

RESEARCH ARTICLE

# Factorial validity and invariance of the Patient Health Questionnaire (PHQ)-9 among clinical and non-clinical populations

Satomi Doi<sup>1,2\*</sup>, Masaya Ito<sup>2</sup>, Yoshitake Takebayashi<sup>2</sup>, Kumiko Muramatsu<sup>3</sup>, Masaru Horikoshi<sup>2</sup>

**1** Department of Global Health Promotion, Tokyo Medical and Dental University, Tokyo, Japan, **2** National Center of Neurology and Psychiatry, National Center for Cognitive Behavior Therapy and Research, Tokyo, Japan, **3** Graduate School of Clinical Psychology, Niigata Seiryō University, Niigata, Japan

\* doi.hlth@tmd.ac.jp



**OPEN ACCESS**

**Citation:** Doi S, Ito M, Takebayashi Y, Muramatsu K, Horikoshi M (2018) Factorial validity and invariance of the Patient Health Questionnaire (PHQ)-9 among clinical and non-clinical populations. PLoS ONE 13(7): e0199235. <https://doi.org/10.1371/journal.pone.0199235>

**Editor:** Camillo Gualtieri, North Carolina Neuropsychiatry Clinics, UNITED STATES

**Received:** September 8, 2017

**Accepted:** April 23, 2018

**Published:** July 19, 2018

**Copyright:** © 2018 Doi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The current study was supported by a Grant-in-Aid for Research Activity start-up (24830127, <http://www.jpsps.go.jp/english/e-grants/index.html>) awarded to MI from the Japan Society for the Promotion of Science, National Center of Neurology and Psychiatry Intramural Research Grant (24-4, <http://www.ncnp.go.jp/guide/cost.html>) for Neurological and Psychiatric Disorders.

## Abstract

The Patient Health Questionnaire-9 (PHQ-9) is commonly used to screen for depressive disorder and for monitoring depressive symptoms. However, there are mixed findings regarding its factor structure (i.e., whether it has a unidimensional, two-dimensional, or bi-factor structure). Furthermore, its measurement invariance between non-clinical and clinical populations and that between patients with major depressive disorder (MDD) and MDD with comorbid anxiety disorder (AD) is unknown. Japanese adults with MDD ( $n = 406$ ), MDD with AD ( $n = 636$ ), and no psychiatric disorders (non-clinical population;  $n = 1,163$ ) answered this questionnaire on the Internet. Confirmatory factor analyses showed that the bi-factor model had a better fit than the unidimensional and two-dimensional factor models did. The results of a multi-group confirmatory factor analysis indicated scalar invariance between the non-clinical and only MDD groups, and that between the only MDD and MDD with AD groups. In conclusion, the bi-factor model with two specific factors was supported among the non-clinical, only MDD, and MDD with AD groups. The scalar measurement invariance model was supported between the groups, which indicated the total or sub-scale scores were comparable between groups.

## Introduction

Depression is an exceedingly common comorbid condition in several mental disorders. To date, numerous self-report measures of depression have been developed, many of which are commonly used in clinical practice and research [1]. In particular, the Patient Health Questionnaire (PHQ) is one of the most useful measures for monitoring depressive symptoms and for screening for major depressive disorder (MDD) [2]. The 9-item version of the PHQ (PHQ-9) [3] is a brief and simple self-administered measure that corresponds with the nine diagnostic criteria for depressive disorder of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). The PHQ-9 has been found to have high reliability and validity in Western populations, and can be used as a one- or two-item measure. The

The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

guidelines of the National Institute for Health and Clinical Excellence recommend the PHQ-9 for assessing the severity of depressive symptoms in clinical practice [4].

Despite its numerous advantages, the PHQ-9 has at least two main issues stemming from a lack of previous research, namely, the unclear factor structure and measurement invariance. First, there are mixed findings regarding the factor structure of the PHQ-9 in Western populations. Some studies using clinical samples (both in primary care and psychiatric settings) have determined that the PHQ-9 has a unidimensional factor structure [5]. In contrast, other studies on psychiatric patients or individuals with physical illness and depression have determined that a two-dimensional factor structure (including somatic and cognitive/affective symptom factors) had a better fit [6, 7, 8]. However, no studies have yet examined a possible bi-factor model that represents the existence of both a general factor and specific sub-factors called as group factors [9, 10]. For example, similar mixed results regarding the factorial unidimensionality of the Symptom Checklist-90-Revised has been resolved by testing a bi-factor model [11]. Hence, we hypothesized that the bi-factor model will show the best fit against the mere unidimensional or two-factor model.

