Article

What Are The Core Symptoms of Antenatal Depression? A Study Using Patient Health Questionnaire-9 among Japanese Pregnant Women in the First Trimester

Toshinori Kitamura [1,2,3,4,*], Yuriko Usui [1,2,5], Mikiyo Wakamatsu [6], Mariko Minatani [7] and Ayako Hada [1,2,8]

[1] Kitamura Institute of Mental Health Tokyo, Tokyo 104-0061, Japan; yusui@g.ecc.u-tokyo.ac.jp (Y.U.); hada@institute-of-mental-health.jp (A.H.)
[2] Kitamura KOKORO Clinic Mental Health, Tokyo 151-3306, Japan
[3] T. and F. Kitamura Foundation for Studies and Skill Advancement in Mental Health, Tokyo 151-0063, Japan
[4] Department of Psychiatry, Graduate School of Medicine, Nagoya University, Nagoya 464-8601, Japan
[5] Department of Midwifery and Women's Health, Division of Health Sciences and Nursing, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan
[6] Department of Reproductive Health Care Nursing, School of Health Sciences, Faculty of Medicine, Kagoshima University, Kagoshima 890-8544, Japan; mikiwaka@health.nop.kagoshima-u.ac.jp
[7] Life Value Creation Unit, NTT DATA Institute of Management Consulting, Inc., Tokyo 102-0093, Japan; minatanim@nttdata-strategy.com
[8] Department of Community Mental Health & Law, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo 187-8551, Japan
[*] Correspondence: kitamura@institute-of-mental-health.jp

Abstract: Background: Depression is frequently seen among pregnant women. This is called antenatal depression (AND). **Aim:** Our aim was to identify clusters of AND and its core symptoms. **Methods:** The Patient Health Questionnaire-9 (PHQ-9), Pregnancy-Unique Quantification of Emesis and Nausea (PUQE-24), and Nausea and Vomiting of Pregnancy Quality of Life Questionnaire (NVP-QOL) were distributed to 382 pregnant women with a gestational age of 10 to 13 weeks who were attending antenatal clinics. The two PHQ-9 subscale scores were entered into a 2-step cluster analysis. The PHQ-9 items' capacity to identify AND were examined in terms of the area under curve (AUC) of a receiver operating characteristic (ROC) analysis. The selected symptom items were examined for their diagnostic capability in terms of the graded response model (GRM) in the item response theory (IRT) analysis. **Results:** Three clusters emerged. Cluster 3 scored highly in the scores of the two PHQ-9 subscales and the two emesis scales. In the ROC, five items showed an AUC > 0.80. The GRM identified four items with high information: 'loss of interest', 'depressed mood', 'self-esteem', and 'poor concentration'. **Conclusions:** The core symptoms of antenatal depression were four non-somatic symptoms; particularly, 'depressed mood' and 'loss of interest'. AND did not exist alone, but was accompanied by nausea and vomiting. Hence, we propose a new category: emesis–depression complex among pregnant women.

Keywords: antenatal depression; emesis; diagnostic entity

1. Introduction

Depression during pregnancy is a major health issue. The pooled prevalence estimate of antenatal depression is 21%, and that of major antenatal depression is 15% [1]. Usually, studies use a standard diagnostic criterion such as the Diagnostic and Statistical Manual of Mental Disorders (DSM). Such definitions of depressive episodes (depressive disorder) date back to Feighner et al.'s [2] St. Louis criteria, in which depression was identified when five or more items were identified out of the following nine (at least one of the first two was indispensable): depressed mood, lack of interest, appetite problems, sleep problems, psychomotor symptoms, fatigue, reduced self-esteem, reduced concentration,

