# Psychometric Properties of the Patient Health Questionnaire-9 in Elderly Chilean Primary Care Users

Joseph Aslan[1], Félix Cova[2], Sandra Saldivia[3]*, Claudio Bustos[2], Carolina Inostroza[2], Paulina Rincón[2], Camila Ortiz[4] and Vasily Bühring[5]

[1] Doctoral Program in Psychology, Universidad de Concepción, Concepción, Chile, [2] Department of Psychology, Faculty of Social Sciences, Universidad de Concepción, Concepción, Chile, [3] Department of Psychiatry and Mental Health, Faculty of Medicine, Universidad de Concepción, Concepción, Chile, [4] Master Program in Politics and Government, Universidad de Concepción, Concepción, Chile, [5] Master Program in Psychology, Universidad de Concepción, Concepción, Chile

**Background:** This study aimed to assess the measurement properties (reliability, factor structure, and criterion validity) of the Patient Health Questionnaire (PHQ-9) as an instrument for screening major depressive disorder (MDD) in elderly primary care users in Chile.

**Method:** About 582 participants aged between 65 and 80 years were enrolled from primary care centers. They completed the Composite International Diagnostic Interview (CIDI), a survey with sociodemographic characteristics and the PHQ-9.

**Results:** The PHQ-9 revealed an acceptable internal consistency ($\omega = 0.79$ [95% CI: 0.75–0.80] and $\alpha = 0.78$ [95% CI: 0.75–0.81]); confirmatory factor analysis demonstrated a good fit for both 1- and 2-factor solutions. The chi-square difference test ($\chi^2 = 0.61$, $gl = 1$, $p = 0.43$) and correlation between the somatic and the cognitive-effective latent factors were very high ($r = 0.97$, $p < 0.001$), indicating that the 1 factor model was more parsimonious. Utilizing the CIDI as the gold standard, the area under the curve (AUC) was 0.88 (SE = 0.04, 95% CI: 0.84–0.90). The optimal cut-off score of $\geq 6$ yielded good sensitivity and specificity for detecting MDD (0.95 and 0.76, respectively). However, considering the clinical utility index, the cut-off score of $\geq 9$ proved to be a more effective marker for discarding cases of MDD.

**Conclusion:** The PHQ-9 has adequate psychometric properties for elderly primary care users. In clinical settings, it showed its greatest utility in ruling out the presence of an MDD, however, its clinical value for identifying possible cases of MDD is limited. In cases above the cut-off point, it is recommended to perform a more thorough evaluation.

Keywords: depression, patient health questionnaire, primary health care, sensitivity, specificity

## INTRODUCTION

Late-life depression is one of the most frequent mental health disorders in older adults and is associated with poorer physical health, more limited social functioning, increased suicide risk, and higher overall mortality. It is also associated with other mental health problems and disorders such as anxiety disorders, substance and medication abuse as well as cognitive deficits (1). Unfortunately,

it is frequently under-recognized, mainly because its clinical indicators, such as depression, sadness, tiredness, anhedonia, sleep disturbances, hypochondriasis, subjective cognitive complaints (e.g., poor concentration and memory) are confounded with normal expressions related to the aging process (2).

Preventing, identifying, and treating depression in a timely and adequate manner in older adults are relevant health objectives. Primary health care (PHC) plays an important role in the early detection and treatment of depressive disorders. Chilean national studies report a 24.4% annual prevalence for common mental disorders in PHC center users in the south-central area of the country, where major depressive disorder (MDD) was present in 13.4% of users (3). However, few studies estimate the prevalence of depression in older adults in PHC and, although this would be somewhat lower than in other age groups, it is still high (4).

Although clinical interviews are indispensable for proper diagnosis and evaluation of the symptoms associated with depressive disorders, simple and brief questionnaires are beneficial for screening and monitoring the trajectory of symptoms over time (5). Screening instruments need to demonstrate evidence of their reliability and validity in various contexts of use, without which their scores are unintelligible (6). The Patient Health Questionnaire (7) (PHQ-9) is one of the most popular depression measures internationally, in clinical and population based settings (8). Despite its brevity it has proven its ability to identify symptoms of depressive disorders, high screening capacity, and sensitivity to change in monitoring the treatment response (9, 10).

Various studies have analyzed the psychometric properties of the PHQ-9 in the elderly. There are studies in the United States (11, 12), China (13–15), Taiwan (16), United Kingdom (17, 18), Germany (19), Netherlands (20), Australia (21), and Brazil (22).

The PHQ-9 uses as an original recommendation—a cut-off score of 10—to detect a depressive disorder in the general population. Nevertheless, recent evidence regarding the optimal cut-off score for screening depression found that scores between 8 and 11 had satisfactory properties (8). In Chile, the PHQ-9 has been used in PHC contexts, showing evidence of good reliability and screening accuracy, with the original and lower cut-off scores (23, 24). Some authors argue that in the elderly, a lower cut-off score in the PHQ-9 would augment both sensitivity and specificity (20).

Concerning the structural validity of the PHQ-9, there is evidence corroborating the unidimensional structure of the scale (11, 24–30). However, some evidence has been found supporting a two-dimensional solution (16, 18, 19, 31). In these studies, one factor of the PHQ-9 represents non-somatic (cognitive and affective) depression symptoms, and another the somatic symptoms.

