

The Patient Health Questionnaire-9 among First-Trimester Pregnant Women in Japan: Factor Structure and Measurement and Structural Invariance between Nulliparas and Multiparas and across Perinatal Measurement Time Points

Mikiyo Wakamatsu¹, Mariko Minatani², Ayako Hada^{3,4}, Toshinori Kitamura^{3,4,5,6*}

¹Department of Reproductive Health Care Nursing, Kagoshima University Faculty of Medicine School of Health Sciences, Kagoshima, Japan

²Life Value Creation Unit, NTT Data Institute of Management Consulting, Inc., Tokyo, Japan

³Kitamura Institute of Mental Health Tokyo, Tokyo, Japan

⁴Kitamura KOKORO Clinic Mental Health, Tokyo, Japan

⁵T. and F. Kitamura Foundation for Studies and Skill Advancement in Mental Health, Tokyo, Japan

⁶Department of Psychiatry, Graduate School of Medicine, Nagoya University, Nagoya, Japan

Email: *kitamura@institute-of-mental-health.jp

How to cite this paper: Wakamatsu, M., Minatani, M., Hada, A., & Kitamura, T. (2021). The Patient Health Questionnaire-9 among First-Trimester Pregnant Women in Japan: Factor Structure and Measurement and Structural Invariance between Nulliparas and Multiparas and across Perinatal Measurement Time Points. *Open Journal of Depression, 10*, 121-137.

<https://doi.org/10.4236/ojd.2021.103008>

Received: June 17, 2021

Accepted: August 7, 2021

Published: August 10, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Background: Depression is frequently seen among pregnant women. The PHQ-9 is a diagnostic criteria-based screening instrument. **Objective:** The objective is to identify the factor structure of the Patient Health Questionnaire-9 (PHQ-9) and to confirm its measurement and structural invariance regarding parity and observation time, as well as to examine reliability by calculating omega indices in a population of pregnant Japanese women. **Methods:** The PHQ-9 was distributed to 382 pregnant women with a gestational age of 10 to 13 weeks who were attending antenatal clinics. Of these women, 129 were tested again with the PHQ-9 one week later. **Results:** Exploratory factor analyses (EFAs) of the PHQ-9 items were conducted using a halved sample. Confirmatory factor analyses (CFAs) among the other halved sample compared the single-, two-factor models derived from the EFAs, literature model and bifactor model. The bifactor model was significantly better in terms of goodness-of-fit indices: $\chi^2/df = 1.030$ (18), CFI = 0.999, RMSEA = 0.012. ECV suggested multidimensionality. Internal consistency was excellent: $\omega = 0.868$, $\omega_{\text{Somatic factor}} = 0.756$, $\omega_{\text{Non-somatic factor}} = 0.817$.

= 0.795, ω_{HS} Somatic factor = 0.330, ω_{HS} Non-somatic factor = 0.032. The bifactor structure model showed configural, measurement, and structural invariances between primiparas and multiparas as well as at two observation occasions. **Conclusion:** The findings suggested that the PHQ-9 used among pregnant women was robust in its factor structure.

Keywords

Patient Health Questionnaire-9, Factor Structure, Pregnancy, Measurement and Structural Invariance, Parity

1. Introduction

Depression is a mental disorder commonly seen during pregnancy. The incidence of a DSM-IV Major Depressive Episode (MDE) during pregnancy is about 5% (Kitamura et al., 2006). If less severe depression such as Minor Depression and Intermittent Depression as defined by the Research Diagnostic Criteria (RDC: Spitzer et al., 1978) is included, its incidence may go beyond 15% (Kitamura, Shima, Sugawara, & Toda, 1993, 1996). The onset of antenatal depression is often characterised by psychosocial correlates such as a lack of support from a partner, poor accommodational conditions, an undesirable rearing experience in childhood, and the woman's unstable personality (e.g., Kitamura, Shima, Sugawara, & Toda, 1993; Kitamura, Toda, Shima, & Sugawara, 1994b; Kitamura, Toda, Shima, Sugawara, & Sugawara, 1998). Therefore, antenatal depression is an important health issue that should be recognised by nurses and midwives.

Quite a few instruments have been used as screening devices to identify antenatal depression. They include the General Health Questionnaire (GHQ: Goldberg, 1972) (e.g., Kitamura, Sugawara, Aoki, & Shima, 1989; Kitamura, Toda, Shima, & Sugawara, 1994a), the Self-Rating Depression Scale (SDS: Zung, 1965) (e.g., Kitamura, Shima, Sugawara, & Toda, 1994c; Kitamura, Sugawara, Shima, & Toda, 1999), the Beck Depression Inventory (BDI: Beck et al., 1961) (e.g., Salameiro et al., 1994), and the Edinburgh Postnatal Depression Scale (EPDS: Cox et al., 1987). Although the EPDS was originally developed as a screening tool to identify postnatal depression, it is often used to screen antenatal depression (Matthey et al., 2016; Wickberg et al., 2005). A drawback of these screening instruments is their relative inability to accurately identify cases that meet diagnostic criteria such as MDE. While its negative predictive value is usually high, its positive predictive value is poor. The Patient Health Questionnaire-9 (PHQ-9: Spitzer et al., 1999) is a self-report measure that can easily be converted to the diagnostic criteria of MDE. This is because all of the items of the PHQ-9 are derived from items of MDE. The PHQ-9 is a diagnostic criteria-based screening instrument. The psychometric properties of the PHQ-9 as a screening for MDE have been widely reported (Beard et al., 2016; Gilbody et al., 2007; Inoue et al., 2012; Manea et al., 2012; Wittkamp et al., 2007). Another strength of the PHQ-9

is its ability to predict depression severity. Among psychiatric inpatients (Kitamura, Nakagawa, & Machizawa, 1993), the number of MDE items identified as present was moderately correlated with the depression severity rated by interviewer-rated scales such as Hamilton's Rating Scale for Depression (Hamilton, 1960) and the Global Assessment Scale (GAS: Endicott et al., 1976). PHQ-9 scores were also associated with severity ratings of other measures including those from the Center for the Epidemiologic Studies of Depression (Radloff, 1977; Beard et al., 2016) and functional difficulty (Kroenke et al., 2001).

When using a psychological scale as either a screening instrument or a severity measure, it is of great importance to examine its factor structure. For example, if it consists of more than one factor, then it is recommended to rate severity using subscales rather than a total score. If the factor structure differs, the scores on the subscales will be different and the care of interventions for the participants will be different.

