



BRIEF COMMUNICATION

Accuracy of three depression screening scales to diagnose major depressive episodes in older adults without neurocognitive disorders

Mônica V. Costa,¹ Maissa F. Diniz,¹ Kenia K. Nascimento,¹ Kelly S. Pereira,¹ Natalia S. Dias,¹ Leandro F. Malloy-Diniz,^{1,2} Breno S. Diniz^{1,2,3}

¹Laboratório de Investigação em Neurociências Clínicas (LINC), Faculdade de Medicina, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brazil. ²Departamento de Saúde Mental, Faculdade de Medicina, UFMG, Belo Horizonte, MG, Brazil. ³Department of Psychiatry and Behavioral Sciences, The University of Texas Health Science Center at Houston, Houston, TX, USA.

Objective: To determine the sensitivity and specificity of three depression screening scales to diagnose major depressive episodes in the elderly.

Methods: Participants (n=129, 88% female) answered a semi-structured psychiatric interview (Mini International Neuropsychiatric Interview) to determine the diagnosis of major depressive disorder. After this, depressive symptoms in depressed and non-depressed subjects were assessed by independent administration of the 15-item Geriatric Depression Scale (GDS-15), Patient Health Questionnaire-9 (PHQ-9), and 17-item Hamilton Rating Scale for Depression (HDRS-17).

Results: Patients with major depression and controls did not differ in age and gender distribution. The sensitivity and specificity of all scales to identify a major depressive episode in older adults were $\geq 90\%$. There were no significant differences between the areas under the curve for PHQ-9 vs. HDRS-17 ($z = 1.2, p = 0.2$), PHQ-9 vs. GDS-15 ($z = 0.26, p = 0.8$), or HDRS-17 vs. GDS-15 ($z = 1.2, p = 0.2$).

Conclusion: This study provides evidence supporting the use of PHQ-9 and GDS-15, both of which are simple to administer and easy to interpret, to diagnose major depressive episodes in older adults without neurocognitive disorders.

Keywords: Late-life depression; diagnosis; screening; geriatric assessment

Introduction

Late-life major depression (LLMD) is one of the most common psychiatric conditions in the elderly.¹ LLMD has been associated with increased risk of dementia, medical comorbidities, mortality, and significant impact on quality of life.²⁻⁴

Several scales are used to assess depressive symptoms in the elderly. The Geriatric Depression Scale (GDS),⁵ developed in 1983, had its short version translated and validated for Brazilian Portuguese.⁶ However, the GDS has moderate sensitivity and specificity for the identification of major depression in older adults.⁶⁻⁸

Other scales are also commonly used in clinical practice, among which the Patient Health Questionnaire-9 (PHQ-9). This short, self-administered scale has been used to screen for depressive symptoms in clinical and population-based studies. Despite having good sensitivity and specificity for the identification of major depression in young adults, few studies have evaluated its accuracy for identification of LLMD.⁹ Similarly, the Hamilton Depression Rating Scale (HDRS) is often employed in the assessment of depressive symptoms in clinical populations, with good

sensitivity and specificity to identify major depressive episodes in young adults; however, few studies have addressed the accuracy of the HDRS for LLMD.¹⁰

Because the diagnosis of depression is often overlooked in the elderly, it is essential to have accurate instruments to identify this disorder in clinical practice. Therefore, we aimed to evaluate the sensitivity and specificity of the 15-item Geriatric Depression Scale (GDS-15), PHQ-9, and the 17-item Hamilton Rating Scale for Depression (HDRS-17) for the diagnosis of major depressive episodes in the elderly. We further compared the accuracy of these three scales to identify major depression in this population. Based on clinical observation, we hypothesized that the HDRS-17 would have higher accuracy to identify LLMD compared to the PHQ-9, and that the GDS-15 would have the lowest accuracy among the three scales.

