



Psychometric properties of the PHQ-9 measure of depression among Brazilian older adults

Darío Moreno-Agostino, Kia-Chong Chua, Tim J. Peters, Marcia Scazufca & Ricardo Araya

To cite this article: Darío Moreno-Agostino, Kia-Chong Chua, Tim J. Peters, Marcia Scazufca & Ricardo Araya (2021): Psychometric properties of the PHQ-9 measure of depression among Brazilian older adults, *Aging & Mental Health*, DOI: [10.1080/13607863.2021.1963951](https://doi.org/10.1080/13607863.2021.1963951)

To link to this article: <https://doi.org/10.1080/13607863.2021.1963951>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 19 Aug 2021.



[Submit your article to this journal](#)



Article views: 63



[View related articles](#)



[View Crossmark data](#)

Psychometric properties of the PHQ-9 measure of depression among Brazilian older adults

Darío Moreno-Agostino^{a,b,c} , Kia-Chong Chua^a , Tim J. Peters^d , Marcia Scazufca^e and Ricardo Araya^a

^aDepartment of Health Service and Population Research; Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; ^bCentre for Longitudinal Studies, UCL Social Research Institute, University College London, London, UK; ^cESRC Centre for Society and Mental Health, King's College London, London, UK; ^dPopulation Health Sciences; Bristol Medical School, University of Bristol, Bristol, UK; ^eHospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, Brazil

ABSTRACT

Objectives: To obtain evidence on the psychometric properties of the Patient Health Questionnaire – 9 (PHQ-9, one of the most extensively used tools for assessing depression) in the Brazilian older population.

Method: Data on 3,356 Brazilian adults aged 60+ years living in Guarulhos, São Paulo state were used. The factor structure of the questionnaire was analysed using a factor analysis approach. The questionnaire's measurement equivalence was tested across gender, age, personal income, and education level groups. The scores were compared across groups based on the highest level of equivalence achieved. The questionnaire's internal consistency was analysed considering its factor structure.

Results: A one-factor solution was identified as the most adequate factor structure, with the factor explaining 57.6% of the items' variance. The correlation of the resulting latent score with the overall raw sum score in the PHQ-9 was $r=0.96$. Measurement equivalence regarding thresholds and loadings was achieved for all tested groups. On average, women, older, less educated, and poorer people had higher latent scores on the depression factor. The measure showed a good internal consistency with Revelle's omega total $\omega_t=0.92$.

Conclusion: The results suggest that, among Brazilian older adults living in Guarulhos, São Paulo state, the PHQ-9 measures depressive symptomatology equivalently across different sociodemographic subgroups. Moreover, it can be scored using the raw sum of the item scores to adequately reflect different levels of depressive symptomatology.

ARTICLE HISTORY

Received 7 May 2021
Accepted 29 July 2021

KEYWORDS

Psychometrics;
depression;
exploratory factor analysis;
confirmatory factor analysis;
measurement invariance

Introduction

Depression is one of the leading causes of disability and global burden worldwide (Vigo et al., 2016; World Health Organization, 2017). A recent meta-analysis (Moreno-Agostino et al., 2021) shows that there is a predominantly increasing trend in the prevalence of depression in population-based studies using equivalent sampling and depression assessment procedures across time points. This interacts with a global shift in demographics, characterised by an increase in the proportion of the older population worldwide (United Nations, 2017). In low- and middle-income countries (LMICs), evidence suggests that depression in older adulthood is both common and burdensome (Guerra et al., 2016). This is the case in Brazil, where the estimated prevalence of depression in adults aged 60–64 years is 11.1%, 9.9% in those aged 65–74, and 6.9% among those aged 75+ (Stopa et al., 2015).

Efforts are being made to develop feasible and scalable interventions to address late life depression in Brazil (Scazufca et al., 2019, 2020), considering the limited resources and reduced ratios of mental health professionals. Adequate measures of depression in older adulthood are needed to ensure the clinical relevance of the assessments (i.e. in relation to depressive symptomatology levels), as well as their comparability across relevant subgroups (e.g. different genders, age groups, and educational levels), both in interventional and observational designs.

The Patient Health Questionnaire – 9 (PHQ-9) (Kroenke et al., 2001) is an extensively used tool for assessing depression. It can be employed as a screening tool for depression, and also to ascertain levels of and changes in depressive symptomatology (Kroenke et al., 2010). There is an extensive body of literature providing evidence on the psychometric properties of the PHQ-9. Some studies have focused on the diagnostic validity of the PHQ-9 (i.e. its ability to differentiate between people with and without depression), supporting the use of a score of 10+ as an adequate classification threshold for major depression (Kroenke et al., 2010; Manea et al., 2015). Other studies have analysed the PHQ-9 factor structure, showing the existence of a single underlying factor structure (Cameron et al., 2008; González-Blanch et al., 2018; Keum et al., 2018; Kocalevent et al., 2013; Titov et al., 2011) or a two-factor structure (Beard et al., 2016; Chilcot et al., 2013; Elhai et al., 2012; Krause et al., 2010; Richardson & Richards, 2008) with a somatic factor (typically accounting for sleep, energy, and appetite disturbances) and a cognitive/affective factor (usually accounting for the depressed mood, loss of interest in activities, feelings of worthlessness, and thoughts of death and self-harm).

