The Patient Health Questionnaire–4: Factor Structure, Measurement Invariance, Latent Profile Analysis of Anxiety and Depressive Symptoms and Screening Results in Polish Adults

Paweł Larionow and *Karolina Mudło-Głagolska*

Kazimierz Wielki University, Faculty of Psychology, 1 Leopolda Staffa street, 85–867 Bydgoszcz, Poland

ABSTRACT

A screening assessment of anxiety and depressive symptoms is of great importance for preventing mental health problems. The current study aimed to (a) examine the factor structure, measurement invariance, reliability, and temporal stability of the Polish version of the Patient Health Questionnaire–4 (PHQ–4) developed for measuring anxiety and depressive symptoms, (b) estimate the prevalence of these symptoms in different age and gender groups from February to July 2022, and (c) identify latent subpopulations of females and males based on their anxiety and depressive symptom levels. The sample included 2557 Polish adults (1730 females, 811 males, and 16 non-binary individuals) aged 18–78 (M = 27.18, SD = 12.29). The Polish version of the PHQ–4 was used. Multi-group confirmatory factor analysis and latent profile analysis (LPA) were carried out. The Polish version of the PHQ–4 had an intended two-factor structure, was invariant across age and gender categories, and in general, displayed strong psychometric properties. Based on the cut-off scores for the anxiety and depression subscales of ≥3, more than half of females and males in three age groups (aged 18–29, 30–44, and 45–59) were screened positively for anxiety and/or depression, and more than one-third of females aged 60–78 and males aged 60–76 were screened positively for both disorders. We distinguished (by LPA) and described four subpopulations common for both females and males. We identified risk groups for mental disorders (females, males aged 30–44, non-binary and younger people in general, as well as unemployed, less educated, and single people). The prevalence of probable anxiety and depressive disorders in the Polish population is extremely high. It is necessary to provide further mental health monitoring.

KEYWORDS

anxiety
depression
factor structure
measurement invariance
latent profile analysis
mental health screening
Patient Health Questionnaire–4.

INTRODUCTION

Living in uncertain times due to the geopolitical crisis near the borders of Poland and the COVID-19 pandemic (Larionow & Mudło-Głagolska, 2021) may negatively impact the mental health of Polish citizens. According to the Institute for Health Metrics and Evaluation, Poland was the country with the lowest percentage of people suffering from depression (2.8%) among European Union countries in 2010 and 2017 (Narodowy Fundusz Zdrowia, 2020). However, current data indicates a high prevalence of depressive symptoms in Polish adults. For example, Gawrych et al. (2022) analyzed the prevalence of possible depression among 1500 young Poles aged 18–23 from 27 November 2020 to 8 January 2021 using the Kutcher Adolescents Depression Scale. They showed that 56.5% of the respondents had possible depression, and thus, that the majority of young Polish adults met the criteria for depressive disorders (Gawrych et al., 2022). Due to this, a screening assessment of mental health problems (i.e., anxiety or depressive symptoms) and their monitoring have great importance for prevention in the general population. For this purpose, short and validated measures (especially ones characterized by measurement invariance across age or gender background) are needed. Our first and main aim in the current study was to introduce and examine the psychometric properties of the Polish version of the Patient Health Questionnaire–4 (PHQ–4).

The PHQ–4 is one of the most used freely available ultra-brief screening tools for anxiety and depression (Kroenke et al., 2009; Löwe et al., 2010). The PHQ–4 has four items and consists of the Patient

Corresponding author: Paweł Larionow, Kazimierz Wielki University, Faculty of Psychology, 1 Leopolda Staffa street, 85–867 Bydgoszcz, Poland

Email: pavel@ukw.edu.pl

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Health Questionnaire-2 (PHQ–2) for measuring depressive symptoms and the Generalized Anxiety Disorder–2 (GAD–2) for measuring anxiety symptoms (Kroenke et al., 2009; Löwe et al., 2010). The PHQ–4 and its components (the PHQ–2 and the GAD–2) showed good validity in different community samples, that is, in the general population in Colombia (Kocalevent et al., 2014), Germany (Löwe et al., 2010), Greece (Christodoulaki et al., 2022), or Spain (Cano-Vindel et al., 2018), as well as in clinical ones, that is, in infertile patients (Ghaheri et al., 2020), in patients with intracranial tumors (Renovanz et al., 2019), in coronary heart disease patients (Thombs et al., 2008), and in adults living with HIV/AIDS (Monahan et al., 2009). Its validity has also been confirmed in samples comprising older people (Li et al., 2007; Wild et al., 2014). Web-based versions of the PHQ–4 and its components have shown good validity as well (Donker et al., 2011) and were used in intervention studies (Cavanagh et al., 2013).

The PHQ–4 provides fast and valid screening for anxiety and depression. According to the Diagnostic Algorithms for the PHQ (Löwe et al., 2010; Pfizer, 2022), a total score for the PHQ–4 of ≥ 6, or anxiety and depression subscales scores of ≥ 3, may suggest anxiety or a depressive disorder. These results can be considered "yellow flags" for anxiety and/or depression, whereas a total score for the PHQ–4 of ≥ 9, and anxiety and depression subscales scores of ≥ 5 may be considered "red flags" (Löwe et al., 2010). In clinical routine settings, the authors of the PHQ–4 recommend using the total score and calculating the anxiety and depression scores for the assessment of the predominant disorder (Löwe et al., 2010). Kroenke et al. (2007) noted that the anxiety subscale score (the GAD–2) with a cut-off score of ≥ 3 has good sensitivity and specificity for generalized anxiety disorder (86% and 83%, respectively), panic disorder (76% and 81%), social anxiety disorder (70% and 81%), posttraumatic stress disorder (59% and 81%) and any anxiety disorder (65% and 88%). The depression subscale (the PHQ–2) score of ≥ 3 has a sensitivity of 83% and a specificity of 92% for major depressive disorder (Kroenke et al., 2003). The positive predictive values of the GAD–2 with the cut-off score of ≥ 3 ranged from 28% (Donker et al., 2011) to 86.6% (García-Campayo et al., 2012) in detecting generalized anxiety disorder, whereas the positive predictive values of the PHQ–2 with the cut-off score of ≥ 3 ranged from 23.1 to 66.0% in detecting depressive disorders (Miller et al., 2021).

Confirmatory factor analysis (CFA) has supported that the PHQ–4 has a two-factor solution with two highly correlated individual anxiety (the GAD–2) and depression (the PHQ–2) subscales as well as its invariance across age, gender, and cross-cultural groups (Löwe et al., 2010; Mendoza et al., 2022; Tibubos et al., 2020). Different studies supported empirically the concurrent validity of the questionnaire and its connection with life satisfaction, self-esteem, and resilience (Löwe et al., 2010), as well as its positive correlations with stress and negative affect, along with its negative association with positive affect (Mendoza et al., 2022). Thus, the PHQ–4 is a valid screening tool for anxiety and depression with promising psychometric properties (factor structure, measurement invariance, reliability, sensitivity, and specificity) in different clinical and nonclinical samples. Despite the existence of a Polish version of the Patient Health Questionnaire–9 used for in-depth measuring of depressive symptoms (Kokoszka et al., 2016; Ślusarska et al., 2019; Tomaszewski et al., 2011), there is a need for a validated Polish version of the PHQ–4 with its two intended scales (the PHQ–2 and the GAD–2) for fast mental health screening of both anxiety and depression in the general population. To summarize, developing the Polish version of the PHQ–4 as a screening measure for anxiety and depressive symptoms and analyzing their prevalence in Polish adults is one of the important tasks in medical research.

Based on our previous studies (Larionow, Mudło-Głagolska et al., 2022; Larionow, Preece et al., 2022), there is evidence of good convergent and discriminant validity of the Polish version of the PHQ–4 administered in the current study. Larionow, Mudło-Głagolska et al.'s. (2022) study was conducted on a sample of 846 Polish adults. It was shown that the anxiety and depression subscales were highly positively correlated with such somatic symptoms as exhaustion (r = .52 and .59, respectively), and moderately with gastrointestinal (r = .29 and .25), musculoskeletal (r = .29 and .26), and cardiovascular (r = .42 and .37) somatic complaints (all ps < .001) measured by the Giessen Subjective Complaints List (GBB–8). Moreover, a second-order exploratory factor analysis of anxiety and depressive symptoms (the two PHQ–4 subscales) and these four somatic symptoms subscales extracted two factors (i.e., Factor 1, mental health symptoms, and Factor 2, somatic symptoms). The two PHQ–4 subscales loaded precisely on the mental health symptoms factor and did not load on the somatic symptoms factor, thus supporting good discriminant validity of the PHQ–4.

The results of our second study (Larionow, Preece et al., 2022) conducted on a sample of 944 Polish adults also revealed good convergent validity of the PHQ–4. In that study, the correlations between the anxiety and depression subscales and alexithymia (as an established transdiagnostic risk factor for affective disorders; Preece et al., 2022) were assessed. The anxiety and depression subscales were moderately positively correlated (all ps < .001) with difficulty identifying negative feelings (r = .40 and r = .40, respectively), difficulty identifying positive feelings (r = .29 and r = .34), difficulty describing negative feelings (r = .35 and r = .37), difficulty describing positive feelings (r = .26 and r= .33), externally oriented thinking (r = .16 and r = .25), and the total alexithymia score (r = .32 and r = .38). Therefore, the results of these studies (Larionow, Mudło-Głagolska et al., 2022; Larionow, Preece et al., 2022) supported good convergent and discriminant validity of the Polish version of the PHQ–4 administered in this study.

