Contents lists available at ScienceDirect

Journal of Psychiatric Research

Validation of the Patient Health Questionnaire-9 and the Generalized Anxiety Disorder-7 in Lithuanian individuals with anxiety and mood disorders

Agne Stanyte [a,*], Naomi A. Fineberg [b,c,d], Aurelija Podlipskyte [a], Julija Gecaite-Stonciene [a], Jurate Macijauskiene [e], Vesta Steibliene [a], Julius Burkauskas [a]

[a] Laboratory of Behavioral Medicine, Neuroscience Institute, Lithuanian University of Health Sciences, Palanga-Kaunas, Lithuania
[b] University of Hertfordshire, Hatfield, UK
[c] Hertfordshire Partnership University NHS Foundation Trust, Hatfield, UK
[d] University of Cambridge Clinical Medical School, Cambridge, UK
[e] Department of Geriatrics, Faculty of Nursing, Lithuanian University of Health Sciences, Kaunas, Lithuania

ARTICLE INFO

ABSTRACT

The Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7) are short self-report questionnaires used to screen and assess depression and anxiety severity in medical and community samples. However, little is known about their psychometric properties in individuals with anxiety and mood disorders (AMD) This study evaluated the psychometric properties of the PHQ-9 and GAD-7 in individuals with AMD. Individuals (n = 244, mean age 39.9 ± 12.3 years) with AMD completed the PHQ-9, GAD-7, as well as other measures of depression, anxiety, and a structured diagnostic interview. The PHQ-9 and GAD-7 demonstrated good internal consistency (Cronbach's alpha 0.87 and 0.84, respectively). The PHQ-9 and GAD-7 showed a weak correlation with clinician-rated scales HAM-D and HAM-A (r = 0.316, p < 0.01, r = 0.307, p < 0.01, respectively). For the PHQ-9, a cut score of ≥11 resulted in 72% sensitivity and 72% specificity at recognizing depression symptoms. For the GAD-7, a cut score ≥7 resulted in 73% sensitivity and 54% specificity at recognizing any anxiety disorders. The confirmatory factor analysis suggested a two-factor structure ("cognitive/affectional", "somatic") for both the PHQ-9 and GAD-7. In conclusion, the PHQ-9 and GAD-7 have adequate formal psychometric properties as severity measures for symptoms of anxiety and depression in individuals with AMD. The PHQ-9 performs well as a screener using a cut score of ≥11. However, the clinical utility of the GAD-7 as a diagnostic tool for recognition of anxiety disorders is limited.

1. Introduction

Anxiety and mood disorders (AMD)[1] are two of the most common groups of psychiatric disorders in the world (Baxter et al., 2013; Ferrari et al., 2013; Stein et al., 2017). According to the World Health Organization (WHO) 300 million people experience depression (WHO, 2017), with the number of incident cases increasing by 49.9% from 1990 to 2017 worldwide (Liu et al., 2020). Considering anxiety disorders, it is estimated that around 264 million people live with this condition (WHO, 2017), with a 50% increase in the number of anxiety disorders worldwide from 1990 (Yang et al., 2021). In fact, the numbers could be even higher if we were to consider many cases of undiagnosed and untreated AMD (Kasper, 2006; Williams et al., 2017). Nowadays, the rates of AMD have been rapidly getting higher due to the coronavirus disease (COVID-19) pandemic. Since the start of COVID-19 pandemic, the global prevalence of anxiety and depression has increased by 25% (Collaborators, 2021).

If left untreated, AMD tend to develop chronically, with multiple recurrences throughout the lifetime (Yang et al., 2021; Ten Have et al., 2018), which can cause a great strain on medical resources, create socioeconomic burden. Therefore, increasing recognition and symptom severity measures could have a wide range of benefits in providing an

---

[1] In our work the term anxiety and mood disorders (AMD) refers to individuals with major depressive disorder, generalized anxiety disorder, panic disorder, agoraphobia, and social anxiety disorder.

adequate treatment.

Structured or semi-structured clinician-rated diagnostic interviews are considered to be the golden standard and intended to accurately determine AMD diagnoses, but they require significant time and resources to complete. In contrast, self-report measures, although sometimes seen not as accurate as clinician-rated measures, demand fewer resources and are easier to administer and score. Brief self-report questionnaires are an efficient way to screen individuals in need of specific clinical attention (Wisting et al., 2021). Given the high comorbidity rates between anxiety and mood disorders (Olfson et al., 2017), reliable and valid instruments are needed to assess depressive and anxiety symptoms in individuals with AMD for measuring the treatment process and timely recognition of the need for additional interventions.

Two of the most commonly used screening and severity measure scales for depression and anxiety are the Patient Health Questionnaire-9 (PHQ-9) and the Generalized Anxiety Disorder Questionnaire-7 (GAD-7) (Levis et al., 2019; Plummer et al., 2016). The PHQ-9 and GAD-7 are scales from a larger Patient Health Questionnaire (PHQ), created to efficiently assess the most common types of mental disorders presenting in primary care (Kroenke et al., 2010).

The PHQ-9 is a brief, nine-item self-report questionnaire, consisting of the nine criteria on which the diagnosis of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) 4th edition (DSM-IV) depressive disorder was based on (Kroenke and Spitzer, 2002). However, the PHQ-9 is still theoretically consistent with the 5th version of the DSM (DSM-5) (Spitzer et al., 2014). The PHQ-9 can be used in two ways: as a screening tool for depressive disorder or as a severity measure of depressive symptoms (Spitzer et al., 2014). The scale was developed in a primary care sample, with high internal consistency and high test-retest reliability (Kroenke et al., 2001). Overall, the PHQ-9 is considered to be a suitable screening tool and severity measure in a range of different populations and countries (Gilbody et al., 2007).

The GAD-7 is a brief, seven-item self-report questionnaire, created based on the symptom criteria for generalized anxiety disorder (GAD) described in the DSM-IV (Spitzer et al., 2006). Although originally targeted towards the assessment of GAD, the GAD-7 was extended to screen for other anxiety disorders, such as social anxiety and panic disorders (Kroenke et al., 2007). The scale can be used as a screener for anxiety disorders or serve as a severity measure of anxiety symptoms (Spitzer et al., 2006). Although the GAD-7 has been studied to a lesser extent than the PHQ-9, it is also considered to be a reliable questionnaire for screening and measuring the severity of anxiety symptoms in different environments and cultures (Parkerson et al., 2015; Plummer et al., 2016).

The reliability of the PHQ-9 and GAD-7 as measures of depression and anxiety is well established in primary care (Kroenke et al., 2001; Spitzer et al., 2006). However, the same characteristics that apply to primary care may not be appropriate in secondary care psychiatric settings, where patients have more serious, complex or concurrent psychiatric disorders. There have been some studies regarding the validation of the PHQ-9 and GAD-7 in secondary care psychiatric samples: the questionnaires both show high internal consistency in individuals attending inpatient and outpatient psychiatric units, with Cronbach's alpha ranging from 0.87 to 0.88 for the PHQ-9; from 0.85 to 0.95 for the GAD-7 (Beard and Bjorgvinsson, 2014; Beard et al., 2016; Dadfar et al., 2018; Feng et al., 2016; Johnson et al., 2019; Rutter and Brown, 2017; Sawaya et al., 2016; Shin et al., 2019).