The factor structure of the nine items of the PHQ-9 has been studied by several researchers. There are reports supporting a single-factor model (González-Blanch et al., 2018). In patients with spinal cord injury and depression, a comparison of different factor structure models indicated that a two-factor structure was the best (Krause et al., 2011). In this model, one factor included items of "sleep change", "fatigue", "appetite change", "psychomotor agitation/retardation", and "concentration difficulties". In another study, one factor included *somatic* items of "sleep change", "fatigue", "appetite change", the other factor including *Non-somatic* items (Hall et al., 2021; Keum et al., 2018). Using both patients with psychiatric issues and non-clinical populations, Doi et al. (2018) asserted that the two-factor bifactor model best fit the data. In this study, three items including "sleep change", "fatigue", and "appetite change" had high factor loadings on the first factor. However, no consensus was reached about the item configuration. Recently, the bifactor model has attracted researchers and clinicians. The bifactor model has a general factor and several group factors (subscales). Research and clinical implications of a bifactor model should be considered whether the model is basically unidimensional (therefore effects of group factors are negligible) or multidimensional (therefore effects of the general factor are negligible). The explained common variance (ECV) and a group of omega (ω) coefficients are useful indicators here (Reise, 2013; Rodriguez et al., 2016). The ECV is the proportion of common variance across all items that are explained only by the general dimension. Stucky and Edelen (2014) suggested that ECV values of approximately 0.85 or higher are needed to consider a set of items sufficiently unidimensional to warrant a one-factor model. Higher ω_{FB} the more is suggested unidimensionality (Rodriguez et al., 2016).

Because of the worldwide use of the PHQ-9, configural, measurement, and factor invariance has attracted researchers' concerns (van de Vijver & Leung, 2000). This is because the instrument will cast doubt about its validity if its factor structure (configural invariance), factor loadings of indicators/items (metric

invariance), indicators' intercepts (scalar invariance), and residuals of the indicators (residual invariance) differ significantly between those belonging to different countries (and cultures and languages). Measurement invariance includes metric, scalar, and residual invariances. Moreover, factor variances, factor covariances, and factor means of a psychological measure should be equivalent (structural invariance) between those belonging to different demographic characteristics if used as a means of comparison (Vandenberg & Lance, 2000). González-Blanch et al. (2018) used a primary care patient population and found that a two-factor structure model better fit with the data than a single-factor model. However, because of a strong correlation between the two factors on the two-factor structure model, they preferred the single-factor model. In addition, they confirmed measurement invariance of the single-factor model between men and women, age groups, marital status, level of education, and employment status. In addition, invariance of the factor structure should also be confirmed between two observation times (Widaman et al., 2010). Comparing English-speaking and Spanish-speaking women in the U.S., Merz et al. (2011) reported that a single-factor PHQ-9 factor structure demonstrated configural and factor variance equivalence. However, this study failed to confirm residual invariance. Factor mean invariance was not reported. Doi et al. (2018) studied patients with MDE and reported that a two-factor bifactor model was the best fit over other models and this model satisfied scalar invariance.

The use of the PHQ-9 in a population of pregnant women was reported in a few studies. Woldetensay et al. (2018) reported the validity of the PHQ-9 as a screening for antenatal depression, which was validated by a mental health specialist interview-derived diagnosis of MDE. In a population of pregnant Peruvian women (Zhong et al., 2014), exploratory factor analysis (EFA) indicated a two-factor structure that was supported by confirmatory factor analysis (CFA). In a population of pregnant Spanish women (Marcos-Nájera et al., 2018), however, a three-factor model fits better with the data than a two-factor model. Invariance of the PHQ-9 factor structure has not been reported. Even when used in pregnant women, the invariance of the factor structure of the scale (configural invariance), the factor loadings of the indicators and items (metric invariance), the intercept of the indicators (scalar invariance), and the residuals of the indicators (residual invariance) must be demonstrated. Invariance in terms of parity is important when taking into consideration that nulliparas and multiparas showed differences in many studies using structural equation modelling analyses (e.g., Kitamura, Ohashi, Murakami, & Goto, 2019; Kitamura, Ohashi, Sakanashi, & Tanaka, 2019).

To the best of our knowledge, the PHQ-9 has never been used in a population of pregnant Japanese women. Taking into account a high incidence of antenatal depression, it may be of clinical and research importance to study the psychometric properties of the PHQ-9 in a population of pregnant women including EFA and CFA, as well as the invariance of the identified factor structure.

2. Methods

2.1. Study Procedures and Participants

The target of this study was pregnant women at 10 to 13 weeks' gestational age. Approximately 1500 pregnant women were recruited at the antenatal clinic of one general hospital and five private clinics located in Tokyo, Chiba, Ibaraki, and Kagoshima Prefectures in Japan. Exclusion criteria included women who: 1) were not fluent in Japanese; 2) were aged under 20; 3) had eating disorders; 4) had symptoms of vaginal bleeding or abdominal pain; 5) had subchorionic haematoma; or 6) had experienced recurrent miscarriages. We administered a set of questionnaires on two occasions, 1 week apart. The total sample consisted of 382 pregnant women (approximately 25% of those who were solicited) who participated in this study. Of these, 129 women responded to the retest 1 week later. Participation was voluntary and anonymity was assured. Written informed consent was obtained from each participant. The mean age of the participants was 31.9 (SD 4.9) years old and that of their partners was 33.5 (SD 5.5) years old (Table 1). Most of them were married (94.5%), 44.0% of the participants were nulliparas and 55.0% were multiparas. The recruitment for this study was conducted from January 2017 to May 2019.

2.2. Measurements

PHQ-9 (Spitzer et al., 1999): We used the Japanese version of the PHQ-9 (Inagaki et al., 2013; Muramatsu, & Kamijima, 2009). This is a nine-item self-report depression scale based on the MDE criteria in the Diagnostic and Statistical Manual of Mental Disorder-IV (DSM-IV). Each item checks for frequency of depressive symptoms over the previous two weeks with a 4-point Likert scale from 0 to 3.

Sheehan Disability Scale (SDS: Sheehan, 1983): We used the Japanese version of the SDS (Yoshida et al., 2004). The SDS is a three-item self-report scale. Three disabilities in the domains of 1) work and school work; 2) social and leisure activities; and 3) family life and home responsibility are measured. Each item is rated from 0 to 10. Its psychometric validation has been reported previously (Arbuckle et al., 2009). The psychometric properties of the SDS among the present

Table 1. Demographic features of the participants.