In the current study, we aimed at a more in-depth examination of the factorial validity and reliability of the questionnaire. Thus, the first and main aim of the current study was to present the psychometric properties (factor structure, test-retest and internal consistency reliabilities) of the PHQ–4 in Polish adults. Moreover, we examined the measurement invariance for different age and gender groups, which provides empirically based evidence for comparing different PHQ–4 scores across age and gender backgrounds.

Our second aim was to estimate the prevalence of anxiety and depressive symptoms (based on the cut-off scores for the GAD–2 and the PHQ–2 of ≥ 3, according to the Diagnostic Algorithms for the PHQ) in different age and gender groups, including non-binary individu-

als. Gender identity is an understudied determinant of mental health (Reisner et al., 2016). Therefore, we considered this issue in our study. Previous studies showed that females and non-binary individuals had more symptoms of anxiety or depression (Herrera-Añazco et al., 2022). Thus, we predicted similar patterns in our study.

Our third aim was to identify latent subpopulations based on the anxiety and depressive symptoms by latent profile analysis (LPA). The LPA is a powerful statistical classification technique that allowed us to identify profiles having similar patterns across a set of variables (Spurk et al., 2020). In our study, we examined and described latent subpopulations that had similar patterns across anxiety and depression symptoms, and we did it separately for females and males.

We predicted that the Polish version of the PHQ–4 would be characterized by an intended two-factor structure and would be invariant across different gender and age categories. We also anticipated good test-retest and internal consistency reliabilities, and a higher prevalence and intensity of anxiety and depressive symptoms in younger people, as well as in females and non-binary individuals. We had no a priori hypotheses regarding how many profiles would be distinguished, but we predicted that there would be at least two profiles, including a subpopulation with low levels of anxiety and depressive symptoms as well as a subpopulation with high levels. To the best of our knowledge, the current study is be the first in the world in which individual latent subpopulations with different anxiety and depressive symptoms levels were identified by LPA.

MATERIALS AND METHODS

Participants and Procedure

The sample included 2557 Polish adults in three gender groups (1730 females, 811 males, and 16 non-binary individuals) aged 18–78 years (M = 27.18, SD = 12.29). People with higher education made up 27.30% of respondents, whereas 72.70% had lower educational levels. 36.14% of the respondents lived in large cities (above 100000 inhabitants), 23.19% in medium-sized towns (from 20000 to 100000), 13.73%, small towns (up to 20000), and 26.95% in villages. 53.15% of the participants were single, whereas 46.85% were married or living in common-law relationships. 35.31% of the respondents worked professionally, 18.65% were unemployed. People working and studying at the same time made up 19.28% of the sample, 23.74% were students, whereas 3.01% were retired.

The current data on the PHQ–4 were derived from our research projects which were conducted from February to July 2022 via Facebook and Instagram, where we published a link to an online anonymous survey by Google Forms. In these projects, the purposeful sampling method with maximum variation design (Palinkas et al., 2015) was used to recruit the participants, who completed the PHQ–4 as a part of a short battery of psychological questionnaires on emotion processing. All the projects were conducted in accordance with the Declaration of Helsinki Ethical Principles and were approved by the Kazimierz Wielki University Research Ethics Committee (no. 1/13.06.2022). All the respondents provided their informed consent digitally before they answered the questions. There was no reimbursement for the participants.

The current data on the PHQ–4 were derived from our published (Larionow, Mudło-Głagolska et al., 2022; Larionow, Preece et al., 2022) and unpublished but ongoing research projects. The scope of these projects is emotion processing, where the PHQ–4 was used as a correlate of different emotion variables. All the projects are independent and very short (with no more than 40 questions). For the current study, we combined the data on the PHQ–4 from these projects to provide a more in-depth examination of the factorial validity and reliability of the questionnaire in a large and diverse sample of Polish citizens.

Measures

The PHQ–4 is a four-item questionnaire for measuring anxiety and depressive symptoms experienced in the previous two weeks. The PHQ–4 has two two-item subscales, namely, anxiety (Item 1, "Feeling nervous, anxious or on edge" and Item 2, "Not being able to stop or control worrying") and depression (Item 3, "Little interest or pleasure in doing things" and Item 4 "Feeling down, depressed, or hopeless"). The PHQ–4 uses a four-point Likert scale from 0 (*not at all*) to 3 (*nearly every day*). The total PHQ–4 score can be calculated. A total score for the PHQ–4 of ≥ 6, or the anxiety and the depression subscales scores of ≥ 3 suggest depression or an anxiety disorder. Respondents with these results can be identified as positively screened and should be encouraged to undergo a more detailed assessment of anxiety and depression.

The original English version of the PHQ–4 was translated into Polish by three independent translators, and a common Polish translation was developed. Then, it was translated back into English, and this back-translation was compared with the original version. The necessary minor corrections were made resulting in the final Polish version of the PHQ–4 (see Appendix).

Statistical Analysis

We tested a one-factor model as well as an intended two-factor model of the PHQ–4 consisting of the anxiety and depression subscales. Considering some critical issues with using the maximum likelihood (ML) estimation method with ordinal data (e.g., in questionnaires with a four-point Likert scale, see Li, 2016), we used both the ML and the weighted least squares (WLS) methods to perform the CFA. The fit was assessed based on the following fit index values: root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), comparative fit index (CFI), and the Tucker–Lewis index (TLI). RMSEA and SRMR values below .08, and CFI and TLI values greater than .9 indicate an acceptable fit (Hu & Bentler, 1999). Measurement equivalence analysis was performed in configural, metric, and scalar levels across gender and age groups. While testing metric and scalar invariance, the ΔCFI of < .01 and ΔRMSEA of < .015 indicates invariance (Chen, 2007). For assessing internal consistency reliability, McDonald's ωcoefficients were calculated for the two PHQ–4 subscales and the total score.

For assessing test-retest reliability of the PHQ–4, we calculated intraclass correlation coefficients (two-way mixed method with absolute

agreement type; Koo & Li, 2016) between the PHQ–4 scores at baseline and a three-week follow-up. For intraclass correlation coefficients, values less than .50, between .50 and .75, between .75 and .90, and greater than .90 indicate poor, moderate, good and excellent reliability, respectively (Koo & Li, 2016).

A series of one-way analyses of variance (ANOVAs) with a Bonferroni correction (post hoc comparisons) was used to compare the levels of anxiety and depressive symptoms among different age groups in females and males separately. We calculated the total effect size for ANOVAs (partial eta-squared, η_p^2; interpretation: negligible < .01 < small < .06 < medium < .14 < large) and for post hoc comparisons (Cohen's *d*; interpretation: negligible < .20 < small < .50 < medium < .80 < large). When examining differences in the PHQ–4 scores by sex, education, marital status, residence, and employment, a series of one-way analyses of covariance (ANCOVAs) with the Bonferroni correction (post hoc comparisons) was used. This analysis provided the control of age influence on the PHQ–4 scores. We also reported the effect size for ANCOVAs (η_p^2). Due to an unequal sample size of the non-binary group (N = 16; all people aged 18–24) compared to other age-gender groups, we compared the non-binary group with appropriate age groups (aged 18–29) of females and males using a non-parametric Kruskal–Wallis H test, which is usually applied for unbalanced sample sizes (Macunluoglu & Ocakoğlu, 2022). For this test, we calculated the epsilon-squared (ε^2; with the same interpretation as indicated above for η_p^2) estimate of the effect size (Tomczak & Tomczak, 2014). Then, we compared separate groups of females and males aged 18–29 with the non-binary group using the Mann-Whitney U test and presented effect sizes (eta-squared, η^2; with the same interpretation as indicated above for η_p^2, McCall, 2018) for these differences. We did not use the Bonferroni correction for these three comparisons because we had a small number of planned comparisons (Armstrong, 2014).

For the LPA, different profile models (from 1 to 6 classes; equal variances, covariances fixed to 0) were tested to identify the best fitting model. To identify the best fiting model, we used the Bayesian information criteria (BIC) and the Akaike information criteria (AIC), the appropriate weight of evidence criterion (AWE), the consistent Akaike information criterion (CAIC), the classification likelihood criterion (CLC), the Kullback information criterion (KIC), the sample size-adjusted Bayesian information criterion (SABIC), and the integrated completed likelihood (ICL). Lower values of these indices indicate a better fitting model. We also used entropy values, with higher values (being acceptable ≥.60 or good ≥.80) indicating a higher certainty of classifying participants into the empirically extracted profiles correctly. The bootstrap likelihood ratio (BLRT) test was used to determine if there was a statistically significant improvement in model fit providing the inclusion of one more class. We also analyzed the size of the smallest class with acceptable values of ≥ 5%. When identifying an optimal solution, we relied on all of these criteria (Spurk et al., 2020). We also analyzed the theoretical and practical premises along with the meaningfulness of the extracted profiles.

