



Universiteit
Leiden
The Netherlands

Psychometric properties of the Patient Health Questionnaire-4 (PHQ-4) in 9230 adults across seven European countries: findings from the ESTSS ADJUST study

Kazlauskas, E.; Gelezelyte, O.; Kvedaraite, M.; Ajdukovic, D.; Johannesson, K.B.; Böttche, M.; ... ; Lotzin, A.

Citation

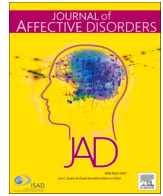
Kazlauskas, E., Gelezelyte, O., Kvedaraite, M., Ajdukovic, D., Johannesson, K. B., Böttche, M., ... Lotzin, A. (2023). Psychometric properties of the Patient Health Questionnaire-4 (PHQ-4) in 9230 adults across seven European countries: findings from the ESTSS ADJUST study. *Journal Of Affective Disorders*, 335, 18-23. doi:10.1016/j.jad.2023.05.007

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/3656963>

Note: To cite this publication please use the final published version (if applicable).



Research paper

Psychometric properties of the Patient Health Questionnaire-4 (PHQ-4) in 9230 adults across seven European countries: Findings from the ESTSS ADJUST study

Evaldas Kazlauskas^{a,*}, Odeta Gelezelyte^a, Monika Kvedaraite^a, Dean Ajdukovic^b, Kerstin Bergh Johannesson^c, Maria Böttche^d, Kristina Bondjers^{c,e}, Małgorzata Dragan^f, Margarida Figueiredo-Braga^{g,h}, Piotr Grajewski^f, Xenia Anastassiou-Hadjicharalambousⁱ, Jana Darejan Javakhishvili^j, Chrysanthi Lioupiⁱ, Brigitte Lueger-Schuster^k, Joanne Mouthaan^l, Ines Rezo Bagaric^m, Luisa Sales^{h,n}, Ingo Schäfer^o, Suzan Soydas^p, Lela Tsiskarishvili^{j,q}, Irina Zrnic Novakovic^k, Annett Lotzin^o

^a Center for Psychotraumatology, Institute of Psychology, Vilnius University, Lithuania

^b Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia

^c National Centre for Disaster Psychiatry, Department of Neuroscience, Uppsala University, Uppsala, Sweden

^d Division of Clinical Psychological Intervention, Freie Universität Berlin, Germany

^e Norwegian Center for Violence and Traumatic Stress Studies, Oslo, Norway

^f Faculty of Psychology, University of Warsaw, Poland

^g Department of Clinical Neurosciences and Mental Health, Faculty of Medicine, University of Porto, Portugal

^h Trauma Observatory, Centre for Social Studies (CES), University of Coimbra, Portugal

ⁱ Psychology Program, School of Humanities, Social Sciences and Law, University of Nicosia, Cyprus

^j Faculty of Arts and Science, Institute of Addiction Studies, Ilia State University, Tbilisi, Georgia

^k Unit of Psychotraumatology, Department of Clinical and Health Psychology, Faculty of Psychology, University of Vienna, Austria

^l Department of Clinical Psychology, Institute of Psychology, Faculty of Social and Behavioural Sciences, Leiden University, the Netherlands

^m Department of Social Work, Faculty of Law, University of Zagreb, Croatia

ⁿ Unit of Psychiatry, Hospital Militar, Coimbra, Portugal

^o Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

^p Department of Clinical Psychology, Faculty of Social and Behavioral Sciences, Utrecht University, the Netherlands

^q The Georgian Centre for Psychosocial and Medical Rehabilitation of Torture Victims, Tbilisi, Georgia

ARTICLE INFO

Keywords:

PHQ-4

Anxiety

Depression

Screening

Confirmatory factor analysis

Measurement invariance

ABSTRACT

Background: The four-item Patient Health Questionnaire-4 (PHQ-4) is a widely used screening measure for depression and anxiety.

Objectives: This study aimed to test factor structure and measurement invariance in an adult sample of the general population across seven European countries.

Method: A total sample of 9230 adults, 71.3 % female, $M_{age} = 44.35$ ($SD = 14.11$) from seven countries (Austria, Croatia, Georgia, Germany, Lithuania, Portugal, and Sweden) participated in the study. We applied confirmatory factor analysis (CFA) to examine the factor structure and measurement invariance testing to evaluate measurement equivalence across countries, gender, and age groups.

Results: The CFA yielded that a two-factor PHQ-4 model with separate depression and anxiety factors had the best fit. Partial scalar measurement invariance was established across different groups based on gender, age, and country.

Conclusions: The PHQ-4 is a valid and reliable measure that can be applied to screen for depression and anxiety in the general population.

Limitations: The limitation of the study includes the sampling, which resulted in the sample structure with the majority of females, predominantly of high education and from urban communities.

* Corresponding author at: Center for Psychotraumatology, Institute of Psychology, Vilnius University, Ciurlionio str. 29, Vilnius 03100, Lithuania.

E-mail address: evaldas.kazlauskas@fsf.vu.lt (E. Kazlauskas).