Demographic features	Mean (SD)/n (%) Total N = 382
Own age	31.9 (4.9) years
Partner's age	33.5 (5.5)
Married	361 (94.5%)
Nulliparae	168 (44.0%)
Followed up 1-week later	129 (33.8%)

sample were reported elsewhere (Hada et al., 2021).

2.3. Data analysis

The whole sample was randomly divided into two groups: Group A ($n = 184$) for EFA and Group B ($n = 198$) for CFA. Among Group A, we calculated mean, SD, skewness, and kurtosis of each PHQ-9 item. As seen in the results, excessive skewness > 2.0 required log transformation of the PHQ-9 items. These log-transformed PHQ-9 items were subjected to EFA. Factorability of the items was checked by the Kaiser-Meyer-Olkin (KMO) index and Bartlett's sphericity test (Burton & Mazerolle, 2011). We then performed a series of exploratory factor analyses (EFAs). This was done by the maximum-likelihood method with PROMAX rotation starting from a single-factor structure model and progressing to models with a greater number of factors (i.e., two- and three-factor structures, and so on). These models were compared with goodness-of-fit in a series of confirmatory factor analyses (CFAs) among Group B. The fit of the models was examined in terms of chi-squared, comparative fit index (CFI), and root mean square of error approximation (RMSEA). A good fit would be indicated by $\chi^2/df < 2$, CFI > 0.97 , and RMSEA < 0.05 , and an acceptable fit by $\chi^2/df < 3$, CFI > 0.95 , and RMSEA < 0.08 (Bentler, 1990; Schermelleh-Engel et al., 2003). A model was judged better than another if its Akaike information criterion (AIC; Akaike, 1987) was lower than the other.

Comparison of factor structure models derived from EFAs, literature model and bifactor model were performed as cross validation (Cliff, 1983; Cudeck, & Browne, 1983; Romera et al., 2008) using the second halved sample, Group B. Starting with the single-factor model, the subsequent model was judged as "accepted" if χ^2 decreased significantly for the difference of df . This was repeated until we reached the best model. After deciding on the best-fit model among first-order models, we built a bifactor model. Then we calculated ECV and ω indices of this model to determine whether the model was basically unidimensional or multidimensional. The ECV is the proportion of common variance across all items that are explained only by the general dimension. Stucky and Edelen (2014) suggested that ECV values of approximately 0.85 or higher are needed to consider a set of items sufficiently unidimensional to warrant a one-factor model. Omega (ω) indicates the proportion of variance of the whole measurement explained by the general and all the group factors. Omega subscale (ω_s) indicates the proportion of the variance among items of each specific group factor explained by both the general and the group factor. The proportion of the variance of the whole measurement explained only by the general factor is termed omega hierarchical (ω_H). Omega hierarchical subscale (ω_{HS}) indicates the proportion of the variance among items of each specific group factor explained by the group factor. Higher ω_{HS} the more is suggested unidimensionality (Rodriguez et al., 2016).

The bifactor model's configural, measurement, and structural invariance was

examined across different attributes (parity and observation occasions) among the whole sample. This started from configural invariance, through metric, scalar, residual, and factor variance invariances to factor covariance invariances. The progress from one step to the next was judged as “accepted” if 1) the χ^2 decrease was not significant for the *df* difference; 2) the decrease of CFI was less than 0.01; or 3) the increase of RMSEA was less than 0.01 (Chen, 2007; Cheung & Rensvold, 2002). This procedure was applied because a χ^2 decrease is strongly sensitive to the sample size (*N*) and, particularly in the case of a large sample, produces an unreasonable rejection of invariance.

2.4. Ethical Consideration

This study was approved by the Institutional Review Board (IRB) of the Kitamura Institute of Mental Health Tokyo (No. 2015052301) and Kagoshima University (No. 170247).

3. Results

Mean, SD, skewness, and kurtosis of each PHQ-9 item in Group A is in **Table 2**. Two items (item 8 “psychomotor symptoms” and item 9 “suicidality”) were excessively skewed. Hence, all of the PHQ-9 items were log transformed for the subsequent analyses.

After finding the data was factorable, KMO = 0.816 and Bartlett’s sphericity test $\chi^2 (36) = 465.199$ $p < 0.001$, we performed EFAs (**Table 3**). In the single-factor model, all the items showed factor loading > 0.30 (Costello & Osborne, 2005). The two-factor model looked similar to previous studies in which the first factor loaded somatic items including “Sleep change”, “Fatigue”, and “Appetite change”, together with “Concentration difficulty”. In the three-factor model, the third factor included only two items, “Psychomotor symptoms” and “Suicidality”. The factor structure in which has factor(s) including only two indicators is

Table 2. Mean, SD, and skewness values for each PHQ-9 item ($n = 184$).

No.	Content	<i>N</i>	Original				After log transformation				
			M	SD	skewness	kurtosis	M	SD	skewness	kurtosis	communality
1	Loss of interest	182	1.70	0.87	1.18	0.74	0.42	0.46	0.53	-1.01	0.50
2	Feeling depressed	183	1.52	0.73	1.53	2.37	0.33	0.41	0.78	-0.61	0.55
3	Sleep change	183	2.36	1.11	0.21	-1.29	0.73	0.52	-0.29	-1.31	0.28
4	Fatigue	183	2.45	0.95	0.16	-0.89	0.81	0.44	-0.62	-0.49	0.38
5	Appetite change	183	2.60	1.01	0.04	-1.09	0.86	0.45	-0.71	-0.45	0.30
6	Self-blame	183	1.46	0.76	1.50	1.21	0.27	0.43	1.12	-0.32	0.44
7	Concentration difficulty	183	1.55	0.85	1.35	0.69	0.32	0.46	0.97	-0.66	0.39
8	Psychomotor symptoms	183	1.21	0.59	3.01	9.34	0.13	0.32	2.48	5.04	0.14
9	Suicidality	183	1.03	0.21	7.03	54.03	0.02	0.13	6.32	40.79	0.11

weak structure. Hence, the three-factor model was rejected as a model of the PHQ-9 factor model.

We then compared models derived from EFAs, two-factor somatic-non-somatic model and bifactor model in CFAs among Group B women. The two-factor somatic-non-somatic model is superior and good (Table 4). In addition to a significant decrease of χ^2 ($\Delta\chi^2(df) = 23.137$ (1), $p < 0.001$), both CFI (from 0.914 to 0.977) and RMSEA (from 0.104 to 0.055) were better in the two-factor somatic-non-somatic model. We further built a bifactor model (Figure 1). The model fit in terms of the goodness-of-fit was improved: $\chi^2/df = 1.030$ (18), CFI = 0.999, RMSEA = 0.012. We thought that the bifactor model was the best.