The data were screened for accuracy (minimum and maximum range of each variable). There was no missing data. Statistical analysis was carried out using *Statistica* (version 13.3), *SPSS* (version 28) and *R* (version 4.2.1). In *R*, the following packages were used: *lavaan* and *semTools* (for the CFA and measurement invariance), *psych* (for the reliability analysis) as well as *tidyLPA* (for the LPA).

RESULTS

Table 1 presents the descriptive statistics for the PHQ–4 scores in different gender groups (i.e., females, males, and non-binary individuals). Skewness and kurtosis values ranged from -1.34 to .19, indicating the fact that the distribution of the PHQ–4 subscales in these groups was normal.

Factor Structure and Measurement Invariance

In the total sample, all PHQ–4 items were normally distributed (skewness values ranged from -.05 to .09, whereas kurtosis values ranged from -1.28 to -1.25). The one-factor model showed a satisfactory fit to the data, whereas the intended two-factor model was an excellent fit to the data using both the ML and WLS estimation methods (see Table 2). The results obtained with these two methods were highly similar. Therefore, we based our findings and measurement invariance on the ML estimation method, which was used in the majority of validation studies on the PHQ–4, including the original one by Löwe et al.

TABLE 1.
Descriptive Statistics for the PHQ–4 Scores in Different Gender Groups and Prevalence of Positively Screened Participants

Scales	Anxiety	Depression	Total score	Anxiety	Depression	Total score	Anxiety	Depression	Total score
Gender	Females (N = 1730)			Males (N = 811)			Non-binary (N = 16)		
M	3.54	3.16	6.71	2.96	2.99	5.95	4.56	4.25	8.81
SD	1.80	1.92	3.42	1.86	2.02	3.56	1.50	1.53	2.74
Median	4.00	3.00	7.00	3.00	3.00	6.00	4.50	4.00	9.50
Skewness	−.10	.05	.01	.19	.08	.15	−.47	−.22	−.44
Kurtosis	−1.14	−1.16	−1.06	−1.03	−1.22	−1.04	−1.21	−1.34	−1.20
McDonald's ω	.75	.81	.85	.73	.81	.85	.80	.35	.79
% of positively screened	66.9	57.20	60.20	53.8	54.3	52.3	87.50	81.20	81.20

Note. Percentage of positively screened participants represents the percentage of participants with a score ≥ 3 for anxiety and depression subscales as well as ≥ 6 for the total PHQ–4 score.

(2010). All factor loadings were high and statistically significant (p < .001; anxiety: Item 1 = .732, Item 2 = .817; depression: Item 3 = .739, Item 4 = .920). The estimated covariance between the anxiety and the depression subscales was .870 (p < .001). The anxiety and depression subscales were highly correlated, but there is theoretical and statistical value in separating them.

Table 2 presents the goodness-of-fit indices for the two-factor model across gender and age categories. In each group, the intended two-factor model had excellent fit to the data. It should be stressed that in two age groups, the RMSEA values were unacceptable. However, this is a common issue with the RMSEA in models with small degrees of freedom (Kenny et al., 2015). The analyses showed that the two-factor model had configural, metric, and scalar invariance across different gender and age categories. In summary, the intended PHQ–4 factor structure and its invariance were supported empirically.

Test-Retest Reliability and Internal Consistency Reliability

Seventy-two participants (57 females, 14 males and 1 non-binary individual, aged 20–51 (M = 24.29, SD = 6.53) filled out the PHQ–4 two times with an interval of approximately three weeks between each test. Intraclass correlation coefficients of all the PHQ–4 scores between the two measurements were high (≥ .59) and statistically significant (all ps < .001; anxiety = .59; depression = .62; total PHQ–4 score = .68), thus supporting test-retest reliability of the PHQ–4.

The reliability of the PHQ–4 was high (McDonald's ω ≥ .70) in all age groups among females and males (see Table 3). In the group of non-binary individuals, only the depression subscale had low reliability (McDonald's ω= .35; see Table 1). Pearson correlation between two items of the anxiety subscale was .60 (p < .001), whereas between two items of the depression subscale, it was .68 (p < .001).

Age and Gender Differences in Anxiety and Depressive Symptoms

Table 3 presents descriptive statistics for the PHQ–4 scores in different age groups in females (aged 18–29, 30–44, 45–59, and 60–78) and males (aged 18–29, 30–44, 45–59, and 60–76) as well as the prevalence of positively screened participants in each of these groups. Among females and males, different age groups differed in anxiety (females: F[3,1726] = 29.66, η_p^2 = .049, p < .001; males: F(3,807) = 6.17, p < .001, η_p^2 = .022), depression (females: F[3,1726] = 28.49, p < .001, η_p^2 = .047; males: F[3,807] = 9.13, p < .001, η_p^2 = .033), and total scores (females: F[3,1726] = 34.22, p < .001, η_p^2 = .056; males: F[3,807] = 8.77, p < .001, η_p^2 = .032; see Table 3).

The ANOVA post hoc comparison analysis revealed that females aged 18–29 had higher levels of anxiety, depressive, and total anxiety-depressive symptoms than females aged 30–44 (anxiety: p < .001, d = .39; depression: p = .003, d = .33; total score: p < .001, d = .39), 45–59 (anxiety: p = .004, d = .26; depression: p < .001, d = .39; total score: p < .001, d = .35) and 60–78 years (anxiety: p < .001, d = .65; depression: p < .001, d = .59; total score: p < .001, d = .68). Females aged 60–78 had significantly lower anxiety scores and total scores than females aged 45–59 years (anxiety: p < .001, d = .37; total score: p = .004, d = .31).

TABLE 2.
Goodness-of-Fit Indices for the PHQ–4 Models Together With a Measurement Invariance Analysis of the PHQ–4 Across Gender and Age Categories

Samples	χ^2/df	CFI	TLI	RMSEA (90% confidence interval)	SRMR
One-factor model					
Total sample (N = 2557; ML estimation)	130.343/2	.971	.914	.158 (.136; .182)	.032
Total sample (N = 2557; WLS estimation)	83.540/2	.961	.883	.126 (.104; .150)	.035
Two-factor model (correlated anxiety and depression subscales)					
Total sample (N = 2557; ML estimation)	5.685/1	.999	.994	.043 (.014; .080)	.006
Total sample (N = 2557; WLS estimation)	4.505/1	.998	.990	.037 (.008; .075)	.006
Females (N = 1730; ML estimation)	6.873/1	.998	.988	.058 (.023; .103)	.008
Males (N = 811; ML estimation)	.081/1	1.000	1.004	.000 (.000; .062)	.001
People aged 18–29 (N = 1715; ML estimation)	1.867/1	1.000	.998	.022 (.000; .073)	.005
People aged 30–44 (N = 234; ML estimation)	4.461/1	.992	.949	.122 (.027; .246)	.019
People aged 45–59 (N = 303; ML estimation)	4.143/1	.995	.971	.102 (.016; .211)	.013
People aged 60–78 (N = 305; ML estimation)	.190/1	1.000	1.007	.000 (.000; .115)	.002

Invariance testing level (ML estimation)	Gender (females and males)					Age groups (18–29, 30–44, 45–59, 60–78)				
Model	χ^2 (df)	CFI	ΔCFI	RMSEA	ΔRMSEA	χ^2 (df)	CFI	ΔCFI	RMSEA	ΔRMSEA
Configural	6.954 (2)	.999	—	.044	—	10.662 (4)	.998	—	.051	—
Metric	7.780 (4)	.999	0	.027	-.017	16.822 (10)	.998	0	.033	-.018
Scalar	13.403 (6)	.998	-.001	.031	.004	28.951 (16)	.997	-.001	.036	.003

Note. ML = maximum likelihood, WLS = weighted least squares, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual, CFI = comparative fit index, TLI = Tucker–Lewis index.

The ANOVA post hoc comparison analysis revealed that males aged 60–76 had statistically significantly lower anxiety and depressive symptoms than males aged 18–29 (p = .003, d = .42 and p < .001, d = .59, respectively), 30–44 (anxiety: p < .001, d = .60; depression: d = .65), and 45–59 years (anxiety: p = .012, d = .50; depression: p = .019, d = .44), as well as lower levels of anxiety-depressive symptoms than males aged 18–29 (p < .001, d = .56), 30–44 (p < .001, d = .67) and 45–59 years (p = .006, d = .48; see Table 3).

Non-binary individuals were aged 18–24, and therefore, this group was compared with female and male samples aged 18–29 using the Kruskal–Wallis H test. These three groups differed significantly in anxiety, $H(2) = 76.77$, p < .001, ε^2 = .045, and depressive symptoms, $H(2) = 13.74$, p = .001, ε^2 = .008, and total scores, $H(2) = 45.20$, p < .001, ε^2 = .026. Post hoc comparisons using the Mann-Whitney U test revealed that the non-binary group had significantly higher anxiety, $U(N_{male} = 531, N_{non\text{-}binary} = 16) = 2187.50$, z = 3.31, p = .001, η^2 = .02, and depressive symptoms, $U(N_{male} = 531, N_{non\text{-}binary} = 16) = 2833.00$, z = 2.27, p = .023, η^2 = .009, and total scores, $U(N_{male} = 531, N_{non\text{-}binary} = 16) = 2331.50$, z = 3.08, p = .002, η^2 = .017) than males aged 18–29. There were no significant differences between the non-binary group and females aged 18–29 (ps from .06 to .09).