Second, previous studies have not examined the possible measurement invariance [12] in the PHQ-9 between non-clinical and clinical populations. Additionally, no studies have yet examined whether the factor structure of the PHQ-9 could be assumed equivalently between patients with only MDD and those with MDD who comorbid anxiety disorder (AD). Given the high co-occurrence rate of depression and anxiety [13], it is necessary to examine the factor structure of the PHQ-9 between patients with only MDD and those with MDD who have comorbid AD.

In the present study, using an existing large dataset of non-clinical and clinical populations in Japan, we (1) compared the fit of the unidimensional, two-dimensional, and bi-factor models of the PHQ-9 via a confirmatory factor analysis; and (2) examined the measurement invariance between non-clinical, only MDD, and MDD with AD groups using a multi-group confirmatory factor analysis.

## Methods

### Participants and procedure

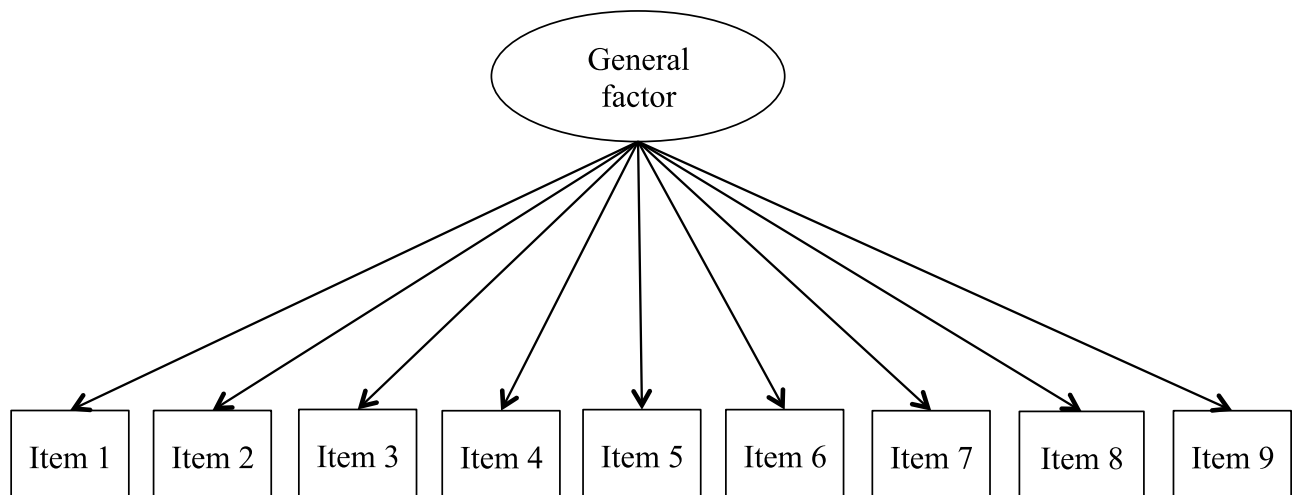
This study was part of a larger web-based survey for examining the emotions and psychopathology of Japanese clinical and non-clinical populations [14, 15]. We recruited participants in this study from panelists registered with Macromill Incorporation. This company is one of the largest Japanese internet marketing research company and has been used in previous studies [16]. A total of 2,830 individuals (1,547 females, 1,283 males; mean age = 42.44 years;  $SD = 10.39$  years; range = 19–79 years) were selected randomly according to their age, gender, and living area from each population, including 619 individuals with MDD, 576 with social anxiety disorder, 619 with panic disorder, 645 with obsessive-compulsive disorder, and 371 without any psychiatric disorder (i.e., non-clinical population). The participants self-reported their own diagnoses by answering the following items regarding their current diagnoses and treatment of mental disorders: “Are you currently diagnosed as having Major Depressive Disorder and being treated for the problem in a medical setting?” Similarly, they were asked to respond “yes” or “no” to the question of their own diagnoses of social anxiety disorder, panic disorder, and obsessive-compulsive disorder. This study was approved by the institutional review board of the National Center of Neurology and Psychiatry (approval number: A2013-002).