Methods

Participants

We studied a convenience sample of 129 outpatient older adults (50 controls and 79 individuals with LLMD). We excluded individuals with history of or current neurological conditions including stroke, epilepsy, Parkinson's disease, and dementia or major psychiatric disorders (e.g., schizophrenia, bipolar disorder). We also excluded

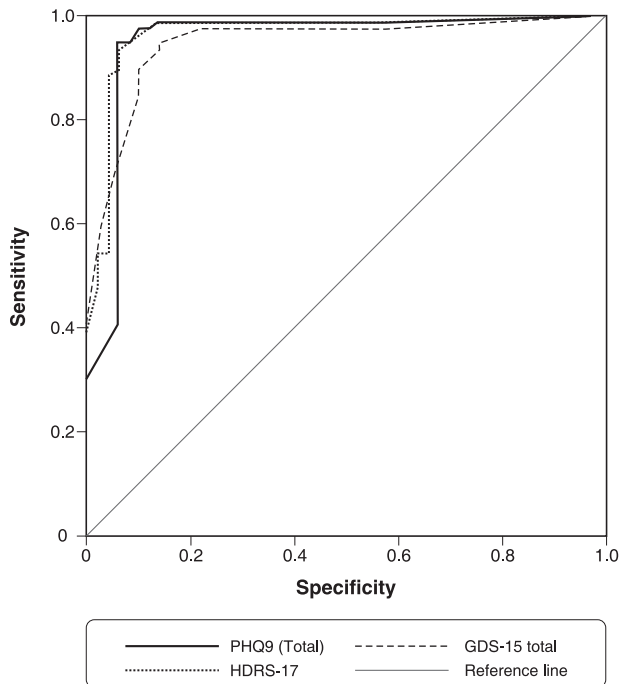


Figure 1 Receiver operating characteristics (ROC) curves comparing the accuracy of the 15-item Geriatric Depression Scale (GDS-15), Patient Health Questionnaire-9 (PHQ-9), and 17-item Hamilton Rating Scale for Depression (HDRS-17) to identify a major depressive episode.

subjects with Alzheimer's disease and mild cognitive impairment after neuropsychological assessments. This study was approved by the Universidade Federal de Minas Gerais Ethics Committee and all subjects signed an informed consent before inclusion in the study.

Assessment of depressive symptoms

Participants answered a semi-structured psychiatric interview which included administration of the Brazilian Portuguese version of the Mini International Neuropsychiatric Interview (MINI).¹¹ The diagnosis of major depressive disorder was based on MINI scores and DSM-5 criteria. Individuals without major depressive disorder were included as controls if they did not have evidence of major psychiatric or neurological disorders.

The GDS-15 and PHQ-9 were self-administered before the psychiatric interview. Because some older adults had

no formal educational, an independent investigator helped subjects if assistance was requested. After the psychiatric interview, the HDRS-17 was administered by a trained geriatric psychiatrist who also conducted the clinical interview. The geriatric psychiatrist was not informed of GDS-15 and PHQ-9 scores before or during the clinical interview.

Statistical analysis

Mann-Whitney or chi-square analyses were used to evaluate differences in demographic and clinical characteristics between LLMD individuals and controls. The diagnostic accuracy, sensitivity, and specificity of the GDS-15, PHQ-9, and HDRS-17 to identify older adults with LLMD were assessed using receiver operating characteristic (ROC) analysis. Statistical analysis was performed using SPSS version 20.0.

Results

The age of participants ranged from 60 to 92 years, with 0 to 26 years of formal education. There was no significant difference between LLMD patients and controls regarding age (depressed, 71 ± 8 years vs. controls, 71 ± 8 years; $z = -0.146$, $p = 0.884$) and gender distribution (depressed, 91% female vs. controls, 86% female; $\chi^2 = 0.790$, $p = 0.374$).

As expected, depressed patients scored significantly higher in the GDS-15 (LLMD: median = 11, 25th-75th percentile⁸⁻¹³; controls: 1, [0-2]; $z = -8.574$, $p < 0.001$), PHQ-9 (LLMD: median = 18, 25th-75th percentile [14-21]; controls: 0, [0-3]; $z = -8.669$, $p < 0.001$), and HDRS-17 (LLMD: median = 20, 25th-75th percentile [14-24]; controls: 1, [0-3]; $z = -8.931$, $p < 0.001$).

