ORIGINAL ARTICLE

Validation of the Whooley questions and the Beck Depression Inventory in older adults

KADRI SUIJA¹, ULLA RAJALA², JARI JOKELAINEN^{2,3}, TIMO LIUKKONEN^{2,4},
PIRJO HÄRKÖNEN^{3,5}, SIRKKA KEINÄNEN-KIUKAANNIEMI² & MARKKU TIMONEN^{2,6}

¹Department of Family Medicine, University of Tartu, Estonia, ²Institute of Health Sciences (General Practice), University of Oulu, Finland, ³Unit of General Practice, Oulu University Hospital, Finland, ⁴Savonlinna Central Hospital, Finland, ⁵Oulu Deaconess Institute, Department of Sport Medicine, Finland, and ⁶Oulu Health Centre, Finland

Abstract

Objective. To analyse the psychometric properties of the Whooley questions and the 21-item Beck Depression Inventory (BDI-21) in older adults with depression and chronic health problems. **Design.** A population-based study. **Setting.** Community. **Subjects.** 474 adults, aged 72–73 years, living in the city of Oulu, Finland. **Main outcome measures.** The screening parameters of the Whooley questions and the BDI-21 for detecting major depression. **Results.** The prevalence of major depression according to the DSM-IV was 5.3% (single or recurrent episode) obtained by the Mini Neuropsychiatric Interview (MINI). The BDI-21 was best able to identify a current episode of major depression with a cut-off point of 11. The sensitivity and specificity of this cut-off point were 88.0% (95% confidence interval (95% CI) 68.8–97.5) and 81.7% (95% CI 77.8–85.2), respectively. The area under the receiver operating characteristics (ROC) curve was 0.89 (95% CI 0.83–0.96). The two Whooley screening questions had a sensitivity of 62.5% (95% CI 40.6–81.2) and either screening question plus the help question had a sensitivity of 66.7% (44.7–84.4). **Conclusions.** The Beck Depression Inventory is a valid instrument for the diagnosis of depression in older adults. As a screening measure, the optimal cut-off score should be 11 or higher. Our results indicate that the sensitivity of the Whooley questions is not high enough to be used as a screening scale among the elderly.

Key Words: Depression, elderly, Finland, general practice, screening instruments, validation studies

Introduction

The recent guideline on depression in adults with chronic health problems, developed by the National Institute for Health and Clinical Excellence (NICE), recommends use of two questions from the Patient Questionnaire of the Primary Care Evaluation of Mental Disorders (PRIME-MD) for screening depression [1,2]. These questions, known as the Whooley questions, may be used “as the first stage of case identification for depression” [3,4]. However, according to a systematic review and a meta-analysis conducted during the development process of the above guideline, the evidence regarding the psychometric properties of the Whooley questions in older (over 65 years) adults is scanty. To our knowledge, the only earlier study comparing the Whooley questions with the structured psychiatric diagnosis according to the Diagnostic and Statistical Manual

of Mental Disorders (DSM) was conducted among patients with coronary heart disease: the Whooley questions had a sensitivity of 91% and 88%, and a specificity of 72% and 72% in patients aged 65–75 years ($n = 357$) and over 75 years ($n = 245$), respectively [5,6].

The other widely used instrument for screening depression is the 21-item version of the Beck Depression Inventory (BDI-21) [7]. In the Finnish guideline for management of depression, both the BDI-21 and the Whooley questions are recommended for use as screening tools in patient populations where the prevalence of depression is expected to be high, e.g. in patients with physical disorders [8,9]. Physical disorders are inevitably more prevalent in older adults. Surprisingly, validation studies of the use of the BDI-21 in older adults with a proper gold standard (DSM or International Classification of Diseases),

Current guidelines recommend use of screening tests in patient populations where the prevalence of depression is high, such as older adults or patients with chronic physical illnesses.