To our knowledge, the PHQ-9 screening characteristics have not yet been validated with the Chilean elderly in PHC. The objective of this study is to analyse the psychometric properties of the PHQ-9 in a Chilean elderly sample, in a PHC setting. Specifically, this investigation focuses on reliability among the PHQ-9 items (internal consistency), the dimensionality of the

test (structural validity), and its convergence with a reference to a gold standard in identifying elderly patients with MDD (criterion validity).

## METHODS

## Subjects and Procedure

This study adopted a descriptive, cross-sectional design. The study subject data were obtained to perform a sub analysis of the baseline sample of a larger study (32). The sample was recruited from July to August 2018, consisting of self-dependent men and women between the ages of 65 and 80, attending 15 PHC centers in two catchment areas in the central-southern part of Chile (Concepción and Talcahuano). The potential participants ($N = 1,220$) were contacted and invited during a home visit; 582 accepted (a 52.3% rejection rate), and an interview was coordinated for data collection. Those who agreed to participate provided written informed consent.

The inclusion criteria of our study were as follows: (1) aged between 65 and 80, (2) self-dependent and (3) user of a PHC center. Exclusion criteria were the following: (1) presence of a severe mental disorder (psychosis or bipolar disorder), (2) mental disability or dementia or (3) a disability preventing communication. The assessment of these criteria was made from the health records using a standardized instrument (EMPAM) (33).

Data collection was done over a 1-month period, through 40-min-long face-to-face interviews by trained research assistants at home visits. The survey included sociodemographic characteristics, the PHQ-9, the Composite International Diagnostic Interview (CIDI) and other measures not related to this study. The participants fulfilling the diagnostic criteria for MDD were referred to their PHC center.

## MEASURES

## Patient Health Questionnaire (PHQ-9)

The PHQ-9 is a self-questionnaire consisting of nine items that assess the presence and severity of depressive symptoms based on the DSM-IV criteria for MDD (34). It refers to symptoms experienced by patients during the 2 weeks preceding to the interview. In this study, the Spanish version of the scale was employed (35). The PHQ-9 scoring consists of a Likert scale: 0 (not at all), 1 (several days), 2 (more than half the days), and 3 (nearly every day). Total scores range from 0 to 27. The severity of symptoms can be organized into four categories: 0–4 (minimum), 5–9 (mild) 10–14 (moderate), 15–19 (moderate to severe), 20–27 (serious) (7). The PHQ-9 was developed as a screening tool, with recommended cut-off scores between 8 and 11 for a probable case of MDD (8).

## Composite International Diagnostic Interview (CIDI)

The CIDI (36) is a structured diagnostic interview developed by the World Health Organization (WHO). It has been used in epidemiological studies in the general population, with high

inter-rater and test-retest reliability, and is a valid gold standard for MDD diagnosis (37, 38), with evidence of validity in multiple international studies, including Chile (4, 39, 40). The CIDI can be administered by lay interviewers, thus overcoming the limitations of interviews conducted only by professionals, maintaining the objective that the diagnoses strictly adhere to the established diagnostic criteria (41). CIDI 2.1 delivers diagnoses following the DSM-IV and ICD-10 criteria, for a life-long disorder, for last 12 months, and for the last 30 days (36, 42). For this study, the diagnoses were used according to DSM-IV criteria and their presence was evaluated during the last 6 months. Sections A (sociodemographic data) and E (depression) were used in this study.

## Data Analysis

For descriptive data and internal consistency analyses (measured by Cronbach's alpha and McDonald's omega coefficient) JASP software (43) was utilized. For criterion validity, we assessed the PHQ-9 performance in comparison to the gold standard using MedCalc 14.8 (44). The CIDI, which is used for the diagnosis of an MDD, was used as a criterion standard. The receiver operating characteristic curve (ROC) and area under the curve (AUC) were constructed against the presence of MDD by the CIDI. The following test characteristics of the PHQ-9 were compared with the CIDI: sensitivity, specificity, predictive values, likelihood ratios, clinical utility index, and the Youden J index. For the confirmatory factorial analysis (CFA), Mplus 8.4 software (45) was used. A unifactorial and bifactorial model were postulated, using the robust weighted least squares estimator (WLSMV), which does not assume the normality of the variables and is considered the best option to model categorical or ordinal data (46). A chi-square difference test was performed to determine which of the models has a best fit to the data using the DIFFTEST command.

## RESULTS

### Participant Characteristics

We excluded from the study 5 of the 582 respondents due to incomplete data. The remaining 577 (99.14%) cases were included in the analyses. The mean age of the 577 participants was 71.77 years (SD = 4.13). **Table 1** ilustrates the demographic characteristics of the sample. There were 374 women (64.8%) and 203 men (35.2%); 302 respondents were married (52.3%), 213 were bereaved or divorced (36.9%), and 62 (10.7%) were never married; 114 (19.9%) of the participants had worked in the last 12 months, 84 (14.6%) lived alone and 366 (63.7%) had <12 years of education.

The mean score of the 577 participants was 3.99 (SD = 4.31), with a range of 0 to 24. The median score was 3, with a skewness of 1.51 (SE = 0.10) and kurtosis of 2.20 (SE = 0.20). The distribution of the PHQ-9 scores is shown in **Table 2**. Most of the participants (87.9%) had a PHQ-9 score under the most common cut-off score (< 10), and only 4.3% had a score of 15 or higher.