## Measurements

**Japanese version of the Patient Health Questionnaire-9 (J-PHQ-9).** The J-PHQ-9 assesses the frequency with which the nine symptoms of depression occurred over the last two weeks [17]. The participants rate each of the nine items on a scale ranging from 0 (*not at all*) to 3 (*nearly every day*). The reliability of the English version of PHQ-9 is excellent, as evidenced by the previous reports of an internal reliability Cronbach's  $\alpha$  of .86 to .91 [3, 5] and test-retest reliability [3]. The construct validity of the English version of the PHQ-9 confirmed by findings of previous studies which reported that increasing PHQ-9 scores were associated with worsening function [3, 18], increasing depression assessed using other measures [6, 18], increasing anxiety [6], and decreasing psychology well-being [6]. Additionally, in the present study, the internal reliability of the J-PHQ-9 was excellent, as evidenced by a Cronbach's  $\alpha$  .93, .84, and .91 for the total score, somatic score, and cognitive/affective score, respectively. The J-PHQ-9 also had good convergent validity as it was associated with the Japanese versions of the Kessler Psychological Distress Scale (K6) [19] ( $r = .81$ ) and Center for Epidemiologic Studies Depression Scale [20] ( $r = .86$ ). In this study, we used the sum of the item scores of this scale for our analyses.

## Statistical analysis

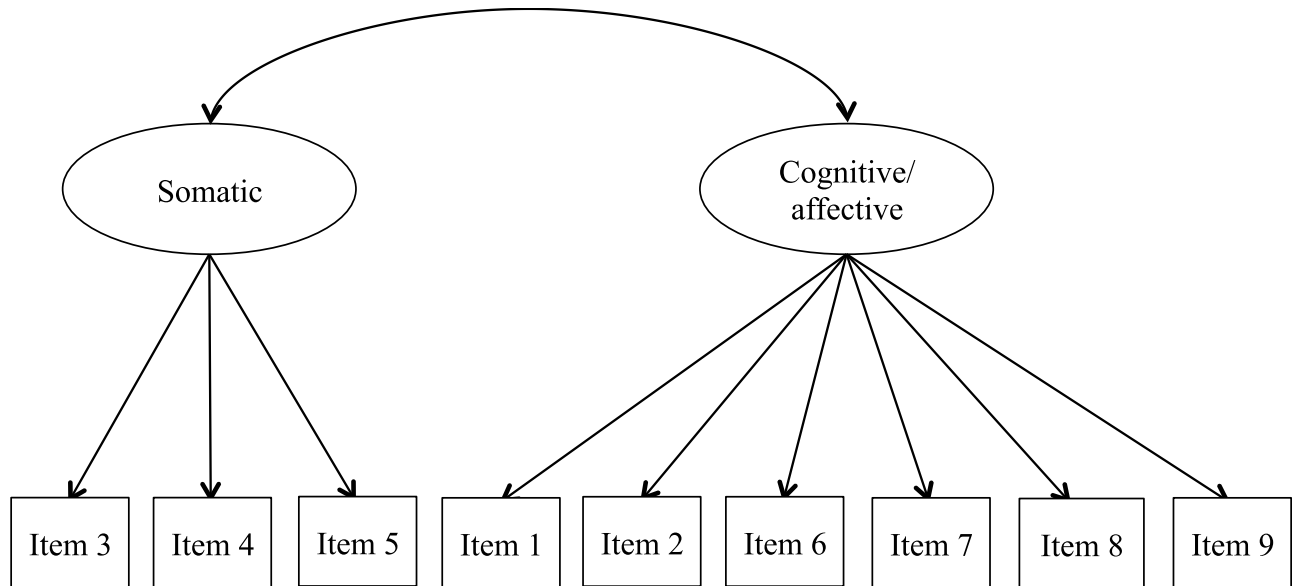
First, we conducted a confirmatory factor analysis of the PHQ-9 using the data collected from the entire sample ( $n = 2,205$ ). In this analysis, we determined and compared the fit of the above-stated three factor models to the data using the full information maximum likelihood method. In the unidimensional factor model, each item was represented by a single factor (Fig 1) [21]. In the two-dimensional factor model, items loaded onto one of the latent factors of somatic and cognitive/affective symptoms (Fig 2) [8]. Finally, in the bi-factor model, we designated somatic and cognitive/affective symptoms as specific group factors, and the sum of the item scores as the general factor (Fig 3).



