, Russo J, Unutzer J. Increased medical costs of a population-based sample of depressed elderly patients. Arch Gen Psychiatry 2003;60:897–903.
[16] Unutzer J, Patrick DL, Simon G, Grembowski D, Walker E, Rutter C, et al. Depressive symptoms and the cost of health services in HMO patients aged 65 years and older. A 4-year prospective study. Jama 1997;277:1618–23.
[17] Sullivan M, Simon G, Spertus J, Russo J. Depression-related costs in heart failure care. Arch Intern Med 2002;162:1860–6.
[18] John-Baptiste AA, Li L, Isaranuwatchai W, Osuch E, Anderson KK. Healthcare utilization costs of emerging adults with mood and anxiety disorders in an early intervention treatment program compared to a matched cohort. Early Interv Psychiatry 2019:1–8.
[19] Chisholm D, Diehr P, Knapp M, Patrick D, Treglia M, Simon G. Depression status, medical comorbidity and resource costs. Evidence from an international study of major depression in primary care (LIDO). The British Journal of Psychiatry: The Journal of Mental Science 2003;183:121–31.
[20] Rentsch D, Dumont P, Borgacci S, Carballeira Y, deTonnac N, Archinard M, et al. Prevalence and treatment of depression in a hospital department of internal medicine. Gen Hosp Psychiatry 2007;29:25–31.
[21] Kessler RC, Wittchen HU, Abelson JM, McGonagle K, Schwarz N, Kendler KS, et al. Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US national comorbidity survey (NCS). Int J Methods Psychiatr Res 1998;7:33–55.
[22] Ferrando L, Bobes J, Gibert J, Soto M, Soto O. Mini-International Neuropsychiatric Interview. Versión en Español 5.0. 0. 2005.
[23] Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. Jama 1999;282:1737–44.
[24] Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16:606–13.
[25] Beck AT, Steer RA, Brown GK. BDI-II: Beck Depression Inventory: manual, svensk version. Psykologiförlaget; 2006.
[26] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67:361–70.
[27] WHO. International statistical classification of diseases and related health problems. World Health Organization; 2004.
[28] Butman J, Arizaga R, Harris P, Drake M, Baumann D, De Pascale A, et al. El “Mini Mental State Examination” en Español. Normas para Buenos Aires. Rev Neurol Arg 2001;26:11–5.
[29] Michal M, Wiltink J, Lackner K, Wild PS, Zwiener I, Blettner M, et al. Association of hypertension with depression in the community: results from the Gutenberg Health Study. J Hypertens 2013;31:893–9.
[30] van Dooren FE, Denollet J, Verhey FR, Stehouwer CD, Sep SJ, Henry RM, et al. Psychological and personality factors in type 2 diabetes mellitus, presenting the rationale and exploratory results from The Maastricht Study, a population-based cohort study. BMC Psychiatry 2016;16:17.
[31] Elperin DT, Pelter MA, Deamer RL, Burchette RJ. A large cohort study evaluating risk factors associated with uncontrolled hypertension. J Clin Hypertens (Greenwich) 2014;16:149–54.
[32] El-Den S, Chen TF, Gan YL, Wong E, O’Reilly CL. The psychometric properties of depression screening tools in primary healthcare settings: a systematic review. J Affect Disord 2018;225:503–22.
[33] First MB, Frances A, Pincus H. DSM-IV: Manual diagnóstico y estadístico de los trastornos mentales. Masson; 1995. p. 401–56.
[34] Mitchell AJ, Yadegarfar M, Gill J, Stubbs B. Case finding and screening clinical utility of the Patient Health Questionnaire (PHQ-9 and PHQ-2) for depression in primary care: a diagnostic meta-analysis of 40 studies. BJPsych Open 2016;2:127–38.
[35] Urtasun M, Daray FM, Teti GL, Coppolillo F, Herlax G, Saba G, et al. Validation and calibration of the patient health questionnaire (PHQ-9) in Argentina. BMC Psychiatry 2019;19:291.
[36] Brenlla M, Rodríguez C. Adaptación argentina del Inventario de Depresión de Beck (BDI-II). BDI-II Inventario de Depresión de Beck Segunda Edición Manual. Buenos Aires: Paidós; 2006. [Links].
[37] Herrero MJ, Blanch J, Peri JM, De Pablo J, Pintor L, Bulbena A. A validation study of the hospital anxiety and depression scale (HADS) in a Spanish population. Gen Hosp Psychiatry 2003;25:277–83.
[38] Gercovich D, Torrente F, López P, Bortolato D, Margiolakis P, Morgenfeld M, et al. Evaluación de propiedades psicométricas de la escala de ansiedad y depresión hospitalaria en pacientes oncológicos de Buenos Aires. Acta Psiquiatr Psicol Am Lat 2009;55:84–91.