<https://doi.org/10.1016/j.jad.2023.05.007>

Received 28 October 2022; Received in revised form 1 May 2023; Accepted 3 May 2023

Available online 8 May 2023

0165-0327/© 2023 Elsevier B.V. All rights reserved.

1. Introduction

The 4-item Patient Health Questionnaire-4 (PHQ-4) is a brief screening measure for depression and anxiety (Kroenke et al., 2009). Empirical studies support the validity and reliability of the PHQ-4 in general population samples (Christodoulaki et al., 2022; Kocalevent et al., 2014; Lenz and Li, 2022; Löwe et al., 2010; Wicke et al., 2022), as well as in clinical samples (Khubchandani et al., 2016; Xiong et al., 2015). Psychometric properties of the PHQ-4 have been reported in multiple languages, such as English, German, Greek, Mandarin, Spanish, and others, revealing a wide use of this screening measure. The majority of studies using confirmatory factor analysis (CFA) suggest that the two-factor structure of the PHQ-4, with two items measuring depression and two items measuring anxiety, best mirrors the instrument's latent structure. Furthermore, several studies provided tentative findings of measurement invariance of the PHQ-4 across different gender and age groups (Lenz and Li, 2022; Mendoza et al., 2022).

While there is a growing number of validation studies of the PHQ-4, more empirical data is still needed to test the psychometric properties of the PHQ-4 across multiple countries. So far, psychometric properties of the PHQ-4 have mostly been tested in a single country, with different sampling approaches and data collection methods used across different studies. Cross-cultural comparison studies of the PHQ-4 usually compare data from different groups collected within one country, for example, migrant vs. host population (Tibubos and Kröger, 2020). Furthermore, little research has been conducted to test the measurement invariance of the PHQ-4 across gender, age groups, or multiple countries. Therefore there is a need for research and clinical practice of large multi-country studies of the PHQ-4, which would provide information on the comparability of the PHQ-4 data obtained in different countries. The present study aimed to evaluate the psychometric properties, factor structure, and measurement invariance of the PHQ-4 in a large general population sample, including individuals from seven European countries.

2. Materials and method

2.1. Participants and procedure

Data for this study was extracted from the dataset of the first wave of the longitudinal European Society of Traumatic Stress Studies (ESTSS) pan-European study ADJUST (Lotzin et al., 2020). The methods and procedures of the ADJUST study have been published previously (Lotzin et al., 2020, 2021). In brief, the ADJUST study was launched in 2020 by the ESTSS to investigate how stressors, risk, and protective factors contribute to the longitudinal course of various mental health indicators, with a primary focus on adjustment disorder and posttraumatic stress disorder symptoms, in response to the COVID-19 pandemic across eleven European countries: Austria, Croatia, Georgia, Germany, Greece, Italy, Lithuania, the Netherlands, Poland, Portugal, and Sweden. Inclusion criteria were (1) at least 18 years of age, (2) the ability to read and write in the respective native language, and (3) willingness to participate in the study. The study has been approved by the relevant ethics committees in all participating countries. All participants provided informed consent prior to participation in the study.

Based on the study protocol (Lotzin et al., 2020), measures of depression and anxiety in the ADJUST study were optional for each participating country; therefore, data on depression and anxiety symptoms were not available in all ADJUST countries. In the current study, we utilized data from seven countries that have included the PHQ-4 items in their national surveys: Austria, Croatia, Georgia, Germany, Lithuania, Portugal, and Sweden. All participants were recruited from the general population from June to November 2020; the sample size varied from 571 to 2245 across the studied countries. Recruitment was predominantly conducted online due to the COVID-19 restrictions. Multiple recruitment strategies were used to reach various age and professional groups, such as promoting the study via messages on social

networks, advertisements in media, or disseminating invitations to various professional organizations and interest groups (Lotzin et al., 2021). In total, data from 9230 adults, 71.4 % female, with a mean age of 44.35 ($SD = 14.11$), were used in this study. The sociodemographic characteristics of the study sample are presented in Table 1.

2.2. Measures

2.2.1. The Patient Health Questionnaire-4 (PHQ-4)

The symptoms of depression and anxiety were assessed using the self-report Patient Health Questionnaire-4 (PHQ-4) (Löwe et al., 2010). The PHQ-4 consists of two Patient Health Questionnaire items (PHQ-2) used to assess the core symptoms of depression (Kroenke et al., 2003), and two Generalized Anxiety Disorder Screener items (GAD-2) used to measure the core symptoms of a generalized anxiety disorder (Kroenke et al., 2007). Participants were asked to indicate how often they were bothered by the listed symptoms during the past two weeks. All the PHQ-4 items are rated on a four-point Likert scale, ranging from 0 = "not at all" to 3 = "nearly every day". The total score of the PHQ-4 is a sum of responses to all four items, ranging from 0 to 12. The sum of both subscales, PHQ-2, and GAD-2 scores, can range from 0 to 6. For the PHQ-2 and GAD-2, a cut-off score of ≥ 3 , indicating a risk of depression or anxiety, respectively (Löwe et al., 2010), was used in the study. The

Table 1
Characteristics of the sample ($N = 9230$).