Notwithstanding these studies, there is a paucity of evidence on the PHQ-9 psychometric properties in the Brazilian population and no evidence focused on older Brazilian adults. To our knowledge, the only two studies that have been conducted cover validity in the general population (Santos et al.,

2013) and among women in primary care settings (de Lima Osorio et al., 2009), supporting the use of a 9+ and 10+ cut-off score respectively for assessing depression. Thus, there is no available evidence on the factor structure of the PHQ-9 and the degree to which it measures depression equivalently across different subgroups of Brazilian older adults.

This lack of evidence hinders the interpretation of the resulting PHQ-9 scores as an adequate reflection of the depression levels. Moreover, it precludes confidence in the comparability across older adults with different genders, age groups, or educational levels, since different scores may be distorted by differences in other aspects rather than in the underlying levels of depression (for instance, differences across groups in the relative relevance of the different items).

This study aims to provide evidence of the factor structure of the PHQ-9 questionnaire in a sample of Brazilian older adults, as well as to analyse potential measurement differences across older adults of different genders, age groups, and educational levels.

Material and methods

Sample and procedure

The sample included the 3,356 adults aged 60+ years old who participated in the screening stage of the first wave of the PROACTIVE cluster randomised controlled trial (Scazufca et al., 2020), interviewed between May 2019 and February 2020. Participants lived in Guarulhos, in the São Paulo metropolitan region. Among the 69 existing Unidades Básicas de Saúde (UBSs, Basic Health Units) in Guarulhos, the 24 containing at least four Family Health Teams were selected.

After randomly selecting four of them as reserve (two from each of the strata defined by the median percentage of individuals either with no education or only a literacy programme for adults), those aged 60+ years enrolled in the remaining 20 UBSs (10 in each of the intervention and control arms of the trial) were contacted using a randomly ordered list. In the first contact, a screening questionnaire including the PHQ-9, as well as basic sociodemographic information including age, gender, income, and educational level, was administered to all participants. The aim of this screening phase was to identify people with depression (operationalised as a PHQ-9 score of 10+) for the PROACTIVE trial. Those participants who fulfilled the inclusion criteria and did not meet any of the exclusion criteria (i.e. individuals whose partner was already included in the study, individuals with acute suicidal risk, hearing or vision loss, or unable to communicate, and individuals unable to engage in the trial for a 12-month period) were asked to participate in the trial. Additional information was then collected in a baseline questionnaire.

All participants provided either oral or written consent to participate. The study was approved by the Ethics Committee of the University of São Paulo Medical School (CEP FMUSP number 2.836.569), and authorised by the Guarulhos Health Secretary.

Measures

PHQ-9

All participants completed the PHQ-9 questionnaire (Kroenke et al., 2001), which is a development of the PRIME-MD patient

health questionnaire (Spitzer et al., 2000). The version of the questionnaire used in this study was originally translated into Brazilian Portuguese by Fraguas et al. (2006) and is available online (<https://www.phqscreeners.com/>). The PHQ-9 comprises nine items, each of them covering one of the symptoms of depression according to the DSM-IV criteria (American Psychiatric Association, 1994). The participants were asked how often they experienced each of those symptoms over the last two weeks. The response options to these questions were 'Not at all' (0), 'Several days' (1), 'More than half the days' (2), and 'Nearly every day' (3). We used the number of days response set ('0–1 day', '2–6 days', '7–11 days', '12–14 days') whenever the participant had any difficulty understanding the standard verbal response options. As the level of literacy among older adults in Brazil is generally low, the research assistants read aloud the PHQ-9 questions to all study participants. These alternative ways of applying the PHQ-9 have been reported previously and have been found acceptable (Kroenke et al., 2010). The responses to all nine items are then summed, resulting in an overall score that ranges from 0 to 27. Scores of 10+ are commonly used as a screening criterion for depression.

Sociodemographic variables. Information on the age, gender, education, and personal income was collected as part of the screening questionnaire. Education was assessed as the grade studied and recoded into the following groups reflecting the number of years studied: 'None' (including those who did not study or just attended literacy for adults), '1–4 years' (including those who studied primary school up to the 4th grade), '5–8 years' (including those who attended primary school up to the 8th grade), and 'More than 8 years' (including those who attended secondary or technical school, or higher education). Personal income was assessed as the monthly number of minimum wages (corresponding to R\$998.00 at the time of the study), and recoded into the following groups: 'Up to 1', 'Between >1 and 2', and 'More than 2'.