- Based on our results, the Beck Depression Inventory is a valid instrument for the diagnosis of depression among older adults. As a screening measure, the optimal cut-off score would be 11 or higher. The sensitivity of the Whooley questions seems not to be high enough to use it as a screening scale among elderly.

are lacking [10]. The systematic review carried out in the course of the development of Depression in Chronic Health Problems Guideline reported one study in which the BDI-21 was used against the DSM among nursing home residents in Canada [11]. However, this study was excluded from final analysis as an outlier. Besides, we found only three other studies conducted among older patient populations with chronic physical health problems, but these studies involved quite specific patient groups: patients with stroke or with Parkinson's disease [12–14]. Thus, there exists an evident need to evaluate the psychometric properties of the two widely used questionnaires, the Whooley questions and the BDI-21, in older adults.

In the present study, we investigated the psychometric properties of the Whooley questions and the BDI-21 using the DSM-IV diagnosis of depression as a gold standard in a community sample of older adults.

Material and methods

We invited all persons ($n = 1008$) born in 1935 and living on 1 October 1990 in the city of Oulu, Northern Finland, to participate in the study. Altogether 831 persons attended the first follow-up. All participants signed an informed consent. A detailed description of the enrolment procedure and the objectives of the study was presented earlier [15,16]. This study is based on the second follow-up of the initial study population, which took place in 2007–2008. Altogether 667 persons were alive and asked to attend the study. The total number of participants at the second follow-up was 474 and their age was 72–73 years.

The participants underwent an interview in which standardized questionnaires were used to obtain self-reported data on gender, education, marital status, medical conditions diagnosed by a physician, and

Mini Mental State Examination (MMSE) [15,16]. For assessment of the patient's subjective view of his/her depressive symptoms, we included the BDI-21 and the Whooley questions in the questionnaires. The BDI-21 contains items that reflect the cognitive, affective, somatic, and vegetative symptoms of depression. Altogether there are 21 items with a score range of 0 to 63. Various cut-off points for major depression have been proposed, depending on the purpose of the study (e.g. to assess depressive symptoms or to measure severity) and the study population [7,17,18].

The Whooley questionnaire consists of two questions. A "yes" answer to either of the following two questions was considered a positive test: (i) "During the past month, have you often been bothered by feeling down, depressed, or hopeless?" and (ii) "During the past month, have you often been bothered by little interest or pleasure in doing things?" In the case where a subject answered "yes" to either of the above questions, the "help" question, as proposed by Arroll and colleagues, was asked: "Is this something with which you would like help?", with three possible responses, "no", "yes, but not today", "yes" [19]. In this study a "yes" or "yes, but not today" answer was considered a positive test.

Both questionnaires were filled in by the participants in the study centre before the psychiatric interview. The research assistant checked the questionnaires, and in the case of missing data the participants were asked to complete the questionnaires. The diagnosis of current or recurrent major depression according to the DSM-IV was made, using a short structured diagnostic interview, the Mini Neuropsychiatric Interview (MINI), by two separately working professional psychiatrists [20]. The psychiatrists were blinded to the results of the above questionnaires while conducting the interview.

Statistical analyses

We compared the baseline characteristics of the participants according to their depression status (major depression diagnosed or not according to the MINI) using the Pearson chi-squared statistics. We calculated sensitivity, specificity, and positive and negative likelihood ratios (LR+ and LR-), as well as positive and negative predictive value with 95% confidence intervals (CI) for the BDI-21, and for the Whooley questions for detecting major depression [21]. To measure the effectiveness of the scale and to select an optimal threshold value (cut-off point), we calculated Youden's index and generated a receiver operating characteristics (ROC) curve [22]. All statistical analyses were performed using the statistical program Stata [23].