Sample characteristics ^a	N	%
Gender		
Female	6585	71.4
Male	2604	28.2
Other	40	0.4
Age		
Mean (SD)	44.35 (14.11)	–
Range	18–96	–
Age group		
18–29	1548	16.8
30–45	3381	36.7
46–64	3496	37.9
65+	800	8.7
Country		
Austria	809	8.8
Georgia	775	8.4
Germany	2245	24.3
Croatia	1942	21.0
Lithuania	571	6.2
Portugal	728	7.9
Sweden	2160	23.4
Residence		
Rural area	735	10.1
Small city or town	1530	21.0
Suburb near a large city	896	12.3
Large city	4127	56.6
Education		
Less than 10 years of schooling	117	1.3
10 or more years of schooling	1553	16.8
Vocational studies	1413	15.3
University degree	6147	66.6
Employment		
Student	987	10.7
Employed part-time	1353	14.7
Employed full-time	5406	58.6
Self-employed	502	5.4
Freelancer	243	2.6
Retired	884	9.6
Seeking work	436	4.7
Relationship status		
Single	2410	26.2
Temporary relationship(s)	213	2.3
Stable relationship, living separately	893	9.7
Stable relationship, living together	5679	61.8

Note. ^a Missing data for gender (1 case), age (5), residence (1942), relationship status (35).

PHQ-4 items in various languages used in the current study are presented in Supplementary materials.

2.2.2. The World Health Organization Well-Being Index (WHO-5)

The subjective psychological well-being was measured using the self-report World Health Organization Well-Being Index (WHO-5, [Topp et al., 2015](#)). The WHO-5 comprises five items, each rated on a six-point scale, ranging from 0 = “at no time” to 5 = “all of the time”. The sum of responses to each item multiplied by 4 derives a total score of the WHO-5, ranging from 0 to 100. Higher scores indicate higher levels of psychological well-being. McDonald’s omega ([McDonald, 1978](#)) of the WHO-5 in the overall sample was 0.90, and in different countries, as follows: Austria $\omega = 0.90$, Croatia $\omega = 0.92$, Georgia $\omega = 0.89$, Germany $\omega = 0.89$, Lithuania $\omega = 0.90$, Portugal $\omega = 0.93$, Sweden $\omega = 0.90$.

2.3. Data analysis

The data analyses were conducted using IBM SPSS version 28.0 and Mplus version 8.2. Confirmatory Factor Analysis (CFA) was conducted to test the factor structure of the PHQ-4. In this analysis, we tested two CFA models – a one-factor model of the total PHQ-4 score and a two-factor model with separate latent depression and anxiety factors comprising two items each. The CFA models were estimated using the Robust Maximum Likelihood (MLR) estimator. The model fit of the CFA models was evaluated by using the Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA), following the goodness of fit recommendation provided by [Kline \(2011\)](#). Namely, CFI values higher than 0.90 indicate an acceptable fit, and values <0.95 represent a good fit; RMSEA values below 0.08 indicate an acceptable fit, and values <0.05 suggest a good fit.

The measurement invariance analysis was applied to estimate whether the PHQ-4 demonstrates comparable psychometric characteristics in terms of configural, metric, and scalar invariance across (1) three gender groups of female vs. male vs. other; (2) four age groups, 18–29, 30–45, 46–64, ≥ 65 years; and (3) seven countries. Model comparisons were conducted by examining the changes in fit indices, where $\Delta CFI \geq 0.010$ supplemented by $\Delta RMSEA \geq 0.015$ were indicative of the significant difference between models for metric and scalar measurement invariance ([Chen, 2007](#)). If scalar measurement invariance was not supported, we tested partial scalar measurement invariance by allowing up to two intercepts of the PHQ-4 items to differ between the models in tested groups. The chi-square test is sensitive to large sample sizes, so we did not apply chi-square tests to evaluate the significance in CFA and measurement invariance models.

3. Results

3.1. Structural validity of the PHQ-4

The estimates of McDonald’s omega demonstrated good reliability of the PHQ-4 scale in the total sample (0.88) and across the studied countries (ranging from 0.83 to 0.92): Austria $\omega = 0.83$, Croatia $\omega = 0.89$, Georgia $\omega = 0.91$, Germany $\omega = 0.85$, Lithuania $\omega = 0.89$, Portugal $\omega = 0.92$, Sweden $\omega = 0.88$.