*Note.* Item 1 = little interest or pleasure, Item 2 = depressed mood, Item 3 = sleep difficulties, Item 4 = fatigue, Item 5 = appetite changes, Item 6 = feeling of worthlessness, Item 7 = trouble concentrating, Item 8 = moving or speaking slowly, Item 9 = suicidal thoughts.

**Fig 1. Unidimensional factor model.**

<https://doi.org/10.1371/journal.pone.0199235.g001>



*Note.* Item 1 = little interest or pleasure, Item 2 = depressed mood, Item 3 = sleep difficulties, Item 4 = fatigue, Item 5 = appetite changes, Item 6 = feeling of worthlessness, Item 7 = trouble concentrating, Item 8 = moving or speaking slowly, Item 9 = suicidal thoughts.

**Fig 2. Two-dimensional factor model.**

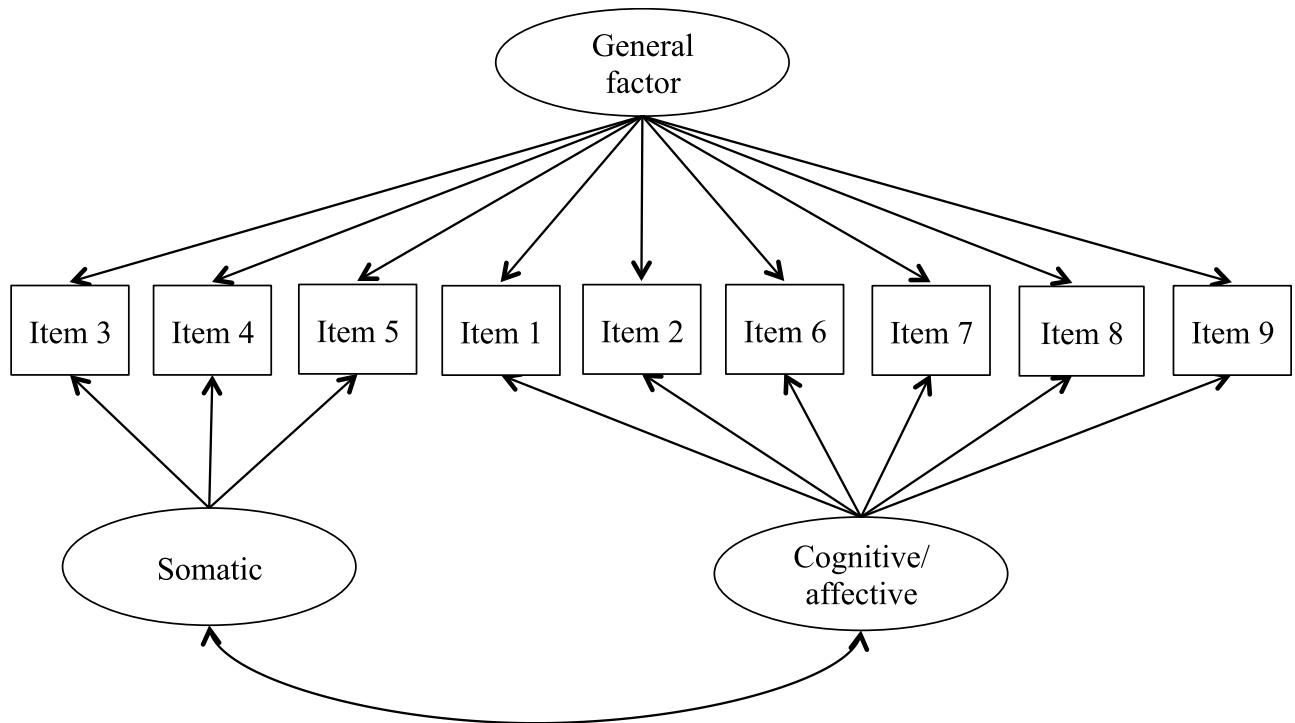
<https://doi.org/10.1371/journal.pone.0199235.g002>