[39] ZJFip D. Using Bayes to get the most out of non-significant results. 5. 2014. p. 781.
[40] Akobeng AKJA. Understanding diagnostic tests 3: Receiver operating characteristic

curves. 96. 2007. p. 644–7.
[41] Elhai JD, Contractor AA, Tamburrino M, Fine TH, Prescott MR, Shirley E, et al. The factor structure of major depression symptoms: a test of four competing models using the Patient Health Questionnaire-9. Psychiatry Res 2012;199:169–73.
[42] Tavakol M, Dennick RJI. Making sense of Cronbach's alpha. 2. 2011. p. 53.
[43] Löwe B, Spitzer RL, Williams JB, Mussell M, Schellberg D, Kroenke KJG. Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. 30. 2008. p. 191–9.
[44] Stafford L, Berk M, Jackson HJ. Validity of the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. Gen Hosp Psychiatry 2007;29:417–24.
[46] Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 2012;184:E191-6.
[47] Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. Gen Hosp Psychiatry 2015;37:567–76.
[48] Milette K, Hudson M, Baron M, Thombs BD. Comparison of the PHQ-9 and CES-D depression scales in systemic sclerosis: internal consistency reliability, convergent validity and clinical correlates. Rheumatology (Oxford) 2010;49:789–96.
[49] Diez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer RL. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. Psychosom Med 2001;63:679–86.
[50] Rogers WH, Adler DA, Bungay KM, Wilson IB. Depression screening instruments made good severity measures in a cross-sectional analysis. J Clin Epidemiol 2005;58:370–7.
[51] Kneipp SM, Kairalla JA, Stacciarini JM, Pereira D, Miller MD. Comparison of depressive symptom severity scores in low-income women. Nurs Res 2010;59:380–8.
[52] Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. J Gen Intern Med 2006;21:547–52.
[53] Hepner KA, Hunter SB, Edelen MO, Zhou AJ, Watkins K. A comparison of two depressive symptomatology measures in residential substance abuse treatment clients. J Subst Abuse Treat 2009;37:318–25.
[54] Lai BP, Tang AK, Lee DT, Yip AS, Chung TK. Detecting postnatal depression in Chinese men: a comparison of three instruments. Psychiatry Res 2010;180:80–5.
[55] Streiner DL, Cairney J. What's under the ROC? An introduction to receiver operating characteristics curves. Canadian Journal of Psychiatry Revue Canadienne de Psychiatrie 2007;52:121–8.
[56] Arrieta J, Aguerrebere M, Raviola G, Flores H, Elliott P, Espinosa A, et al. Validity and utility of the patient health questionnaire (PHQ)-2 and PHQ-9 for screening and diagnosis of depression in rural Chiapas, Mexico: a cross-sectional study. J Clin Psychol 2017;73:1076–90.
[57] Guo B, Kaylor-Hughes C, Garland A, Nixon N, Sweeney T, Simpson S, et al. Factor structure and longitudinal measurement invariance of PHQ-9 for specialist mental health care patients with persistent major depressive disorder: Exploratory Structural Equation Modelling. J Affect Disord 2017;219:1–8.
[58] Hanlon C, Medhin G, Selamu M, Breuer E, Worku B, Hailemariam M, et al. Validity of brief screening questionnaires to detect depression in primary care in Ethiopia. J Affect Disord 2015;186:32–9.
[59] Petersen JJ, Paulitsch MA, Hartig J, Mergenthal K, Gerlach FM, Gensichen J. Factor structure and measurement invariance of the Patient Health Questionnaire-9 for female and male primary care patients with major depression in Germany. J Affect Disord 2015;170:138–42.
[60] Miranda CAC, Scoppetta O. Factorial structure of the Patient Health Questionnaire-9 as a depression screening instrument for university students in Cartagena, Colombia. Psychiatry Res 2018;269:425–9.
[61] Bambauer KZ, Locke SE, Aupont O, Mullan MG, McLaughlin TJ. Using the Hospital Anxiety and Depression Scale to screen for depression in cardiac patients. Gen Hosp Psychiatry 2005;27:275–84.
[62] Chilcot J, Hudson JL, Moss-Morris R, Carroll A, Game D, Simpson A, et al. Screening for psychological distress using the Patient Health Questionnaire Anxiety and Depression Scale (PHQ-ADS): initial validation of structural validity in dialysis patients. Gen Hosp Psychiatry 2018;50:15–9.
[63] Levis B, Benedetti A, Riehm KE, Saadat N, Levis AW, Azar M, et al. Probability of major depression diagnostic classification using semi-structured versus fully structured diagnostic interviews. The British Journal of Psychiatry: The Journal of Mental Science 2018;212:377–85.
[64] Jeffreys H. The theory of probability. Oxford: OUP; 1998.