**TABLE 1 |** Socio-Demographic characteristics of the sample (N = 577).

| | | N | % |
|---|---|---|---|
| **Demographics** | | | |
| Gender | Men | 203 | 35.0 |
| | Women | 374 | 65.0 |
| Age Group | 64–69 | 206 | 36.0 |
| | 70–79 | 349 | 61.0 |
| | 80+ | 22 | 4.0 |
| Marital Status | Married | 302 | 52.3 |
| | Widow | 138 | 23.9 |
| | Divorced | 75 | 13.0 |
| | Never Married | 62 | 10.7 |
| Studies | Basic Education or incomplete secondary education | 366 | 63.0 |
| | Complete secondary education | 104 | 18.0 |
| | Higher Education | 105 | 18.0 |
| | Missing Data | 2 | 0.3 |
| Employment Situation | Full time work | 32 | 5.5 |
| | Part time work | 66 | 11.4 |
| | No work | 479 | 83.0 |

**TABLE 2 |** Distribution of PHQ-9 scores and 6-CIDI MDD Diagnosis (n =577).

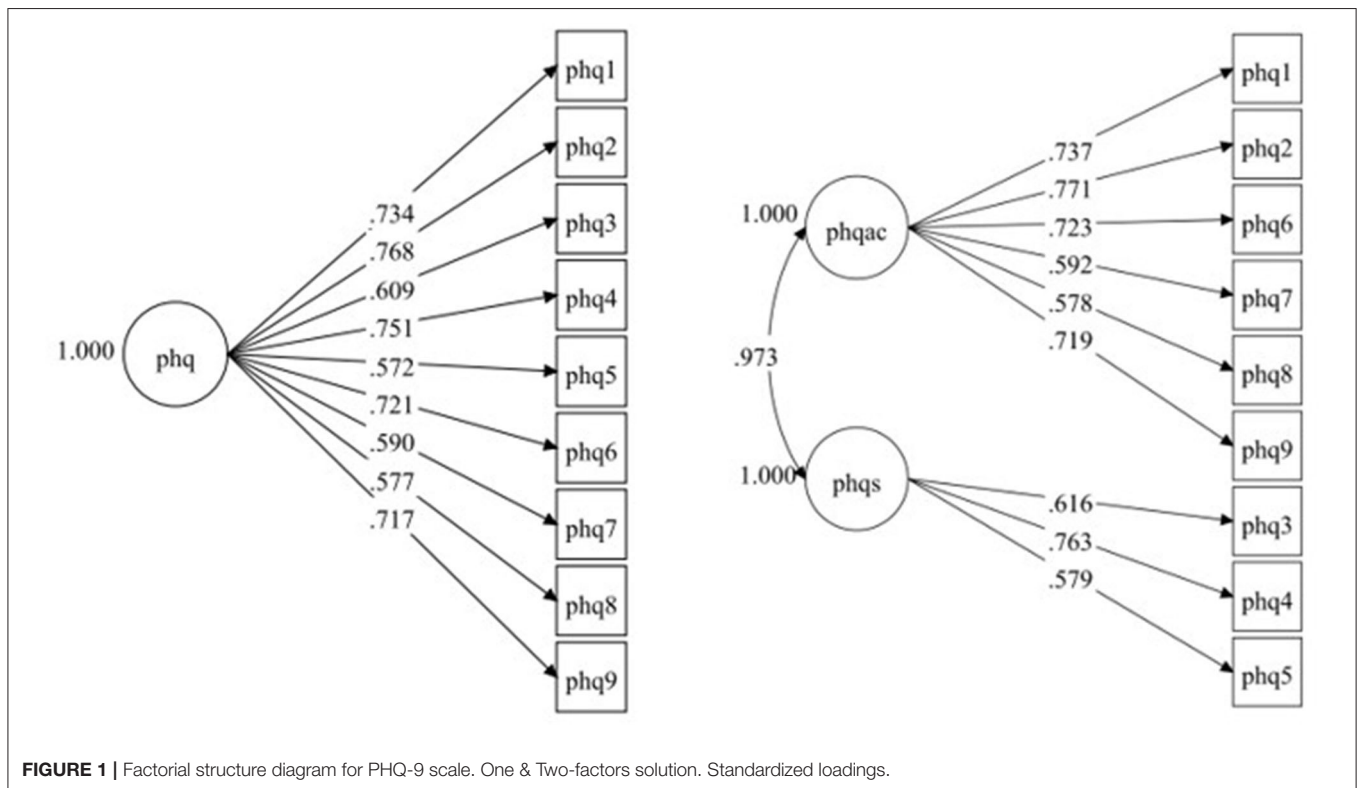| | MDD Diagnosis (Last 6 months) | | | |
|---|---|---|---|---|
| Depression level (PHQ-9) | CIDI negative | CIDI positive | Total | % |
| Minimum (0–4) | 385 | 1 | 386 | 66.9 |
| Mild (5–9) | 113 | 8 | 121 | 21.0 |
| Moderate (10–14) | 41 | 4 | 45 | 7.8 |
| Moderate to severe (15–19) | 15 | 8 | 23 | 4.0 |
| Severe (20–23, 25–28) | 2 | 0 | 2 | 0.3 |
| Total | 556 | 21 | 577 | 100.0 |

MDD, Major depressive disorder; PHQ-9, Patient Health Questionnaire-9; CIDI, Composite international Diagnostic Interview.