Summarizing the results, females, non-binary individuals, and younger people in general experienced higher symptoms of mental health conditions. A series of one-way ANCOVAs was conducted to examine whether people with different education, marital status, residence, and occupation differed concerning the PHQ–4 results (age was used as a covariate for controlling its influence; see Table 4).

The results showed that participants with lower educational levels and unemployed participants scored higher on anxiety-depressive symptoms. As for marital status, being single was related to a significantly higher depression scores. Place of residence did not alter the PHQ–4 scores. Detailed results are presented in Table 4.

Latent Profile Analysis for Anxiety and Depression Symptoms

The AIC, BIC, AWE, CAIC, CLC, KIC, SABIC, and ICL indices decreased with adding a new class (see Table 5). In general, in Profiles 2 to 5, the entropy values were comparable in the samples of females and males, with the highest value observed in the six-profile solution. However, the least numerous profile of the six-profile solution in the sample of males had less than 5% of participants, which indicates the rejection of this solution. Thus, based on all the fit indices and theoretical and practical premises along with the meaningfulness of the extracted profiles, we identified four profile solutions in the samples of females and males as optimal and succinct.

The description of the four distinguished profiles in females and males is presented in Table 6. Figures 1 and 2 present the profiles graphically.

The distinguished profiles in females and males were very similar according to their clinical meaningfulness and prevalence. We identified four common subpopulations among females and males with similarities in clinical meaningfulness. The first and the most numerous subpopulation was represented by Profile 3 in females and Profile 1 in males. It represented the subpopulation with low anxiety and depressive symptoms. The mean scores for anxiety and depression suggested negative screening results in both anxiety and depression. The second subpopulation was represented by Profile 1 in females and Profile 4 in males. It represented a subpopulation with very high anxiety and very high depressive symptoms. The mean scores for anxiety and depression suggested positive screening results for both anxiety and depression. The third subpopulation was represented by Profile 2 in females and Profile 3 in males. It represented a subpopulation with high anxiety symptoms and moderate depressive ones. The mean scores for anxiety and depression suggested positive screening results for anxiety and negative ones for depression. The fourth subpopulation was represented by Profile 4 in females and Profile 2 in males. It represented a subpopulation with moderate anxiety symptoms and high depressive ones. The mean scores of anxiety and depression in Profile 4 in females suggested positive screening results for both anxiety and depression. The mean scores for anxiety and depression in Profile 2 in males suggested negative screening results for anxiety and positive ones for depression.

Summarizing the LPA results, four subpopulations were distinguished among females and males. There were (a) a non-anxious and non-depressed subpopulation, (b) a highly anxious and highly depressed subpopulation, (c) a highly anxious and moderately depressed subpopulation, and (d) a moderately anxious and highly depressed subpopulation.

DISCUSSION

Factor Structure, Measurement Invariance, and Reliability

The first aim of the current study was to examine the factor structure of the PHQ–4 and its measurement invariance for different age and gender groups. The analysis showed that the Polish version of the PHQ–4 is characterized by the intended factor structure. Measurement invariance of the PHQ–4 regarding its configural, metric, and scalar invariance across age and gender categories was supported empirically. The obtained CFA results were consistent with other validation studies (Löwe et al., 2010; Mendoza et al., 2022; Tibubos et al., 2020). The internal consistency reliability of all the PHQ–4 subscales and total scores was high for all age groups in females and males, and test-retest reliability was also supported. In sum, the Polish version of the PHQ–4 has an intended two-factor structure and is invariant across age and gender categories. This demonstrates the possibility of comparing PHQ–4 scores across different age and gender groups. Taking together these results with our previous findings (Larionow, Mudło-Głagolska et al., 2022; Larionow, Preece et al., 2022) supporting good convergent and discriminant validity of the Polish version of the PHQ–4, the questionnaire appears to have strong psychometric properties.

TABLE 3.
Descriptive Statistics of Anxiety, Depression, and the Total PHQ–4 Scores in Different Age Groups by Gender

Age groups	Females 18–29 (*N* = 1168; 67.51% of all females)			Females 30–44 (*N* = 126; 7.28%)			Females 45–59 (*N* = 216; 12.49%)			Females 60–78 (*N* = 220; 12.72%)		
Scales	A	D	T	A	D	T	A	D	T	A	D	T
M	3.79	3.44	7.23	3.12	2.83	5.95	3.35	2.72	6.06	2.65	2.33	4.98
SD	1.71	1.86	3.24	1.78	1.81	3.28	1.85	1.90	3.52	1.91	1.94	3.57
Skewness	−.18	−.10	−.09	.08	.29	.17	−.01	.32	.21	.47	.62	.55
Kurtosis	−1.10	−1.12	−1.03	−1.02	−.95	−.80	−1.19	−.98	−1.07	−.95	−.76	−.75
McDonald's ω	.70	.79	.83	.73	.76	.84	.79	.82	.88	.85	.86	.89
% of positively screened	73.10	64.40	66.60	61.80	47.60	54	57.90	46.80	51.90	44.10	34.60	37.8
Age groups	Males 18–29 (*N* = 531; 65.47% of all males)			Males 30–44 (*N* = 108; 13.32%)			Males 45–59 (*N* = 87; 10.73%)			Males aged 60–76 (*N* = 85; 10.48%)		
Scales	A	D	T	A	D	T	A	D	T	A	D	T
M	2.98	3.12	6.09	3.33	3.29	6.62	3.09	2.86	5.95	2.22	1.96	4.10
SD	1.82	1.96	3.42	1.93	2.09	3.71	1.93	2.09	3.77	1.77	1.97	3.51
Skewness	.20	.04	.15	.02	−.08	−.07	.11	.08	.15	.51	.76	.57
Kurtosis	−1.00	−1.19	−.98	−1.27	−1.27	−1.22	−1.12	−1.31	−1.16	−.55	−.57	−.65
McDonald's ω	.71	.78	.83	.75	.83	.87	.74	.89	.89	.73	.85	.88
% of positively screened	54.30	57.20	53.70	61.10	57.40	58.70	51.80	51.90	51.90	36.40	35.4	36.50

Note. A = anxiety score, D = depression score, T = total score, % of positively screened participants represents the percentage of participants with a score of ≥ 3 for anxiety and depression subscales as well as ≥ 6 for the total PHQ–4 score.

Screening Results of Anxiety and Depressive Symptoms in Different Age-Gender Groups

The second aim of the current study was to estimate the prevalence of anxiety and depressive symptoms according to the Diagnostic Algorithms for the PHQ (Pfizer, 2022) in different age and gender groups. The best mental health state observed in females aged 60–78 and males aged 60–76. Females aged 18–29 had the worst mental health among the age groups in the sample of females. In this group, about 65% of females were screened positively for depression and/or anxiety-depressive disorders. In contrast, males aged 30–44 had the worst mental health (about 57% of them were screened positively for depression) compared to other age groups in the sample of males. We should also note that individuals who identified as non-binary had extremely high intensity and prevalence of anxiety and depressive symptoms (more than 80% of them were screened positively). However, due to the small number of people in this group (*N* = 16), we cannot generalize these findings. To synthesize the results, more than half of Poles were screened positively for anxiety and/or depression. Females experienced more anxiety and anxiety-depressive symptoms than males. Younger people (especially non-binary individuals, females aged 18–29, and males aged 30–44) experienced more anxiety and depressive symptoms. Our results are in line with Gawrych et al. (2022) who revealed that the majority of young Poles aged 18–23 (56.5%) met the criteria (of a screening questionnaire) for depressive disorders. In general, the obtained results are consistent with the conclusions presented in other Polish studies (e.g., Gambin et al., 2021) and indicate the need for further monitoring of mental health in these risk groups.

Based on the cut-off scores for the PHQ–2 and the GAD–2 of ≥ 3, we compared our results with the German ones, because the German population is similar to the Polish one from a sociocultural point of view. In the nationally representative German sample in the years 2012–2014, 9.8% of the participants had probable anxiety, whereas 10.4% had probable depression (Hajek & König, 2020). The newest German group norms of the PHQ–4 (years 2020–2021) suggest that 6.5 and 7.0% of the participants had probable anxiety and depression, respectively (Wicke et al., 2022). By comparing these German results with ours, it can be concluded that the prevalence of probable anxiety and depressive disorders in the Polish population is extremely high. Wicke et al. (2022) also indicated that the prevalence of possible anxiety and depression is very similar in different age groups among females and males in the German population. In contrast, in our sample, the prevalence of probable anxiety and depression decreased with age in females (prevalence of positive screening results from the highest to lowest in age groups: 18–29 > 30–44 > 45–59 > 60–78), whereas a somewhat different trend was observed in males (30–44 > 18–29 > 45–59 > 60–76), indicating the specific age-gender risk groups for mental health problems. The prevalence of possible anxiety and depression in our Polish sample was much higher than in a Korean general public sample in Busan, since 30.7% of Korean respondents had possible depression and 22.6% had possible anxiety (Kim et al., 2021).

All things considered, more than one-half of our Polish participants in three age groups (i.e., 18–29, 30–44, and 45–59) were screened positively for anxiety and depression (except depression in females aged 30–44 [47.6%] and 45–59 [46.8%], see Table 3). More than one-third of females aged 60–78 and males aged 60–76 were screened positively

TABLE 4.