Second, to examine the measurement invariance across non-clinical, only MDD, and MDD with AD populations, we conducted a multi-group confirmatory factor analysis [22]. We examined the measurement invariance of the PHQ-9 scores between the non-clinical and only MDD groups, and between the only MDD and MDD with AD groups. We constructed the following five increasingly restrictive models: where all parameters were free (Model 1: configural invariance); where loadings were invariant (Model 2: metric invariance); where loadings and intercepts were invariant (Model 3: scalar invariance); where loadings, intercepts, and residuals were invariant (Model 4: error variance invariance); and where loadings, intercepts, residuals, and factor means were invariant (Model 5: factor variance invariance). We used the following fit indices to evaluate the models: chi-square, root mean square error of approximation (RMSEA), Akaike information criterion (AIC), Bayesian information criterion (BIC), comparative fit index (CFI), and standardized root mean square residual (SRMR). Goodness-of-fit indices were examined in light of the following standards used in past literature [23]: the chi-square test ( $\chi^2$ ) should not be significant; the RMSEA should be  $< .10$  for acceptable fit and  $< .06$  for good fit; the CFI should be  $\geq .90$  for acceptable fit and  $> .95$  for good fit; and the SRMR should be  $< .10$  for acceptable fit and  $< .08$  for good fit. The following criterion was used to adopting the model: a difference of less than  $.01$  in the  $\Delta$ CFI index supports the less parameterized model [24].

## Results

### Distribution of the PHQ-9 score

Mean values of the total, somatic, and cognitive/affective scores on the PHQ-9 were as follows: non-clinical group: 6.96 (Standard Deviation: SD = 6.46), 3.10 (SD = 2.62), and 3.86



*Note.* Item 1 = little interest or pleasure, Item 2 = depressed mood, Item 3 = sleep difficulties, Item 4 = fatigue, Item 5 = appetite changes, Item 6 = feeling of worthlessness, Item 7 = trouble concentrating, Item 8 = moving or speaking slowly, Item 9 = suicidal thoughts.

**Fig 3. Bi-factor model.**

<https://doi.org/10.1371/journal.pone.0199235.g003>

(SD = 4.26); only MDD group: 12.42 (SD = 7.57), 5.16 (SD = 2.80), and 7.26 (SD = 5.27); MDD with AD group: 15.86 (SD = 7.20), 6.17 (SD = 2.58), and 9.69 (SD = 5.10).

### Confirmatory factor analysis

We compared the fit indices of the three models (unidimensional, two-dimensional, and bi-factor models) using the entire sample. The fit indices of the unidimensional model ( $\chi^2(27) = 1171.93, p < 0.001$ ; RMSEA = .122; CFI = .936 SRMR = .037) and two-dimensional model ( $\chi^2(26) = 745.14, p < 0.001$ ; RMSEA = .098; CFI = .960; SRMR = .029) were poorer than those of the bi-factor model ( $\chi^2(18) = 373.05, p < 0.001$ ; RMSEA = .083; CFI = .980; SRMR = .020) (Table 1). Therefore, we selected the bi-factor model for subsequent analyses. In addition, Table 1 shows the fit indices of three models using the non-clinical, only MDD, and MDD with AD groups.

Table 2 shows the standardized factor loadings for the bi-factor model using the entire sample. The general factor, the group cognitive/affective factor, and the group somatic factor accounted for 56.9%, 25.5%, and 17.6% of the common variance, respectively.