### Item Analysis and Reliability

Mean scores for all PHQ-9 items are depicted in **Table 3**. The two items reported most frequently were sleep problems and low energy. The least-recognized item was suicidal ideation, which also had the least corrected item-total correlation (0.35). Corrected item-total correlation was in the range of 0.35–0.59. As indicators of internal consistency, both McDonald's coefficient (ω), more suitable for scales with few response options, and Cronbach's coefficient (α), were calculated. Both indicators were acceptable: ω = 0.79 [95% CI: 0.75–0.80] and α = 0.78 [95% CI: 0.75–0.81], respectively. All items, if deleted, would decrease both coefficients for the total scale.

**TABLE 3 |** PHQ-9 item level values and item-total correlations ($n = 577$).

| Questionnaire Item | Mean | SD | Item-total correlation | If item dropped | |
|---|---|---|---|---|---|
| | | | | McDonald's $\omega$ | Cronbach's $\alpha$ |
| 1. Little interest or pleasure in doing things | 0.37 | 0.72 | 0.52 | 0.77 | 0.75 |
| 2. Feeling down, depressed, or hopeless | 0.55 | 0.85 | 0.59 | 0.76 | 0.74 |
| 3. Trouble falling or staying asleep, or sleeping too much | 0.89 | 1.11 | 0.48 | 0.78 | 0.76 |
| 4. Feeling tired or having little energy | 0.77 | 0.95 | 0.59 | 0.75 | 0.74 |
| 5. Poor appetite or overeating | 0.52 | 0.90 | 0.42 | 0.78 | 0.77 |
| 6. Feeling bad about yourself or that you are a failure | 0.28 | 0.67 | 0.51 | 0.77 | 0.76 |
| 7. Trouble concentrating on things | 0.31 | 0.72 | 0.41 | 0.78 | 0.77 |
| 8. Moving or speaking so slowly that other people could have noticed | 0.26 | 0.67 | 0.41 | 0.78 | 0.77 |
| 9. Thoughts that you would be better off dead or hurting yourself | 0.06 | 0.31 | 0.35 | 0.79 | 0.78 |



**FIGURE 1 |** Factorial structure diagram for PHQ-9 scale. One & Two-factors solution. Standardized loadings.
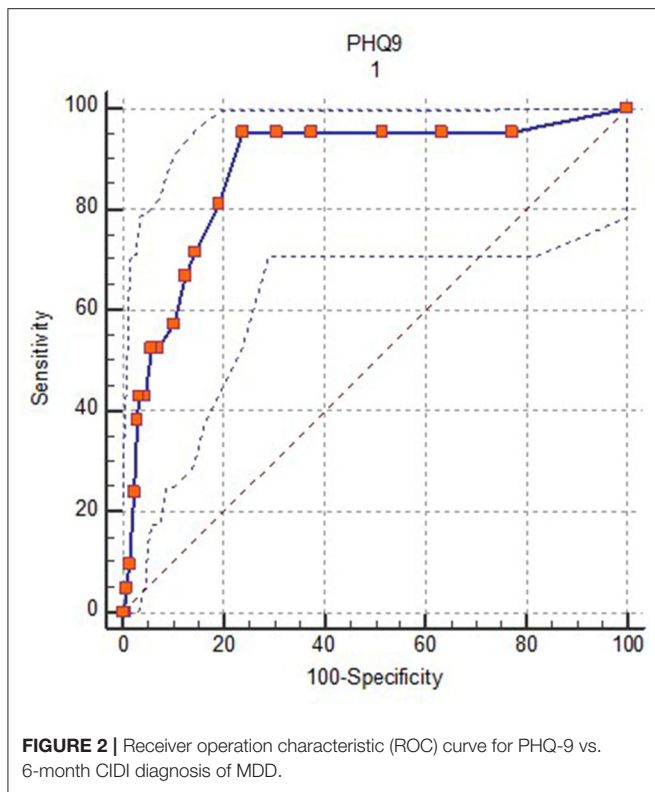
## Structural Validity

One- and two-factor models of the PHQ-9 were assessed using CFA; the 1-factor solution indicated a good fit of the model to the data, ($\chi^2 = 51.91$, $gl = 27$, $p < 0.001$; CFI = 0.99; TLI = 0.98; RMSEA = 0.04 [90% I.C. 0.02–0.06]); the 2-factor solution with three somatic items and six cognitive-affective items also indicated a good fit ($\chi^2 = 51.12$, $gl = 26$, $p < 0.001$; CFI = 0.99; TLI = 0.98; RMSEA = 0.04 [90% I.C. 0.02–0.06]). In both models, although the chi-square value was significant, given the sample size, it is an expected result; CFI, TLI, and RMSEA values are within the recommended standards (i.e., CFI > 0.95, TLI > 0.95, RMSEA < 0.08) (47). A chi-square difference test was also performed to determine which of the two models has the best fit to the data. The nested comparison shows that the

difference between both models was not significant ($\chi^2 = 0.61$, $gl = 1$, $p = 0.43$), suggesting that the more parsimonious 1 factor solution should be retained. **Figure 1** shows the PHQ-9 factor structure for the 1-factor solution, where all the factor loadings were significant ($p < 0.001$) and >0.572. **Figure 1** also reveals the PHQ-9 factor structure for the 2-factor solution, where all factor loadings were also significant ($p < 0.001$) and >0.578. The correlation between the somatic and the cognitive-affective latent factors was very high ($r = 0.973$, $p < 0.001$), suggesting that both factors would be overlapping.