The CFA results yielded that a two-factor PHQ-4 model with PHQ-2 and GAD-2 as separate correlated latent factors had a better fit ($\chi^2(1) = 2.03$, $p = .155$; CFI/TLI = 1.000/1.000; RMSEA [90 % CI] = 0.011 [0.000–0.032]; SRMR = 0.001) than a single factor PHQ-4 CFA model ($\chi^2(1) = 941.40$, $p = .000$; CFI/TLI = 0.987/0.962; RMSEA [90 % CI] = 0.226 [0.214–0.236]). All factor loadings of the PHQ-4 items in the two-factor CFA model were significant at $p < .001$ level and ranged from 0.81 to 0.96, specifically, item1 = 0.81 (S.E. = 0.01), item2 = 0.96 (0.01), item3 = 0.90 (0.01), item4 = 0.94 (0.01). The standardized factor correlation between PHQ-2 and GAD-2 latent factors was 0.87 ($p < .001$). Furthermore, we found a strong positive correlation between PHQ-4

total score and depression ($r = 0.92$, $p < .01$) as well as anxiety ($r = 0.93$, $p < .01$), and a strong correlation between anxiety and depression scores ($r = 0.71$, $p < .01$).

3.2. Measurement invariance

Measurement invariance was tested for a two-factor PHQ-4 model, which demonstrated the best fit in the CFA analysis. Configural and metric measurement invariance of the PHQ-4 was established across all gender and age groups and countries (see [Table 2](#)). However, scalar measurement invariance was not supported based on the criteria of $\Delta RMSEA > 0.15$. Therefore, we tested partial scalar measurement invariance based on modification indices to identify the most non-invariant item intercept. Partial scalar measurement invariance was established across the tested age groups by freeing the intercept of the GAD-2 item (“not being able to stop or control worrying”) to vary across the analyzed age groups (see [Table 2](#)). Partial scalar measurement invariance by gender and country was established by freeing the intercepts of two items, one anxiety GAD-2 item (“not being able to stop or control worrying”) and one depression PHQ-2 item (“bothered by feeling down, depressed, or hopeless”) to vary across gender groups and countries (see [Table 2](#)).

3.3. Anxiety and depression across study groups

After confirming the partial scalar measurement invariance of the PHQ-4 across the studied groups, we analyzed group differences in PHQ-4 scores. We also estimated the prevalence of the risk for anxiety and depression in the total sample and the studied groups. The total score of the PHQ-4 in the sample was 3.00 ($SD = 3.03$), with PHQ-2 $M = 1.55$ ($SD = 1.61$), and GAD-2 $M = 1.77$ ($SD = 1.67$). In the total sample, 20.9 % of the participants were at risk for depression (PHQ-2 cut-off score ≥ 3), and 20.2 % of the sample was at risk for anxiety disorder (GAD-2 cut-off score ≥ 3). Detailed statistics for the PHQ-4 total, depression, and anxiety scores, as well as WHO-5 scores across gender, age groups, and country, are presented in [Table 3](#).

We found significant gender effects on depression, anxiety, and well-being index (see [Table 3](#)). Compared to male participants, female participants tend to have higher levels of depression and anxiety. Moreover, participants who identified themselves as of the other gender had the highest levels of anxiety and depression and the lowest psychological well-being (see [Table 3](#)).

Furthermore, we found a significant age group effect on depression and anxiety. In our study, emerging adults (18–29 years) reported the highest levels of depression (32.8 %) and anxiety (28.4 %) and the lowest psychological well-being in comparison to other age groups. Participants of older age (65+) reported lower rates of depression (16.5 %) and anxiety (12.4) in comparison to other age groups and had the highest psychological well-being (see [Table 3](#)).

Depression and anxiety scores differed considerably across the study countries. The prevalence of risk for depression ranged from 15.7 % to 31.9 % (PHQ-2 range 1.34–2.05), and the risk for anxiety varied from 14.6 % to 31.0 % (GAD-2 range 1.17–2.04) across the different countries. Southern European country Croatia demonstrated significantly lower levels of depression, and Austria had the lowest anxiety levels compared to other countries. The Northern Eastern European country Lithuania had higher rates of anxiety and depression compared to all other countries.

To test the validity of the PHQ-4 further, we computed correlations between the observed scores of the PHQ-4 and psychological well-being (WHO-5). All correlations were negative, with $r = -0.67$ ($p < .01$) for the total PHQ-4 score, $r = -0.67$ ($p < .01$) for depression, and $r = -0.59$ ($p < .01$) for anxiety scores.

Table 2

Measurement invariance of the PHQ-4 by gender, age, and country.

	Model fit indices				Model comparisons		
	χ^2 (df)	CFI	RMSEA [90 % CI]	SRMR	Δ CFI	Δ RMSEA	Δ SRMR
Gender							
Configural	3.405 (3)	1.000	0.007 [0.000; 0.032]	0.002			
Metric	6.535 (7)	1.000	0.000 [0.000; 0.021]	0.004	<0.001	0.007	0.002
Scalar	33.005 (11)	0.998	0.026 [0.016; 0.036]	0.009	0.002	0.026	0.005
Partial scalar	6.536 (7)	1.000	0.000 [0.000; 0.021]	0.004	<0.001	<0.001	<0.001
Age							
Configural	1.890 (4)	1.000	0.000 [0.000; 0.022]	0.001			
Metric	14.177 (10)	1.000	0.013 [0.000; 0.028]	0.009	<0.001	0.013	0.008
Scalar	52.662 (16)	0.997	0.032 [0.022; 0.041]	0.013	0.003	0.019	0.004
Partial scalar	17.086 (13)	1.000	0.012 [0.000; 0.025]	0.010	<0.001	0.001	0.001
Country							
Configural	15.792 (7)	0.999	0.031 [0.010; 0.051]	0.004			
Metric	68.376 (19)	0.996	0.044 [0.033; 0.056]	0.021	0.003	0.013	0.017
Scalar	362.473 (31)	0.971	0.090 [0.082; 0.099]	0.038	0.025	0.046	0.017
Partial scalar	68.376 (19)	0.996	0.044 [0.033; 0.056]	0.021	<0.001	<0.001	<0.001