We then calculated ECV and omega indices (Table 5). Since ECV was 0.761, we considered that the items for PHQ-9 were reasonable to consider as multidimensional (Stucky & Edelen, 2014). Similarly, ω coefficients suggested multidimensionality: $\omega = 0.868$, ω . Somatic factor = 0.756, ω . Non-somatic factor = 0.817. $\omega_H = 0.795$, ω_{HS} . Somatic factor = 0.330, ω_{HS} . Non-somatic factor = 0.032.

We then tested whether the bifactor model was invariant between nulliparas ($n = 168$) and multiparas ($n = 210$) using the whole sample. Five cases were not known for parity. The bifactor model showed configural, measurement, and structural invariance (Table 6). Factor means were significantly different between the

Table 3. EFA of the PHQ-9 ($n = 184$).

Item	1-factor	2-factor		3-factor		
	I	I	II	I	II	III
1: Loss of interest	0.71	0.25	0.50	0.30	0.39	-0.11
2: Feeling depressed	0.74	0.04	0.79	0.01	1.01	-0.13
3: Sleep change	0.53	0.67	-0.61	0.63	-0.05	0.03
4: Fatigue	0.61	0.84	-0.12	0.78	-0.02	-0.07
5: Appetite change	0.55	0.62	-0.01	0.69	0.02	-0.14
6: Self-blame	0.66	0.04	0.68	0.09	0.48	0.25
7: Concentration difficulty	0.63	0.40	0.26	0.42	0.00	0.39
8: Psychomotor symptoms	0.37	-0.65	0.31	0.08	-0.02	0.75
9: Suicidality	0.33	-0.19	0.53	-0.18	0.34	0.31

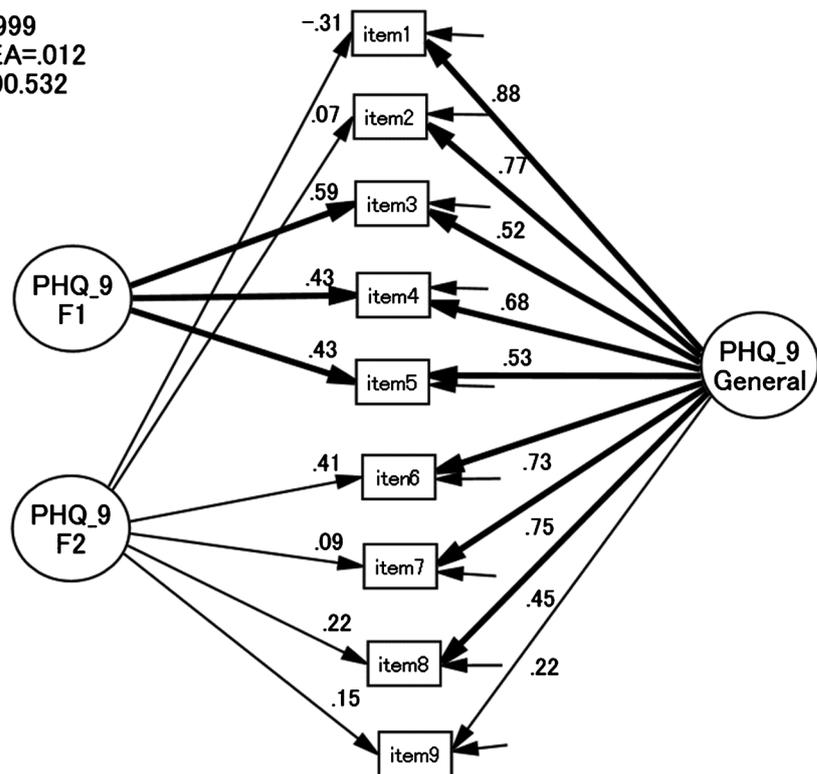
Note: Factor loading > 0.3 in bold.

Table 4. Comparison of factor structure models by CFA ($n = 198$).

Model	χ^2	df	χ^2/df	$\Delta \chi^2(df)$	CFI	ΔCFI	RMSEA	$\Delta RMSEA$	AIC
1-factor	84.480	27	3.129	Ref	0.914	Ref	0.104	Ref	138.480
2-factor (EFA-derived)	79.225	26	3.047	5.255 (1)	0.920	0.006	0.102	0.002	135.225
somatic-non-somatic model	41.669	26	1.603	42.811 (1)***	0.977	0.057	0.055	0.047	97.669
bifactor	18.532	18	1.030	23.137 (8)**	0.999	0.028	0.012	0.043	90.532

Note: Indices of Somatic-Non somatic model was compared with 1-factor model; ** $p < 0.01$; *** $p < 0.001$; CFI: Comparative fit index; RMSEA: Root mean square of error approximation; AIC: Akaike information criterion.

chi-squared=18.532
 df=18
 CFI=.999
 RMSEA=.012
 AIC=90.532



PHQ-9: Patient Health Questionnaire-9; CFI: Comparative fit index; RMSEA: Root mean square error of approximation; AIC: Akaike information criteria. Paths are standardised. Significant paths are in bold. The names of error variables are deleted. Paths with significant ($p < 0.001$) coefficients are in bold.

Figure 1. Confirmative factor analysis of the PHQ-9 ($n = 198$).

Table 5. Omega indices for the bifactor model.

Factors	ω/ω_s	ω_H/ω_{HS}
General	0.868	0.795
Somatic factor	0.756	0.330
Non-somatic factor	0.817	0.032

Note: ECV = 0.761.

Table 6. Configural, measurement, and structural invariance of the two-factor model between nulliparas ($n = 168$) and multiparas ($n = 210$).

	χ^2	df	χ^2/df	$\Delta\chi^2(df)$	CFI	ΔCFI	RMSEA	$\Delta RMSEA$	AIC	Judgement
Configural	50.347	36	1.399	Ref	0.987	Ref	0.033	Ref	194.347	ACCEPT
Metric	74.729	51	1.465	24.382 (15)NS	0.978	0.009	0.035	0.002	188.729	ACCEPT
Scalar	91.086	60	1.518	16.357 (9)NS	0.971	0.007	0.037	0.002	187.086	ACCAPT
Residual	98.993	69	1.435	7.907 (9)NS	0.972	+0.001	0.034	+0.003	176.993	ACCEPT
Factor variance	102.230	72	1.420	3.237 (3)NS	0.972	0.000	0.033	+0.001	174.230	ACCEPT

NS: Not significant; CFI: Comparative fit index; RMSEA: Root mean square of error approximation; AIC: Akaike information criterion.

nulliparas and multiparas only for Non-somatic factor (Somatic factor, difference (SE) = -0.024 (0.038), NS; Non-somatic factor, difference (SE) = -0.107 (0.036), $p < 0.01$; General factor, difference (SE) = -0.015 (0.046), NS).