## Criterion Validity

According to the CIDI, 21 patients (3.6%) met the criteria for diagnosis of MDD in the last 6 months. Significant

**FIGURE 2 |** Receiver operation characteristic (ROC) curve for PHQ-9 vs. 6-month CIDI diagnosis of MDD.

differences were observed in the average PHQ-9 scores between the participants with and without diagnosis ($M = 11.19$, $DT = 4.89$; $M = 3.72$, $DT = 4.05$, $t$ $(575) = 8.22$, $p < 0.001$, $r = 0.32$). **Figure 2** illustrates the ROC curve showing PHQ-9 performance in the identification of patients with MDD. The AUC was 0.88 (SE = 0.04, 95% CI: 0.85–0.90), which accounts for moderate accuracy [48].

**Table 4** shows the sensitivity (Se), specificity (Sp), Youden J index, positive predictive value (PPV), negative predictive value (NPV), clinical utility index (CUI) and the likelihood ratio (LR) of different cut-off PHQ-9 scores in the diagnosis of MDD. The six-point cut-off score maximizes sensitivity and specificity values based on the Youden J index (sensitivity of 0.95 and specificity of 0.76). With this cut-off score, the PPV is 0.13 and the NPV is 0.99, and the positive and negative likelihood ratios are 3.95 and 0.06, respectively. Some authors [49–51] have pointed out limitations of the usual indicators used to estimate cut-off scores and have suggested alternatives such as the clinical utility index, which combines measures of occurrence and discrimination [52]. This clinically relevant rule in accuracy is estimated by the clinical utility index positive (Se × PPV), and rule out accuracy is estimated through the clinical utility index negative (Sp × NPV). These measure the clinical value of a diagnostic test when applied in a specific setting, and can be graded using the following scale: < 0.5 poor, ≥ 0.5 < 0.64 fair, ≥ 0.64 < 0.81 good and ≥ 0.81 ≤ 1 excellent.

## DISCUSSION

In this study, the psychometric properties of the PHQ-9 and their usefulness as a depressive disorder screening instrument are analyzed in a sample of older Chilean adults. Although this is not the only study of the functioning of the PHQ-9 that has been organized in Chile [23, 24], to our knowledge, it is the first study targeted at the older adult population, using a large sample of older adults who attend PHC centers. In addition, it has the advantage of having used the CIDI as a gold standard, which is an instrument widely applied in epidemiological studies due to its rigorous use of diagnostic criteria for established mental disorders.

The results show an adequate psychometric behavior of the PHQ-9 in the sample studied. Although the one dimension and two-dimensional solutions showed good a fit and high factor loadings, the scale appears to be one-dimensional given that the result of the chi-square difference test and the correlation between the non-somatic and somatic items was very high ($r = 0.97$), suggesting an overlap and supporting the notion of depression being a coherent unidimensional construct [11, 19]. All items demonstrated loads >0.57, and corrected item-total correlations were >0.35. The internal consistency observed was also acceptable.

The PHQ-9 also reflected adequate sensibility and specificity values, comparable to those obtained in other studies [13, 16, 20]. As is common with screening instruments used in contexts with low prevalence rates of the disorder (3.6% in this study), the best outcome of the PHQ-9 was in ruling out the presence of a depressive disorder rather than to positively confirm its presence. The PPV and NPV of scores between 6 and 11 are in the range of 0.13–0.22 and 0.99–0.98, respectively. These results suggest that in the population studied, a PHQ-9 high score should not be considered an indicator of probable a positive diagnosis but as an indication of the suitability of a new and more thorough evaluation; conversely, a low score confirms that the presence of an MDD is unlikely.

The cut-off score must be based on the objectives of those who use the instrument and the need to maximize detection or reduce the number of false positives [53, 54]. Various indicators are considered for these purposes. One of the most used is the Youden index, which seeks to estimate the score that best combines sensitivity and specificity. In this study, the highest Youden J value is for the 6 cut-off score, which is in the lower range compared to other studies in elderly people [10, 13, 14, 17, 22]. Nonetheless, some studies have presented similar results in elderly populations [16, 20, 21]. It is not clear to us what to attribute this result to. Most of the studies mentioned above have been done with more specific samples of chronic patients than in this study, and with a higher prevalence of depressive disorder; these could be influencing factors. Cultural aspects may also play a relevant role, but it is unclear to which degree they affect this result. It should be considered that the Youden index is not necessarily the best criterion to select the cut-off point, since it does not consider the effect of prevalence and, therefore, the predictive value of the scores [52]. Considering the clinical utility index, which combines measures of occurrence

**TABLE 4 |** Performance of PHQ-9 cut-off scores for detecting major depression.

| Cut-Off score | Sensitivity | Specificity | Youden J index | PPV | NPV | CUI+ | CUI- | LR+ | LR– |
|---|---|---|---|---|---|---|---|---|---|
| ≥ 6 | 0.95 | 0.76 | 0.71 | 0.13 | 0.99 | 0.12 | 0.75 | 3.95 | 0.06 |
| ≥ 7 | 0.81 | 0.81 | 0.62 | 0.14 | 0.99 | 0.11 | 0.80 | 4.25 | 0.24 |
| ≥ 8 | 0.71 | 0.85 | 0.56 | 0.16 | 0.99 | 0.11 | 0.84 | 4.90 | 0.33 |
| ≥ 9 | 0.67 | 0.87 | 0.54 | 0.17 | 0.99 | 0.11 | 0.86 | 5.3 | 0.38 |
| ≥ 10 | 0.57 | 0.90 | 0.47 | 0.17 | 0.98 | 0.10 | 0.88 | 5.48 | 0.48 |
| ≥ 11 | 0.52 | 0.93 | 0.45 | 0.22 | 0.98 | 0.11 | 0.91 | 7.47 | 0.51 |
| ≥ 12 | 0.52 | 0.94 | 0.46 | 0.26 | 0.98 | 0.14 | 0.92 | 9.10 | 0.51 |
| ≥ 13 | 0.43 | 0.95 | 0.38 | 0.26 | 0.98 | 0.11 | 0.93 | 9.16 | 0.6 |
| ≥ 14 | 0.43 | 0.97 | 0.40 | 0.32 | 0.98 | 0.14 | 0.95 | 12.54 | 0.59 |
| ≥ 15 | 0.38 | 0.97 | 0.35 | 0.32 | 0.98 | 0.12 | 0.95 | 12.46 | 0.64 |