Table I. General characteristics of the study group.¹

Characteristics	Total n = 474	Non-depressed* n = 449 (94.7%)	Depressed* n = 25 (5.3%)	p-value
Gender:				0.157
Male	197	190 (96.5)	7 (3.6)	
Female	277	259 (93.5)	18 (6.5)	
Marital status:				0.721
Married/cohabiting	322	308 (95.7)	14 (4.4)	
Single	20	19 (95.0)	1 (5.0)	
Widower/widow	112	105 (93.8)	7 (6.3)	
Missing data	20			
MMSE, mean (standard error) [†]	26.6 (2.3) [†]	26.6 (2.3) [†]	27.0 (1.8) [†]	0.485
Any chronic diseases:				0.965
No	37	35 (94.6)	2 (5.4)	
Yes	420	398 (94.8)	22 (5.2)	
Missing data	17			

Notes: ¹Values are total numbers (percentages) of participants unless stated otherwise. *According to the MINI. [†]Mean (standard error).

Results

Table I presents the general characteristics of the study population. The prevalence of major depression according to the DSM-IV among our study population was 5.3% (single or recurrent episode) obtained by the MINI.

Table II presents the measures of validity (sensitivity, specificity, Youden's index, and likelihood ratios) for the different cut-off points of the BDI-21. The specificity increased and the sensitivity decreased with higher cut-off points. Youden's index was the highest (0.70) for cut-off point 11. The sensitivity and specificity parameters for this cut-off point were 88.0% (95% CI 68.8–97.5) and 81.7% (95% CI 77.8–85.2), respectively; its positive predictive value was 21.2% (95% CI 13.8–30.3) and negative predictive value was 99.2% (95% CI 97.6–99.8).

Figure 1 shows the ROC curve of the BDI for detecting the presence of major depression. In ROC analysis, the area under the curve was 0.89 (95% CI 0.83–0.96).

Table III presents the measures of validity of the Whooley questions. A "yes" answer to either of the screening questions was 62.5% sensitive (95% CI 40.6–81.2) and its specificity was 88.9% (95% CI 85.6–91.7) among this population. When the help question was included, the instrument was 66.7% sensitive (95% CI 44.7–84.4) and 85.9% specific (95% CI 82.3–89.1).

Discussion

Statement of principal findings

The results of this study conducted with a community sample of older adults with chronic health problems

Table II. Screening parameters of the BDI-21 for detecting major depression among the community sample of older adults (n = 474).

Cut-off point	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Youden's index	LR+ (95% CI)	LR- (95% CI)
6	96.0 (79.6–99.9)	47.2 (42.5–52.0)	0.43	1.8 (1.6–2.1)	0.1 (0.01–0.6)
7	96.0 (79.6–99.9)	57.2 (52.5–61.9)	0.53	2.3 (2.0–2.6)	0.1 (0.01–0.5)
8	96.0 (79.6–99.9)	64.8 (60.2–69.2)	0.61	2.7 (2.4–3.2)	0.1 (0.01–0.4)
9	96.0 (79.6–99.9)	72.4 (68.0–76.5)	0.68	3.5 (2.9–4.1)	0.1 (0.01–0.4)
10	88.0 (68.8–97.5)	77.3 (73.1–81.1)	0.65	3.9 (3.1–4.9)	0.2 (0.05–0.5)
11	88.0 (68.8–97.5)	81.7 (77.8–85.2)	0.70	4.8 (3.8–6.2)	0.2 (0.05–0.4)
12	84.0 (63.9–95.5)	85.1 (81.4–88.2)	0.69	5.60 (4.3–7.4)	0.2 (0.1–0.5)
13	76.0 (54.9–90.6)	86.6 (83.1–89.6)	0.63	5.7 (4.1–7.9)	0.3 (0.1–0.6)
14	68.0 (46.5–85.1)	89.3 (86.1–92)	0.57	6.4 (4.4–9.3)	0.4 (0.2–0.6)
15	60.0 (38.7–78.9)	92.0 (89.1–94.3)	0.52	7.5 (4.8–11.7)	0.4 (0.3–0.7)
16	44.0 (24.4–65.1)	94.0 (91.4–96.0)	0.38	7.3 (4.1–13.0)	0.6 (0.4–0.8)
17	40.0 (21.1–61.3)	95.5 (93.2–97.3)	0.36	9.0 (4.7–17.1)	0.6 (0.5–0.9)
18	32.0 (14.9–53.5)	96.9 (94.8–98.3)	0.29	10.3 (4.8–22.2)	0.7 (0.5–0.9)
19	32.0 (14.9–53.5)	97.1 (95.1–98.4)	0.29	11.1 (5.1–24.2)	0.7 (0.5–0.9)
20	24.0 (9.4–45.1)	97.1 (95.1–98.4)	0.21	8.3 (3.4–20.0)	0.8 (0.6–1.0)
21	20.0 (6.8–40.7)	97.6 (95.7–98.8)	0.18	8.2 (3.1–21.7)	0.8 (0.7–1.0)