The bifactor structure model was also examined in terms of its invariance between the initial ($n = 377$) and follow-up ($n = 126$) occasions (Table 7). Here again, the bifactor model showed configural, measurement, and structural invariance. Factor means were not significantly different between time 1 and time 2 (Table 8).

In the CFA, Somatic factor and the General factor were significantly correlated with the SDS factor ($r = 0.24$ and 0.71 , respectively), whereas Non-somatic factor was not. It suggested that the severity of depression was associated with disability (Figure 2).

4. Discussion

Our study demonstrated that the PHQ-9 among pregnant women had a bifactor structure. This is in line with the previous literature (Doi et al., 2018; Hall et al., 2021; Keum et al., 2018). A different pattern of a two-factor structure of the PHQ-9 was reported by Krause et al. (2011); however, that sample included those suffering from spinal cord injuries and was, therefore, more likely to complain about physical-disease-derived symptoms such as moving and concentration difficulty that may have mingled with the somatic factor of PHQ-9. González-Blanch et al. (2018) claimed a single-factor model. However, in their study, the CFI was substantially better for the somatic-non somatic model (0.97) than the single-factor model (0.91). Here, the same three PHQ-9 items (items 3, 4, and 5) were grouped in the same factor. Their argument rejecting the two-factor model

Table 7. Configural, measurement, and structural invariance of the two-factor model between two observation occasions.

	χ^2	df	χ^2/df	$\Delta\chi^2(df)$	CFI	ΔCFI	RMSEA	$\Delta RMSEA$	AIC	Judgement
Configural	72.645	36	2.018	Ref	0.976	Ref	0.045	Ref	216.645	ACCEPT
Metric	86.843	51	1.703	14.20 (15)NS	0.976	0.000	0.037	+0.008	200.841	ACCEPT
Scalar	85.955	60	1.433	-0.888 (9)NS	0.983	+0.007	0.029	+0.008	181.955	ACCEPT
Residual	95.871	69	1.389	9.916 (9)NS	0.982	0.001	0.028	+0.001	173.871	ACCEPT
Factor variance	97.651	72	1.356	1.780 (3)NS	0.983	+0.001	0.026	+0.002	169.651	ACCEPT

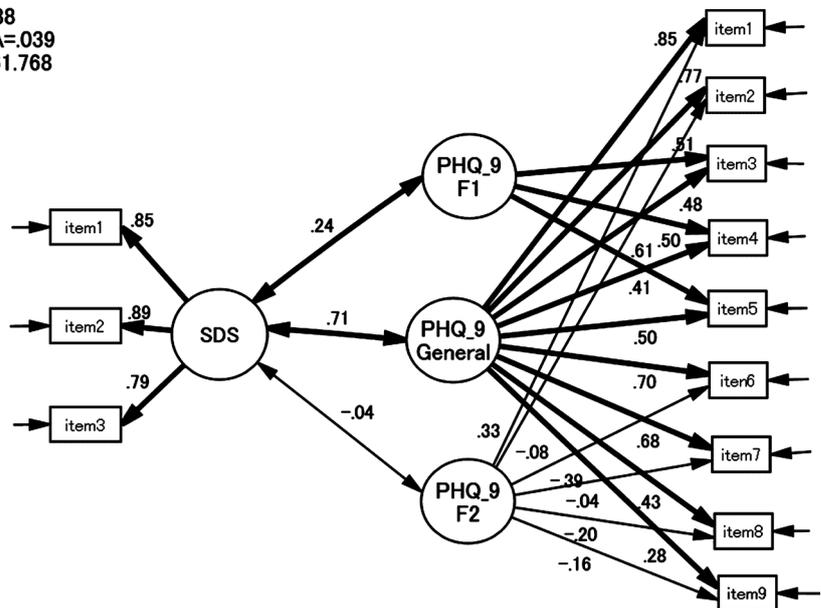
NS: Not significant; CFI: Comparative fit index; RMSEA: Root mean square of error approximation; AIC: Akaike information criterion.

Table 8. Factor mean for the PHQ-9.

	factor mean (SE)		
	Somatic factor	Non-somatic factor	General factor
Nulliparae ($n = 168$) compared with multiparae ($n = 210$)	-0.024 NS (0.038)	-0.107 ** (0.036)	-0.015 NS (0.046)
time1 compared with time 2	0.010 NS (0.038)	0.067 NS (0.037)	-0.026 NS (0.044)

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; NS: Not significant; SE: Standard error.

chi-squared=65.768
df=42
CFI=.988
RMSEA=.039
AIC=161.768



SDS: Sheehan Disability Scale; PHQ-9: Patient Health Questionnaire-9; SDS: Sheehan Disability Scale; CFI: Comparative fit index; RMSEA: Root mean square error of approximation; AIC: Akaike information criteria. Paths are standardised. Significant paths are in bold. The names of error variables are deleted. Paths with significant ($p < 0.001$) coefficients are in bold.

Figure 2. Confirmative factor analysis of the PHQ-9 and SDS ($N = 382$).

was based on a strong correlation ($r = 0.86$) between the two factors. However, our results indicated the bifactor model of the PHQ-9 was the best. ECV of our data suggested a set of items for PHQ-9 was reasonable to consider as multidimensional. ω_{HS} showed low values: ω_{HS} . Non-somatic factor was extremely low (0.032). Item 1, 2, 6, 7, 8, and 9 are little explained by Non-somatic factor but by the General factor. All items are more strongly explained by the General factor than each group factor (subscale). This is the character of construction of factor for the PHQ-9.

Our results also supported invariance of the model in terms of parity as well as observation time. Invariance of the factor structure was present at the configural, measurement, and structural levels. On the factor mean level, there was no difference between nulliparae and multiparae in General factor. Regarding general severity of depression, this implies that the PHQ-9 measures the same phenomena during pregnancy regardless of parity and observation time. Disability was significantly correlated with general factor and somatic factor of PHQ-9, but Non-somatic was not collated. This implied that disabilities in pregnancy are affected by somatic symptoms (i.e., sleep change, fatigue, and appetite change) and general severity of depression.