However, there have been some inconsistencies reported in previous studies that need to be considered before applying the PHQ-9 and GAD-7 to individuals with AMD.

Starting with the PHQ-9, a common problem seems to be the optimal cut score and optimal balance between sensitivity and specificity. In primary care a cut score of 10 has been recommended for detecting cases of current depressive disorder (Kroenke and Spitzer, 2002) but the same cut score has not been optimal when used in psychiatric settings (Beard et al., 2016; Inoue et al., 2012; Johnson et al., 2019). A study of

Lebanese outpatients attending psychiatric unit found that the PHQ-9 with a cut score of 10 was sensitive but not specific at capturing depressive symptoms (Sawaya et al., 2016), meaning it could not be used for diagnostic purposes. Beard et al. (2016) found the optimal balance between sensitivity and specificity using a cut score of 13. A higher cut score may be needed due to the specificities of secondary care compared to primary care: psychiatric samples in secondary care have higher rates of comorbidity, symptom severity, as well as the fact that most psychiatric disorders have depressive symptoms. Another issue that needs to be considered is the mixed factor structure of the PHQ-9. Different factors could help conceptualize the results better, reflecting on independent but inter-correlated groups of symptoms. However, there has been evidence of one-factor (Dadfar et al., 2018; Rutter and Brown, 2017) as well as two-factor (Beard et al., 2016) structure of the PHQ-9.

Similar inconsistencies can be spotted in the GAD-7. Regarding sensitivity and specificity, Beard and Bjorgvinsson (2014) found that the GAD-7 yielded adequate sensitivity but poor specificity when using the recommended cut score of 10 for primary care samples, meaning that the GAD-7 was not specific at capturing anxiety symptoms. Similar findings of good sensitivity but poor specificity were found by Rutter and Brown (2017), while another study by Sawaya et al. (2016) found the GAD-7 to be neither sensitive nor specific at capturing anxiety symptoms. A study by Johnson et al. (2019) in a heterogeneous psychiatric sample found the optimal balance between sensitivity and specificity at the cut score of 8. Not many studies have looked into the factor structure of the GAD-7, however some claim that the GAD-7 fits into the one-factor (Johnson et al., 2019), while others see the GAD-7 as having a two-factor structure (Beard and Bjorgvinsson, 2014; Rutter and Brown, 2017).

While both the PHQ-9 and GAD-7 have been translated for use in various countries and settings, no studies have investigated the psychometric properties of the Lithuanian version of the PHQ-9 and GAD-7 in individuals with AMD. Therefore, the aim of this study is to evaluate psychometric properties of the PHQ-9 and GAD-7 in a sample of Lithuanian individuals with AMD, who are attending a secondary care treatment clinic. Specifically, we investigated internal consistency, convergent validity, factor structure as well as sensitivity and specificity of both the PHQ-9 and GAD-7. We hypothesized that the Lithuanian versions of the PHQ-9 and GAD-7 would exhibit acceptable psychometric properties in individuals with AMD.

2. Methods

2.1. Study participants and procedure

Participants included in the study were receiving treatment for the first time at the Stress-related Disorders Department of Lithuanian University of Health Sciences Neuroscience Institute Palanga Clinic, Lithuania from April of 2018 to September of 2022. Individuals included in the study had a current diagnosis of anxiety and/or mood disorder, established according to the DSM-5 diagnostic criteria (American Psychiatric Association, 2013), using the Mini-International Neuropsychiatric Interview (M.I.N.I. 7.0.2) (Sheehan et al., 1998). The participants had to be older than 18 years old. The exclusion criteria were current severe somatic illness (e.g., oncological illness, thyroid related disorders), cognitive impairment, psychotic symptoms, high suicidal risk or individuals' inability to speak Lithuanian fluently. All participants received standard treatment for AMD consisting of psychopharmacological and psychotherapeutic interventions based on their clinical needs.

The study protocol was approved by the Lithuanian Biomedical Research Ethics Committee (reference No. B-2-38), which is in line with the principles outlined by the Declaration of Helsinki. Before inclusion into the study, each participant signed informed consent forms. All participants were evaluated within the first 5 days of the admission to

the clinic. Participants were interviewed by a trained clinical psychologist and completed self-evaluation questionnaires.

2.2. Measures

The M.I.N.I. 7.0.2 (Sheehan et al., 1998) was used to screen for DSM-5 disorders. The M.I.N.I. is organized into modules, with each module involving one to four screening questions, which are used to rule out the diagnosis when answered negatively. Positive answers to screening questions are explored by further investigation of diagnostic criteria. The permission to use M.I.N.I. was gathered from the original authors.

The PHQ-9 is a brief nine-item self-report measure used to assess depression symptoms and severity (Kroenke et al., 2001). The individuals had to rate the severity of their symptoms over the past two weeks using a four-point Likert scale with possible responses ranging from "not at all" to "nearly every day". The total scores range from 0 to 27 with higher scores indicating higher prevalence of depressive symptoms. Originally, a cut score of ≥ 10 has been recommended for detecting cases of current major depressive episode (Kroenke and Spitzer, 2002). The Lithuanian version of the PHQ-9 which is available at the PHQ Screeners website https://www.phqscreeners.com/was used for the current research.

The GAD-7 is a seven-item self-report scale used to measure the symptoms and severity of GAD in the past two weeks (Spitzer et al., 2006). Severity of each symptom is scored on a scale from 0 "not at all" to 3 "nearly every day". The sum of scores ranges from 0 to 21, with higher scores showing greater anxiety symptom severity. Originally, a cut score of ≥ 10 has been recommended for identifying cases of GAD (Spitzer et al., 2006). The Lithuanian version of the GAD-7 which is available at the PHQ Screeners website https://www.phqscreeners.com/was used for the current study.

The Hamilton Depression Rating Scale (HAM-D) is a clinician-administered depression assessment scale (Hamilton, 1960); it includes 17 questions about symptoms of depression experienced over the past week. The assessment of HAM-D in the current study was completed by trained clinical psychologists. Cronbach's alpha for the measure in the current study was 0.80, while McDonald's Omega was 0.81.

The Hamilton Anxiety Rating Scale (HAM-A) is a clinician-rated scale developed to measure the severity of anxiety symptoms (Hamilton, 1959); it is a 14-item scale measuring different series of anxiety symptoms, including psychological anxiety and somatic anxiety. In this study the HAM-A was completed by trained clinical psychologists. Cronbach's alpha for the measure in the current study was 0.88, while McDonald's Omega was 0.88.