Statistical analyses

To allow for the use of a cross-validation approach, the overall sample was randomly split into two equal-sized groups. The first half of the sample was used to implement an exploratory factor analysis (EFA) approach. First, we explored the adequacy of the data to perform the EFA by means of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity, as well as the visual inspection of the correlation matrix among the items. Due to the nature of the PHQ-9 response options (ordered categories), the polychoric correlation matrix was used. Provided the preliminary results were satisfactory, a parallel analysis was performed using 50 simulated datasets, extracting the eigenvalues with a principal components analysis. We compared the 95th quantile of the simulated eigenvalues with the observed eigenvalues, retaining those factors whose observed eigenvalues were higher than the simulated ones (Hayton et al., 2004). A scree plot with the observed and simulated eigenvalues was also used to guide the decision of the optimal number of factors to initially extract, retaining the number of factors up to the point where the eigenvalues begin to level out.

Then, EFAs with increasing number of factors were computed over the polychoric correlation matrix, using a weighted

least squares (WLS) extraction method and an oblique rotation method (PROMAX). The different solutions were then compared based on their fit indices [including χ^2 , Root Mean Square Residuals (RMSR), Root Mean Square Error of Approximation (RMSEA), Tucker-Lewis Index (TLI), and Bayesian Information Criterion (BIC)], the meaningfulness of the factors (their interpretability and the percentage of items' variance accounted for) and the parsimony of the model. Models with lower χ^2 , RMSR, and BIC values would reflect a better fit to the data, whereas RMSEA values below 0.060 and TLI values above 0.950 are often considered indicators of good fit based on commonly-used guidelines (Hu & Bentler, 1999).

Once the most adequate factor solution was identified, it was cross-validated in the second half of the sample by means of a confirmatory factor analysis (CFA) approach. Again, due to the nature of the items, a Diagonally Weighted Least Squares (DWLS) estimation was used, obtaining mean and variance adjusted (WLSMV) fit indices. The goodness of fit of the CFA model was assessed by means of the rescaled fit indices [χ^2 , Comparative Fit Index (CFI), TLI, and RMSEA], which were interpreted using Hu and Bentler (1999) guidelines.

We then proceeded to analyse the measurement equivalence of the PHQ-9 across genders, age groups (60–69, 70–79, and 80+ years), and education and personal income levels. To achieve this, we used a measurement equivalence (measurement invariance) strategy implemented by means of multiple group CFA. This strategy is based on the comparison of models with increasing numbers of constraints to ascertain the extent to which the measures obtained across groups can be considered equivalent. Due to the ordered nature of the PHQ-9 items, we based our analytical strategy on the recent developments and recommendations on measurement equivalence testing with ordered categorical data (Svetina et al., 2020; Wu & Estabrook, 2016). Details on the procedure can be found in the **Appendix S1** (supplementary material).

The resulting factor scores were compared across categories of the different groups and interpreted considering the highest level of measurement equivalence achieved in the previous analyses. Finally, a measure of the internal consistency of the PHQ-9 questionnaire was obtained in the overall sample, considering the underlying factor structure found in the EFA analyses (Revelle & Zinbarg, 2009).

All analyses were performed using R version 3.6.3 (R Core Team, 2020). Parallel analysis and EFA models were performed using the *psych* package version 2.0.9 (Revelle, 2020). CFA models were performed using the *lavaan* package version 0.6-7 (Rosseel, 2012) and *semTools* package version 0.5-3 (Jorgensen et al., 2020).

Results

The sample comprised $n = 3,356$ participants with a mean age of 68.26 years ($SD = 6.71$). Of these, 63.9% were 60–69 years old, 28.7% were aged 70–79 years, and 7.4% were 80+ years old. The majority of participants were women (63.3%, $n = 2,125$). Regarding educational level, 15.9% ($n = 535$) had no education, 46.4% ($n = 1,556$) had studied 1–4 years, 22.5% ($n = 754$) between 5 and 8 years, and 14.6% ($n = 491$) more than 8 years; 0.60% of participants ($n = 20$) did not provide information regarding educational level. Regarding income, most participants (62.6%, $n = 2,100$) had up to 1 minimum wage of personal income, 19.3% ($n = 648$) up to 2 minimum wages, 11.9% ($n = 398$) more

than 2 minimum wages; 6.3% ($n = 210$) participants did not disclose the information on their personal income level. The mean PHQ-9 total score was 6.94 ($SD = 6.96$), being positively skewed (skewness = 0.94) and leptokurtic (kurtosis = 2.92).

The random subsample for the EFA comprised 1,623 participants (48.4% of the overall sample). Bartlett's sphericity test was $\chi^2(36) = 9,192.35$ ($p < 0.001$), whereas the KMO measure of sampling adequacy was 0.93. Both these values, along with the visual inspection of the polychoric correlation matrix (Table S1, supplementary material), which displayed positive correlations ranging from 0.40 to 0.76, suggested the adequacy of the data for performing EFA. The results of the parallel analysis are shown in Table S2 (supplementary material), and the resulting scree plot is shown in Figure S1 (supplementary material). A first component explained 62.14% of the variance, the ratio of the first to the second observed eigenvalues was 8.34, and the size of the observed eigenvalues dropped below the 95th percentiles of the simulated eigenvalues after the first component, plateauing towards the maximum number of components. According to these results, a single factor should be retained.