The robust factor structure regardless of parity and measurement time reported in this study suggests that the PHQ-9 may be used among pregnant women as a tool for screening antenatal depression as well as a measure of depression severity. Depression during pregnancy is not infrequent, causes psychological and functional

maladjustment, and is often associated with a variety of social hardships such as narrow accommodation and the lack of a partner's support (Kitamura, Shima, Sugawara, & Toda, 1996). Accurate identification of cases of depression during pregnancy can lead to better antenatal psychological care by midwives.

Precise identification of cases of antenatal depression may lead to early intervention by perinatal mental health professionals. Because of possibility of foetal malformation by drug therapy, main means of therapeutic approaches are psychotherapies. These include interpersonal therapy (Stuart & Koleva, 2014). Taking into consideration the importance of the pregnant woman's partner, it may be recommended to apply partner-assisted psychotherapy (Brandon et al., 2012).

Our study was not without limitations. First, the participants were outpatients who were pregnant women. They were unlikely to have serious psychiatric disorders. The sample size was relatively small. Moreover, our sample was women at 10 to 13 weeks' gestational age. It is necessary to replicate the study among women in the later stages of pregnancy. Validation of the results should have been conducted with the use of a structured diagnostic interview.

Despite these drawbacks, we think that the PHQ-9 is a promising tool as a measure of depression severity in pregnancy.

Acknowledgements

We are grateful for all of the participants and the Japanese Red Cross Medical Centre, Endou Ladies Clinic, Kubonoya Women's Hospital, Tsuchiya Obsteric & Gynaecology Clinic, Aiiku Hospital, and Nakae Obstetiric & Gynaecology Clinic.

Author Contributions

MM and TK set up the research design. MM, AH, and MW collected data. MW, AH and TK analysed data. MW, AH and TK wrote the manuscript.

Data Availability

The datasets used and/or analysed during the current study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Akaike, H. (1987). Factor Analysis and AIC. *Psychometrika*, 52, 317-332.
<https://doi.org/10.1007/BF02294359>
- Arbuckle, R., Frye, M., Brecher, M., Paulsson, B., Rajagopalam, K., Palmer, S., & Innocenti, A. D. (2009). The Psychometric Validation of the Sheehan Disability Scale (SDS) in Patients with Bipolar Disorder. *Psychiatric Research*, 165, 163-174.
<https://doi.org/10.1016/j.psychres.2007.11.018>

- Beard, C., Hsu, K. J., Rifkin, L. S., Busch, A. B., & Björgvinsson, 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>
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An Inventory for Measuring Depression. *Archives of General Psychiatry, 4*, 561-571. <https://doi.org/10.1001/archpsyc.1961.01710120031004>
- Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin, 107*, 238-246. <https://doi.apa.org/doi/10.1037/0033-2909.107.2.238>
- Brandon, A. R., Ceccotti, N., Hynan, L. S., Shivakumar, G., Johnson, N., & Jarrett, R. B. (2012). Proof of Concept: Partner-Assisted Interpersonal Psychotherapy for Perinatal Depression. *Archives of Women's Mental Health, 15*, 469-480. <https://doi.org/10.1007/s00737-012-0311-1>
- Burton, L. J., & Mazerolle, S. M. (2011). Survey Instrument Validity Part I: Principles of Survey Instrument Development and Validity in Athletic Training Education Research. *Athletic Training Education Journal, 6*, 27-35. <https://doi.org/10.4085/1947-380X-6.1.27>
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling, 14*, 464-504. <https://doi.org/10.1080/10705510701301834>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling, 9*, 233-255. https://doi.org/10.1207/S15328007SEM0902_5
- Cliff, N. (1983). Some Cautions Concerning the Application of Causal Modelling Methods. *Multivariate Behavioral Research, 18*, 115-126. https://doi.org/10.1207/s15327906mbr1801_7
- Costello, A. B., & Osborne, J. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis. *Practical Assessment, Research, and Evaluation, 10*, Article No. 7.
- Cox, J., Holden, J. M., & Sagovsky, R. (1987). Detection of Postnatal Depression: Development of the 10-Item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry, 150*, 782-786. <https://doi.org/10.1192/bjp.150.6.782>
- Cudeck, R., & Browne, M. W. (1983). Cross-Validation of Covariance Structure. *Multivariate Behavioral Research, 18*, 147-167. https://doi.org/10.1207/s15327906mbr1802_2
- Doi, S., Ito, M., Takebayashi, Y., Muramatsu, K., & Horikoshi, M. (2018). Factor Validity and Invariance of the Patient Health Questionnaire (PHQ)-9 among Clinical and Non-Clinical Populations. *PLoS ONE, 13*, e0199235. <https://doi.org/10.1371/journal.pone.0199235>
- Endicott, J., Spitzer, R. L., Fleiss, J. F., & Cohen, J. (1976). The Global Assessment Scale: A Procedure for Measuring Overall Severity of Psychiatric Disturbance. *Archives of General Psychiatry, 33*, 766-771. <https://doi.org/10.1001/archpsyc.1976.01770060086012>
- Gilbody, S., Richards, D., Brealey, S., & Hewitt, C. (2007). Screening for Depression in Medical Settings with the Patient Health Questionnaire (PHQ): A Diagnostic Meta-Analysis. *Journal of General Internal Medicine, 22*, 1596-1602. <https://doi.org/10.1007/s11606-007-0333-y>
- Goldberg, D. P. (1972). *The Detection of Psychiatric Illness by Questionnaire: A Technique for the Identification and Assessment of Non-Psychotic Psychiatric Illness (Maudsley Monograph, 21)*. Oxford University Press.
- González-Blanch, C., Medrano, L. A., Muñoz-Navarro, R., Ruíz-Rodríguez, P., Moriana, J. A., Limonero, J. T., Schmitz, F., Cano-Vindel, A., & PsicAP Research Group (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*, e0193356. <https://doi.org/10.1371/journal.pone.0193356>
- Hada, A., Minatani, M., Yamagishi, Y., Wakamatsu, M., & Kitamura, T. (2021). *Is the Disability Coming into Existence in Early Pregnancy? The Sheehan Disability Scale among First-Trimester Pregnant Women in Japan (Under Review)*. Open Nursing.
- Hall, B. J., Patel, A., Lao, L., Liem, A., Mayawati, E. H., & Tjipto, S. (2021). Structural Validation of the Patient Health Questionnaire-9 (PHQ-9) among Filipina and Indonesian Female Migrant Domestic Workers in Macao: Structural Validation of PHQ-9. *Psychiatry Research*, *295*, Article ID: 113575. <https://doi.org/10.1016/j.psychres.2020.113575>
- Hamilton, M. (1960). A Rating Scale for Depression. *Journal of Neurology, Neurosurgery and Psychiatry*, *23*, 56-62. <https://doi.org/10.1136/jnnp.23.1.56>
- Inagaki, M., Ohtsuki, T., Yonemoto, N., Kawashima, Y., Saitoh, A., Oikawa, Y., Kurosawa, M., Muramatsu, K., Furukawa, T. A., & Yamada, M. (2013). Validity of the Patient Health Questionnaire (PHQ)-9 and PHQ-2 in General Internal Medicine Primary Care at a Japanese Rural Hospital: A Cross-Sectional Study. *General Hospital Psychiatry*, *35*, 592-597. <https://doi.org/10.1016/j.genhosppsych.2013.08.001>
- Inoue, T., Tanaka, T., Nakagawa, S., Nakato, Y., Kameyama, R., Boku, S., Toda, H., Kurita, T., & Koyama, T. (2012). Utility and Limitations of PHQ-9 in a Clinic Specializing in Psychiatric Care. *BMC Psychiatry*, *12*, Article No. 73. <https://doi.org/10.1186/1471-244X-12-73>
- 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*, 1096-1106. <https://doi.apa.org/doi/10.1037/pas0000550>
- Kitamura, T., Nakagawa, Y., & Machizawa, S. (1993). Grading Depression Severity by Symptom Scores: Is It a Valid Way of Subclassifying Depressive Disorders? *Comprehensive Psychiatry*, *34*, 280-283. [https://doi.org/10.1016/0010-440X\(93\)90011-R](https://doi.org/10.1016/0010-440X(93)90011-R)
- Kitamura, T., Ohashi, Y., Murakami, M., & Goto, Y. (2019). Bonding Disorders, Current Adult Attachment, and Perceived Rearing in Childhood in Parents of Children Aged 10 Years Old or Younger: A Structural Equation Model Approach. In T. Kitamura, & Y. Ohashi (Eds.), *Perinatal Bonding Disorders: Causes and Consequences* (pp. 50-81). Cambridge Scholars Publishing.
- Kitamura, T., Ohashi, Y., Sakanashi, K., & Tanaka, T. (2019). Do You Want Another Baby? Bonding Disorder, Dysphoric Mood and Demographic Determinants of Desire for Next Baby among Japanese Women with a 1-Month Old Child. In T. Kitamura, & Y. Ohashi (Eds.), *Perinatal Bonding Disorders: Causes and Consequences* (pp. 212-228). Cambridge Scholars Publishing.
- Kitamura, T., Shima, S., Sugawara, M., & Toda, M. (1993). Psychological and Social Correlates of the Onset of Affective Disorders among Pregnant Women. *Psychological Medicine*, *23*, 967-975. <https://doi.org/10.1017/S003329170002643X>
- Kitamura, T., Shima, S., Sugawara, M., & Toda, M. A. (1996). Clinical and Psychosocial Correlates of Antenatal Depression: A Review. *Psychotherapy and Psychosomatics*, *65*, 117-123. <https://doi.org/10.1159/000289062>
- Kitamura, T., Sugawara, M., Aoki, M., & Shima, S. (1989). Validity of the Japanese Version of the GHQ among Antenatal Clinic Attendants. *Psychological Medicine*, *19*, 507-511. <https://doi.org/10.1017/S0033291700012538>
- Kitamura, T., Sugawara, M., Shima, S., & Toda, M. A. (1999). Temporal Variation of Validity of Self-Rating Questionnaires: Improved Validity of Repeated Use of Zung's Self-Rating Depression Scale among Women during Perinatal Period. *Journal of Psychosomatic*