and suicidality. These criteria have stood the test of history, remaining in the Research Diagnostic Criteria [3] and DSM-III to the current DSM-V. However, the validity of the above set of diagnostic items has long been debatable. Even in 1984, Endicott [4] cast doubt on the application of the DSM-III criterion of a major depressive episode (MDE) among patients with cancer. Intuitively, we presume that somatic symptoms such as insomnia, anorexia, fatigue, and retardation among cancer patients are likely somatic (physiological) symptoms (particularly under chemotherapy) and should not be treated as MDE indicators. This was supported by a study using an item response theory analysis [5]. A similar argument was made by Cavanaugh [6], who reviewed the diagnostic validity of the DSM-IV MDE criteria and suggested modifications to the descriptions in the texts. In the same vein, some researchers have avoided somatic symptoms when creating an assessment inventory (e.g., the Hospital Anxiety and Depression Scale (HADS) [7] and Beck Depression Inventory (BDI) [8]). Expectant women often experience emesis/hyperemesis gravidarum (HG). Emesis/HG is one of the uncomfortable symptoms experienced by pregnant women. HG affects approximately 1.1% of pregnancies [9]. It is doubtful to apply the "five-out-of-nine rule" to depressive states experienced by pregnant women without modification. For example, Matthey and Ross-Hamid [10] interviewed 118 pregnant women, asking about symptoms of depression and anxiety together with an additional enquiry about whether the endorsement of a symptom was due to their physical changes in pregnancy. More than a half of them who met the criteria for MDE answered that a sufficient number of their symptoms were due to the normal physical changes of their pregnancy, so that they no longer met the criteria for a diagnosis of MDE. A shortcoming of Matthey and Ross-Hamid's [10] study was that they relied on the participants' opinion about whether symptoms experienced were due to the 'normal' process of pregnancy or mental illness. Thus, we are still uncertain about which symptoms are free from the effects of physical changes and which are 'real' (core) symptoms of MDE during pregnancy.

Several previous studies have reported an association between emesis/HG and depression or other mental illnesses [11–17], and it is clear that emesis/HG and depression coexist [18–21]. Even though a decrease or an increase in appetite and weight loss or gain due to emesis/HG and pregnancy-related sleep disturbances are also present in normal pregnant women, the diagnostic criteria for depression in non-pregnant women have automatically been applied to pregnant women. As a result, common discomforts of pregnancy may also be diagnosed as depression. This may have led to an inflation in the prevalence of antenatal depression (AND).

A possible means to disentangle this complicated symptomatology and diagnostic categorization is a cluster analysis accompanied by item response theory (IRT). A cluster analysis is a statistical method for the classification of similar objects into groups based on the observed values of several variables for each individual. Classification has always played an essential role in science [22]. A cluster analysis is useful to identify subgroups. It may be feasible to expect that expectant women with AND constitute a category that is discrete from those who are normal or with only nausea and vomiting (pure emesis/HG). There may appear to be a cluster of women who have somatic symptoms such as nausea and vomiting, but without the core symptoms of AND. We presume that there are symptoms that distinguish the AND cluster from both emesis/HG and healthy women. Such symptoms should be regarded as core indicators (symptoms) of AND. Therefore, after identifying a cluster of pregnant women with AND, we planned to examine the contribution of *each* diagnostic item (indicator/symptom) so that we could select the best set of symptom items to diagnose AND. Here, we used IRT. When measuring depression, the greater the amount of information obtained from the item score, the more useful the item score for distinguishing between people in terms of the severity of depression [23–25]. We expected that some items that were indicators of MDE among non-perinatal women might lose their diagnostic specificity when applied to expectant women.

2. Methods

2.1. Study Procedures and Participants

This was a secondary analysis of our previously reported epidemiological dataset [26–29]. Approximately 1500 pregnant women in the period of 10 to 13 weeks' gestational age 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 those (a) who were not fluent in Japanese, (b) who were aged under 20, (c) who had eating disorders, (d) who had symptoms of vaginal bleeding or abdominal pain, (e) who had subchorionic hematomas, or (f) who had experiences of recurrent miscarriages. We administered a set of questionnaires on 2 occasions, 1 week apart. The total sample consisted of 382 pregnant women (approximately 25% of those who were solicited) who participated in this study. Participation was voluntary and anonymity was assured. The mean (SD) age of the participants was 31.9 (4.9) years old and that of their partners was 33.5 (5.5) years old. Most of them were married (94%); 44% of the participants were nulliparas and 55% were multiparas. The recruitment of this study was conducted from January 2017 to May 2019.