2.3. Statistical analyses

Data was analyzed using IBM SPSS Statistics for Windows (version 28) (SPSS Inc, Chicago, IL, USA) and IBM SPSS AMOS 28 (IBM Corp., Armonk, NY, USA). The recommended sample size for the psychometric validation was considered to be 10:1 respondent-to-item ratio (Kline, 1979) with 200 participants considered to be an adequate sample size (Tsang et al., 2017). Descriptive statistics were used to explore the samples' clinical and sociodemographic characteristics.

The scales' internal consistency was estimated by both Cronbach's alpha coefficient and McDonald's omega. According to the recent literature, the McDonald's omega will take precedence over the Cronbach's alpha (Hayes and AuthorAnonymous, 2020; McNeish, 2018). Values between 0.80 and 0.95 will be considered acceptable (Boateng, 2018). Convergent validity was assessed using correlations with well-established measures of depression (HAM-D) and anxiety (HAM-A).

Screening parameters including sensitivity, specificity, and positive predictive values were calculated for the PHQ-9 and GAD-7. Several criteria were used to calculate sensitivity, specificity and positive predictive values of the PHQ-9 and GAD-7. First, Receiver operating

characteristics (ROC) analysis was used to calibrate a screening measure (the PHQ-9 and the GAD-7) against a diagnostic instrument M.I.N.I. By using this curve, it is possible to evaluate the rate of true positive value, which represents the sensitivity of the test, against the rate of a false positive value, representing the specificity of the scale. Therefore, an optimum cut score of the screening scale is determined (Hosmer et al., 2013). The utility of the questionnaires was based on the area under the curve (AUC), which is an overall index of the accuracy of the discrimination provided by the questionnaires, with an AUC of >0.50 indicating good chances of discrimination (Hajian-Tilaki, 2013).

Finally, a confirmatory factor analyses (CFA) were used to identify if the models fit the two-factor predicted by other studies. The model fit was evaluated using the Chi-Square test and the following indices: comparative fit index (CFI; ≥ 0.90 adequate, ≥ 0.95 good), Tucker-Lewis index (TLI; ≥ 0.90 adequate, ≥ 0.95 good), normed fit index (NFI; ≥ 0.90 adequate, ≥ 0.95 good), the goodness of fit index (GFI; ≥ 0.90 adequate), root mean square error of approximation with its 90% confidence interval (RMSEA; $0.10 \leq$ acceptable. ≤ 0.08 adequate, and ≤ 0.05 good), and standardized root mean square residual (SRMR; ≤ 0.08 good) (Sahoo, 2019).

3. Results

3.1. Study sample characteristics

A total of 244 individuals (22.1% men, 77.9% women; mean age 39.9 ± 12.3 years) participated in the study. Out of all the participants, 54.9% had a comorbid AMD diagnosis, while the most common current diagnosis was major depressive disorder (61.9%), followed by GAD (39.3%). Other comorbidities noted in the study were alcohol use disorder, bulimia nervosa, and substance abuse disorder. Detailed characteristics of study participants are shown in Table 1.

Table 1
Baseline characteristics of study participants.

	Total (n = 244)
Age, mean (SD)	39.9 (12.3)
Gender, n (%)	
Men	54 (22.1)
Women	190 (77.9)
Education, n (%)	
Tertiary education	74 (30.3)
College/University degree	170 (69.7)
Anxiety and mood disorders, n (%)	
Mood	47 (19.3)
Anxiety	63 (25.8)
Anxiety and mood	134 (54.9)
Diagnosis, n (%)	
Major depressive disorder	151 (61.9)
Generalized anxiety disorder	96 (39.3)
Panic disorder	85 (34.8)
Agoraphobia	61 (25.0)
Social anxiety disorder	56 (23.0)
Substance abuse disorder	6 (2.5)
Bulimia nervosa	7 (2.9)
Alcohol use disorder	31 (12.7)
Current medication use, n (%)	
Antidepressants	176 (72.1)
Benzodiazepines	91 (37.3)
Mood stabilizers	7 (2.9)
Antipsychotics	58 (23.8)
PHQ-9 scores, mean (SD)	12.5 (6.0)
GAD-7 scores, mean (SD)	8.1 (3.6)
HAM-D scores, mean (SD)	14.6 (7.0)
HAM-N scores, mean (SD)	23.2 (9.8)

Note: PHQ-9 – Patient Health Questionnaire-9; GAD-7 – Generalized Anxiety Disorder-7; HAM-D – Hamilton Depression Rating Scale; HAM-N – Hamilton Anxiety Rating Scale.

3.2. Psychometric characteristics and factor structure of the PHQ-9

Reliability analysis demonstrated adequate internal consistency for the PHQ-9. The PHQ-9 had a Cronbach's alpha of 0.87 and a McDonald's Omega of 0.87.

Analysing convergent validity, higher scores on the PHQ-9 were associated with higher scores on measures of depression, as evaluated by the HAM-D scale. The strength of the association was weak ($r = 0.316$, $p < 0.01$).

Sensitivities, specificities and positive predictive values for varying cut scores were determined for the PHQ-9 and are illustrated in Table 2. Analysing the results, we found a good discriminatory power recognizing individuals possibly having major depressive episode (AUC = 0.767, $p < 0.001$; Fig. 1). The ROC curve suggested a cut score of ≥11. Using a cut score of 11, the PHQ-9 showed good psychometric characteristics identifying current major depressive disorder: at this score, the PHQ-9 resulted in 72% sensitivity, 72% specificity and 81% positive predictive value identifying the individuals with current major depressive disorder.

Finally, factorial loadings for different dimensional models for the PHQ-9 are depicted in Fig. 2. The two-factor model was found to be a better fit than a single factor model (Table 3). The two factors that emerged were related to cognitive/affective symptoms (Cronbach's alpha = 0.798) and somatic symptoms (Cronbach's alpha = 0.782). The two subscales showed strong positive correlation with each other ($r = 0.700$, $p < 0.01$).

3.3. Psychometric characteristics and factor structure of the GAD-7

Reliability analysis demonstrated adequate internal consistency for the GAD-7. The GAD-7 had a Cronbach's alpha of 0.84 and a McDonald's Omega of 0.83.

Analysing convergent validity, higher scores on the GAD-7 scale were associated with higher scores on measures of anxiety, as evaluated by the HAM-A scale. The strength of the association was weak ($r = 0.307$, $p < 0.01$).

Sensitivities, specificities and positive predictive values for varying cut scores were determined for the GAD-7 and are illustrated in Table 2. Analysing the results for the GAD-7 in predicting GAD, panic disorder, agoraphobia or social anxiety disorder, the analysis indicated that a cut score of ≥7 displayed the most balance between sensitivity and specificity measures (AUC = 0.670, $p < 0.001$; Fig. 3). Using a cut score of 7, the GAD-7 showed good sensitivity (73%) but poor specificity (54%) in predicting those who currently met diagnostic criteria for GAD, panic disorder, agoraphobia or social anxiety disorder.