Figure 1. Receiver operating characteristic (ROC) curve of the BDI-21 for predicting major depression.

showed that the BDI-21 is a valid screening tool and it is best able to identify a current episode of major depression with a cut-off point of 11. However, the sensitivity of the Whooley questions was not sufficiently high to be recommended as a screening instrument among older adults.

Strengths and weaknesses of the study

The main strengths of our study are methodological: the homogeneous study group (72–73 years) which represent well the population of this age in Oulu, use of a psychiatric interview (MINI) as the comparator, and the arrangement that the psychiatrists conducting the interview were blinded to the results of the screening questionnaires.

A limitation of the study is that the screening tools, the BDI-21 and the Whooley questions, were introduced to the participants among a large number of questions. In the event that a patient had been asked the Whooley questions by his or her own physician during a medical consultation, the sensitivity of these questions may possibly have been higher. Moreover, the low sensitivity of the Whooley questions in our study could be related to the fact that the questions themselves are quite lengthy and complicated. Also, the 21 questions of the BDI can be difficult to complete for older people. The scales developed for elderly people, such as the Geriatric

Depression Scale, make use of a rather short and simple format of questions [24]. However, based on our experience the BDI-21 can be used in older adults.

Comparison with existing literature

There is evidence that major depression is prevalent and often associated with disability and increased risk of mortality in the elderly [14,25]. Although there are effective treatments available, depression in older adults often remains under-diagnosed and under-treated [10,26,27]. To improve detection of depression, guidelines recommend use of screening tests among populations where its prevalence is high, such as older people with physical illnesses [1,8]. Several scales have been specifically developed for older people. On the other hand, many commonly used scales are not validated in older people [10].

One of the shortest depression measurement scales consists of two questions. Whooley and colleagues reported its sensitivity to be 89% to 96% and specificity to be 51% to 72% for diagnosing major depression in patients visiting an urgent care clinic [3]. Arroll et al. (2005) found that the response to both the two screening questions and to either screening question plus the help question was 96% sensitive [19]. According to a recent study conducted among primary care patients in Switzerland, the sensitivity of the two-question method was 91.3% but addition of the “help” question decreased the sensitivity to 59.4% [28]. However, we found only one earlier study conducted in older adults, with a sensitivity of 88% to 91% and a specificity of 72% for detecting depression in elderly patients with coronary heart disease [6]. Thus, the need to validate the Whooley questions in older adults was clearly evident.

The sensitivity of the Whooley questions in our study population was only 62.5% and inclusion of the help question increased it to 66.7%. On the other hand, the specificity was much higher: 88.9% and 85.9%, respectively. This discrepancy may be ascribed to differences in the study setting and population. For example, the questionnaires working well in a general practice setting may not be sufficiently valid or suitable for use in population studies [28]. Also Lombardo et al. (2011) reported recently that the

Table III. Screening parameters of the Whooley questions for detecting major depression among the community sample of older adults (n = 474).

Variable	Sensitivity, % (95% CI)	Specificity, % (95% CI)	LR+ (95% CI)	LR- (95% CI)
Either of the two screening questions alone	62.5 (40.6–81.2)	88.9 (85.6–91.7)	5.6 (3.7–8.5)	0.4 (0.2–0.7)
Help question alone	47.8 (26.8–69.4)	92.1 (89.2–94.5)	6.1 (3.4–10.7)	0.6 (0.5–0.7)
Either screening question plus the help question	66.7 (44.7–84.4)	85.9 (82.3–89.1)	4.7 (3.29–6.8)	0.4 (0.2–0.7)

sensitivity of the Whooley questions varied in different patient groups [29].