2.2. Measurements

For major depressive episode (MDE) and depression symptoms, we used the Patient Health Questionnaire-9 (PHQ-9) [30], a self-reported nine-item depression scale based on the criteria of MDE of DSM-IV. Each item checks the frequency of depressive symptoms over the previous two weeks. Each item is calculated with a 4-point Likert scale from 0 to 3. The psychometric properties' validity and reliability of the Japanese version of PHQ-9 were reported [31,32]. We already reported a result of a factor analysis of the PHQ-9 items among current expectant women that yielded two subscales, somatic and non-somatic [26]. The somatic subscale included the items 'sleep', 'fatigue', and 'appetite', whereas the non-somatic subscale included the items 'loss of interest', 'depressed mood', 'self-esteem', 'poor concentration', 'psychomotor', and 'suicidality'. In addition, we used the indicator of MDE according to Spitzer et al.'s [30] criteria. This was a dichotomous measure of MDE based on the DSM-IV diagnostic criteria. Here, each DSM item was rated as 'present' if it was answered more than 'more than half the days' and, as the MDE criterion, 'present' items at least were required to diagnose MDE. If the number of 'present' items was less than 5, the case was judged as non-MDE.

The nausea and vomiting symptoms were rated by two scales, the Pregnancy-Unique Quantification of Emesis and Nausea (PUQE-24) [33] and the Nausea and Vomiting of Pregnancy Quality of Life Questionnaire (NVP-QOL) [34]. PUQE-24 is a scale that rates (a) nausea (the length of nausea in hours for the last 24 h), (b) vomiting (number of vomiting episodes in the last 24 h), and (c) retching (the number of retching episodes in the last 24 h), each with a 5-point scale. Hada et. al. [27] validated its psychometric properties in Japanese. NVP-QOL is a questionnaire with 30 items to evaluate nausea and vomiting of pregnancy (NVP) and related QOL in the previous week with a 7-point Likert scale (1 = none of the time to 7 = all of the time). Its 1-factor structure was identified in Japan [29]. In addition, we created an ad hoc item, 'nausea therapy', which rated outpatient visits or inpatient admissions for emesis/HG (1 = neither, 2 = outpatient visit, and 3 = inpatient admission).

For social disability, we used the Sheehan Disability Scale (SDS (Copyright 1983–2020, Sheehan, D.V. All rights reserved. May be reproduced only with the permission of Dr David V. Sheehan, copyright holder. For permission, contact davidsheehan@gmail.com)), which contains three items that are simple questions to assess disability covering (a) work and schoolwork, (b) social and leisure activities, and (c) family life and home responsibilities [35]. The validation of the Japanese version was reported [28].

For insomnia, we used the Insomnia Severity Index (ISI) [36,37], which is a questionnaire to assess the nature, severity, and impact of insomnia by asking about sleep for the last two weeks and consists of seven items with a 5-point Likert-type scale (from 0 = no problem to 4 = very severe problem). The 2-factor structure of the Japanese version

of the ISI consists of Severity (severity of sleep onset, sleep maintenance, early morning awakening problems, and sleep satisfaction) and Impact (interference in sleep difficulties with daytime functioning, noticeability of sleep problems by others, and distress caused by the sleep difficulties) [38].

2.3. Data Analysis

First, we calculated the mean, SD, skewness, and kurtosis for the scores of somatic and non-somatic subscales of PHQ-9. Four missing cases in PHQ-9 were treated with a listwise deletion. Then, a 2-step cluster analysis was conducted to classify cases depending on the two PHQ-9 subscale scores. One-way ANOVAs with Tukey post hoc comparisons were conducted to compare sociodemographic and other characteristics. After the identification of clusters, the screening performance of *each* item of PHQ-9 was estimated as an area under curve (AUC) by a receiver operating characteristic (ROC) curve [39].