Analysing the results for the GAD-7 in predicting GAD, the analysis indicated that using cut score of ≥7 displayed the most balance between sensitivity and specificity numbers, however, the measure was not

Fig. 1. Receiver Operating Characteristic Curve of the Patient Health Questionnaire-9 compared with M.I.N.I. 7.0.2 as the reference standard for major depression episode

Note: PHQ-9 – Patient Health Questionnaire-9; AUC – area under the curve.

statistically significant (AUC = 0.504, $p = 0.924$; Fig. 4). Using a cut score of 7, the GAD-7 showed poor sensitivity (55%) and specificity (39%) in predicting those who currently met diagnostic criteria for GAD.

Finally, factorial loadings for different dimensional models for the GAD-7 can be seen in Fig. 5. The CFA for the GAD-7 suggested a two-factor structure and indicated adequate fit to the model (Table 3). The two factors that emerged were related to cognitive/affective symptoms (Cronbach's alpha = 0.807) and somatic symptoms (Cronbach's alpha = 0.647). The two factors showed strong positive correlations with each other ($r = 0.688$, $p < 0.01$).

4. Discussion

Our study evaluated psychometric properties of the PHQ-9 and GAD-7 questionnaires in a sample of individuals with AMD in Lithuania. Our research proposed that the PHQ-9 and GAD-7 are adequate instruments for measuring depression and anxiety severity in individuals with AMD.

Our study revealed that the PHQ-9 and GAD-7 showed good internal

Table 2

Sensitivity, specificity, and positive predictive values for the PHQ-9 and the GAD-7 cut score levels.

Cut score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)
PHQ-9 for major depression episode			
PHQ-9≥10	76.2 (68.6–82.7)	66.7 (56.1–76.1)	78.8 (73.3–83.4)
PHQ-9≥11	72.2 (64.3–79.2)	72.0 (61.8–80.9)	80.7 (74.9–85.5)
PHQ-9≥12	66.2 (58.1–73.7)	75.3 (65.2–83.6)	81.3 (75.0–86.3)
GAD-7 for GAD, PA, agoraphobia, or SAD			
GAD-7≥6	60.3 (47.2–83.4)	65.2 (57.8–72.1)	37.6 (31.3–44.4)
GAD-7≥7	73.0 (60.3–83.4)	53.6 (46.0–61.0)	35.4 (30.6–40.5)
GAD-7≥8	77.8 (65.5–87.3)	44.8 (37.4–52.3)	32.9 (28.9–37.1)
GAD-7 for GAD			
GAD-7≥6	70.8 (60.7–79.7)	29.1 (21.9–37.1)	39.3 (35.5–43.3)
GAD-7≥7	55.2 (44.7–65.4)	39.2 (31.3–47.5)	37.1 (32.1–42.4)
GAD-7≥8	45.8 (35.6–56.3)	52.7 (44.3–61.0)	38.6 (32.3–45.3)

Note: PHQ-9 – Patient Health Questionnaire-9; GAD-7 – Generalized Anxiety Disorder-7; GAD – generalized anxiety disorder, PA – panic disorder, SAD – social anxiety disorder.
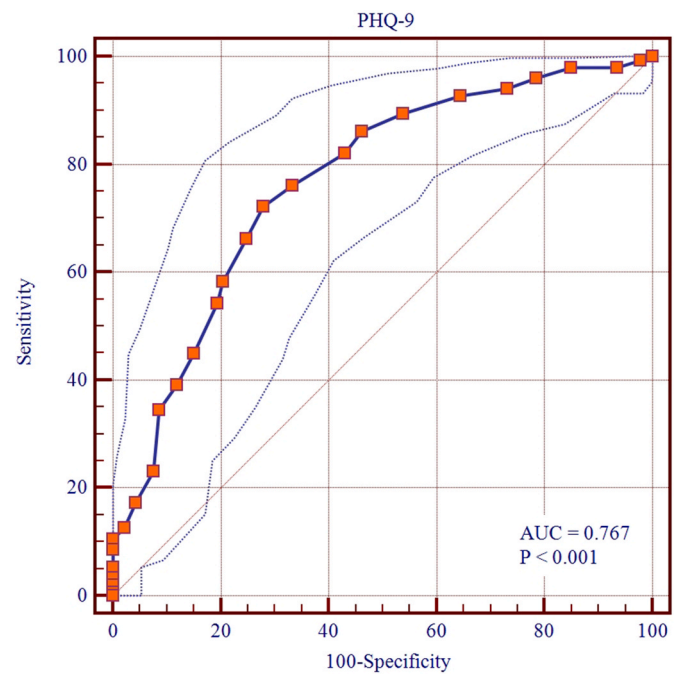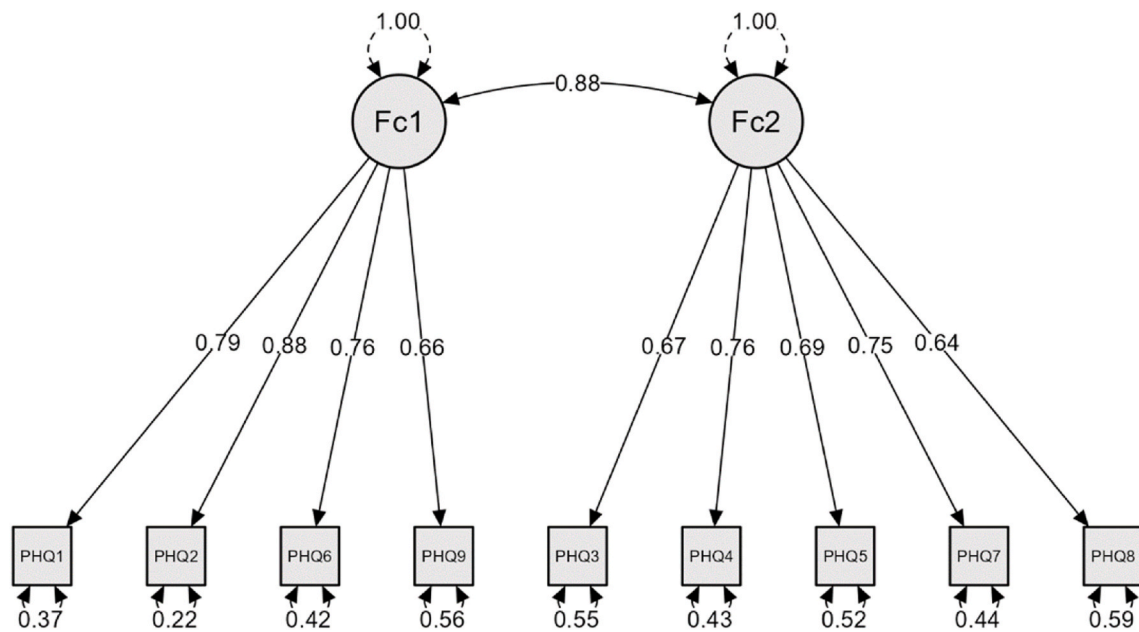
Fig. 2. Factorial Loadings for Different Dimensional Models for the PHQ-9
Note: Fc – Factor; PHQ1 – loss of interest; PHQ2 – feeling depressed; PHQ6 – feeling bad about yourself; PHQ9 – suicidal thoughts; PHQ3 – sleep problems; PHQ4 – loss of energy; PHQ5 – poor appetite or overeating; PHQ7 – trouble concentrating; PHQ8 – being slower or more restless.