*PPV, Positive predictive value; NPV, Negative predictive value; CUI+, Clinical utility index positive; CUI–, Clinical utility index negative; LR+, Likelihood ratio positive; LR–, Likelihood ratio negative.*

(Se/Sp) and discrimination (PPV/NPV), the best cut-off score should be higher. The results of this study suggest that the clinical utility index+ (adequate identification of cases) of the PHQ-9 in this simple is very low; in contrast, the clinical utility index- (discarding non-cases) is excellent, especially from 8 points onwards.

This study has some limitations. First, the performance of the PHQ-9 was limited by the low prevalence rate of MDD in the target population. However, it was slightly higher than that observed in the only Chilean epidemiological study of national scope, where the 12-month prevalence of MDD in the general population was 2.9% in those over 65 years of age, 1.6% among those of 65–75 and higher than 5.2% after 75 (4). The lower prevalence a of depressive disorder in older people, compared to that observed in other ages, has been confirmed in various studies (55). However, this does not diminish its clinical relevance, but, on the contrary, accentuates the challenge of its adequate detection and treatment. Second, the analysis in this study was done by contrasting two specific measures over time (PHQ-9 and CIDI) and assuming that the use of the CIDI as a gold standard and the DSM-IV diagnostic criteria in older adults results in the detection of "authentic" depressive disorders. This is questionable. The DSM-IV criteria may overestimate if they are used out of context (56, 57). On the other hand, the DSM-IV criteria (and those of the current DSM-V) do not facilitate the recognition of the specificities that depression can manifest in elderly people. There is some debate about the PHQ-9's applicability in persons suffering from multimorbidity, such as the elderly, where the somatic symptoms (fatigue, sleep disturbance, and poor appetite) may also be attributable to pain or another disease, and as a result may produce an overestimation of depression because of symptom overlap (58). The differential diagnosis between medical comorbidities and depressive symptoms was not considered in this study.

In summary, the results obtained indicate that the PHQ-9 is an instrument with adequate psychometric properties in elderly PHC users. In clinical settings, it should be considered that its greatest utility in this study population was to rule out the presence of depressive disorders rather than to identify possible cases. In people with scores above the cut-off point, it is essential to perform a new and more thorough evaluation.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee of Health Services of Concepción and Talcahuano. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

SS, JA, FC, CI, and PR prepared the study design. SS, FC, JA, and VB were involved in the selection of measurements. JA and CB prepared the data set, performed statistical analysis, and prepared the tables. CO polished and checked the manuscript. All authors have approved the final version of this manuscript.

## FUNDING

# REFERENCES

1. Power C, Greene E, Lawlor BA. Depression in late life: etiology, presentation, and management. In: H Chiu, K. Shulman, editors. *Mental Health and Illness of the Elderly. Mental Health and Illness Worldwide.* Singapore: Springer (2016). doi: 10.1007/978-981-10-0370-7_10-1

2. Hegeman JM, Kok RM, Van Der Mast RC, Giltay EJ. Phenomenology of depression in older compared with younger adults: meta-analysis. *Br J Psychiatr.* (2012) 200:275–81. doi: 10.1192/bjp.bp.111.095950

3. Saldivia S. *Prevalencia y Variables Asociadas a Trastornos Mentales Comunes En Centros De Atención Primaria De La Provincia De Concepción [Tesis de Magíster]*, Universidad de Chile] (2016).

4. Kohn R, Vicente B, Saldivia S, Rioseco P, Torres S. Psychiatric epidemiology of the elderly population in Chile. *Am J Geriatr Psychiatr.* (2008) 16:1020–8. doi: 10.1097/JGP.0b013e31818a0e1c

5. Schantz K, Reighard C, Aikens JE, Aruquipa A, Pinto B, Valverde H, et al. Screening for depression in Andean Latin America: factor structure and reliability of the CES-D short form and the PHQ-8 among Bolivian public hospital patients. *Int J Psychiatr Med.* (2017) 52:315–27. doi: 10.1177/0091217417738934

6. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use.* Oxford University Press (2015). Available online at: https://books.google.es/books?id=JH3OBAAAQBAJ

7. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J General Int Med.* (2001) 16:606–13. doi: 10.1046/j.1525-1497.2001.016009606.x

8. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *Can Med Association J.* (2012) 184:E191–6. doi: 10.1503/cmaj.110829

9. Löwe B, Schenkel I, Carney-Doebbeling C, Göbel C. Responsiveness of the PHQ-9 to psychopharmacological depression treatment. *Psychosomatics.* (2006) 47:62–7. doi: 10.1176/appi.psy.47.1.62

10. Löwe B, Spitzer RL, Gräfe K, Kroenke K, Quenter A, Zipfel S, et al. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians diagnoses. *J Affective Disord.* (2004) 78:131–40. doi: 10.1016/S0165-0327(02)00237-9