Figure 1 shows the ROC analysis of the accuracy of GDS-15, PHQ-9, and HDRS-17 to identify a major depressive episode in this sample of older adults without neurocognitive disorders. Sensitivity and specificity were $\geq 90\%$ for all the scales (Table 1). There were no significant differences in the area under the curve (AUC) between PHQ-9 and HDRS-17 ($z = 1.239$, $p = 0.2$), PHQ-9 and GDS-15 ($z = 0.265$, $p = 0.8$), and HDRS-17 and GDS-15 ($z = 1.202$, $p = 0.2$).

Discussion

Few studies so far have evaluated the diagnostic performance of different depression assessment scales in older

Table 1 Diagnostic properties of GDS-15, PHQ-9 and HDRS-17

Scale	Cutoff	Sensitivity (%)	Specificity (%)	AUC	PPV	NPV	PLR	NLR
GDS-15	6	90 (79-97)	90 (79-95)	0.946* (0.906-0.986)	84 (71-92)	93 (85-98)	8.0 (4.3-15.0)	0.11 (0.05-0.25)
PHQ-9	10	94 (81-97)	94 (84-97)	0.951* (0.905-0.997)	89 (77-96)	95 (87-96)	12.3 (5.7-27)	0.08 (0.03-0.21)
HDRS-17	9	94 (84-99)	94 (86-98)	0.966* (0.932-1.000)	90 (79-97)	96 (89-99)	14.8 (6.3-34.8)	0.06 (0.02-0.19)

AUC = area under the curve; GDS-15 = 15-item Geriatric Depression Scale; HDRS-17 = 17-item Hamilton Rating Scale for Depression; NLR = negative likelihood ratio; NPV = negative predictive value; PHQ-9 = Patient Health Questionnaire-9; PLR = positive likelihood ratio; PPV = positive predictive value.

* Significant at $p < 0.001$.

Numbers in parentheses show 95% confidence interval.

adults with depression. We found that the PHQ-9, the GDS-15, and the HDRS-17 had excellent diagnostic accuracy for the identification of a major depressive episode in older adults. Moreover, the cutoff values for PHQ-9 and HDRS-17 are similar to those previously reported in the literature for both young and older adults with depression.

The GDS was designed specifically for the assessment of depressive complaints in older adults.⁷ Despite its widespread use in clinical practice, the GDS has some negative points, such as low test-retest reliability and low diagnostic accuracy for depression in subjects with neurocognitive disorders.^{12,13} However, in this sample of older adults without evidence of neurocognitive disorders, the GDS-15 showed excellent accuracy for the identification of a major depressive episode.

The PHQ-9 is widely used as a screening tool for depression in clinical and research settings.¹⁴ It is based on DSM-IV criteria for major depressive episode. There are few studies with older adults.⁹ In a study with older adults, the PHQ-9 (cutoff ≥ 10) had sensitivity of 63% and specificity of 82%.¹⁵ In contrast, our study showed a much better diagnostic profile for the PHQ-9, since the same cutoff value had sensitivity and specificity of 94% for the identification of a major depressive episode. The diagnostic performance of the HDRS-17 was similar to that of the PHQ-9. Nonetheless, the former scale is not suitable for screening purposes, as it is long and requires specific training for administration and interpretation of the test scores. Finally, it should be noted that even though the PHQ-9 performed better than the GDS-15, the difference was not statistically significant. Thus, both scales are suitable for screening and identification of major depressive episode in older adults without neurocognitive disorders.