Table 3
Goodness-of-Fit Indices for the PHQ-9 and the GAD-7 models.

Model	χ^2	df	χ^2/df	RMSEA	CFI	NFI	TLI	GFI	SRMR
Recommended			≤5	<0.08	≥0.90	≥0.95	≥0.95	≥0.90	≤0.08
PHQ-9									
One factor [a]	64.3	27	2.38	0.071	0.955	0.926	0.941	0.942	0.057
Two factors [b]	44.7	26	1.72	0.051	0.978	0.949	0.969	0.960	0.047
GAD-7									
One factor [c]	23.8	14	1.70	0.054	0.983	0.960	0.975	0.974	0.035
Two factors [d]	22.3	13	1.72	0.054	0.984	0.963	0.974	0.977	0.033

Note: PHQ-9 – Patient Health Questionnaire-9; GAD-7 – Generalized Anxiety Disorder-7; df – degrees of freedom; RMSEA – root-mean square error of approximation; CFI – comparative fit index; NFI – normed fit index; TLI – Tucker-Levis index; GFI – Goodness of Fit Index; SRMR – Standardized Root Mean Square Residual.
[a] One-factor model by sum of all nine items of the PHQ-9.
[b] Two-factor model of the PHQ-9 measures cognitive and affective symptoms (items 1, 2, 6, 9) and somatic symptoms (items 3, 4, 5, 7, 8).
[c] One-factor model by sum of all seven items of the GAD-7.
[d] Two-factor model of the GAD-7 measures cognitive and affective symptoms (items 1, 2, 3, 7) and somatic symptoms (items 4–6).

consistency. This is in line with previous studies in similar psychiatric samples (Dadfar et al., 2018; Shin et al., 2019). In our research, both questionnaires were weakly correlated with other measures of depression and anxiety, measured by the HAM-D and HAM-A scales respectively. Studies that have used HAM-D as a measure for construct validity have found moderate correlations between the HAM-D and PHQ-9 (Feng et al., 2016; Shin et al., 2019). Dissimilarities between the questionnaires can arise from different means of assessment, as the HAM-D and HAM-A both are observer-rated scales, while the PHQ-9 and GAD-7 are self-rating scales. There has been some critique for observer-rated scales and their suitability for clinical evaluation (Uher et al., 2012). For example, a study by Ma et al. (2021) has found that the PHQ-9 was better than HAM-D in distinguishing the severity of depression in individuals with major depressive disorder. Overall, self-rated and clinician-rated scales can differ from one another in the information gathered and are not completely equivalent to each other, so low correlations should not be considered as a limitation of the instruments.

Summarising the results of diagnostic properties of the scales, the PHQ-9 showed good sensitivity, specificity and positive predictive value at a cut score of 11 points as a screening tool for the presence or absence of major depressive episode. This cut score showed more balanced results between sensitivity and specificity when compared to the recommended cut score of 10 (Kroenke and Spitzer, 2002). The optimal cut score in our study was lower in comparison to other studies, where the optimal cut score was found to be 13 points (Beard et al., 2016; Inoue et al., 2012). The main difference between these studies is that they included wide heterogeneous psychiatric samples, while our study sample consisted only of individuals with AMD. This result highlights the importance of adjusting cut score points for particular psychiatric samples.

Analysing the results for the GAD-7, the questionnaire displayed good sensitivity but poor specificity with the optimal cut score of 7 as a screening tool for the presence or absence of GAD, panic disorder, agoraphobia or social anxiety disorder. This is lower than the originally recommended cut score of 10 (Spitzer et al., 2006). However, when screening only for GAD, the questionnaire did not have an optimal cut score and displayed poor sensitivity and specificity at the cut score of 10. Other research on this topic have shown mixed results. In a study by Johnson et al. (2019) a cut score of 8 was the optimal point for sensitivity and specificity, while Rutter and Brown (2017) found that no cut scores demonstrated adequately balanced sensitivity and specificity. However, there have been some studies reporting good sensitivity and
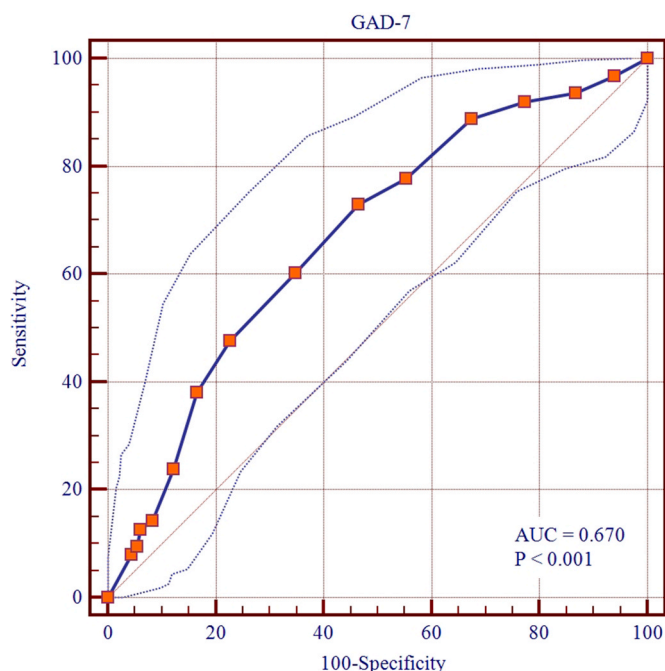
Fig. 3. Receiver Operating Characteristic Curve of the Generalized Anxiety Disorder-7 compared with M.I.N.I. 7.0.2 as the reference standard for generalized anxiety disorder, panic disorder, agoraphobia or social anxiety disorder Note: GAD-7 – Generalized Anxiety Disorder-7; AUC – area under the curve.
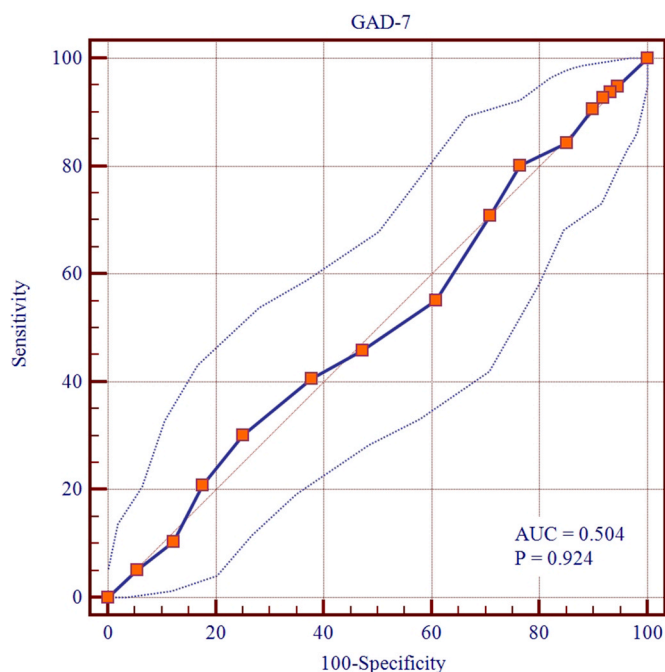