The identification of the PHQ-9 items that had a substantial capacity to detect AND cases led to IRT analyses of such items. The graded response model (GRM) [40,41] was fitted with item response theory (IRT) to confirm the item discrimination parameters and item information parameters for latent AND traits. Item information curves (IICs) were also graphed. IICs show the amount of information about an ability level that an item score provides at each point on the ability scale [24]. As we will discuss later, non-somatic symptoms were more reflective of AND than somatic symptoms. Item 9, 'suicidality', was included as a non-somatic factor; however, its AUC was low (probably due to its very low basic rate). Although 'suicidality' is an extremely important symptom in depression, it appears only in particularly severe cases. 'Suicidality' should be treated with another independent domain as an important task in perinatal mental health. Therefore, we excluded item 9, 'suicidality', as well as somatic items from IRT to identify the core symptoms of AND.

Before fitting the GRM, the IRT assumptions were examined. The 1-factor model of the non-somatic domain was fitted to the data by a confirmatory factor analysis (CFA). An exploratory factor analysis (EFA) was performed if the CFA showed a poor fit [42,43]. The 1-factor model was assured if the greatest factor accounted more than 20% of the variance as well as more than 4 of the ratio of the first and second factors' eigenvalues (PROPMIS Standard 2.0). Local independence was examined by residual correlations from the 1-factor CFA where any item pairs with >0.2 residual correlation were flagged [42,44]. Monotonicity was evaluated using scalability coefficients (coef_*Hi*) (>0.3) [44,45] using the Mokken scale.

Analyses were performed by SPSS® ver.28 (IBM. Inc., Armonk, NY, USA). In addition, we used R Studio V.07.2-576. EFAs were conducted using psych version 2.1.3 [46], residual correlations were calculated using lavaan version 0.6-11 [47], monotonicity was examined using mokken version 3.0.6 [48], and the GRM was examined via ltm version 1.2-0' [49] packages in R.

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

Both somatic and non-somatic subscale scores showed skewness < 2.0 and kurtosis < 4.0, thus suggesting a normal distribution (Table 1). The 2-step cluster analysis using the two PHQ-9 subscale scores revealed 3 clusters. Its silhouette measure of cluster cohesion was >0.5. Clusters 1, 2, and 3 consisted of 135, 144, and 98 cases, respectively (Table 2). Both somatic and non-somatic subscale scores were statistically significantly higher in the order of Clusters 3, 2, and 1. The prevalence of MDE (according to the algorithm of Spitzer et al. [30]) was 0%, 0.7%, and 46% among Clusters 1, 2, and 3, respectively. A noteworthy fact was that the cases of Cluster 3 were more likely to reflect non-somatic symptoms

than Cluster 2 (see Figure 1). Hence, Cluster 3 was characterized by a predominance of non-somatic items as well as the presence of MDE compared with the other two clusters.

Table 1. Mean, SD, skewness, and kurtosis for the subscales of PHQ-9 ($n = 378$).

Subscales	Mean	SD	Skewness	Kurtosis
Somatic	7.63	2.55	0.80	−0.99
Non-somatic	8.79	3.18	1.40	1.60

Table 2. Cluster characteristics.

Cluster	Cluster 1 ($n = 135$)	Cluster 2 ($n = 144$)	Cluster 3 ($n = 98$)	F	Post Hoc Comparison
Somatic subscale (z score)	−2.92	3.64	5.63	382.9 ***	1 < 2 < 3
Non-somatic subscale (z score)	−2.53	0.96	9.79	458.3 ***	1 < 2 < 3
MDE, *n* (%)	0 (0.0)	1 (0.7)	45 (45.9)	140.6 ***	−
PUQE-24 (z score)	2.51	4.13	5.23	30.8 ***	1 < 2 < 3
NVP-QOL (z score)	73.59	105.10	137.76	88.8 ***	1 < 2 < 3
Nausea therapy (z score)	−0.20	−0.16	0.03	1.7 ^{NS}	−
SDS (z score)	1.99	6.90	12.98	88.1 ***	1 < 2 < 3
ISI severity (z score)	3.02	6.53	8.46	63.5 ***	1 < 2 < 3
ISI impact (z score)	1.19	3.95	4.68	38.0 ***	1 < 2 < 3
Own age	25.48	24.75	25.84	0.8 ^{NS}	−
Partner's age	27.33	28.88	28.17	3.7 *	1, 3 < 2
Nulliparae, *n* (%)	59 (44.0)	66 (46.5)	41 (41.8)	$\chi^2 = 0.52$ ^{NS}	−
Multiparae, *n* (%)	75 (56.0)	76 (53.5)	57 (58.2)		−