We therefore extracted a single factor in the first EFA, and performed additional EFAs with increasing number of factors up to five. The fit indices of the EFA models are included in Table 1. Models with up to four factors showed increasing levels of fit to the data, with that fit decreasing from the inclusion of a fifth factor. The models with three and four factors showed an adequate fit based on the usual guidelines (TLI higher than 0.950, RMSEA lower than 0.060, in this last case only slightly for the three-factor model). However, the resulting factors in both the three- and four-factor solutions were highly correlated with each other (all correlations above 0.73), and the RMSR was acceptable in all solutions. Moreover, the increased complexity of these models compared with the one-factor solution did not translate into a greater proportion of variance explained (57.6%, 48.1%, and 55.2% in the one-, three-, and four-factor solutions, respectively). Additionally, all factor loadings in the one-factor solution were substantively high, with a minimum loading of 0.64. Considering these statistics, and to use the most parsimonious and generalisable model, the one-factor model was selected as the optimal one.

The random sample for the CFA analyses included 1,733 participants. The unidimensional CFA model showed acceptable fit with the following mean- and variance- adjusted fit indices: $\chi^2(27) = 202.73$ ($p < 0.001$); CFI = 0.985; TLI = 0.980; RMSEA = 0.061 ($p = 0.009$). The correlation of the latent score with the overall raw sum score in the PHQ-9 in this subsample was $r = 0.96$.

The results of the multigroup CFA models performed to analyse measurement invariance of the PHQ-9 are shown in Table 2. Invariance of the thresholds and loadings was achieved in all cases (across different genders, age groups, and education and

Table 1. Fit indices of the exploratory factor analysis models.

Factors	χ^2	df	RMSR	BIC	RMSEA	TLI
1	174.6	27	0.039	354.6	0.110	0.923
2	94.1	19	0.028	174.9	0.098	0.939
3	25.6	12	0.015	-7.1	0.060	0.977
4	10.4	6	0.009	-7.8	0.056	0.980
5	7	1	0.008	28	0.146	0.865

Note. The results are based on exploratory factor analysis models using the polychoric correlation matrix and a Weighted Least Squares (WLS) extraction method with oblique rotation (PROMAX).

df: degrees of freedom; BIC: Bayesian Information Criterion; RMSEA: Root Mean Square Error of Approximation; RMSR: Root Mean Square Residual; TLI: Tucker-Lewis Index.

Table 2. Fit indices of the multigroup confirmatory factor analysis models performed to analyse measurement equivalence across gender, age, education, and income level groups.

Model		χ^2	df _{χ^2}	$\Delta\chi^2$	df _{$\Delta\chi^2$}	$p_{\Delta\chi^2}$	TLI	CFI	Δ CFI	RMSEA	Δ RMSEA
Gender (<i>n</i> = 1,733)	Baseline	228.48***	54	–	–	–	0.979	0.985	–	0.061	–
	Equal thresholds	252.23***	63	14.01	9	0.122	0.981	0.983	–0.002	0.059	–0.002
	Equal thresholds and loadings	263.09***	71	7.07	8	0.529	0.983	0.983	0.000	0.056	–0.003
Education (<i>n</i> = 1,723)	Baseline	273.58***	108	–	–	–	0.981	0.985	–	0.060	–
	Equal thresholds	318.21***	135	27.37	27	0.444	0.983	0.984	–0.001	0.056	–0.004
	Equal thresholds and loadings	350.56***	159	27.05	24	0.302	0.985	0.983	–0.001	0.053	–0.003
Age group (<i>n</i> = 1,733)	Baseline	235.39***	81	–	–	–	0.983	0.987	–	0.057	–
	Equal thresholds	268.16***	99	17.73	18	0.474	0.985	0.986	–0.001	0.054	–0.003
	Equal thresholds and loadings	285.35***	115	15.00	16	0.525	0.987	0.986	0.000	0.051	–0.003
Personal income (<i>n</i> = 1,634)	Baseline	239.19***	81	–	–	–	0.980	0.985	–	0.060	–
	Equal thresholds	270.66***	99	16.12	18	0.584	0.982	0.984	–0.001	0.056	–0.004
	Equal thresholds and loadings	287.34***	115	16.63	16	0.410	0.985	0.984	0.000	0.053	–0.003

Note. The results are based on a Diagonally Weighted Least Squares (DWLS) estimation, with mean and variance adjusted (WLSMV) fit indices.