11. Bélanger E, Thomas KS, Jones RN, Epstein-Lubow G, Mor V. Measurement validity of the Patient-Health Questionnaire-9 in US nursing home residents. *Int J Geriatr Psychiatr.* (2019) 34:700–8. doi: 10.1002/gps.5074

12. Bernd L, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring Depression Treatment Outcomes with the Patient Health Questionnaire-9. *Med Care.* (2004) 42:1194–201. doi: 10.1097/00005650-200412000-00006

13. Chen S, Chiu H, Xu B, Ma Y, Jin T, Wu M, et al. Reliability and validity of the PHQ-9 for screening late-life depression in Chinese primary care. *Int J Geriatr Psychiatr.* (2010) 25:1127–33. doi: 10.1002/gps.2442

14. Liu Z-W, Yu Y, Hu M, Liu H-M, Zhou L, Xiao S-Y. PHQ-9 and PHQ-2 for screening depression in chinese rural elderly. *PLoS ONE.* (2016) 11:e0151042. doi: 10.1371/journal.pone.0151042

15. Zhang H, Wang S, Wang L, Yi X, Jia X, Jia C. Comparison of the geriatric depression scale-15 and the patient health questionnaire-9 for screening depression in older adults. *Geriatr Gerontol Int.* (2019) 20:138–43. doi: 10.1111/ggi.13840

16. Chen IP, Liu S-I, Huang H-C, Sun F-J, Huang C-R, Sung M-R, et al. Validation of the patient health questionnaire for depression screening among the elderly patients in Taiwan. *Int J Gerontol.* (2016) 10:193–7. doi: 10.1016/j.ijge.2016.05.002

17. Chagas MHN, Tumas V, Rodrigues GR, Machado-de-Sousa JP, Filho AS, Hallak JEC, et al. Validation and internal consistency of Patient Health Questionnaire-9 for major depression in Parkinson's disease. *Age Ageing.* (2013) 42:645–9. doi: 10.1093/ageing/aft065

18. Chilcot J, Rayner L, Lee W, Price A, Goodwin L, Monroe B, et al. The factor structure of the PHQ-9 in palliative care. *J Psychosom Res.* (2013) 75:60–4. doi: 10.1016/j.jpsychores.2012.12.012

19. Forkmann T, Gauggel S, Spangenberg L, Brähler E, Glaesmer H. Dimensional assessment of depressive severity in the elderly general population: psychometric evaluation of the PHQ-9 using Rasch Analysis. *J Affect Disord.* (2013) 148:323–30. doi: 10.1016/j.jad.2012.12.019

20. Lamers F, Jonkers CCM, Bosma H, Penninx BWJH, Knottnerus JA, van Eijk JTM. Summed score of the patient health questionnaire-9 was a reliable and valid method for depression screening in chronically ill elderly patients. *J Clin Epidemiol.* (2008) 61:679–87. doi: 10.1016/j.jclinepi.2007.07.018

21. 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. *General Hospital Psychiatr.* (2007) 29:417–24. doi: 10.1016/j.genhosppsych.2007.06.005

22. Costa MV, Diniz MF, Nascimento KK, Pereira KS, Dias NS, Malloy-Diniz LF, et al. Accuracy of three depression screening scales to diagnose major depressive episodes in older adults without neurocognitive disorders. *Brazil J Psychiatr.* (2016) 38:154–6. doi: 10.1590/1516-4446-2015-1818

23. Baader T, Molina JL, Venezian S, Rojas C, Farías R, Fierro-Freixenet C, et al. Validación y utilidad de la encuesta PHQ-9 (Patient Health Questionnaire) en el diagnóstico de depresión en pacientes usuarios de atención primaria en Chile. *Revista chilena de neuro-psiquiatría.* (2012) 50:10–22. doi: 10.4067/S0717-92272012000100002

24. Saldivia S, Aslan J, Cova F, Vicente B, Inostroza C, Rincón P. Propiedades psicométricas del PHQ-9 (Patient Health Questionnaire) en centros de atención primaria de Chile. *Rev Med Chile.* (2019) 147:53–60. doi: 10.4067/S0034-98872019000100053

25. Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *Br J General Pract.* (2008) 58:32–6. doi: 10.3399/bjgp08X263794

26. Hansson M, Chotai J, Nordström A, Bodlund O. Comparison of two self-rating scales to detect depression: HADS and PHQ-9. *Br J General Pract.* (2009) 59:e283–e288. doi: 10.3399/bjgp09X454070

27. 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 General Int Med.* (2006) 21:547–52. doi: 10.1111/j.1525-1497.2006.00409.x

28. Merz EL, Malcarne VL, Roesch SC, Riley N, Sadler GR. A multigroup confirmatory factor analysis of the Patient Health Questionnaire-9 among english- and spanish-speaking Latinas. *Cultural Diversity Ethnic Minority Psychol.* (2011) 17:309–16. doi: 10.1037/a0023883

29. Mewes R, Christ O, Rief W, Braehler E, Martin A, and Glaesmer H. Are depression and somatisation equivalent for migrants and native Germans? An investigation of measurement invariance for the PHQ-9 and PHQ-15. *Diagnostica.* (2010) 56:230–9. doi: 10.1026/0012-1924/a000026

30. Yu X, Tam WWS, Wong PTK, Lam TH, Stewart SM. The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. *Comprehens Psychiatr.* (2012) 53:95–102. doi: 10.1016/j.comppsych.2010.11.002