* $p < 0.05$; *** $p < 0.01$; ^{NS}, not significant; ISI: Insomnia Severity Index; MDE: major depressive episode; NVP-QOL: Nausea and Vomiting of Pregnancy Quality of Life Questionnaire; PUQE-24: Pregnancy-Unique Quantification of Emesis and Nausea; SDS: Sheehan Disability Scale.
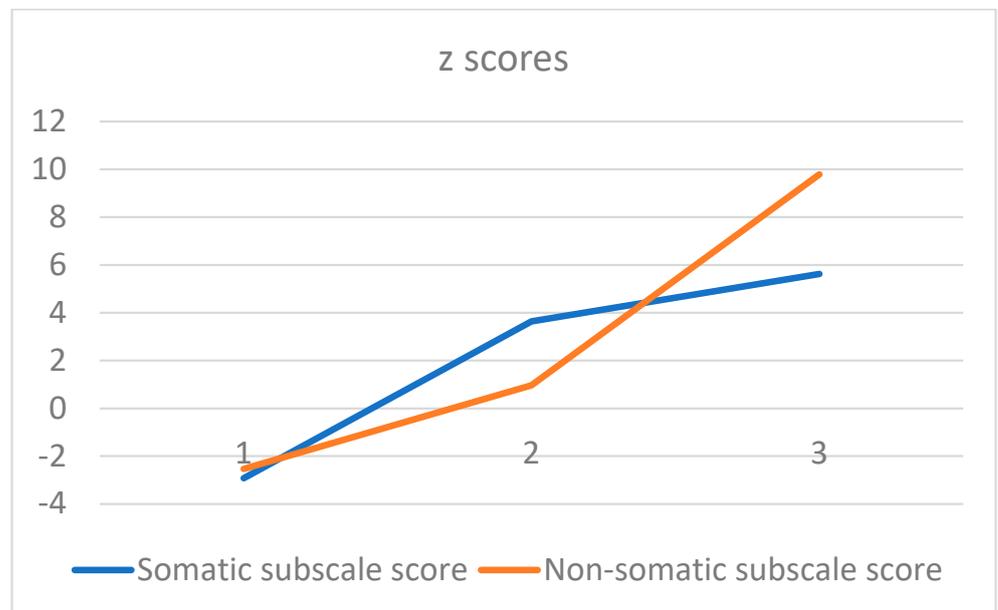

Figure 1. Somatic subscale and non-somatic subscale scores (z score) for each cluster.

The scores of PUQE-24 and NVP-QOL were again statistically significantly higher in the order of Clusters 3, 2, and 1 (Table 2). Although not statistically significant, the nausea therapy score showed a similar trend across the three clusters. Cluster 2 was, therefore, characterized by nausea and vomiting and PHQ-9 somatic subscale scores compared with Cluster 1. Therefore, we considered that those women in Cluster 1 were normal, those

in Cluster 2 had nausea and vomiting, and those in Cluster 3 had AND together with nausea and vomiting. The last group may have indicated the existence of nausea/vomiting and depression. We failed to identify a cluster characterized only by depression (with an absence of nausea and vomiting).

The scores of the SDS and ISI subscales were again statistically significantly higher in the order of Clusters 3, 2, and 1 (Table 2). There was no significant difference among the cases belonging to all clusters in terms of own age and parity. Partner's age was statistically higher among Cluster 2, but its difference was slight.