Several studies have used the BDI-21 for screening depression, with different proposed cut-off points (10 to 18) [17,18]. Yet only a few studies have assessed the BDI-21 in comparison with a diagnostic interview in older adults: Laprise and Vezina (1998) proposed a cut-off score of 10 (96% sensitivity and 45% specificity), Lincoln et al. (2003) suggested a cut-off score of 15/16 (91% sensitivity and 56% specificity), and Leentjens et al. (2000) recommended a cut-off score of 13/14 [11,12,14]. We found that a cut-off point of 11 had the best screening parameters. However, as the study population in the above-mentioned studies was different from our study group, it is difficult to compare the results. However, our results are in line with other studies showing that the BDI can be used in patients with somatic diseases [11–14,30].

Implications for clinical practice and future research

In clinical practice, it is convenient to use one measurement scale which is as short as possible, even if the cut-off point for different patient groups is different. Therefore, the BDI-21 would be one of the best options. When using self-report instruments we should be aware that scales containing somatic symptoms might be misleading, as both depression and physical disease have somatic symptoms. According to our study, the sensitivity of the BDI-21 in older adults with physical illnesses was satisfactory.

We also wanted to assess the influence of accompanying somatic diseases on the screening parameters of the BDI-21. But as chronic somatic diseases were so prevalent in our patients (almost 90% of them had at least one chronic disease) we were not able to compare the validity of the BDI-21 in patients with and without chronic somatic diseases. Future studies in this field are necessary.

Conclusions

Based on our results, the BDI is a valid instrument for the diagnosis of depression among older adults. As a screening measure, the optimal cut-off score would be 11 or higher. The sensitivity of the Whooley questions seems not to be high enough to use it as a screening scale among the elderly.

Acknowledgments

The Ethics Committee of the Faculty of Medicine, University of Oulu, Finland approved the study.

No additional funding was received.

Declaration of interest

The authors report no conflict of interest. The authors alone are responsible for the content and writing of the paper.

References

- [1] Depression in Adults with a Chronic Physical Health Problem: Treatment and management. National Collaborating Centre for Mental Health. Commissioned by the National Institute for Health and Clinical Excellence; 2009.
- [2] Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV III, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care: The PRIME-MD study. *JAMA* 1994;272:1749–56.
- [3] Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression: Two questions are as good as many. *J Gen Intern Med* 1997;12:439–45.
- [4] Pilling S, Anderson I, Goldberg D, Meader N, Taylor C. Two Guideline Development Groups. Depression in adults, including those with a chronic physical health problem: Summary of NICE guidance. *BMJ* 2009;339:b4108.
- [5] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed. Washington, DC: American Psychiatric Association; 1994.
- [6] McManus D, Pipkin SS, Whooley MA. Screening for depression in patients with coronary heart disease (data from the heart and soul study). *Am J Cardiol* 2005;96:1076–81.
- [7] Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;4:53–63.
- [8] Working group appointed by the Finnish Medical Society Duodecim and the Finnish Psychiatric Association Current Care Guideline. Depression Current Care Summary 6.8.2009.
- [9] Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic disease and decrements in health: Results from the World Health Survey. *Lancet* 2007;370:851–8.
- [10] Rodda J, Walker Z, Carter J. Depression in older adults. *BMJ* 2011;343:d5219.
- [11] Laprise R, Vezina J. Diagnostic performance of the Geriatric Depression Scale and the Beck Depression Inventory with nursing home residents. *Can J Aging* 1998;17:401–13.
- [12] Aben I, Verhey F, Lousberg R, Lodder J, Honig A. Validity of the Beck depression inventory, hospital anxiety and depression scale, SCL-90, and Hamilton depression rating scale as screening instruments for depression in stroke patients. *Psychosomatics* 2002;43:386–93.
- [13] Lincoln NB, Nicholl CR, Flannaghan T, Leonard M, Van der Gucht E. The validity of questionnaire measures for assessing depression after stroke. *Clin Rehabil* 2003; 17:840–6.
- [14] Leentjens AF, Verhey FR, Luijckx GJ, Troost J. The validity of the Beck Depression Inventory as a screening and diagnostic instrument for depression in patients with Parkinson's disease. *Mov Disord* 2000;15:1221–4.
- [15] Rajala U, Laakso M, Paivansalo M, Pelkonen O, Suramo I, Keinänen-Kiukaanniemi S. Low insulin sensitivity measured by both quantitative insulin sensitivity check index and homeostasis model assessment method as a risk factor of increased intima-media thickness of the carotid artery. *J Clin Endocrinol Metab* 2002;87:5092–7.
- [16] Timonen M, Laakso M, Jokelainen J, Rajala U, Meyer-Rochow VB, Keinänen-Kiukaanniemi S. Insulin resistance