Note. χ^2 = chi-square, df = degrees of freedom, CFI = comparative fit index, RMSEA = root mean square error of approximation, CI = confidence interval, Δ = change in the parameter.

Table 3Scores of PHQ-4 and WHO-5 across study groups ($N = 9230$).

	Depression, PHQ-2			Anxiety, GAD-2			Total, PHQ-4		Psychological well-being, WHO-5	
	M (SD)	% (a.s.r.)	F/χ^2 [†]	M (SD)	% (a.s.r.)	F/χ^2 [†]	M (SD)	F [†]	M (SD)	F [†]
Gender										
Female ^a	1.62 (1.63) ^{b,c}	22.0 (3.95)*	34.15	1.58 (1.71) ^{b,c}	22.2 (7.27)***	78.62	3.20 (3.08) ^{b,c}	62.45	53.93 (22.04) ^b	30.67
Male ^b	1.35 (1.54) ^{a,c}	18.0 (−4.50)***	/	1.11 (1.5) ^{a,c}	15.0 (−7.88)***	/	2.46 (2.82) ^{a,c}		57.88 (22.71) ^{a,c}	
Other ^c	2.58 (2.11) ^{a,b}	40.0 (2.97)	27.13	2.20 (1.95) ^{a,b}	45.0 (3.91)*	74.88	4.78 (3.66) ^{a,b}		49.03 (24.57) ^b	
Age group										
18–29 ^a	2.11 (1.73) ^{b,c,d}	32.8 (12.52)***	93.92	1.90 (1.79) ^{b,c,d}	28.4 (8.80)***	69.37	4.01 (3.24) ^{b,c,d}	93.44	49.52 (21.13) ^{b,c,d}	84.08
30–45 ^b	1.56 (1.58) ^{a,c,d}	20.1 (−1.49)	/	1.50 (1.65) ^{a,c,d}	20.8 (1.08)	/	3.06 (2.97) ^{a,c,d}		53.29 (21.75) ^{a,c,d}	
46–64 ^c	1.35 (1.54) ^{a,b}	17.5 (−6.29)***	165.86	1.31 (1.61) ^{a,b,d}	17.8 (−4.50)***	108.31	2.66 (2.92) ^{a,b,d}		57.42 (22.69) ^{a,b,d}	
65+ ^d	1.22 (1.53) ^{a,b}	16.5 (−3.24)		0.98 (1.55) ^{a,b,c}	12.4 (−5.79)***		2.20 (2.85) ^{a,b,c}		62.59 (21.73) ^{a,b,c}	
Country										
Austria ^a	1.42 (1.42) ^{d,e}	16.9 (−2.93)	20.45	1.17 (1.43) ^{b,e,f,g}	14.6 (−4.18)	20.86	2.60 (2.59) ^{d,e,g}	16.02	55.86 (22.68) ^{c,d,g}	67.06
Croatia ^b	1.34 (1.50) ^{c,d,e,g}	15.7 (−6.38)***	/	1.48 (1.58) ^{a,d,e}	18.0 (−2.79)	/	2.82 (2.87) ^e		56.73 (21.87) ^{c,d,e,g}	
Georgia ^c	1.61 (1.83) ^{b,e}	24.4 (2.46)	103.20	1.31 (1.76) ^{e,f}	18.8 (−1.01)	97.64	2.93 (3.41) ^e		44.46 (20.54) ^{a,b,d,e,f,g}	
Germany ^d	1.69 (1.47) ^{a,b,e,f,g}	21.9 (1.24)		1.31 (1.56) ^{b,e,f,g}	17.7 (−3.39)		3.00 (2.76) ^{a,e}		51.31 (22.17) ^{a,b,c,f,g}	
Lithuania ^e	2.05 (1.72) ^{a,b,c,d,f,g}	31.9 (6.63)***		2.04 (1.78) ^{a,b,c,d,f,g}	31.0 (6.62)***		4.09 (3.24) ^{a,b,c,d,f,g}		52.26 (21.31) ^{b,c,f,g}	
Portugal ^f	1.39 (1.66) ^{d,e}	16.9 (−2.80)		1.58 (1.70) ^{a,c,d,e}	20.5 (0.17)		2.97 (3.19) ^e		57.64 (21.92) ^{c,d,e,g}	
Sweden ^g	1.52 (1.74) ^{b,d,e}	23.4 (3.24)		1.52 (1.81) ^{a,d,e}	24.5 (5.70)***		3.04 (3.28) ^{a,e}		60.48 (21.69) ^{a,b,c,d,e,f}	

Note. [†] F , χ^2 all values are significant at $p < .001$; % = percentage of depression and anxiety within each subgroup; a.s.r. = adjusted standardized residuals; * $p < .05$, *** $p < .001$ with applied Bonferroni corrections for multiple tests; ^{a,b,c,d,e,f,g} indicates significant differences at $p < .001$ within gender, age groups and country; χ^2 = chi-square.