The screening performances of *each* item of PHQ-9 for Cluster 3 (against Clusters 1 and 2 combined) were estimated (Table 3) in terms of the AUC of the ROC. 'Loss of interest', 'depressed mood', 'self-esteem', and 'poor concentration' showed an AUC > 0.8. These all belonged to the non-somatic subscale. Of the somatic subscale items, only 'fatigue' had an AUC > 0.8. However, fatigue is a very common symptom for pregnant women. The mean score of 'fatigue' was the highest among the women belonging to Clusters 1 or 2 combined (Table 3). Therefore, we identified the above four non-somatic symptoms (together with one more non-somatic item, 'psychomotor') as a latent trait of core symptoms of AND, proceeding to an item analysis using IRT.

Table 3. Mean SD and AUC of ROC (Cluster 3 vs. the others).

No.	Item	Clusters 1 + 2	Cluster 3	AUC
1	Loss of interest	1.42 (0.58)	2.84 (0.92)	0.883
2	Depressed mood	1.31 (0.49)	2.47 (0.88)	0.858
3	Sleep	2.24 (1.06)	3.16 (0.89)	0.735
4	Fatigue	2.25 (0.86)	3.32 (0.73)	0.810
5	Appetite	2.36 (1.00)	3.34 (0.81)	0.761
6	Self-esteem	1.16 (0.41)	2.39 (0.95)	0.858
7	Poor concentration	1.23 (0.49)	2.64 (0.88)	0.896
8	Psychomotor	1.10 (0.35)	1.78 (0.96)	0.701
9	Suicidality	1.01 (0.09)	1.12 (0.36)	0.553

The 1-factor model consisting of the above five items did not show an acceptable fit with the data in the CFA (CFI = 0.929; RMSEA = 0.155) (Supplementary Table S1). An EFA was conducted for the examination of unidimensionality. The scree test suggested a 1-factor model. Residual correlations for each item ranged between −0.069 and 0.088 (<0.2), and coefficient *H* ranged between 0.399 and 0.584 (>0.3) (Supplementary Table S2). The IRT assumptions became assured, and the GRM was fitted to the data. The item discrimination parameter was higher in the order of 'loss of interest' (2.722), 'depressed mood' (2.714), 'self-esteem' (2.365), 'poor concentration' (2.171), and 'psychomotor'(1.315). Item information was higher in the order of 'depressed mood' (6.37), 'loss of interest' (6.34), 'self-esteem' (5.18), 'poor concentration' (4.70), and 'psychomotor' (2.52) (Table 4). The item information curve of 'psychomotor' showed a substantially lower curve than the other items (Figure 2). Thus, 'psychomotor' did not have a sufficient ability to measure the AND latent trait. Hence, we identified the other four items as core symptoms of AND. As expected, 'loss of interest' and 'depressed mood' were the two most reflecting symptoms of AND.

Table 4. Threshold parameters, discrimination parameters, and item information of GRM.

PHQ-9		b1	b2	b3	a	Information
1	Loss of interest	−0.093	1.060	1.711	2.722	6.34
2	Depressed mood	0.085	1.345	1.974	2.714	6.37
6	Self-esteem	0.572	1.387	2.285	2.365	5.18
7	Poor concentration	0.379	1.215	2.200	2.171	4.70
8	Psychomotor	1.468	2.436	3.724	1.315	2.52

PHQ-9: Patient Health Questionnaire-9; b: item threshold parameter; a: item discrimination parameter.