### Multi-group confirmatory factor analysis

First, we conducted a multi-group confirmatory factor analysis (of the bi-factor model) for the non-clinical and only MDD groups (Table 3). According to the criterion for adopting the

**Table 1. The fit indices of the unidimensional, two-dimensional, and bi-factor models.**

	$\chi^2$ value	df	RMSEA	CFI	SRMR
Entire sample					
Unidimensional model	1171.93	27	.122	.936	.037
Two-dimensional model	745.14	26	.098	.960	.029
Bi-factor model	373.05	17	.083	.980	.020
Non-clinical group					
Unidimensional model	354.40	27	.132	.914	.045
Two-dimensional model	196.88	26	.095	.958	.040
Bi-factor model	83.03	17	.074	.983	.031
Only MDD group					
Unidimensional model	223.03	27	.120	.920	.047
Two-dimensional model	179.54	26	.107	.938	.034
Bi-factor model	95.21	17	.094	.969	.021
MDD with AD group					
Unidimensional model	129.83	27	.105	.940	.041
Two-dimensional model	111.11	26	.097	.951	.037
Bi-factor model	36.32	17	.058	.989	.018

Note. df = degree of freedom, RMSEA = standardized root mean square residual, CFI = comparative fit index, and SRMR = standardized root mean square residual, MDD = Major depressive disorder, AD = Anxiety disorder.

<https://doi.org/10.1371/journal.pone.0199235.t001>

model, Model 3 showed the best fit (scalar invariance), wherein the loadings and intercepts were invariant.

Second, we conducted a multi-group confirmatory factor analysis using only the MDD and MDD with AD groups (Table 3). Similar to the findings pertaining to the non-clinical and only MDD groups, Model 3 showed the best fit (scalar invariance). Scalar invariance indicates that differences in the factor mean lead to differences in item mean.

## Discussion

In this study, we compared a bi-factor model of the PHQ-9 with unidimensional and two-dimensional factor models and examined the measurement invariance of the PHQ-9 across non-clinical, only MDD, and MDD with AD groups. Among both non-clinical and clinical

**Table 2. Standardized factor loadings for the bi-factor model using entire sample.**

	General	Cog/affect	Somatic
1. Little interest or pleasure	.579	.640	
2. Feeling down, depressed, or hopeless	.556	.737	
3. Trouble falling/staying asleep/sleeping too much	.497		.585
4. Feeling tired or having little energy	.530		.705
5. Poor appetite or overeating	.594		.479
6. Feeling bad about yourself/failure	.607	.531	
7. Trouble concentrating	.772	.294	
8. Moving or speaking so slowly	.749	.192	
9. Thoughts that you would be better off dead	.636	.439	
Factor correlation			.772

Note. Cog/affect = Cognitive/affective.

<https://doi.org/10.1371/journal.pone.0199235.t002>

**Table 3. Summary of goodness of fit statistics for tested models in multi-group analyses.**

	$\chi^2$	df	RMSEA	AIC	BIC	SRMR	CFI	$\Delta$ CFI
Non-clinical group vs. Only MDD group								
Model 1	176.55	34	.096	21724.96	22097.37	.025	.976	-
Model 2	225.81	52	.083	21738.21	22020.04	.073	.971	.005
Model 3	230.39	55	.081	21736.79	22003.52	.034	.971	0
Model 4	428.04	64	.102	21916.45	22137.88	.048	.939	.032
Model 5	520.86	67	.111	22003.26	22209.60	.106	.924	.015
Only MDD group vs. MDD with AD group								
Model 1	148.76	36	.089	17203.25	17203.26	.084	.976	-
Model 2	187.79	54	.079	17127.13	17127.13	.077	.969	.007
Model 3	210.92	60	.080	17105.43	17105.43	.076	.966	.003
Model 4	271.15	68	.087	17115.67	17115.67	.086	.950	.016
Model 5	452.50	71	.116	17276.77	17276.77	.116	.905	.045

Model 5 (Factor invariance model) = loadings, intercepts, residuals, and factor means are invariant.

Note. MDD = Major depressive disorder, AD = Anxiety disorder, RMSEA = Root Mean Square Error of Approximation, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, CFI = Comparative Fit Index, SRMR = Standardized Root-Mean-Square Residual.

Model 1 (Configural model) = all parameter free.

Model 2 (Metric model) = loadings are invariant.

Model 3 (Scalar model) = loadings and intercepts are invariant.

Model 4 (Error invariance model) = loadings, intercepts, and residuals are invariant.