4. Discussion

Our study was the first to test the factor structure and measurement invariance of the PHQ-4 in a large European multi-country adult sample. Overall, the study findings support the validity and measurement invariance of the PHQ-4. In line with previous studies (Christodoulaki et al., 2022; Löwe et al., 2010; Wicke et al., 2022), we found strong support for the structure of the PHQ-4 measure comprising depression and anxiety factors. Furthermore, configure, metric, and partial scalar invariance was supported across age, gender groups, and countries. These results corroborate previous findings (Christodoulaki et al., 2022; Fong et al., 2023; Lenz and Li, 2022; Mendoza et al., 2022; Tibubos and Kröger, 2020) and encourage the

application of the PHQ-4 for the screening of anxiety and depression in various adult populations.

The current study did not fully support scalar measurement invariance, indicating that specific groups included in the measurement invariance tests might have had variable responses to the PHQ-4 scale. It is not surprising, as numerous previous studies indicated gender and age differences in mental disorders, particularly depression and anxiety (Altemus et al., 2014; Luppa et al., 2012). Furthermore, studies indicate that individuals with non-binary sexual identities have a higher risk for anxiety and depression (Feinstein et al., 2020). Nevertheless, partial scalar invariance was supported across different gender, age groups, and countries in our study. Measurement invariance can be tested in multiple groups (e.g., 10 or 20) in international surveys (Rutkowski and

Svetina, 2017); however, there is limited knowledge from empirical studies on the impact of analyzing measurement invariance across multiple groups as the majority of studies test measurement invariance in two or three subsamples (Putnick and Bornstein, 2016). We tested measurement invariance across four age groups and seven countries, so multigroup testing with variable sample sizes across the tested groups might have had an impact on our results, and therefore only partial scalar invariance measurement was established.

We found a quite high prevalence of anxiety (20.9 %) and depression (20.2 %) in the total sample. These results were higher than the initial validation study of the PHQ-4, which reported 6.6 % and 4.8 % prevalence of depression and anxiety, using the same cut-off score of ≥ 3 (Löwe et al., 2010). We also found significant gender and age effects on depression and anxiety, which are in line with previous studies (Altemus et al., 2014; Luppá et al., 2012). Furthermore, considering a limited number of pan-European mental health empirical studies, this study provides valuable data on cross-country comparisons revealing significant cross-country differences in the levels of anxiety and depression. However, it is important to note that the data collection of the current study took place during the first year of the COVID-19 pandemic, and the prevalence rates of anxiety and depression in the sample might have been affected by the social restrictions and uncertainty that were predominant at the time. Therefore, the results of our study should be interpreted with caution and warrant further research.

There are several limitations of the study. First, while the ADJUST study was well coordinated among the countries using the same time-frame for the data collection, there were differences in sampling procedures across countries. Convenient sampling via social media or other channels did not result in a representative sample, thus limiting the generalization of the findings (Lotzin et al., 2021). Second, the studied countries significantly vary in population sizes and sample sizes. There are indications that sample size, in particular, evaluating measurement invariance, might impact results (Chen, 2007; Rutkowski and Svetina, 2017). Moreover, the sample structure included the majority of females, predominantly with high education and from urban communities, and relatively low participation of the 65+ age group. Finally, we focused on the PHQ-4 factor structure and measurement invariance in our study; however, the inclusion of other anxiety and depression measures could have provided more information on the divergent validity, sensitivity, or specificity of the PHQ-4.

Despite these limitations, the study provides strong evidence for the structural validity of the PHQ-4 and indicates that the PHQ-4 scale can be used for screening for depression and anxiety in various settings.

CRediT authorship contribution statement

Evaldas Kazlauskas: conceptualization, statistical data analysis, writing the first draft of the manuscript; Monika Kvedaraite: statistical data analysis; Odeta Gelezelyte, Dean Ajdukovic, Kerstin Bergh Johansson, Maria Böttche, Kristina Bondjers, Margarida Figueiredo-Braga, Jana Darejan Javakhishvili, Brigitte Lueger-Schuster, Joanne Mouthaan, Ines Rezo Bagaric, Luisa Sales, Ingo Schäfer, Suzan Soydas, Lela Tsiskarishvili, Irina Zrnic Novakovic, & Annett Lotzin: methodology, investigation, data collection. All the co-authors contributed to the writing and editing of the manuscript.

Declaration of competing interest

None.