Differences in Anxiety and Depression PHQ–4 Scores for Groups in Different Sex, Education, Marital Status, Residence, and Main Activity Categories (Controlling for Age, Bonferroni Corrected)

	Anxiety score				Depression score			Total score		
	N	*M* (*SD*)	Adjusted M^a (*SE*)	Model parameters and significant differences	*M* (*SD*)	Adjusted M^a (*SE*)	Model parameters and significant differences	*M* (*SD*)	Adjusted M^a (*SE*)	Model parameters and significant differences
Sex										
Females	1730	3.54 (1.80)	3.54 (.04)	$F(1, 2538) = 62.12$, $p < .001$, $\eta_p^2 = .024$	3.17 (1.92)	3.17 (.05)	$F(1, 2538) = 5.68$, $p = .017$, $\eta_p^2 = .002$	6.71 (3.42)	6.72 (.08)	$F(1, 2538) = 30.35$, $p < .001$, $\eta_p^2 = .012$
Males	811	2.96 (1.86)	2.95 (.06)	Differences: females > males	2.99 (2.02)	2.98 (.07)	Differences: females > males	5.95 (3.56)	5.92 (.12)	Differences: females > males
Educational categories										
Primary	192	3.73 (1.86)	3.58 (.13)	$F(3, 2552) = 4.01$, $p = .007$, $\eta_p^2 = .005$. Differences: primary > higher; secondary > higher.	3.83 (1.79)	3.68 (.14)	$F(3, 2552) = 12.17$, $p < .001$, $\eta_p^2 = .014$. Differences: primary > secondary; primary > higher; vocational > higher; secondary > higher.	7.56 (3.22)	7.26 (.25)	$F(3, 2552) = 9.01$, $p < .001$, $\eta_p^2 = .010$. Differences: primary > higher; vocational > higher; secondary > higher.
Vocational	130	3.63 (1.87)	3.59 (.16)		3.45 (1.94)	3.41 (.17)		7.08 (3.48)	7.00 (.30)	
Secondary	1537	3.49 (1.83)	3.41 (.05)		3.26 (1.95)	3.18 (.05)		6.75 (3.47)	6.59 (.09)	
Higher	698	2.94 (1.78)	3.15 (.08)		2.54 (1.88)	2.76 (.08)		5.48 (3.36)	5.92 (.14)	
Marital status										
Single	1359	3.40 (1.83)	3.33 (.05)	ns.	3.31 (1.94)	3.25 (.05)	F(1,2554) = 12.71, p < .001, ηp2 = .005. Differences: single > in a relationship.	6.71 (3.45)	6.58 (.09)	ns.
In a relationship	1198	3.32 (1.85)	3.40 (.05)		2.89 (1.95)	2.97 (.06)		6.22 (3.50)	6.37 (.10)	
Residence										
Villages	689	3.35 (1.82)	3.31 (.07)	ns.	3.13 (1.93)	3.08 (.07)	ns.	6.49 (3.40)	6.39 (.13)	ns.
Small towns (up to 20000 inhabitants)	351	3.32 (1.91)	3.31 (.10)		3.13 (1.97)	3.11 (.10)		6.45 (3.56)	6.42 (.18)	
Medium-sized towns (from 20000 to 100000)	593	3.42 (1.87)	3.43 (.07)		3.24 (1.96)	3.25 (.08)		6.67 (3.54)	6.68 (.14)	
Large cities (above 100000)	924	3.34 (1.81)	3.38 (.06)		3.01 (1.96)	3.06 (.06)		6.35 (3.48)	6.44 (.11)	
Main activity										
Unemployed	477	3.76 (1.80)	3.67 (.08)	$F(4, 2551) = 4.57$, $p = .001$, $\eta_p^2 = .007$. Differences: unemployed > employed; unemployed > combining work and studies; unemployed > students.	3.67 (1.96)	3.55 (.09)	$F(4, 2551) = 8.88$, $p < .001$, $\eta_p^2 = .014$. Differences: unemployed > employed; unemployed > combining work and studies; unemployed > students.	7.43 (3.42)	7.22 (.16)	$F(4, 2551) = 7.67$, $p < .001$, $\eta_p^2 = .012$. Differences: unemployed > employed; unemployed > combining work and studies; unemployed > students.
Employed	903	3.19 (1.86)	3.32 (.06)		2.90 (1.95)	3.08 (.07)		6.09 (3.54)	6.40 (.12)	
Combining work and studies	493	3.33 (1.69)	3.22 (.08)		2.99 (1.87)	2.85 (.09)		6.32 (3.22)	6.07 (.16)	
Students	607	3.49 (1.86)	3.34 (.08)		3.25 (1.89)	3.04 (.08)		6.74 (3.42)	6.38 (.15)	
Retired	77	2.14 (1.85)	3.04 (.26)		1.99 (1.92)	3.21 (.27)		4.13 (3.38)	6.25 (.48)	

Note. ns = nonsignificant, η_p^2 = partial eta squared (interpretation: negligible < .01 < small < .06 < medium < .14 < large).[a] = The adjusted means (controlling for the covariate of age). The covariates in the models for sex categories were estimated at the following values: age = 27.22, whereas the covariates in the models for educational categories, marital status, residence and main activity categories were estimated at the following values: age = 27.18.

TABLE 5.
Results of Latent Profile Analysis in Females and Males

Number of classes	AIC	BIC	AWE	CAIC	CLC	KIC	SABIC	ICL	Entropy	prob_ min	prob_ max	n_min	n_max	BLRT *p*-value
Females (N = 1730)														
1	14116.66	14138.49	14178.31	14142.49	14110.66	14123.66	14125.78	-14138.49	1.00	1.00	1.00	1.00	1.00	—
2	12973.14	13011.33	13082.95	13018.33	12960.71	12983.14	12989.09	-13285.20	.79	.92	.95	.46	.54	.01
3	12752.20	12806.76	12909.85	12816.76	12733.67	12765.20	12774.99	-13325.22	.73	.82	.93	.29	.37	.01
4	12614.06	12684.99	12819.39	12697.99	12589.58	12630.06	12643.69	-13245.39	.76	.69	.96	.16	.37	.01
5	12518.26	12605.56	12771.31	12621.56	12487.80	12537.26	12554.72	-13208.09	.77	.70	.94	.10	.33	.01
6	12429.73	12533.39	12730.41	12552.39	12393.36	12451.73	12473.03	-13054.17	.82	.74	.98	.07	.27	.01
Males (N = 811)														
1	6757.95	6776.74	6813.53	6780.74	6751.95	6764.95	6764.04	-6776.74	1.00	1.00	1.00	1.00	1.00	—
2	6235.53	6268.42	6334.71	6275.42	6223.12	6245.53	6246.19	-6380.53	.80	.94	.94	.45	.55	.01
3	6128.83	6175.81	6271.28	6185.81	6110.34	6141.83	6144.05	-6399.65	.76	.82	.92	.28	.39	.01
4	6066.94	6128.01	6252.56	6141.01	6042.47	6082.94	6086.73	-6377.88	.77	.75	.94	.14	.39	.01
5	6032.52	6107.69	6261.38	6123.69	6002.01	6051.52	6056.88	-6424.53	.74	.79	.93	.14	.26	.01
6	5961.96	6051.23	6233.80	6070.23	5925.65	5983.96	5990.89	-6255.81	.85	.46	.99	.02	.26	.01

Note. AIC = the Akaike information criterion, BIC = the Bayesian information criterion, AWE = appropriate weight of evidence criterion, CAIC = consistent Akaike information criterion, CLC = classification likelihood criterion, KIC = Kullback information criterion. SABIC = sample size-adjusted Bayesian information criterion, ICL = integrated completed likelihood, prob_min = minimum of the diagonal of the average latent class probabilities for the most likely class membership, by assigned class; prob_max = maximum of the diagonal of the average latent class probabilities for the most likely class membership, by assigned class; n_min = the proportion of the sample assigned to the smallest class; n_max = the proportion of the sample assigned to the largest class; BLRT_p = p value for the bootstrapped likelihood ratio test (Rosenberg et al., 2018).