- Obstetrics and Gynaecology*, 20, 112-117. <https://doi.org/10.3109/01674829909075584>
- Kitamura, T., Toda, M. A., Shima, S., & Sugawara, M. (1994a). Validity of the Repeated GHQ among Pregnant Women: A Study in a Japanese General Hospital. *International Journal of Psychiatry in Medicine*, 24, 149-156. <https://doi.org/10.2190%2FW6MU-XRC6-16KF-3Q88>
- Kitamura, T., Toda, M. A., Shima, S., & Sugawara, M. (1994b). Early Loss of Parents and Early Rearing Experience among Women with Antenatal Depression. *Journal of Psychosomatic Obstetrics and Gynaecology*, 15, 133-139. <https://doi.org/10.3109/01674829409025638>
- Kitamura, T., Shima, S., Sugawara, M., & Toda, M. A. (1994c). Temporal Variation of Validity of Self-Rating Questionnaires: Repeated Use of the General Health Questionnaire and Zung's Self-Rating Depression Scale among Women during Antenatal and Postnatal Periods. *Acta Psychiatrica Scandinavica*, 90, 446-450. <https://doi.org/10.1111/j.1600-0447.1994.tb01622.x>
- Kitamura, T., Toda, M. A., Shima, S., Sugawara, K., & Sugawara, M. (1998). Social Support and Pregnancy: II. Its Relationship with Depressive Symptoms among Japanese Women. *Psychiatry and Clinical Neurosciences*, 52, 37-45. <https://doi.org/10.1111/j.1440-1819.1998.tb00970.x>
- Kitamura, T., Yoshida, K., Okano, T., Kinoshita, K., Hayashi, M., Toyoda, N., Ito, M., Kudo, N., Tada, K., Kanazawa, K., Sakumoto, K., Satoh, S., Furukawa, T., & Nakano, H. (2006). Multicentre Prospective Study of Perinatal Depression in Japan: Incidence and Correlates. *Archives of Women's Mental Health*, 9, 121-130. <https://doi.org/10.1007/s00737-006-0122-3>
- Krause, J. S., Saunders, L. L., Bombardier, C., & Kalpakjian, C. (2011). Confirmatory Factor Analysis of the Patient Health Questionnaire-9: A Study of the Participants from the Spinal Cord Injury Model System. *Physical Medicine and Rehabilitation*, 3, 533-540. <https://doi.org/10.1016/j.pmrj.2011.03.003>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a Brief Depression Severity Measure. *Journal of General Internal Medicine*, 16, 606-613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Manea, L., Gilbody, S., & McMillan, D. (2012). Optimal Cut-Off Score for Diagnosing Depression with the Patient Health Questionnaire (PHQ-9): A Meta-Analysis. *Canadian Medical Association Journal*, 184, E191-E196. <https://doi.org/10.1503/cmaj.110829>
- Marcos-Nájera, R., Le, H.-M., Rodríguez-Muños, M. F., Crespo, M. E. O., & Mendez, N. I. (2018). The Structure of the Patient Health Questionnaire-9 in Pregnant Women in Spain. *Midwifery*, 62, 36-41. <https://doi.org/10.1016/j.midw.2018.03.011>
- Matthey, S., Souter, K., Mortimer, K., Stephens, C., & Sheridan-Magro, A. (2016). Routine Antenatal Maternal Screening for Current Mental Health: Evaluation of a Change in the Use of the Edinburgh Depression Scale in Clinical Practice. *Archives of Women's Mental Health*, 19, 367-372. <https://doi.org/10.1007/s00737-015-0570-8>
- Merz, E. L., Malcarne, V. L., Roesch, S. C., Riley, N., & Sadler, G. R. (2011). A Multigroup Confirmatory Factor Analysis of the Patient Health Questionnaire-9 among English- and Spanish-Speaking Latinas. *Cultural Diversity and Ethnic Minority Psychology*, 17, 309-316. <https://doi.org/10.1037/a0023883>
- Muramatsu, K., & Kamijima, K. (2009). Puraimarikea shinnryou to utubyou sukuri-ningu tsuru: Patient Health Questionnaire-9 Nihongoban 'Kokoroto Karadano Shitsumonyou' (Primary Care and Depression Screening Tool: The Japanese Version of the Patient Health Questionnaire-9 'Questionnaire of Mind and Body'). *Shindan to Chiryou*, 97, 1465-1473. (In Japanese)