Acknowledgment

We are grateful to the ADJUST consortium members and national teams who contributed to the study. Austria: Ozan Demirok. Croatia: Marina Ajdukovic, Helena Bakic, Tanja Franciskovic, Aleksandra Stevanovic; Georgia: Nino Makhashvili, Sophio Viblinai, Ivane

Javakhishvili. Lithuania: Auguste Nomeikaite. The Netherlands: CON-NECT Research Team: Marloes Eidhof, Marie-José van Hoof, Simon Groen, Lonneke Lenferink, Trudy Mooren, Lyanne Reitsma. Portugal: Aida Dias, Camila Borges, Diana Andringa, Inês Nascimento Rodrigues, Joana P. Becker, João Veloso, Margarida Calafate Ribeiro. Sweden: Filip Arnberg, Ida Hensler, Josefin Sveen & Rakel Eklund, Maria Bragesjö.

Ethics statement

The study was registered prior to the start of data collection (OSF registry, doi:10.17605/OSF.IO/8XHYG). Each country obtained ethical approval. Austria: Ethics Committee of the University of Vienna, 00554. Croatia: Ethics Committee of the Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb: May 21, 2020. Georgia: Ethics Review Board of the Faculty of Social and Behavioural Sciences, Utrecht University, Number 20-360. Ilia State University Faculty of Arts and Science Research Ethics Committee: 12/06/2020. Germany: Local Psychological Ethics Committee at the Center for Psychosocial Medicine (LPEK) at University Medical Center Hamburg-Eppendorf (LPEK-0149). Sweden: The Swedish Ethical Review Authority, 2020-03217. Lithuania: Vilnius University Ethics Committee of Research in Psychology, 2020, 44. Portugal: Ethics Committee of the Medical Faculty, University of Porto and Centro Hospitalar São João, Porto, CE 201-20. All participants provided informed consent before taking part in the study.

Funding source

The ADJUST study was partially funded by the European Society for Traumatic Stress Studies (ESTSS). The study in the Netherlands received funding from CZ Zorgverzekerings and DSW Zorgverzekeraar. The study in Sweden received funding from the National Board of Health and Welfare (No. 10.1-37970/2020). The remaining countries did not receive specific funding for this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jad.2023.05.007>.

References

- Altemus, M., Sarvaiya, N., Neill Epperson, C., 2014. Sex differences in anxiety and depression clinical perspectives. *Front. Neuroendocrinol.* 35 (3), 320–330. <https://doi.org/10.1016/j.yfrne.2014.05.004>.
- Chen, F.F., 2007. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Model.* 14 (3), 464–504. <https://doi.org/10.1080/10705510701301834>.
- Christodoulaki, A., Baralou, V., Konstantakopoulos, G., Touloumi, G., 2022. Validation of the patient health Questionnaire-4 (PHQ-4) to screen for depression and anxiety in the Greek general population. *J. Psychosom. Res.* 160 (June) <https://doi.org/10.1016/j.jpsychores.2022.110970>.
- Feinstein, B.A., Xavier Hall, C.D., Dyar, C., Davila, J., 2020. Motivations for sexual identity concealment and their associations with mental health among bisexual, pansexual, queer, and fluid (Bi+) individuals. *Journal of Bisexuality* 20 (3), 324–341. <https://doi.org/10.1080/15299716.2020.1743402>.
- Fong, T.C.T., Ho, R.T.H., Yip, P.S.F., 2023. Psychometric properties of the Patient Health Questionnaire-4 among Hong Kong young adults in 2021: associations with meaning in life and suicidal ideation. *Frontiers in Psychiatry* 14 (March). <https://doi.org/10.3389/fpsy.2023.1138755>.
- Khubchandani, J., Brey, R., Kotecki, J., Kleinfelder, J.A., Anderson, J., 2016. The psychometric properties of PHQ-4 depression and anxiety screening scale among college students. *Arch. Psychiatr. Nurs.* 30 (4), 457–462. <https://doi.org/10.1016/j.apnu.2016.01.014>.
- Kline, R.B., 2011. Principles and practice of structural equation modeling. In: *Principles and Practice of Structural Equation Modeling*, (3rd ed.). The Guilford Press. <https://doi.org/10.1038/156278a0>.
- Kocalevent, R.D., Finck, C., Jimenez-Leal, W., Sautier, L., Hinz, A., 2014. Standardization of the colombian version of the PHQ-4 in the general population. *BMC Psychiatry* 14 (1), 1–8. <https://doi.org/10.1186/1471-244X-14-205>.
- Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2003. The patient health questionnaire-2: validity of a two-item depression screener. *Med. Care* 41 (11), 1284–1292. <https://doi.org/10.1097/01.MLR.0000093487.78664.3C>.