<https://doi.org/10.1371/journal.pone.0199235.t003>

populations, we found that the bi-factor model had the best fit. This explains the mixed results found in previous studies that reported that the PHQ-9 has either a unidimensional or a two-dimensional factor structure [8, 21]. The bi-factor model allows one to use the unidimensional factor model of the PHQ-9, that is, we can use the cut-off point and the total score as a single variable. Additionally, we can use the two-dimensional factor model of the PHQ-9 for assessing more detailed symptoms. Moreover, the general PHQ-9 factor accounted for over 40% of the common variance. Thus, using both total score and sub-scale scores allows us to assess patients' symptoms more precisely. In addition to assessing patients' symptoms more precisely, we may be able to detect the change in patients' symptoms due to treatment more fully by using the PHQ-9 regularly during treatment. This in turn will aid the implementation of appropriate treatment or the modification of the treatment according to the patients' needs.

According to the results of the measurement invariance, scalar invariance showed best fit between the non-clinical and only MDD groups, and between the only MDD and MDD with AD groups, which means that we can compare the latent mean of the PHQ-9 between these two populations. Although the PHQ-9 total, somatic, and cognitive/affective scores of the MDD with AD group were higher than those of the MDD and non-clinical groups, these populations responded to each item similarly.

This study has several limitations. First, we might have obtained a biased sample because we conducted a web-based survey. For example, patients with more severe depressive symptoms and those who do not use the Internet frequently might have been excluded from this web-based survey. Second, participants were asked to report their own diagnoses and were not interviewed to assess whether they actually had MDD/AD. In other words, some of the participants might not have met the required diagnostic criteria for MDD/AD. This is in part supported by the low mean PHQ-9 score reported in the present study ( $M = 12.42$  for MDD only,  $M = 15.86$  for MDD/AD) as compared to that found in previous studies conducted in Western countries. For example, Kroenke [3] reported a mean score of 17.1 among 41 patients with

MDD, and Petersen [8] reported a mean score of 17.3 among the 626 such patients. Future studies must test the higher-order factorial model and assess measurement invariance with participants diagnosed using a structured interview. Finally, we used only Japanese non-clinical and clinical populations, making it unclear whether these results are applicable to a Western population.

## Supporting information

### S1 Dataset.

(XLSX)

## Acknowledgments

We thank all the participants who contributed to our larger web-based survey.

## Author Contributions

**Conceptualization:** Masaya Ito.

**Data curation:** Masaya Ito.

**Formal analysis:** Satomi Doi, Yoshitake Takebayashi.

**Funding acquisition:** Masaya Ito.

**Investigation:** Masaya Ito.

**Methodology:** Masaya Ito, Yoshitake Takebayashi, Kumiko Muramatsu.

**Project administration:** Masaya Ito.

**Supervision:** Masaya Ito, Kumiko Muramatsu, Masaru Horikoshi.

**Writing – original draft:** Satomi Doi.

**Writing – review & editing:** Masaya Ito, Yoshitake Takebayashi.

## References

1. Furukawa TA. Assessment of mood: guides for clinicians. *Psychosom Res.* 2010; 68(6):581–589.
2. Spitzer RL, Kroenke K, Williams JB. Patient Health Questionnaire Primary Care Study Group. Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA.* 1999; 282(18):1737–1744. PMID: [10568646](https://pubmed.ncbi.nlm.nih.gov/10568646/)
3. Kroenke K, Spitzer RL, Williams JB. The PhQ-9. *J Gen Int Med.* 2001; 16(91):606–613.
4. National Institute for Health and Clinical Excellence 8 Depression The treatment and management of depression in adults (update: Clinical guideline 90) [Internet]. National Institute for Health and Clinical Excellence; 2009 [cited 2016/6/1]. [www.nice.org.uk/CG90](http://www.nice.org.uk/CG90).
5. Hansson M, Chotai J, Nordstöm A, Bodlund O. Comparison of two self-rating scales to detect depression: HADS and PHQ-9. *Br J Gen Pract.* 2009; 59(566):e283–288. <https://doi.org/10.3399/bjgp09X454070> PMID: [19761655](https://pubmed.ncbi.nlm.nih.gov/19761655/)
6. Beard C, Hsu KJ, Rifkin LS, Busch AB, Björgvinsson T. Validation of the PHQ-9 in a psychiatric sample. *J Affect Disord.* 2016; 193:267–273. <https://doi.org/10.1016/j.jad.2015.12.075> PMID: [26774513](https://pubmed.ncbi.nlm.nih.gov/26774513/)
7. Chilcot J, Rayner L, Lee W, Price A, Goodwin L, Monroe B, et al. The factor structure of the PHQ-9 in palliative care. *J Psychosom Res* 2013; 75(1):60–4. <https://doi.org/10.1016/j.jpsychores.2012.12.012> PMID: [23751240](https://pubmed.ncbi.nlm.nih.gov/23751240/)
8. Petersen JJ, Paulitsch MA, Hartig J, Mergenthal K, Gerlach FM, Gensichen J. Factor structure and measurement invariance of the Patient Health Questionnaire-9 for female and male primary care patients with major depression in Germany. *J Affect Disord.* 2015; 170:138–142. <https://doi.org/10.1016/j.jad.2014.08.053> PMID: [25240840](https://pubmed.ncbi.nlm.nih.gov/25240840/)