*** $p < 0.001$. Δ : difference in fit between nested models; df: degrees of freedom; CFI: Comparative Fit Index; RMSEA: Root Mean Square Error of Approximation; TLI: Tucker-Lewis Index. Selected models are highlighted in boldface.

Table 3. Comparison of the depression factor scores across gender, age, education, and income level groups.

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>t</i> / <i>F</i>	Effect size
Gender				–7.58***	–0.37
Men	646	–0.15	0.79		
Women	1087	0.16	0.85		
Age group				0.30	0.02
60–69	1122	0.03	0.84		
70–79	490	0.06	0.84		
80+	121	0.07	0.83		
Education				5.45**	0.10
None	258	0.11	0.84		
1–4 years	821	0.09	0.83		
5–8 years	404	0.01	0.84		
8+ years	240	–0.13	0.82		
Personal income				24.30***	0.17
Less than 1 MW	1101	0.15	0.85		
Up to 2 MW	332	–0.10	0.79		
More than 2 MW	201	–0.21	0.73		

Note. ** $p < 0.01$; *** $p < 0.001$. *F*: ANOVA's *F*-statistic; *M*: mean; MW: minimum wage(s); *N*: number of observations; *SD*: standard deviation; *t*: Student's *t*-statistic. The effect size measure corresponds to Cohen's *d* or *f* for the Student's *t* and ANOVA's *F* statistics, respectively.

personal income levels), meaning that the resulting PHQ-9 scores were comparable across these groups. Additionally, the factor scores resulting from the main CFA model were compared across the groups in the CFA subsample in order to provide evidence of the known-groups validity (i.e. the extent to which scores differ across the groups as would be expected). The results of these comparisons are included in Table 3. On average, women, older, less educated, and poorer people had higher latent scores on the depression factor.

Finally, considering the single factor structure of the PHQ-9 in the sample, Revelle's omega total (ω_t) was computed as a measure of internal consistency, which resulted in $\omega_t = 0.92$.

Discussion

This is the first study to provide evidence on the factor structure, measurement invariance, known-group validity, and internal consistency of the widely used PHQ-9 questionnaire among Brazilian older adults living in urban areas that are predominantly socioeconomically deprived.

Regarding the factor structure, our study suggests that, in this population, the questionnaire has an essentially unidimensional structure. This structure has been previously reported in other populations (Cameron et al., 2008; González-Blanch et al., 2018; Keum et al., 2018; Kocalevent et al., 2013; Titov et al., 2011). A two-factor structure reflecting a difference between somatic and non-somatic (more cognitive and affective) symptomatology has

also been widely reported in the literature (Beard et al., 2016; Chilcot et al., 2013; Elhai et al., 2012; Krause et al., 2010; Richardson & Richards, 2008). However, we found that increasing the number of factors did not result in a substantially better explanation of the shared variance across the items, and thus we selected the simplest solution (one factor) for the sake of parsimony and generalizability of the model. Moreover, we found that the latent factor score resulting from considering a single depression dimension correlated very highly with the raw sum of item responses following the scoring instructions of the questionnaire (Spitzer et al., 2000). Altogether, our findings support the appropriateness of this latter (and simpler) scoring approach, in which all items reflect a single depression dimension and contribute equally to the resulting score.

Regarding the degree of measurement invariance of the PHQ-9 across different demographic and socioeconomic groups, we found that the questionnaire performed equivalently among Brazilian older adults regardless of their gender, education, age group, and personal income. Differences in the resulting scores across these subgroups are therefore due to differences in the underlying depression levels and not to other aspects (e.g. the relative importance of different items or the way in which different symptoms' intensities contribute to the underlying latent factor of depression). These findings are in line with previous studies showing similar invariance features in other populations (González-Blanch et al., 2018; Keum et al., 2018), providing further justification for using this measure to derive meaningful comparisons of the depression levels across these subgroups in the Brazilian older population.

Our study also provides evidence on the known-group validity of the PHQ-9 questionnaire by showing the expected differences in the depression levels across these different demographic and socioeconomic groups, once the potential existence of measurement differences has been ruled out. Our results are in line with previous studies in the general (Alonso et al., 2004; Lorant et al., 2003) and older (Cole & Dendukuri, 2003; Fiske et al., 2009) population by showing that female, less educated and poorer older adults show higher levels of depressive symptomatology than their male, more educated and higher income counterparts.

Strengths and limitations

The results of this study must be interpreted in the light of several strengths and limitations. Among the strengths, we used information on a large sample of Brazilian older adults

living in a relatively socioeconomically deprived area of the São Paulo state. By focusing our study in a LMIC population, we narrow the existing divide in the research (and, more specifically, mental health research), which disproportionately over-represents populations from high-income countries, accounting for less than 20% of the global population (Saxena et al., 2006; World Bank, 2021). Moreover, the random sampling procedures and the large sample size used make our results easier to generalise to the reference population. From a methodological point of view, we used a robust approach to the analysis of tests with ordered categorical items and updated guidelines for the analysis of the measurement invariance of such tests, overcoming the limitations of previous research using less appropriate analytical approaches developed for the analysis of continuous measures (Svetina et al., 2020; Wu & Estabrook, 2016).