TABLE 6.
The Description of the Latent Profile Analysis Profiles

Profiles in females (N = 1730)	Profiles in males (N = 811)	Screening results (based on means)	Clinical meaningfulness	Names of subpopulations	Mental disorder risk
Profile 3 (N = 644, 37%): Anxiety (M = 1.68, SE = .06) Depression (M = 1.33, SE = .05)	Profile 1 (N = 316, 39%): Anxiety (M = 1.28, SE = .08) Depression (M = 1.09, SE = .12)	Screened negatively for both anxiety and depression	Low anxiety and depressive symptoms	Non-anxious and non-depressed subpopulation	Low
Profile 1 (N = 491, 28%): Anxiety (M = 5.55, SE = .05), Depression (M = 5.36, SE = .05)	Profile 4 (N = 197, 24%): Anxiety (M = 5.39, SE = .09) Depression (M = 5.37, SE = .09)	Screened positively for both anxiety and depression	Very high anxiety and depressive symptoms	Highly anxious and highly depressed subpopulation	Very high
Profile 2 (N = 268, 16%): Anxiety (M = 4.27, SE = .10), Depression (M = 2.42, SE = .11)	Profile 3 (N = 186, 23%): Anxiety (M = 3.59, SE = .22) Depression (M = 2.64, SE = .19)	Screened positively for anxiety and negatively for depression	High anxiety symptoms and average depressive symptoms	Highly anxious and moderately depressed subpopulation	High
Profile 4 (N = 327, 19%): Anxiety (M = 3.11, SE = .14), Depression (M = 4.04, SE = .11)	Profile 2 (N = 112, 14%): Anxiety (M = 2.46, SE = .20) Depression (M = 4.56, SE = .20)	Females: screened positively for both anxiety and depression. Males: screened negatively for anxiety and positively for depression	Average anxiety and high depressive symptoms	Moderately anxious and highly depressed subpopulation	High
The variance for anxiety was .738 and .883 for depression	The variance for anxiety was .830 and .943 for depression	—	—	—	—

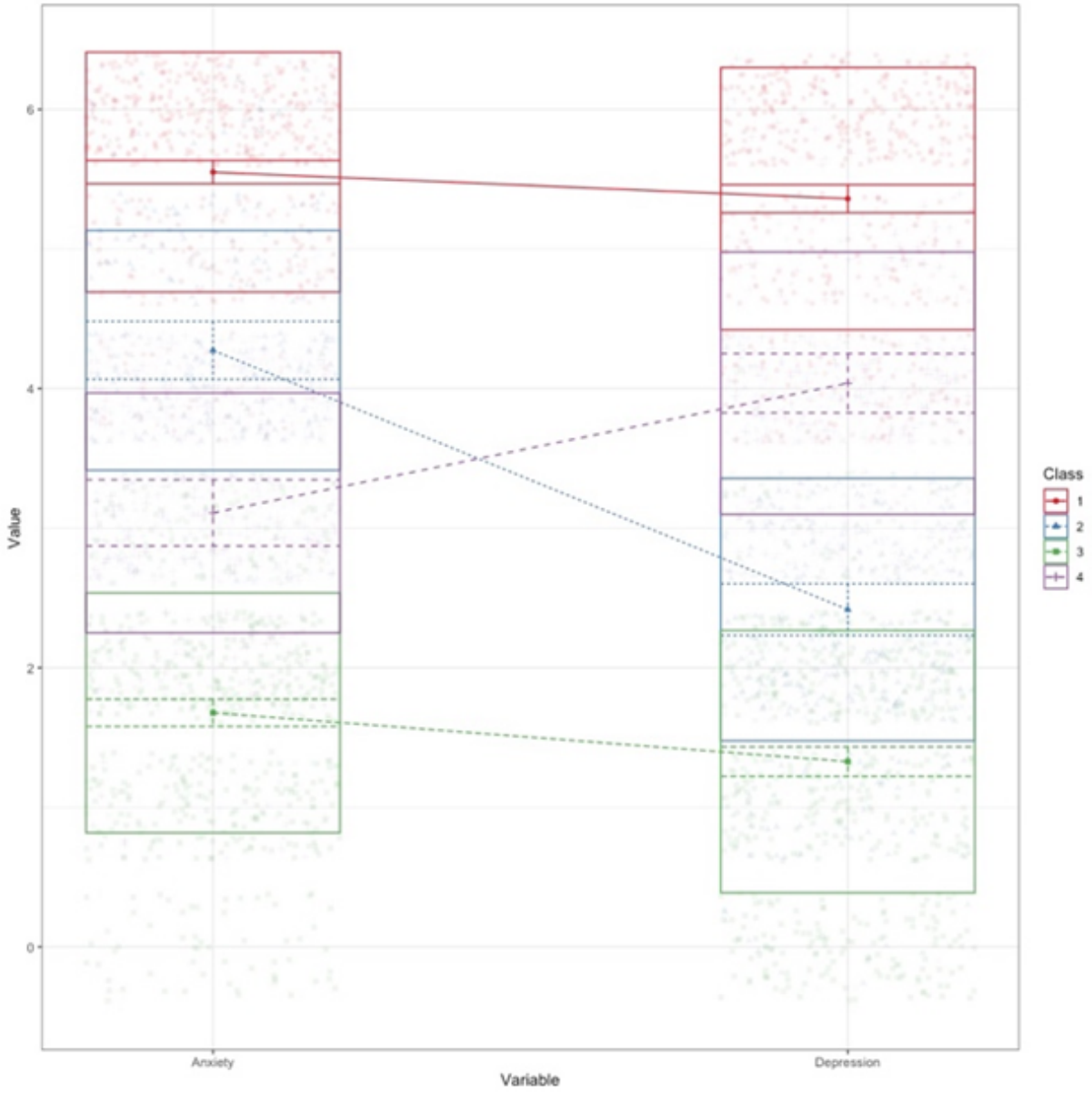


FIGURE 1.
Latent profiles for anxiety and depressive symptoms in females.

for anxiety and depression. People with positive screening results on the PHQ–4 need a more detailed diagnosis for these mental health problems using, for example, the PHQ–9, the Generalized Anxiety Disorder–7, or by visiting a psychiatrist.

Based on our results on the differences in PHQ–4 scores considering different sociodemographic characteristics, we identified high-risk groups for anxiety and depression disorders, namely, females, non-binary individuals, and younger people in general, as well as unemployed, less educated, and single people. Our results on high-risk groups are in line with previous studies (Löwe et al., 2010). Despite the fact that the prevalence of anxiety and depressive symptoms was higher in females than in males, researchers should take into account the fact that mild and moderate depression tends to be reported more often by females, whereas severe depression and suicidal behaviors tend to be reported more often by males (Shi et al., 2021). It should be stressed that males aged 30–44 had the highest prevalence of positive screening results compared to other age groups of males, highlighting the need for mental health diagnosis and psychological support for this group in primary care.

Latent Profile Analysis of Anxiety and Depressive Symptoms

The third aim of the current study was to identify subpopulations based on the anxiety and depressive symptoms in females and males separately. Our exploratory idea in the LPA was to investigate possible configurations of anxiety and depressive symptoms. Four subpopulations were identified: (a) a nonanxious and nondepressed subpopulation (a low-risk group with negative screening results for anxiety and depression), (b) a highly anxious and highly depressed

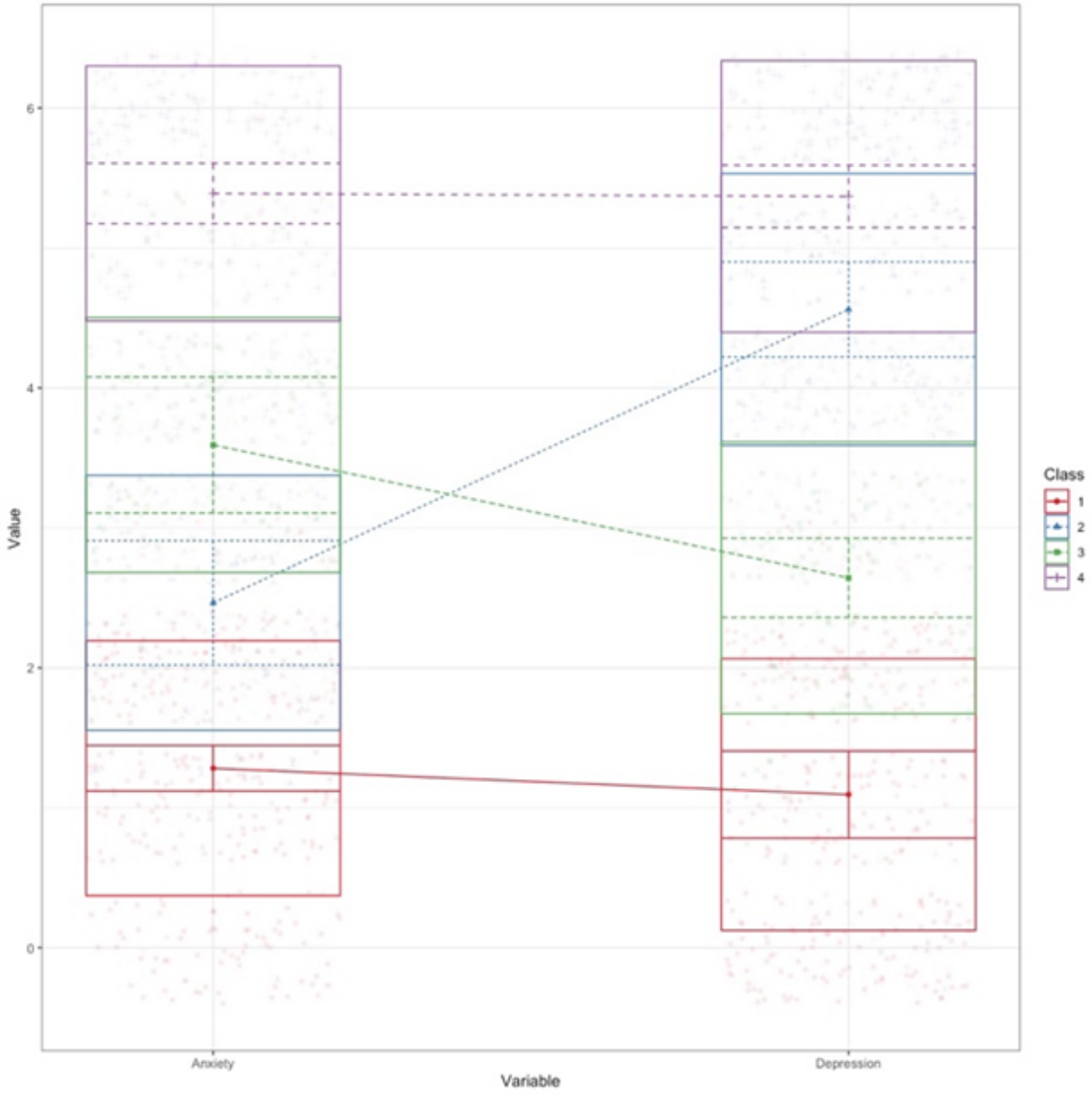


FIGURE 2.
Latent profiles for anxiety and depressive symptoms in males.