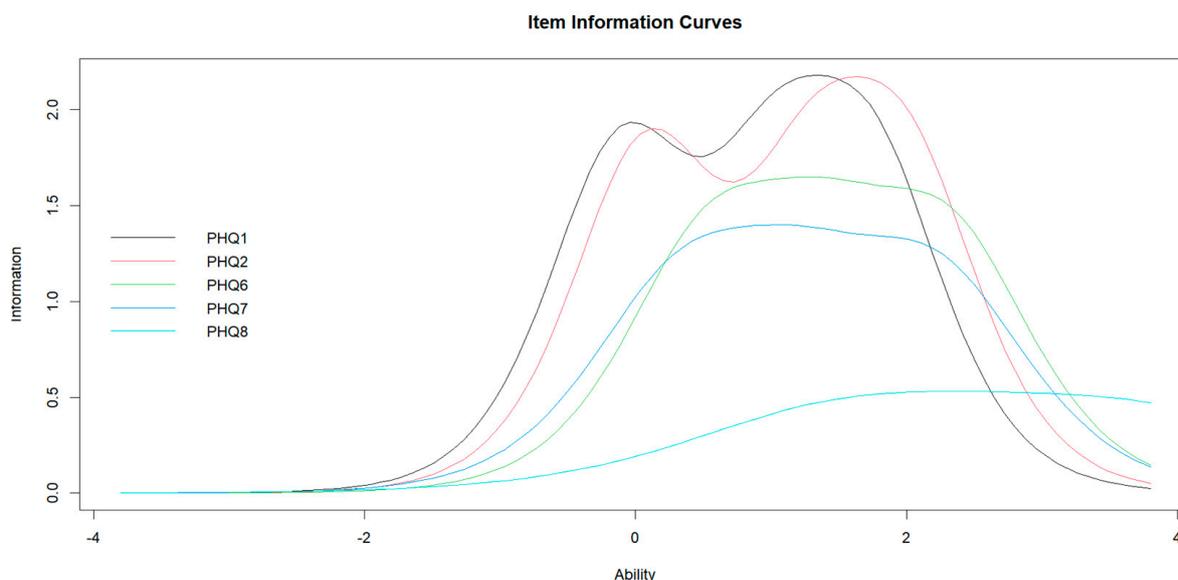**Item Information Curves**

Figure 2. Item information curves of GRM for *non-somatic* items. PHQ: Patient Health Questionnaire.

4. Discussion

In this study, a cluster analysis derived three clusters. Cluster 3 consisted of about a quarter of the sample of women, and was the highest in both the PHQ-9 somatic and non-somatic subscale scores. When the DSM MDE diagnostic criterion (the 5-out-of-9 rule) was strictly applied, the prevalence of MDE was 45% among the women of Cluster 3 compared with virtually naught of those of Clusters 1 and 2. The Cluster 3 women also scored the highest in social disability (SDS) and insomnia (ISI). The SDS scores in this group were higher than those of individuals without mental disorders [50,51]. These findings suggest that women in this cluster experienced daily life difficulty, thus this cluster was that of AND. The fact that only half of the women in this cluster met the MDE criterion indicated that these women included cases of not only MDE but also minor and intermittent depression [19]. The boundary of AND identified by the cluster analysis was wider than that defined by DSM-V. It is also noteworthy that the Cluster 3 women scored the highest in terms of nausea and vomiting. As there appeared to be no cluster characterized only by depressive symptoms without signs of nausea and vomiting, we considered that Cluster 3 women reflected a condition of both AND and emesis. Hence, we named this Cluster 'emesis–depression complex'. In 1993, Kitamura et al. [19] interviewed expectant women, of whom 16% were diagnosed as suffering from depression and reported that such women were more likely to complain about nausea.

The women of Cluster 2 scored higher than Cluster 1 women in terms of the scores of the somatic and non-somatic subscales, but only 2% of them met the MDE criterion. These women scored higher than Cluster 1 women in terms of nausea and vomiting measures (PUQE-24 and NVP-QOL). Hence, we named Cluster 2 'nausea and vomiting'. They scored lower than Cluster 3, but higher than Cluster 1 in terms of the PHQ-9 subscales. This was in line with past investigations that women with emesis/HG had higher depression scores [14,16,18,20]. These past studies were, however, based on instruments of depression severity measures and did not use diagnosis-specific measures (such as PHQ-9) or structured diagnostic interviews, thus we are unaware of the proportion of AND women in these studies.

An issue of clinical importance is the distinction between emesis and HG. Our Cluster 3 may have included cases of HG. We were uncertain about this issue because we did not measure ketonuria or electrolytes. However, a differentiation between emesis and HG is unclear, even among obstetricians. For example, Koot et al. [52], in a systematic review, examined all the RCTs of HG (defined by researchers) and checked the definition

of HG. From 34 reports, components of a definition were vomiting (34; 100%), nausea (30; 88%), weight loss (9; 26%), dehydration (7; 21%), inability to tolerate oral food/water intake (4; 12%), ketonuria or acidosis (19; 56%), electrolyte disturbances (5; 15%), and a need for hospitalization (17; 50%). This suggests a substantial variation in the definition of HG. Therefore, it is obvious that obstetricians have different diagnostic categories for emesis and HG and there are, to the best of our knowledge, no empirical studies (such as cluster analyses and taxometrics) about the distinction of emesis and HG. Future studies are recommended to measure psychological as well as somatic symptoms simultaneously and identify diagnostic categories.