Notwithstanding this, several limitations should be considered when interpreting our results. First, despite the strengths derived from the sampling procedures and resulting sample, the results may not be generalisable to the overall Brazilian older population, which also includes older adults living in different urban and rural areas, where the context characteristics (e.g. socioeconomic conditions or even access to mental health services) may be substantially different. Second, due to the available information in the administered questionnaires, we could not consider alternative economic variables (e.g. household income or household wealth) that may have more adequately reflected the economic conditions of older adults than the personal income. Likewise, we could not perform the measurement invariance analysis by other relevant subgroups (e.g. ethnicity) nor include additional variables such as disability, anxiety, loneliness, or additional measures of depression that may have been relevant for the concurrent and discriminant validity analyses (Cole & Dendukuri, 2003; Hooker et al., 2019). Information on some of these variables was only available for a subset of the participants who presented at least moderate levels of depression symptomatology (PHQ-9 \geq 10) and who consented into the randomised trial; the inevitable lack of generalisability to the wider population meant that these variables were therefore not analysed further for the purposes of this study. Nevertheless, the results found using personal income and other sociodemographic variables are in line with the previous literature, thus supporting the known-group validity of the PHQ-9 questionnaire.

Conclusions

Overall, our study provides evidence of the adequacy of the PHQ-9 questionnaire for assessing the depression levels of Brazilian older adults living in urban areas. Our results support the use of the widely used and relatively simple scoring procedures (the raw sum of item responses) to adequately reflect the depressive symptomatology levels in this population. Moreover, they suggest that the PHQ-9 measures depression equivalently across relevant sociodemographic subgroups, opening the way to the study of differences in the PHQ-9 scores after ruling out the potential effect of the lack of measurement invariance across these groups.

Disclosure statement

The authors report no conflict of interest.

Funding

This work was supported by the São Paulo Research Foundation (FAPESP) under Grant number 17/50094-2; and the Joint Global Health Trial initiative jointly funded by Medical Research Council, Wellcome Trust, and the UK Department for International Development under Grant number MR/R006229/1. DMA is supported by the ESRC Centre for Society and Mental Health at King's College London (ESRC Reference: ES/S012567/1). The views expressed are those of the author(s) and not necessarily those of the ESRC or King's College London. MS was partially funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico-Brazil. This research was funded in whole, or in part, by the Wellcome Trust [grant number MR/R006229/1]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP); Joint Global Health Trial initiative (jointly funded by Medical Research Council, Wellcome Trust, and the UK Department for International Development); ESRC Centre for Society and Mental Health, King's College London.

ORCID

Dário Moreno-Agostino  <http://orcid.org/0000-0003-0459-657X>
Kia-Chong Chua  <http://orcid.org/0000-0002-6693-6333>
Tim J. Peters  <http://orcid.org/0000-0003-2881-4180>

References

- Alonso, J., Angermeyer, M. C., Bernert, S., Bruffaerts, R., Brugha, T. S., Bryson, H., de Girolamo, G., Graaf, R., Demyttenaere, K., Gasquet, I., Haro, J. M., Katz, S. J., Kessler, R. C., Kovess, V., Lepine, J. P., Ormel, J., Polidori, G., Russo, L. J., Vilagut, G., ... Vollebergh, W. A., & ESEMeD/Mhede Investigators European Study of the Epidemiology of Mental Disorders Project. (2004). Prevalence of mental disorders in Europe: Results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica*, 420, 21–27. <https://doi.org/10.1111/j.1600-0047.2004.00327.x>
- American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*, 4th text revision. Washington, DC: American Psychiatric Association.
- Beard, C., Hsu, K. J., Rifkin, L. S., Busch, A. B., & Bjorgvinsson, T. (2016). Validation of the PHQ-9 in a psychiatric sample. *Journal of Affective Disorders*, 193, 267–273. <https://doi.org/10.1016/j.jad.2015.12.075>
- Cameron, I. M., Crawford, J. R., Lawton, K., & Reid, I. C. (2008). Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *British Journal of General Practice*, 58(546), 32–36. <https://doi.org/10.3399/bjgp08X263794>
- Chilcot, J., Rayner, L., Lee, W., Price, A., Goodwin, L., Monroe, B., Sykes, N., Hansford, P., & Hotopf, M. (2013). The factor structure of the PHQ-9 in palliative care. *Journal of Psychosomatic Research*, 75(1), 60–64. <https://doi.org/10.1016/j.jpsychores.2012.12.012>
- Cole, M. G., & Dendukuri, N. (2003). Risk factors for depression among elderly community subjects: A systematic review and meta-analysis. *The American Journal of Psychiatry*, 160(6), 1147–1156. <https://doi.org/10.1176/appi.ajp.160.6.1147>
- Core Team, R. (2020). *R: A language and environment for statistical computing* (Version 3.6.3). R Foundation for Statistical Computing. <https://www.R-project.org/>
- de Lima Osorio, F., Vilela Mendes, A., Crippa, J. A., & Loureiro, S. R. (2009). Study of the discriminative validity of the PHQ-9 and PHQ-2 in a sample of Brazilian women in the context of primary health care. *Perspectives in Psychiatric Care*, 45(3), 216–227. <https://doi.org/10.1111/j.1744-6163.2009.00224.x>
- Elhai, J. D., Contractor, A. A., Tamburrino, M., Fine, T. H., Prescott, M. R., Shirley, E., Chan, P. K., Slembariski, R., Liberzon, I., Galea, S., & Calabrese, J. R. (2012). The factor structure of major depression symptoms: A test of four competing models using the Patient Health Questionnaire-9. *Psychiatry Research*, 199(3), 169–173. <https://doi.org/10.1016/j.psychres.2012.05.018>
- Fiske, A., Wetherell, J. L., & Gatz, M. (2009). Depression in older adults. *Annual Review of Clinical Psychology*, 5, 363–389. <https://doi.org/10.1146/annurev.clinpsy.032408.153621>