subpopulation (a very high-risk group with positive screening results for anxiety and depression), (c) a highly anxious and moderately depressed subpopulation (a high-risk group with positive screening results for anxiety and negative for depression), and (d) a moderately anxious and highly depressed subpopulation (a high-risk group with positive screening results for both disorders in females and positive ones for depression and negative ones for anxiety in males). The main finding was that these subpopulations were common in both females and males. The mean scores of anxiety and depression in these four common subpopulations, in general, were higher in females than in males (see Table 6, Figures 1 and 2). Depression and anxiety disorders are highly comorbid and their symptoms are frequently not separable (Kalin, 2020). However, our LPA allowed us to distinguish not only low and high mental health symptoms groups, but to examine individual profiles within a high symptoms group. This allowed us to provide a more specific and accurate description of subpopulations with various symptom levels. Our study has shown that the LPA can be useful in distinguishing separate, highly specific profiles of anxiety and depressive symptoms. Due to this, we present some new research directions to explore the etiology and maintenance of subthreshold psychological disorders and their clinical forms.

We believe that future studies may focus on (a) investigating profiles of anxiety and depressive symptoms and their prevalence within a general population-based sample of Poles, (b) examining how these subpopulations differ in demographics and physical and/or psychological variables, and (c) investigating psychosocial functioning and the effectiveness of treatment (e.g., psychological support, psychotherapy, or computer-delivered and web-based interventions; see Davies et al.,

2014), which is important for people with different configurations of anxiety and depressive symptoms. For example, it was shown that the presence of significant anxiety symptoms generally predicts worse outcomes for treatment of depression (for a review, see, e.g., Cosci & Fava, 2021; Kalin, 2020). We assume that future data, that is, the prevalence data in particular subpopulations with various symptom levels, may be helpful when developing and providing mental health care programs, which will be individually designed for different profiles of anxiety and depressive symptoms for better effectiveness. Our future studies will focus on anxiety and depression symptom profiles combined with emotion regulation variables that may characterize these profiles in order to determine possible mechanisms underlying the development of anxiety and depression disorders.

Practical Applications

The Polish version of the PHQ–4 (see Appendix) is freely available. It is a very short questionnaire, which can be successively applied in epidemiological screening studies, primary care practice, and in clinical settings (e.g., for investigating anxiety and depressive symptoms and/or their predominant configurations as moderators of treatment, patient adherence, etc.).

Strengths and Limitations of the Study

A broad large sample with a wide range of represented age groups (with sufficient sample sizes for age- and gender-specific subgroups for a PHQ–4 screening analysis) and gender identities were the strengths of the current study. The results of this study, which was conducted online, stress the effectiveness of online mental health screening and provide good opportunities for further research in different non-clinical or clinical settings in large samples (Larionow, 2022). It should be stressed that our sample is not a population-based sample (according to the Polish census data). Moreover, people with poor internet accessibility, especially older people, might not have participated in the study, leading to a selection bias in the sample. However, our sample has sufficient sizes for age- and gender-specific subgroups for conducting a PHQ–4 screening analysis and examining the factorial validity of the questionnaire and its measurement invariance across gender and age categories.

Our analyses suggest that the LPA is a promising approach in terms of separating groups of people based on their anxiety and depressive symptom scores. Thus, future work on LPA will be needed to test the generalizability of our findings on profiles with anxiety and depressive symptoms in a general, representative sample of Poles. The lack of a gold standard for establishing anxiety and depression disorders as well as a low reliability of the depression subscale in the group of non-binary individuals are also the limitations of the current study.

CONCLUSIONS

Overall, the Polish version of the PHQ–4 appears to have strong psychometric properties. It has an intended two-factor structure, is invariant across age and gender categories, and is characterized by good test-retest and internal consistency reliabilities, along with an empirically supported validity. Due to the extremely high prevalence and intensity of anxiety and depressive symptoms in the Polish population from February to July 2022, it is necessary to provide further mental health monitoring and conduct qualitative research on the psychological reasons for their occurrence, especially in high-risk groups (females, males aged 30–44, non-binary individuals, and younger people as well as unemployed, less educated, and single people). The high levels of anxiety and depression symptoms can be considered a precursor to a possible major increase in mental disorders in 2023.

ACKNOWLEDGEMENTS

The authors are grateful to Dr David A. Preece (Curtin University, Perth, Australia) for his advice on the statistical analyses. The authors would also like to thank the reviewers of this paper for their recommendations, and the participants of this study for their effort.

The authors state no conflict of interest and no external sources of funding.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics, 34*(5), 502–508. https://doi.org/10.1111/opo.12131

Cano-Vindel, A., Muñoz-Navarro, R., Medrano, L. A., Ruiz-Rodríguez, P., González-Blanch, C., Gómez-Castillo, M. D., Capafons, A., Chacón, F., Santolaya, F., & PsicAP Research Group (2018). A computerized version of the Patient Health Questionnaire-4 as an ultra-brief screening tool to detect emotional disorders in primary care. *Journal of Affective Disorders, 234*, 247–255. https://doi.org/10.1016/j.jad.2018.01.030

Cavanagh, K., Strauss, C., Cicconi, F., Griffiths, N., Wyper, A., & Jones, F. (2013). A randomised controlled trial of a brief online mindfulness-based intervention. *Behaviour Research and Therapy, 51*(9), 573–578. https://doi.org/10.1016/j.brat.2013.06.003

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 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. *Journal of Psychosomatic Research, 160*, 110970. https://doi.org/10.1016/j.jpsychores.2022.110970

Cosci, F., & Fava, G. A. (2021). When anxiety and depression coexist: the role of differential diagnosis using clinimetric criteria. *Psychotherapy and Psychosomatics, 90*, 308–317. https://doi.org/10.1159/000517518

Davies, E. B., Morriss, R., & Glazebrook, C. (2014). Computer-delivered and web-based interventions to improve depression, anxiety, and

psychological well-being of university students: A systematic review and meta-analysis. *Journal of Medical Internet Research, 16*(5), e130. https://doi.org/10.2196/jmir.3142

Donker, T., van Straten, A., Marks, I., & Cuijpers, P. (2011). Quick and easy self-rating of Generalized Anxiety Disorder: validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. *Psychiatry Research, 188*(1), 58–64. https://doi.org/10.1016/j.psychres.2011.01.016

Gambin, M., Sękowski, M., Woźniak-Prus, M., Wnuk, A., Oleksy, T., Cudo, A., Hansen, K., Huflejt-Łukasik, M., Kubicka, K., Łyś, A. E., Gorgol, J., Holas, P., Kmita, G., Łojek, E., & Maison, D. (2021). Generalized anxiety and depressive symptoms in various age groups during the COVID-19 lockdown in Poland. Specific predictors and differences in symptoms severity. *Comprehensive Psychiatry, 105*, 152222. https://doi.org/10.1016/j.comppsych.2020.152222

García-Campayo, J., Zamorano, E., Ruiz, M. A., Pérez-Páramo, M., López-Gómez, V., & Rejas, J. (2012). The assessment of generalized anxiety disorder: psychometric validation of the Spanish version of the self-administered GAD-2 scale in daily medical practice. *Health and Quality of Life Outcomes, 10*, 114. https://doi.org/10.1186/1477-7525-10-114

Gawrych, M., Cichoń, E., & Kiejna, A. (2022). Depression among young adults – risks and protective factors in the COVID-19 pandemic. *Advances in Psychiatry and Neurology, 31*(2), 52–61. https://doi.org/10.5114/ppn.2022.118265

Ghaheri, A., Omani-Samani, R., Sepidarkish, M., Hosseini, M., & Maroufizadeh, S. (2020). The four-item Patient Health Questionnaire for anxiety and depression: A validation study in infertile patients. *International Journal of Fertility & Sterility, 14*(3), 234–239. https://doi.org/10.22074/ijfs.2020.44412

Hajek, A., & König, H. H. (2020). Prevalence and correlates of individuals screening positive for depression and anxiety on the phq-4 in the german general population: Findings from the Nationally Representative German Socio-Economic Panel (GSOEP). *International Journal of Environmental Research and Public Health, 17*(21), 7865. https://doi.org/10.3390/ijerph17217865

Herrera-Añazco, P., Urrunaga-Pastor, D., Benites-Zapata, V. A., Bendezu-Quispe, G., Toro-Huamanchumo, C. J., & Hernandez, A. V. (2022). Gender differences in depressive and anxiety symptoms during the first stage of the COVID-19 pandemic: A cross-sectional study in latin america and the caribbean. *Frontiers in Psychiatry, 13*, 727034. https://doi.org/10.3389/fpsyt.2022.727034