31. Krause JS, Reed KS, McArdle JJ. Prediction of somatic and non-somatic depressive symptoms between inpatient rehabilitation and follow-up. *Spinal Cord.* (2010) 48:239–44. doi: 10.1038/sc.2009.113

32. Saldivia S, Inostroza C, Bustos C, Rincón P, Aslan J, Bühring V, et al. Effectiveness of a group-based psychosocial program to prevent depression and anxiety in older people attending primary health care centres: a randomised controlled trial. *BMC Geriatr.* (2019) 19:237. doi: 10.1186/s12877-019-1255-3

33. Ministerio de Salud. *Manual de Examen de Medicina Preventiva del Adulto Mayor.* MINSAL (2008).

34. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 4th Edn..* American Psychiatric Publishing (1994).

35. 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. *Psychosomat Med.* (2001) 63:679–86. doi: 10.1097/00006842-200107000-00021

36. World Health Organization. *Composite International Diagnostic Interview (CIDI).* Version 2.1. World Health Organization (1997).

37. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Annals Family Med.* (2010) 8:348–53. doi: 10.1370/afm.1139

38. Pocklington C, Gilbody S, Manea L, McMillan D. The diagnostic accuracy of brief versions of the geriatric depression scale: a systematic review and meta-analysis. *Int J Geriatr Psychiatr.* (2016) 31:837–57. doi: 10.1002/gps.4407

39. Sherina MS, Arroll B, Goodyear-Smith F. Criterion validity of the PHQ-9 (Malay version) in a primary care clinic in Malaysia. *Med J Malaysia.* (2012) 67:309–15. Available online at: http://www.e-mjm.org/2012/v67n3/PHQ-9.pdf

40. Vielma M, Vicente B, Rioseco P, Castro N, Torres S. Validación en Chile de la entrevista diagnóstica estandarizada para estudios epidemiológicos CIDI. *Revista de Psiquiatría.* (1992) 9:1039–49.

41. Gelaye B, Tadesse MG, Williams MA, Fann JR, Stoep AV, Zhou X-HA. Assessing validity of a depression screening instrument in the absence of a gold standard. *Annals Epidemiol.* (2014) 24:527–31. doi: 10.1016/j.annepidem.2014.04.009

42. Robins LN, Wing J, Wittchen H. The composite international diagnostic interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Arch General Psychiatr.* (1988) 45:1069–77. doi: 10.1001/archpsyc.1988.01800360017003

43. JASP Team. *JASP (Version 0.11.1)* (2019).

44. MedCalc Software. *MedCalc Statistical Software version 14.8.1.* MedCalc Software (2014).

45. Muthén LK, Muthén BO. *2017 Mplus User's Guide.* Eighth Edition. Muthén and Muthén (1998).

46. Brown TA. *Confirmatory Factor Analysis for Applied Research (2nd ed.).* The Guilford Press (2015).

47. Hair JF, Black WC, Babin BJ, Anderson RE. *Multivariate Data Analysis.* Pearson Education Limited (2013). Available online at: https://books.google.cl/books?id=VvXZnQEACAAJ

48. Streiner DL, Cairney J. What's under the ROC? *Can J Psychiatr.* (2007) 52:121–8. doi: 10.1177/070674370705200210

49. Mitchell AJ, Bird V, Rizzo M, Meader N. Diagnostic validity and added value of the geriatric depression scale for depression in primary care: A meta-analysis of GDS30 and GDS15. *J Affective Disord.* (2010) 125:10–7. doi: 10.1016/j.jad.2009.08.019

50. 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.* (2018) 2:127–38. doi: 10.1192/bjpo.bp.115.001685

51. Østergaard SD, Dinesen PT, Foldager L. Quantifying the value of markers in screening programmes. *Eur J Epidemiol.* (2010) 25:151–4. doi: 10.1007/s10654-010-9430-z

52. Mitchell AJ, Coyne JC. *Screening for Depression in Clinical Practice: An Evidence-Based Guide.* Oxford: Oxford University Press (2009).

53. Kroenke K, Spitzer RL, Williams JB, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *General Hospital Psychiatr.* (2010) 32:345–59. doi: 10.1016/j.genhosppsych.2010.03.006

54. Patel V, Araya R, Chowdhary N, King M, Kirkwood B, Nayak S, et al. Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. *Psychol Med.* (2008) 38:221–8. doi: 10.1017/S0033291707002334

55. Hybels CF, Blazer DG. Epidemiology of Psychiatric Disorders. In BJ. Sadock, VA. Sadock, P. Ruiz, editors. *Kaplan and Sadock's Comprehensive Textbook of Psychiatry, 10 Edn.* Ostend: Wolters Kluwer Health (2017).

56. Frances A. *¿ Somos todos enfermos mentales?* Manifiesto contra los abusos de la psiquiatría. Ariel (2014).

57. Wakefield JC, Demazeux S. *Sadness or Depression?: International Perspectives on the Depression Epidemic and Its Meaning.* Springer (2016). Available online at: https://books.google.cl/books?id=yFVECwAAQBAJ doi: 10.1007/978-94-017-7423-9

58. Phelan E, Williams B, Meeker K, Bonn K, Frederick J, LoGerfo J, et al. A study of the diagnostic accuracy of the PHQ-9 in primary care elderly [journal article]. *BMC Family Pract.* (2010) 11:63. doi: 10.1186/1471-2296-11-63