Fig. 4. Receiver Operating Characteristic Curve of the Generalized Anxiety Disorder-7 compared with M.I.N.I. 7.0.2 as the reference standard for generalized anxiety disorder
Note: GAD-7 – Generalized Anxiety Disorder-7; AUC – area under the curve.

specificity for the scale (Garcia-Campayo et al., 2010). A study by Delgadillo et al. (2012) found that the GAD-7 had good sensitivity and specificity when finding any anxiety disorders, but had low specificity when screening specifically for GAD in an outpatient substance abuse treatment population. Taking our results and other research into account, it could be hypothesized that the GAD-7 questionnaire measures

negative affect more broadly, rather than specifically differentiating for GAD. Therefore, the GAD-7 is unlikely to be useful as a screening tool for GAD in a sample with AMD. In the future research and clinical practice, the questionnaire could be used with additional assessment to confirm anxiety disorder diagnosis as the GAD-7 may be more useful as a severity measurement rather than a screening tool for the presence or absence of GAD.

Findings regarding the factorial validity confirmed two-factor structure for both the PHQ-9 and GAD-7, consisting of one question group about cognitive/affective symptoms, and the second question group regarding somatic symptoms for both scales. Previous studies regarding the factor structure of the PHQ-9 and GAD-7 have been inconclusive. Some studies suggest the original one-factor model (Dadfar et al., 2018; Johnson et al., 2019; Kroenke et al., 2010; Rutter and Brown, 2017; Spitzer et al., 2006), while others proposed two-factor models for the PHQ-9 (Beard et al., 2016; Petersen et al., 2015), as well as for the GAD-7 (Beard and Bjorgvinsson, 2014). A study by Doi et al. (2018) suggested that the bi-factor model had the best fit for the PHQ-9 in both non-clinical and clinical populations. The bi-factor model allows the use of the cut score and the total score as a single variable, as well as a two-factor model for assessing more detailed symptoms. However, our study proposes that for individuals with AMD, the PHQ-9 and GAD-7 explore two different symptom dimensions and their examination for cognitive/affective and somatic domains separately could yield more comprehensive results in research as well as clinical practice.

Our results contribute an important clinical nuance in the use and interpretation of the PHQ-9 and GAD-7 in individuals with AMD. A healthy control group could help solidify these nuances. Even though our study did not have a healthy control group, there has been a study evaluating the psychometric properties of the PHQ-9 and GAD-7 in a sample of Lithuanian university students (Pranckeviciene et al., 2022). The study found that using a cut score of ≥10 the PHQ-9 resulted in 71% sensitivity and 66% specificity recognizing students with AMD. For the GAD-7, a cut score of ≥9 resulted in 73% sensitivity and 70% specificity recognizing students at risk (Pranckeviciene et al., 2022). These results differ from ours, highlighting the nuances of assessment in different populations.

4.1. Strengths and limitations

The strengths of our current study include an examination of several important psychometric characteristics such as internal consistency, convergent validity, factor structure, as well as sensitivity and specificity, related to the practical and academic use of the PHQ-9 and GAD-7 in a secondary care psychiatric sample of individuals with AMD.

However, our study should be interpreted in the context of its design and limitations. First, our study did not evaluate the test-retest reliability, thus preventing us from making interpretations of the questionnaires' sensitivity to change. Second, the current study was completed in a convenience sample of individuals with AMD from one treatment facility, so the generalizability to other clinical and cultural samples should be considered with caution. Third, our findings were made in a sample of individuals with AMD in secondary care, therefore the results cannot be generalized to other types of psychiatric samples. Finally, due to the relatively small sample size and overrepresentation of female participants in the current sample, we did not compare the psychometric properties of the PHQ-9 and GAD-7 based on gender differences. Future research with larger and more diverse samples may help to further explore the applicability of the PHQ-9 and GAD-7 in individuals with AMD.

5. Conclusions

In conclusion, this is the first study showing the PHQ-9 and GAD-7 as adequate instruments to evaluate depression and anxiety symptoms in Lithuanian individuals with AMD. The PHQ-9 showed adequate
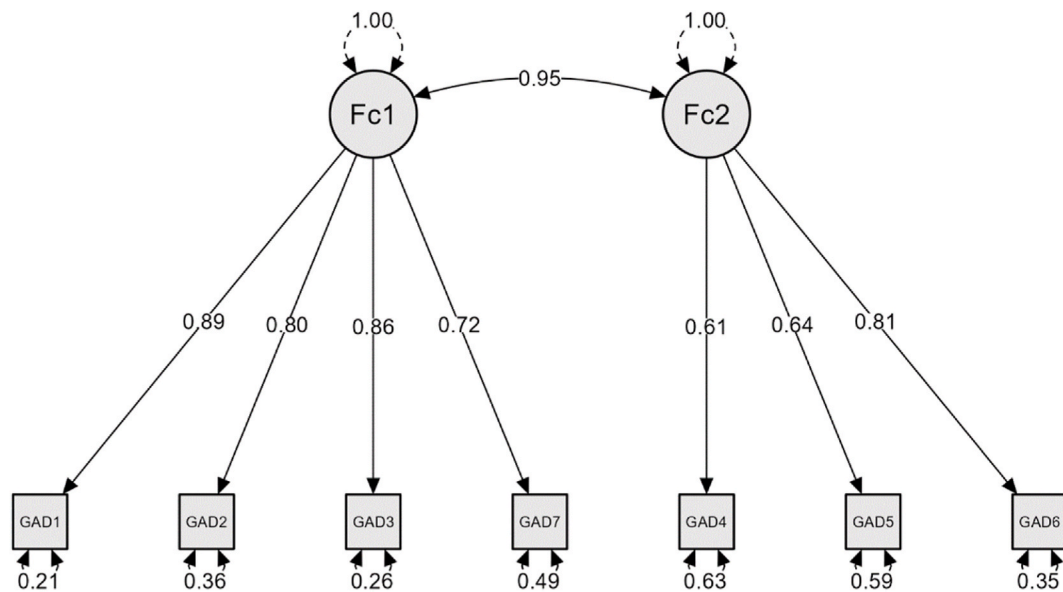
Fig. 5. Factorial Loadings for Different Dimensional Models for the GAD-7
Note: GAD –Fc – Factor; GAD1 – feeling nervous; GAD2 – not able to stop worrying; GAD3 – worry about different things; GAD7 – feeling afraid; GAD4 – trouble relaxing; GAD5 – being restless; GAD6 – being easily annoyed or irritable.

psychometric properties as a measure of symptom severity for depression. The PHQ-9 also performed well as a screening tool for detecting major depressive disorder using a higher cut score than it has been previously recommended in primary care settings. The GAD-7 was a reliable tool for assessing anxiety severity, however, due to low specificity did not have indications for adequately assessing any anxiety disorders or GAD specifically in individuals with AMD in secondary care.