- Fraguas, R., Henriques, S. G., Gonsalves Henriques, S., De Lucia, M. S., Iosifescu, D. V., Schwartz, F. H., Menezes, P. R., Rossi Menezes, P., Gattaz, W. F., Farid Gattaz, W., Martins, M. A., & Arruda Martins, M. (2006). The detection of depression in medical setting: A study with PRIME-MD. *Journal of Affective Disorders*, *91*(1), 11–17. <https://doi.org/10.1016/j.jad.2005.12.003>
- González-Blanch, C., Medrano, L. A., Muñoz-Navarro, R., Ruíz-Rodríguez, P., Moriana, J. A., Limonero, J. T., Schmitz, F., & Cano-Vindel, A. (2018). Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. *PLoS One*, *13*(2), e0193356. <https://doi.org/10.1371/journal.pone.0193356>
- Guerra, M., Prina, A. M., Ferri, C. P., Acosta, D., Gallardo, S., Huang, Y., Jacob, K. S., Jimenez-Velazquez, I. Z., Libre Rodriguez, J. J., Liu, Z., Salas, A., Sosa, A. L., Williams, J. D., Uwakwe, R., & Prince, M. (2016). A comparative cross-cultural study of the prevalence of late life depression in low and middle income countries. *Journal of Affective Disorders*, *190*, 362–368. <https://doi.org/10.1016/j.jad.2015.09.004>
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191–205. <https://doi.org/10.1177/1094428104263675>
- Hooker, K., Phibbs, S., Irvin, V. L., Mendez-Luck, C. A., Doan, L. N., Li, T., Turner, S., & Choun, S. (2019). Depression among older adults in the United States by disaggregated race and ethnicity. *The Gerontologist*, *59*(5), 886–891. <https://doi.org/10.1093/geront/gny159>
- Hu, L. t., & 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), 1–55. <https://doi.org/10.1080/10705519909540118>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., Rosseel, Y. (2020). *semTools: Useful tools for structural equation modeling* (Version 0.5-3). <https://cran.r-project.org/package=semTools>
- Keum, B. T., Miller, M. J., & Inkelas, K. K. (2018). Testing the factor structure and measurement invariance of the PHQ-9 across racially diverse U.S. college students. *Psychological Assessment*, *30*(8), 1096–1106. <https://doi.org/10.1037/pas0000550>
- Kocalevent, R. D., Hinz, A., & Brahler, E. (2013). Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. *General Hospital Psychiatry*, *35*(5), 551–555. <https://doi.org/10.1016/j.genhosppsych.2013.04.006>
- Krause, J. S., Reed, K. S., & McArdle, J. J. (2010). Factor structure and predictive validity of somatic and nonsomatic symptoms from the patient health questionnaire-9: A longitudinal study after spinal cord injury. *Archives of Physical Medicine and Rehabilitation*, *91*(8), 1218–1224. <https://doi.org/10.1016/j.apmr.2010.04.015>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Kroenke, K., Spitzer, R. L., Williams, J. B., & Lowe, B. (2010). The patient health questionnaire somatic, anxiety, and depressive symptom scales: A systematic review. *General Hospital Psychiatry*, *32*(4), 345–359. <https://doi.org/10.1016/j.genhosppsych.2010.03.006>
- Lorant, V., Deliege, D., Eaton, W., Robert, A., Philippot, P., & Anseau, M. (2003). Socioeconomic inequalities in depression: A meta-analysis. *American Journal of Epidemiology*, *157*(2), 98–112. <https://doi.org/10.1093/aje/kwf182>
- Manea, L., Gilbody, S., & McMillan, D. (2015). A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. *General Hospital Psychiatry*, *37*(1), 67–75. <https://doi.org/10.1016/j.genhosppsych.2014.09.009>
- Moreno-Agostino, D., Wu, Y. T., Daskalopoulou, C., Hasan, M. T., Huisman, M., & Prina, M. (2021). Global trends in the prevalence and incidence of depression: systematic review and meta-analysis. *Journal of Affective Disorders*, *281*, 235–243. <https://doi.org/10.1016/j.jad.2020.12.035>
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145–154. <https://doi.org/10.1007/s11336-008-9102-z>
- Revelle, W. (2020). *psych: Procedures for psychological, psychometric, and personality research* (Version 2.0.9). <https://CRAN.R-project.org/package=psych>