In clinical practice, PHQ-9 items follow the proposed criteria for the diagnosis of depression and may provide a more straightforward evaluation of depressive symptoms compared to the GDS-15. Also, the PHQ-9 is easier to administer, its scores are easier to interpret, and it requires less training for administration as compared to the HDRS-17. One additional important aspect when choosing a scale for clinical use is the amount of time required for administration. Administration of the HDRS-17 is time-consuming (usually the length of the interview), and it is thus not suitable for screening purposes. The PHQ-9 and GDS-15 are self-administered and usually take 5 minutes to be completed by a patient. The shorter time of administration facilitates the use of scales in clinical settings. Therefore, given the high accuracy to identify LLMD, as well as the ease and short time of administration, we suggest that the PHQ-9 is a more suitable tool for the screening and assessment of depressive symptoms in older adults in clinical settings.

The current results should be viewed in light of the study limitations. The sample size is relatively small and recruited at a tertiary clinic. Also, there was a significantly higher proportion of women in the depressed group compared to the controls. Thus, our results cannot be generalized to the overall population. In contrast, the careful psychiatric evaluation, the diagnosis of LLMD based on a semi-structured interview, and neuropsychological evaluation as exclusion criteria for neurocognitive disorders are major strengths of our study.

In conclusion, our study provides data that support the use of PHQ-9 and GDS-15 for the screening of major depressive episodes in older adults without neurocognitive disorders. Since both assessment scales are simple to administer and to interpret, they should be included in the routine assessment of older adults, given the relevance and high frequency of depressive episodes in older adults.

Acknowledgements

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; grants 472138/2013-8 and 466623/2014-3).

Disclosure

The authors report no conflicts of interest.

References

- 1 Byers AL, Yaffe K, Covinsky KE, Friedman MB, Bruce ML. High occurrence of mood and anxiety disorders among older adults: the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2010;67:489-96.
- 2 Diniz BS, Reynolds CF 3rd, Butters MA, Dew MA, Firmo JO, Lima-Costa MF, et al. The effect of gender, age, and symptom severity in late-life depression on the risk of all-cause mortality: The Bambuí Cohort Study of Aging. *Depress Anxiety*. 2014;31:787-95.
- 3 Diniz BS, Butters MA, Albert SM, Dew MA, Reynolds CF 3rd. Late-life depression and risk of vascular dementia and Alzheimer's disease: systematic review and meta-analysis of community-based cohort studies. *Br J Psychiatry*. 2013;202:329-35.
- 4 Rapaport MH, Clary C, Fayyad R, Endicott J. Quality-of-life impairment in depressive and anxiety disorders. *Am J Psychiatry*. 2005;162:1171-8.
- 5 Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res*. 1982-1983;17:37-49.
- 6 Almeida OP, Almeida SA. Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. *Int J Geriatr Psychiatry*. 1999;14:858-65.
- 7 Yesavage JA, Sheikh JI. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clin Gerontol*. 1986;5:165-73.
- 8 Castelo MS, Coelho-Filho JM, Carvalho AF, Lima JW, Noleto J, Ribeiro KG, et al. Validity of the Brazilian version of the Geriatric Depression Scale (GDS) among primary care patients. *Int Psychogeriatr*. 2010;22:109-13.
- 9 Manea L, Gilbody S, McMillan D. A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. *Gen Hosp Psychiatry*. 2015;37:67-75.
- 10 Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight?. *Am J Psychiatry*. 2014;161:2163-77.
- 11 Amorin P. Mini International Neuropsychiatric Interview (MINI): validação de entrevista breve para diagnóstico de transtornos mentais. *Rev Bras Psiquiatr*. 2000;22:106-15.
- 12 Lach HW, Chang YP, Edwards D. Can older adults with dementia accurately report depression using brief forms? Reliability and validity of the Geriatric Depression Scale. *J Gerontol Nurs*. 2010;36:30-7.
- 13 Müller-Thomsen T, Arit S, Mann U, Mass R, Ganzer S. Detecting depression in Alzheimer's disease: evaluation of four different scales. *Arch Clin Neuropsychology*. 2005;20:271-6.
- 14 Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16:606-13.
- 15 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. *BMC Fam Pract*. 2010;11:63-63.