Author statement

Category 1.
Conception and design of the study: V. Steibliene, J. Burkauskas.
Acquisition of data: J. Gecaite-Stonciene, J. Burkauskas.
Analysis and/or interpretation of data: A. Podlipskyte, A. Stanyte, J. Burkauskas.
Category 2.
Drafting the manuscript: A. Stanyte.
Revising the manuscript critically for important intellectual content: N.A. Fineberg, J. Gecaite-Stonciene, J. Macijauskiene, V. Steibliene, J. Burkauskas.
Category 3.
Approval of the version of the manuscript to be published: A. Stanyte, N.A. Fineberg, A. Podlipskyte, J. Gecaite-Stonciene, J. Macijauskiene, V. Steibliene, J. Burkauskas.

Role of the funding source

The Funding Source had no involvement in the conduct of the research and preparation of the article.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:
1) Naomi A. Fineberg reports in the past 5 years personal fees from Taylor and Francis, Oxford University Press, Global Mental Health Academy and Elsevier; personal fees and non-financial support from Sun; non-financial support from RCPsych, CINP, WPA, the International Forum of Mood and Anxiety Disorders, ECNP and the Indian Association of Biological Psychiatry; grants from the Wellcome, UKRI, Orchard and

Acknowledgments

References

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, fifth ed. Author.
Baxter, A.J., Scott, K.M., Vos, T., Whiteford, H.A., 2013. Global prevalence of anxiety disorders: a systematic review and meta-regression. Psychol. Med. 43 (5), 897–910. https://doi.org/10.1017/S003329171200147X.
Beard, C., Bjorgvinsson, T., 2014. Beyond generalized anxiety disorder: psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. J. Anxiety Disord. 28 (6), 547–552. https://doi.org/10.1016/j.janxdis.2014.06.002.
Beard, C., Hsu, K.J., Rifkin, L.S., Busch, A.B., Bjorgvinsson, T., 2016. Validation of the PHQ-9 in a psychiatric sample. J. Affect. Disord. 193, 267–273. https://doi.org/10.1016/j.jad.2015.12.075.
Boateng, G.O., Neilands, T.B., Frongillo, E.A., Melgar-Quinonez, H.R., Young, S.L., 2018. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front. Public Health 6 (149). https://doi.org/10.3389/fpubh.2018.00149.
Collaborators, C.-M.D., 2021. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 398 (10312), 1700–1712. https://doi.org/10.1016/S0140-6736(21)02143-7.
Dadfar, M., Kalibatseva, Z., Lester, D., 2018. Reliability and validity of the farsi version of the patient health questionnaire-9 (PHQ-9) with Iranian psychiatric outpatients. Trends Psychiatry Psychother 40 (2), 144–151. https://doi.org/10.1590/2237-6089-2017-0116.
Delgadillo, J., Payne, S., Gilbody, S., Godfrey, C., Gore, S., Jessop, D., Dale, V., 2012. Brief case finding tools for anxiety disorders: validation of GAD-7 and GAD-2 in

addictions treatment. Drug Alcohol Depend. 125 (1–2), 37–42. https://doi.org/10.1016/j.drugalcdep.2012.03.011.

Feng, Y., Huang, W., Tian, T.F., Wang, G., Hu, C., Chiu, H.F., Ungvari, G.S., Kilbourne, A.M., Xiang, Y.T., 2016. The psychometric properties of the quick inventory of depressive symptomatology-self-report (QIDS-SR) and the patient health questionnaire-9 (PHQ-9) in depressed inpatients in China. Psychiatr. Res. 243, 92–96. https://doi.org/10.1016/j.psychres.2016.06.021.

Ferrari, A.J., Somerville, A.J., Baxter, A.J., Norman, R., Patten, S.B., Vos, T., Whiteford, H.A., 2013. Global variation in the prevalence and incidence of major depressive disorder: a systematic review of the epidemiological literature. Psychol. Med. 43 (3), 471–481. https://doi.org/10.1017/S0033291712001511.

Garcia-Campayo, J., Zamorano, E., Ruiz, M.A., Pardo, A., Perez-Paramo, M., Lopez-Gomez, V., Freire, O., Rejas, J., 2010. Cultural adaptation into Spanish of the generalized anxiety disorder-7 (GAD-7) scale as a screening tool. Health Qual. Life Outcome 8, 8. https://doi.org/10.1186/1477-7525-8-8.

Gilbody, S., Richards, D., Brealey, S., Hewitt, C., 2007. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J. Gen. Intern. Med. 22 (11), 1596–1602. https://doi.org/10.1007/s11606-007-0333-y.

Hajian-Tilaki, K., 2013. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 4 (2), 627–635. https://www.ncbi.nlm.nih.gov/pubmed/24009950.

Hamilton, M., 1959. The assessment of anxiety states by rating. Br. J. Med. Psychol. 32 (1), 50–55. https://doi.org/10.1111/j.2044-8341.1959.tb00467.x.

Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23, 56–62. https://doi.org/10.1136/jnnp.23.1.56.

Hayes, A.F.C., J, J., 2020. Use omega rather than cronbach's alpha for estimating reliability. But. Commun. Methods Meas. 14, 1–24. https://doi.org/10.1080/19312458.2020.1718629.

Hosmer, D.W.J., Lemeshow, S., Sturdivant, R.X., 2013. Applied Logistic Regression. John Wiley & Sons.

Inoue, T., Tanaka, T., Nakagawa, S., Nakato, Y., Kameyama, R., Boku, S., Toda, H., Kurita, T., Koyama, T., 2012. Utility and limitations of PHQ-9 in a clinic specializing in psychiatric care. BMC Psychiatr. 12, 73. https://doi.org/10.1186/1471-244X-12-73.

Johnson, S.U., Ulvenes, P.G., Oktedalen, T., Hoffart, A., 2019. Psychometric properties of the general anxiety disorder 7-item (GAD-7) scale in a heterogeneous psychiatric sample. Front. Psychol. 10, 1713. https://doi.org/10.3389/fpsyg.2019.01713.

Kasper, S., 2006. Anxiety disorders: under-diagnosed and insufficiently treated. Int. J. Psychiatr. Clin. Pract. 10 (Suppl. 1), 3–9. https://doi.org/10.1080/13651500600552297.

Kline, P., 1979. Psychometrics and Psychology, vol. 381. Academic Press, London.

Kroenke, K., Spitzer, R.L., 2002. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr. Ann. 32 (9), 509–515.

Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16 (9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Kroenke, K., Spitzer, R.L., Williams, J.B., Monahan, P.O., Lowe, B., 2007. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann. Intern. Med. 146 (5), 317–325. https://doi.org/10.7326/0003-4819-146-5-200703060-00004.

Kroenke, K., Spitzer, R.L., Williams, J.B., Lowe, B., 2010. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen. Hosp. Psychiatry 32 (4), 345–359. https://doi.org/10.1016/j.genhosppsych.2010.03.006.

Levis, B., Benedetti, A., Thombs, B.D., Collaboration, D.E.S.D., 2019. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 365, l1476. https://doi.org/10.1136/bmj.l1476.

Liu, Q., He, H., Yang, J., Feng, X., Zhao, F., Lyu, J., 2020. Changes in the global burden of depression from 1990 to 2017: findings from the global burden of disease study. J. Psychiatr. Res. 126, 134–140. https://doi.org/10.1016/j.jpsychires.2019.08.002.

Ma, S., Yang, J., Yang, B., Kang, L., Wang, P., Zhang, N., Wang, W., Zong, X., Wang, Y., Bai, H., Guo, Q., Yao, L., Fang, L., Liu, Z., 2021. The patient health questionnaire-9 vs. The Hamilton rating scale for depression in assessing major depressive disorder. Front. Psychiatr. 12, 747139 https://doi.org/10.3389/fpsyt.2021.747139.

McNeish, D., 2018. Thanks coefficient alpha, we'll take it from here. Psychol. Methods 23 (3), 412–433. https://doi.org/10.1037/met0000144.

Olfson, M., Mojtabai, R., Merikangas, K.R., Compton, W.M., Wang, S., Grant, B.F., Blanco, C., 2017. Reexamining associations between mania, depression, anxiety and

substance use disorders: results from a prospective national cohort. Mol. Psychiatr. 22 (2), 235–241. https://doi.org/10.1038/mp.2016.64.

Parkerson, H.A., Thibodeau, M.A., Brandt, C.P., Zvolensky, M.J., Asmundson, G.J., 2015. Cultural-based biases of the GAD-7. J. Anxiety Disord. 31, 38–42. https://doi.org/10.1016/j.janxdis.2015.01.005.

Petersen, J.J., Paulitsch, M.A., Hartig, J., Mergenthal, K., Gerlach, F.M., Gensichen, J., 2015. 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. 170, 138–142. https://doi.org/10.1016/j.jad.2014.08.053.

Plummer, F., Manea, L., Trepel, D., McMillan, D., 2016. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen. Hosp. Psychiatry 39, 24–31. https://doi.org/10.1016/j.genhosppsych.2015.11.005.

Pranckeviciene, A., Saudargiene, A., Gecaite-Stonciene, J., Liaugaudaite, V., Griskova-Bulanova, I., Simkute, D., Naginiene, R., Dainauskas, L.L., Ceidaite, G., Burkauskas, J., 2022. Validation of the patient health questionnaire-9 and the generalized anxiety disorder-7 in Lithuanian student sample. PLoS One 17 (1), e0263027. https://doi.org/10.1371/journal.pone.0263027.

Rutter, L.A., Brown, T.A., 2017. Psychometric properties of the generalized anxiety disorder scale-7 (GAD-7) in outpatients with anxiety and mood disorders. J. Psychopathol. Behav. Assess. 39 (1), 140–146. https://doi.org/10.1007/s10862-016-9571-9.

Sahoo, M., 2019. Structural Equation Modeling: Threshold Criteria for Assessing Model Fit. Emerald Publishing Limited.

Sawaya, H., Atoui, M., Hamadeh, A., Zeinoun, P., Nahas, Z., 2016. Adaptation and initial validation of the patient health questionnaire - 9 (PHQ-9) and the generalized anxiety disorder - 7 questionnaire (GAD-7) in an Arabic speaking Lebanese psychiatric outpatient sample. Psychiatr. Res. 239, 245–252. https://doi.org/10.1016/j.psychres.2016.03.030.

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry 59 (Suppl. 20), 22–33 quiz 34-57. https://www.ncbi.nlm.nih.gov/pubmed/9881538.

Shin, C., Lee, S.H., Han, K.M., Yoon, H.K., Han, C., 2019. Comparison of the usefulness of the PHQ-8 and PHQ-9 for screening for major depressive disorder: analysis of psychiatric outpatient data. Psychiatry Investig 16 (4), 300–305. https://doi.org/10.30773/pi.2019.02.01.

Spitzer, R.L., Kroenke, K., Williams, J.B., Lowe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. 166 (10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092.

Spitzer, R.L., Williams, J.B., Kroenke, K., 2014. Test review: patient health questionnaire-9 (PHQ-9). Rehabil. Counsel. Bull. 57 (4), 246–248.

Stein, D.J., Scott, K.M., de Jonge, P., Kessler, R.C., 2017. Epidemiology of anxiety disorders: from surveys to nosology and back. Dialogues Clin. Neurosci. 19 (2), 127–136. https://www.ncbi.nlm.nih.gov/pubmed/28867937.

Ten Have, M., de Graaf, R., van Dorsselaer, S., Tuithof, M., Kleinjan, M., Penninx, B., 2018. Recurrence and chronicity of major depressive disorder and their risk indicators in a population cohort. Acta Psychiatr. Scand. 137 (6), 503–515. https://doi.org/10.1111/acps.12874.

Tsang, S., Royse, C.F., Terkawi, A.S., 2017. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi J. Anaesth. 11 (Suppl. 1), S80–S89. https://doi.org/10.4103/sja.SJA_203_17.

Uher, R., Perlis, R.H., Placentino, A., Dernovsek, M.Z., Henigsberg, N., Mors, O., Maier, W., McGuffin, P., Farmer, A., 2012. Self-report and clinician-rated measures of depression severity: can one replace the other? Depress. Anxiety 29 (12), 1043–1049. https://doi.org/10.1002/da.21993.

WHO, 2017. Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization.

Williams, S.Z., Chung, G.S., Muennig, P.A., 2017. Undiagnosed depression: a community diagnosis. SSM Popul Health 3, 633–638. https://doi.org/10.1016/j.ssmph.2017.07.012.

Wisting, L., Johnson, S.U., Bulik, C.M., Andreassen, O.A., Ro, O., Bang, L., 2021. Psychometric properties of the Norwegian version of the Patient Health Questionnaire-9 (PHQ-9) in a large female sample of adults with and without eating disorders. BMC Psychiatr. 21 (1), 6. https://doi.org/10.1186/s12888-020-03013-0.

Yang, X., Fang, Y., Chen, H., Zhang, T., Yin, X., Man, J., Yang, L., Lu, M., 2021. Global, regional and national burden of anxiety disorders from 1990 to 2019: results from the Global Burden of Disease Study 2019. Epidemiol. Psychiatr. Sci. 30, e36. https://doi.org/10.1017/S2045796021000275.