- Radloff, L. S. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement, 1*, 385-401.
<https://doi.org/10.1177%2F014662167700100306>
- Reise, S. P. (2013). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research, 47*, 667-696.
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating Bifactor Models: Calculating and Interpreting Statistical Indices. *Psychological Methods, 21*, 137-150.
<https://doi.apa.org/doi/10.1037/met0000045>
- Romera, I., Delgado-Cohen, H., Prez, T., Caballero, L., & Gilaberte, I. (2008). Factor Analysis of the Zung Self-Rating Depression Scale in a Large Sample of Patients with Major Depressive Disorder in Primary Care. *BMC Psychiatry, 8*, Article No. 4.
<https://doi.org/10.1186/1471-244X-8-4>
- Salamero, M., Marcos, T., Gutierrez, F., & Rebull, E. (1994). Factorial Study of the BDI in Pregnant Women. *Psychological Medicine, 24*, 1031-1035.
<https://doi.org/10.1017/S0033291700029111>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research, 8*, 23-74.
- Sheehan, D.V. (1983) *The Anxiety Disease*. Scribner.
- Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research Diagnostic Criteria: Rationale and Reliability. *Archives of General Psychiatry, 35*, 773-782.
<https://doi.org/10.1001/archpsyc.1978.01770300115013>
- Spitzer, R. L., Kronke, K., Williams, J. B. W., & Patient Health Questionnaire Study Group (1999). Validity and Utility of a Self-Report Version of PRIME-MD: The PHQ Primary Care Study. *Journal of the American Medical Association, 282*, 1737-1744.
<https://doi.org/10.1001/jama.282.18.1737>
- Stuart, S., & Koleva, H. (2014). Psychological Treatment for Perinatal Depression. *Best Practice & Research Clinical Obstetrics and Gynaecology, 28*, 61-70.
<https://doi.org/10.1016/j.bpobgyn.2013.09.004>
- Stucky, B. D., & Edelen, M. O. (2014). Using Hierarchical IRT Models to Create Unidimensional Measures from Multidimensional Data. In S. P. Reise, & D. A. Revicki (Eds.), *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment* (pp. 183-206). Routledge.
- van de Vijver, F. J. R., & Leung, K. (2000). Methodological Issues in Psychological Research on Culture. *Journal of Cross-Cultural Psychology, 31*, 33-51.
<https://doi.org/10.1177%2F0022022100031001004>
- Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Method, 3*, 4-70.
<https://doi.org/10.1177%2F109442810031002>
- Wickberg, B., Tjus, T., & Hwang, P. (2005). Using the EPDS in Routine Antenatal Care in Sweden: A Naturalistic Study. *Journal of Reproductive and Infant Psychology, 23*, 33-41.
<https://doi.org/10.1080/02646830512331330956>
- Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factor Invariance within Longitudinal Structural Equation Models: Measuring the Same Construct across Time. *Child Development Perspectives, 4*, 10-18. <https://doi.org/10.1111/j.1750-8606.2009.00110.x>
- Wittkamp, K. A., Naeije, L., Schene, A. H., Huyser, J., & van Weert, H. C. (2007). Diagnostic Accuracy of the Mood Module of the Patient Health Questionnaire: A Systemat-

ic Review. *General Hospital Psychiatry*, 29, 388-395.

<https://doi.org/10.1016/j.genhosppsy.2007.06.004>

Woldetensay, Y. K., Belachew, T., Testaye, M., Spielman, K., Biesalski, H. K., Kantelhardi, E. J., & Scherbaum, V. (2018). Validation of the Patient Health Questionnaire (PHQ-9) as a Screening Tool for Depression in Pregnant Women: Afaan Oromo Version. *PLoS ONE*, 13, e0191782. <https://doi.org/10.1371/journal.pone.0191782>

Yoshida, T., Otsubo, T., Tsuchida, H., Wada, Y., Kamijima, K., & Fukui, K. (2004). Sheehan Disability Scale (SDISS) nihongoban no sakusei to shinraisei oyobi datousei no kentou (The Japanese Version of the Sheehan Disability Scale (SDISS): Development, Reliability and Validity). *Japanese Journal of Clinical Psychopharmacology*, 7, 1645-1653.

Zhong, Q., Gelaye, B., Rondon, M., Sánchez, S. E., García, P., Sánchez, E., Barrios, Y. V., Simon, G. E., Henderson, D. C., Cripe, S. M., & Williams, M. A. (2014). Comparative Performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for Screening Antepartum Depression. *Journal of Affective Disorders*, 162, 1-7. <https://doi.org/10.1016/j.jad.2014.03.028>

Zung, W. W. K. (1965). A Self-Rating Depression Scale. *Archives of General Psychiatry*, 12, 63-70. <https://doi.org/10.1001/archpsyc.1965.01720310065008>