- and depression: Cross sectional study. *BMJ* 2005;330:17–18.
- [17] Nuevo R, Lehtinen V, Reyna-Liberato PM, Ayuso-Mateos JL. Usefulness of the Beck Depression Inventory as a screening method for depression among the general population of Finland. *Scand J Public Health* 2009;37:28.
- [18] Viinamäki H, Tanskanen A, Honkalampi K, Koivumaa-Honkanen H, Haatainen K, Kaustio O, et al. Is the Beck Depression Inventory suitable for screening major depression in different phases of the disease? *Nord J Psychiatry* 2004;58:49–53.
- [19] Arroll B, Goodyear-Smith F, Kerse N, Fishman T, Gunn J. Effect of the addition of a “help” question to two screening questions on specificity for diagnosis of depression in general practice: Diagnostic validity study. *BMJ* 2005;331:884.
- [20] Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini International Neuropsychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998;59(Suppl 20):22–33.
- [21] Shapiro DE. The interpretation of diagnostic tests. *Stat Methods Med Res* 1999;8:113–34.
- [22] Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J* 2005;47:458–72.
- [23] Stata Statistical Software: Release 11. College Station, TX: StataCorp LP; 2009.
- [24] Wancata J, Alexandrowicz R, Marquart B, Weiss M, Friedrich F. The criterion validity of the Geriatric Depression Scale: A systematic review. *Acta Psychiatr Scand* 2006;114:398–410.
- [25] Rapp MA, Gerstorf D, Helmchen H, Smith J. Depression predicts mortality in the young old, but not in the oldest old: Results from the Berlin Aging Study. *Am J Geriatr Psychiatry* 2008;16:844–52.
- [26] Taylor D, Meader N, Bird V, Pilling S, Creed F, Goldberg D. Pharmacological interventions for people with depression and chronic physical health problems: Systematic review and meta-analyses of safety and efficacy. *BJP* 2011;198:179–88.
- [27] Kendrick T, Dowrick C, McBride A, Howe A, Clarke P, Maisey S, et al. Management of depression in UK general practice in relation to scores on depression severity questionnaires: Analysis of medical record data. *BMJ* 2009;338:b750.
- [28] Magnil M, Gunnarsson R, Björkelund C. Using patient-centred consultation when screening for depression in elderly patients: A comparative pilot study. *Scand J Prim Health Care* 2011;29:51–6.
- [29] Lombardo P, Vaucher P, Haftgoli N, Burnand B, Favrat B, Verdon F, et al. The “help” question doesn’t help when screening for major depression: External validation of the three-question screening test for primary care patients managed for physical complaints. *BMC Med* 2011;9:114.
- [30] Berg A, Lönnqvist J, Palomäki H, Kaste M. Assessment of depression after stroke: A comparison of different screening instruments. *Stroke* 2009;40:523–9.