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55. https://doi.org/10.1080/10705519909540118

Kalin, N. H. (2020). The critical relationship between anxiety and depression. *The American Journal of Psychiatry, 177*(5), 365–367. https://doi.org/10.1176/appi.ajp.2020.20030305

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research, 44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Kim, D. M., Bang, Y. R., Kim, J. H., & Park, J. H. (2021). The prevalence of depression, anxiety and associated factors among the general public during COVID-19 pandemic: A cross-sectional study in Korea. *The Journal of Korean Medical Science, 36*(29), e214. https://doi.org/10.3346/jkms.2021.36.e214

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*, 205. https://doi.org/10.1186/1471-244X-14-205

Kokoszka, A., Jastrzębski, A., Obrębski, M. (2016). Psychometric properties of the polish version of Patient Health Questionnaire-9. *Psychiatria, 13*(4), 187–193.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2003). The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Medical Care, 41*(11), 1284–1292. https://doi.org/10.1097/01.MLR.0000093487.78664.3C

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics, 50*(6), 613–621. https://doi.org/10.1176/appi.psy.50.6.613

Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine, 146*(5), 317–325. https://doi.org/10.7326/0003-4819-146-5-200703060-00004

Larionow, P. (2022). Wybrane problemy badań w medycynie psychosomatycznej i sposoby na ich rozwiązania: przestrzeń dla nowych osiągnięć. *PsyArXiv*. https://doi.org/10.31234/osf.io/fbaxc

Larionow, P., & Mudło-Głagolska, K. (2022). Mental health risk factors during the COVID-19 pandemic in the Polish population. *Psychiatria, 19*(2), 90–110. https://doi.org/10.5603/PSYCH.a2021.0041

Larionow, P., Mudło-Głagolska, K., & Michalak, M. (2022). Towards psychosomatic medicine: The psychometric properties of the Polish version of the Giessen Subjective Complaints List (GBB-8) and the prevalence of somatic complaints in a Polish community sample. *Annales Universitatis Mariae Curie-Skłodowska. Sectio J, Paedagogia-Psychologia, 35*(4), 117–138. https://doi.org/10.17951/j.2022.35.4.117-138

Larionow, P., Preece, D. A., & Mudło-Głagolska, K. (2022). Assessing alexithymia across negative and positive emotions: Psychometric properties of the Polish version of the Perth Alexithymia Questionnaire. *Frontiers in Psychiatry, 13,* 1047191. https://doi.org/10.3389/fpsyt.2022.1047191

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Li, C., Friedman, B., Conwell, Y., & Fiscella, K. (2007). Validity of the Patient Health Questionnaire 2 (PHQ-2) in identifying major depression in older people. *Journal of the American Geriatrics Society, 55*(4), 596–602. https://doi.org/10.1111/j.1532-5415.2007.01103.x

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. *Journal of Affective Disorders, 122*(1-2), 86–95. https://doi.org/10.1016/j.jad.2009.06.019

Macunluoglu, A. C., & Ocakoğlu, G. (2022). Comparison of the performances of non-parametric k-sample test procedures as an alternative to one-way analysis of variance. *The European Research Journal*, 1–10. https://doi.org/10.18621/eurj.1037546

McCall, G.S. (2018). *Strategies for quantitative research: Archaeology by numbers* (1st ed.). Routledge.

Mendoza, N. B., Frondozo, C. E., Dizon, J., & 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. *Current Psychology, 2022*. https://doi.org/10.1007/s12144-022-02833-5

Miller, P., Newby, D., Walkom, E., Schneider, J., Li, S. C., & Evans, T. J. (2021). The performance and accuracy of depression screening tools capable of self-administration in primary care: A systematic review and meta-analysis. *The European Journal of Psychiatry, 35*(1), 1–18. https://doi.org/10.1016/j.ejpsy.2020.10.002

Monahan, P. O., Shacham, E., Reece, M., Kroenke, K., Ong'or, W. O., Omollo, O., Yebei, V. N., & Ojwang, C. (2009). Validity/reliability of PHQ-9 and PHQ-2 depression scales among adults living with HIV/AIDS in western Kenya. *Journal of General Internal Medicine, 24*(2), 189–197. https://doi.org/10.1007/s11606-008-0846-z

Narodowy Fundusz Zdrowia. (2020). *NFZ o zdrowiu. Depresja.* Narodowy Fundusz Zdrowia.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health, 42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y

Pfizer (2022, August 1). *Instructions for Patient Health Questionnaire (PHQ) and GAD-7 measures.* https://www.phqscreeners.com/images/sites/g/files/g10016261/f/201412/instructions.pdf

Preece, D. A., Mehta, A., Becerra, R., Chen, W., Allan, A., Robinson, K., Boyes, M., Hasking, P., & Gross, J. J. (2022). Why is alexithymia a risk factor for affective disorder symptoms? The role of emotion regulation. *Journal of Affective Disorders, 296*, 337–341. https://doi.org/10.1016/j.jad.2021.09.085

Reisner, S. L., Katz-Wise, S. L., Gordon, A. R., Corliss, H. L., & Austin, S. B. (2016). Social epidemiology of depression and anxiety by gender identity. *The Journal of Adolescent Health, 59*(2), 203–208. https://doi.org/10.1016/j.jadohealth.2016.04.006

Renovanz, M., Soebianto, S., Tsakmaklis, H., Keric, N., Nadji-Ohl, M., Beutel, M., Ringel, F., Wollschläger, D., & Hickmann, A. K. (2019). Evaluation of the psychological burden during the early disease trajectory in patients with intracranial tumors by the ultra-brief Patient Health Questionnaire for Depression and Anxiety (PHQ-4). *Supportive Care in Cancer, 27*(12), 4469–4477. https://doi.org/10.1007/s00520-019-04718-z

Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Van Lissa, C. J., Schmidt, J. A. (2018). tidyLPA: an R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *The Journal of Open Source Software, 3*(30), 978. https://doi.org/10.21105/joss.00978

Shi, P., Yang, A., Zhao, Q., Chen, Z., Ren, X., & Dai, Q. (2021). A hypothesis of gender differences in self-reporting symptom of depression: Implications to solve under-diagnosis and under-treatment of depression in males. *Frontiers in Psychiatry, 12*, 589687. https://doi.org/10.3389/fpsyt.2021.589687

Ślusarska, B. J., Nowicki, G., Piasecka, H., Zarzycka, D., Mazur, A., Saran, T., & Bednarek, A. (2019). Validation of the Polish language version of the Patient Health Questionnaire-9 in a population of adults aged 35-64. *Annals of Agricultural and Environmental Medicine, 26*(3), 420–424. https://doi.org/10.26444/aaem/99246

Spurk, D., Hirschi, A., Wang, M., Valero, D., Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior, 120*, 103445. https://doi.org/10.1016/j.jvb.2020.103445

Thombs, B. D., Ziegelstein, R. C., & Whooley, M. A. (2008). Optimizing detection of major depression among patients with coronary artery disease using the patient health questionnaire: data from the heart and soul study. *Journal of General Internal Medicine, 23*(12), 2014–2017. https://doi.org/10.1007/s11606-008-0802-y

Tibubos, A. N., & Kröger, H. (2020). A cross-cultural comparison of the ultrabrief mental health screeners PHQ-4 and SF-12 in Germany. *Psychological Assessment, 32*(7), 690–697. https://doi.org/10.1037/pas0000814

Tomaszewski, K., Zarychta, M., Bieńkowska, A., Chmurowicz, E., Nowak, W., & Skalska, A. (2011). Walidacja polskiej wersji jezykowej Patient Health Questionnaire-9 w populacji hospitalizowanych osób starszych. *Psychiatria Polska, 45*(2), 223–233.

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences, 21*(1), 19–25.

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. *Journal of Affective Disorders, 312*, 310–314. https://doi.org/10.1016/j.jad.2022.06.054

Wild, B., Eckl, A., Herzog, W., Niehoff, D., Lechner, S., Maatouk, I., Schellberg, D., Brenner, H., Müller, H., & Löwe, B. (2014). Assessing generalized anxiety disorder in elderly people using the GAD-7 and GAD-2 scales: results of a validation study. *The American Journal of Geriatric Psychiatry, 22*(10), 1029–1038. https://doi.org/10.1016/j.jagp.2013.01.076

RECEIVED 31.08.2022 | ACCEPTED 02.03.2023

APPENDIX

Kwestionariusz zdrowia pacjenta–4 The Polish version of the Patient Health Questionnaire–4 (PHQ–4)				
Jak często odczuwałaś/eś następujące problemy w ostatnich 2 tygodniach? Zaznacz "✓" lub kółkiem swoją odpowiedź.	Wcale	Przez kilka dni	Więcej niż połowę dni	Prawie każdego dnia
1. Zdenerwowanie, lęk lub irytację	0	1	2	3
2. Trudności związane z opanowaniem zamartwiania się	0	1	2	3
3. Małe zainteresowanie albo brak przyjemności w robieniu czegokolwiek	0	1	2	3
4. Odczuwanie smutku, przygnębienia lub beznadziejności	0	1	2	3