- Kroenke, K., Spitzer, R.L., Williams, J.B.W., Monahan, P.O., Löwe, B., 2007. Anxiety disorders in primary care: prevalence, annals of internal medicine anxiety disorders in primary care: prevalence, impairment. *Ann. Intern. Med.* 146, 317–325. <https://doi.org/10.7326/003-4819-146-5-200703060-00004>.
- Kroenke, K., Spitzer, R.L., Williams, J.B.W., Löwe, B., 2009. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics* 50 (6), 613–621. [https://doi.org/10.1016/s0033-3182\(09\)70864-3](https://doi.org/10.1016/s0033-3182(09)70864-3).
- Lenz, A.S., Li, C., 2022. Evidence for measurement invariance and psychometric reliability for scores on the PHQ-4 from a rural and predominately hispanic community. *Meas. Eval. Couns. Dev.* 55 (1), 17–29. <https://doi.org/10.1080/07481756.2021.1906157>.
- Lotzin, A., Acquarini, E., Ajdukovic, D., Ardino, V., Böttche, M., Bondjers, K., Bragesjö, M., Dragan, M., Grajewski, P., Figueiredo-Braga, M., Gelezelyte, O., Javakhishvili, J.D., Kazlauskas, E., Knefel, M., Lueger-Schuster, B., Makhshvili, N., Mooren, T., Sales, L., Stevanovic, A., Schäfer, I., 2020. Stressors, coping and symptoms of adjustment disorder in the course of the COVID-19 pandemic—study protocol of the European Society for Traumatic Stress Studies (ESTSS) pan-European study. *Eur. J. Psychotraumatol.* 11 (1) <https://doi.org/10.1080/20008198.2020.1780832>.
- Lotzin, A., Krause, L., Acquarini, E., Ajdukovic, D., Ardino, V., Arnberg, F., Böttche, M., Bragesjö, M., Dragan, M., Figueiredo-Braga, M., Gelezelyte, O., Grajewski, P., Anastassiou-Hadjicharalambous, X., Javakhishvili, J.D., Kazlauskas, E., Lenferink, L., Lioupi, C., Lueger-Schuster, B., Tsiskarishvili, L., Soydas, S., 2021. Risk and protective factors, stressors, and symptoms of adjustment disorder during the COVID-19 pandemic—First results of the ESTSS COVID-19 pan-European ADJUST study. *Eur. J. Psychotraumatol.* 12 (1) <https://doi.org/10.1080/20008198.2021.1964197>.
- Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., Schneider, A., Brähler, E., 2010. A 4-item measure of depression and anxiety: validation and standardization of the patient health Questionnaire-4 (PHQ-4) in the general population. *J. Affect. Disord.* 122 (1–2), 86–95. <https://doi.org/10.1016/j.jad.2009.06.019>.
- Luppa, M., Sikorski, C., Luck, T., Ehreke, L., Konnopka, A., Wiese, B., Weyerer, S., König, H.H., Riedel-Heller, S.G., 2012. Age- and gender-specific prevalence of depression in latest-life - systematic review and meta-analysis. *J. Affect. Disord.* 136 (3), 212–221. <https://doi.org/10.1016/j.jad.2010.11.033>.
- McDonald, R.P., 1978. Generalizability in factorable domains: “Domain validity and generalizability”. *Educ. Psychol. Meas.* 38 (1), 75–79. <https://doi.org/10.1177/001316447803800111>.
- Mendoza, N.B., Frondozo, C.E., Dizon, J.I.W.T., Buenconsejo, J.U., 2022. The factor structure and measurement invariance of the PHQ-4 and the prevalence of depression and anxiety in a Southeast Asian context amid the COVID-19 pandemic. *Curr. Psychol.* 0123456789 <https://doi.org/10.1007/s12144-022-02833-5>.
- Putnick, D.L., Bornstein, M.H., 2016. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev. Rev.* 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>.
- Rutkowski, L., Svetina, D., 2017. Measurement invariance in international surveys: categorical indicators and fit measure performance. *Appl. Meas. Educ.* 30 (1), 39–51. <https://doi.org/10.1080/08957347.2016.1243540>.
- Tibubos, A.N., Kröger, H., 2020. A cross-cultural comparison of the ultrabrief mental health screeners PHQ-4 and SF-12 in Germany. *Psychol. Assess.* 32 (7), 690–697. <https://doi.org/10.1037/pas0000814>.
- Topp, C.W., Østergaard, S.D., Søndergaard, S., Bech, P., 2015. The WHO-5 well-being index: a systematic review of the literature. *Psychother. Psychosom.* 84 (3), 167–176. <https://doi.org/10.1159/000376585>.
- Wicke, F.S., Krakau, L., Löwe, B., Beutel, M.E., Brähler, E., 2022. Update of the standardization of the patient health Questionnaire-4 (PHQ-4) in the general population. *J. Affect. Disord.* 312 (May), 310–314. <https://doi.org/10.1016/j.jad.2022.06.054>.
- Xiong, N., Fritzsche, K., Wei, J., Hong, X., Leonhart, R., Zhao, X., Zhang, L., Zhu, L., Tian, G., Nolte, S., Fischer, F., 2015. Validation of patient health questionnaire (PHQ) for major depression in Chinese outpatients with multiple somatic symptoms: a multicenter cross-sectional study. *J. Affect. Disord.* 174, 636–643. <https://doi.org/10.1016/j.jad.2014.12.042>.