9. Credé M, Harms PD. 25 years of higher-order confirmatory factor analysis in the organizational sciences: A critical review and development of reporting recommendations. *J Organ Behav.* 2015; 36(6):845–872.
10. Yung YF, Thissen D, McLeod LD. On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika.* 1999; 64(2):113–128.
11. Urbán R, Arrindell WA, Demetrovics Z, Unoka Z, Timman R. Cross-cultural confirmation of bi-factor models of a symptom distress measure: Symptom Checklist-90-Revised in clinical samples. *Psychiat Res.* 2016; 239:265–274.
12. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organ Res Methods.* 2000; 3(1):4–70.
13. Sartorius N, Üstün TB, Lecrubier Y, Wittchen HU. Depression comorbid with anxiety: Results from the WHO study on “Psychological disorders in primary health care.”. *Br J Psychiatr.* 1996; 168(30):38–43.
14. Ito M, Bentley KH, Oe Y, Nakajima S, Fujisato H, Kato N, et al. Assessing depression related severity and functional impairment: the overall depression severity and impairment scale (ODSIS). *PloS One.* 2015; 10(4):e0122969. <https://doi.org/10.1371/journal.pone.0122969> PMID: 25874558
15. Ito M, Oe Y, Kato N, Nakajima S, Fujisato H, Miyamae M, et al. Validity and clinical interpretability of Overall Anxiety Severity and Impairment Scale (OASIS). *J Affect Disord.* 2015; 170:217–224. <https://doi.org/10.1016/j.jad.2014.08.045> PMID: 25259673
16. Sawada N, Uchida H, Watanabe K, Kikuchi T, Suzuki T, Kashima H, et al. How successful are physicians in eliciting the truth from their patients? A large-scale Internet survey from patients’ perspectives. *J Clin Psychiatry.* 2012; 73(3): 311–317. <https://doi.org/10.4088/JCP.11m07078> PMID: 22490259
17. Muramatsu K, Kamijima K, Yoshida M, Otsubo T, Miyaoka H, Muramatsu Y, Gejyo F. The Patient Health Questionnaire, Japanese Version: Validity According to the Mini-International Neuropsychiatric Interview–Plus. *Psychol Reports.* 2007; 101(3):952–960.
18. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the brief Patient Health Questionnaire mood scale (PHQ-9) in the general population. *Gen Hosp Psychiat.* 2006; 28:71–77.
19. Furukawa TA, Kawakami N, Saitoh M, Ono Y, Nakane Y, Nakamura Y, et al. The performance of the Japanese version of the K6 and K10 in the World Mental Health Survey Japan. *Int J Methods Psychiatr Res.* 2008; 17(3):152–158. <https://doi.org/10.1002/mpr.257> PMID: 18763695
20. Shima S, Shikano T, Kitamura T, Asai M. Reliability and validity of CES-D. *Jpn J Psychiatry.* 1985; 27:717–723.
21. Yu X, Tam WW, Wong PT, Lam TH, Stewart S. The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. *Compr Psychiatr.* 2012; 53(1):95–102.
22. Gregorich SE. Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Med Care.* 2006; 44(11 Suppl 3):S78. <https://doi.org/10.1097/01.mlr.0000245454.12228.8f> PMID: 17060839
23. Kline RB. Principles and practice of structural equation modeling. New York: Guilford Press; 2015.
24. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equ Modeling.* 2002; 9(2):233–255.