- Richardson, E. J., & Richards, J. S. (2008). Factor structure of the PHQ-9 screen for depression across time since injury among persons with spinal cord injury. *Rehabilitation Psychology*, *53*(2), 243–249. <https://doi.org/10.1037/0090-5550.53.2.243>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Santos, I. S., Tavares, B. F., Munhoz, T. N., Almeida, L. S., Silva, N. T., Tams, B. D., Patella, A. M., & Matijasevich, A. (2013). Sensibilidade e especificidade do Patient Health Questionnaire-9 (PHQ-9) entre adultos da população geral [Sensitivity and specificity of the Patient Health Questionnaire-9 (PHQ-9) among adults from the general population]. *Cadernos de Saude Publica*, *29*(8), 1533–1543. <https://doi.org/10.1590/0102-311x00144612>
- Saxena, S., Paraje, G., Sharan, P., Karam, G., & Sadana, R. (2006). The 10/90 divide in mental health research: Trends over a 10-year period. *The British Journal of Psychiatry: The Journal of Mental Science*, *188*, 81–82. <https://doi.org/10.1192/bjp.bp.105.011221>
- Scazufo, M., de Paula Couto, M. C. P., Henrique, M. G., Mendes, A. V., Matijasevich, A., Pereda, P. C., Franzin, R. M., Seabra, A. C., van de Ven, P., Hollingworth, W., Peters, T. J., & Araya, R. (2019). Pilot study of a two-arm non-randomized controlled cluster trial of a psychosocial intervention to improve late life depression in socioeconomically deprived areas of Sao Paulo, Brazil (PROACTIVE): feasibility study of a psychosocial intervention for late life depression in Sao Paulo. *BMC Public Health*, *19*(1), 1152. <https://doi.org/10.1186/s12889-019-7495-5>
- Scazufo, M., Nakamura, C. A., Peters, T. J., Henrique, M. G., Seabra, A., La Rotta, E. G., Franzin, R. M., Martins, D. F., Van de Ven, P., Hollingworth, W., & Araya, R. (2020). A collaborative care psychosocial intervention to improve late life depression in socioeconomically deprived areas of Guarulhos, Brazil: The PROACTIVE cluster randomised controlled trial protocol. *Trials*, *21*(1), 914. <https://doi.org/10.1186/s13063-020-04826-w>
- Spitzer, R. L., Williams, J. B., Kroenke, K., Hornyak, R., & McMurray, J. (2000). Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: The PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *American Journal of Obstetrics and Gynecology*, *183*(3), 759–769. <https://doi.org/10.1067/mob.2000.106580>
- Stopa, S. R., Malta, D. C., de Oliveira, M. M., Lopes, C. d S., Menezes, P. R., & Kinoshita, R. T. (2015). Prevalence of self-reported depression in Brazil: 2013 National Health Survey results. *Revista Brasileira de Epidemiologia*, *18*(Suppl.2), 170–180. <https://doi.org/10.1590/1980-5497201500060015>
- Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using Mplus and the lavaan/semTools packages. *Structural Equation Modeling: a Multidisciplinary Journal*, *27*(1), 111–130. <https://doi.org/10.1080/10705511.2019.1602776>
- Titov, N., Dear, B. F., McMillan, D., Anderson, T., Zou, J., & Sunderland, M. (2011). Psychometric comparison of the PHQ-9 and BDI-II for measuring response during treatment of depression. *Cognitive Behaviour Therapy*, *40*(2), 126–136. <https://doi.org/10.1080/16506073.2010.550059>
- United Nations. (2017). *World Population Ageing Report*. New York: United Nations.
- Vigo, D., Thornicroft, G., & Atun, R. (2016). Estimating the true global burden of mental illness. *The Lancet Psychiatry*, *3*(2), 171–178. [https://doi.org/10.1016/S2215-0366\(15\)00505-2](https://doi.org/10.1016/S2215-0366(15)00505-2)
- World Bank. (2021). Retrieved October 10, 2020 from <http://data.worldbank.org/country>.
- World Health Organization. (2017). *Depression and other common mental disorders: Global health estimates*. Geneva: World Health Organization.
- Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, *81*(4), 1014–1045. <https://doi.org/10.1007/s11336-016-9506-0>