The AUC showed a screening performance capacity of each depression item. The AND women differentiated from normal and nausea and vomiting women (and normal women) by 'loss of interest', 'depressed mood', 'reduced self-esteem', and 'poor concentration'. This was endorsed by IRT analyses, demonstrating that these four items were core symptoms of AND. Among those symptoms, 'loss of interest' and 'depressed mood' were the most reflecting symptoms of AND. This was in line with Huey et al. [53] who identified core symptoms of MDE among palliative care patients.

Pregnant women with AND may be erroneously treated for emesis or HG in perinatal settings, but it may be necessary to carefully assess them and to aggressively provide psychological care for *depressive* symptoms in addition to treatments for nausea and vomiting. It may be useful to make therapeutic plans based on the possible causal variables described in the vast amount of literature on correlates of AND. Kitamura et al. [54] listed as correlates of AND (1) obstetric factors (first pregnancy, first childbirth, and past history of artificial abortion), (2) early experiences such as a loss of father, (3) personality, including Eysenck's high neuroticism, (4) negative attitudes towards the current pregnancy, (5) poor accommodation such as non-detached housing and expected crowdedness after childbirth, and (6) a lack of social support such as a low level of partner intimacy. Of particular note is that Salomonsson [55] proposed a psychodynamic treatment for pregnancy-related mental disorders. Evans et al.'s [56] systematic review suggested that women with antenatal anxiety perceived a benefit from peer support and individual discussions of their situation.

Our results suggested a new set of symptoms for defining depression. Although the current 5-out-of-9 convention dates back to half a century ago, there have been debates on its validity. For example, Angst and Dobler-Mikola [57], in their epidemiological study on a young adult population, noted that the gender ratio of depression was higher among women when applying DSM-III or RDC, but equal between the two genders when relying only on the existence of dysphoric mood (joyless, depressive, sick of life, loss of interest, loss of efficiency, and feelings of inferiority) and work disability. This was because women with a dysphoric mood reported much greater numbers of depression-related symptoms (such as appetite, weight, sleep, and psychomotor symptoms) than men with dysphoria. Accordingly, Angst and Dobler-Mikola [58] proposed that the number of required depression-related symptoms should be different between the two genders; 3 and 5 for men and women, respectively. Unfortunately, they did not perform IRT analyses on their data. However, their studies suggested that we should be cautious about the definition of the boundary of *clinical* depression. When reporting their monumental paper, Feighner et al. [2] wrote that "This communication is meant to provide common ground for different research groups so that diagnostic definitions can be amended constructively as further studies are completed". They further added that "These criteria are not intended as final for any illness". Our data indicated that 'depressed mood' and 'loss of interest' were the two core symptoms of depression that potently differentiated cases of depression from others, even when colored by nausea and vomiting.

This study had several limitations. First, we only studied pregnant women in the first trimester. Although the worst time for emesis/HG is in the first trimester, pregnant women with emesis also exist in the second and third trimesters. As many pregnant women with emesis/HG through the second and third trimesters have mental health problems, future studies should identify the core symptoms of depression throughout the

pregnancy period. Second, the low response rates limited the generalizability of the results. Pregnant women with more severe emesis/HG or depression may not have been able to answer the questions. Third, our sole reliance on questionnaire data should be amended by interview data. This, however, requires a structured interview specially designed for a pregnant woman population. We also became aware of shortcomings of PHQ-9. Its somatic items lump diametrically opposed content into one item, such as insomnia/hypersomnia, decreased/increased appetite, and psychomotor agitation/retardation. If each direction was asked separately, the results might have been different. Of course, as our sample was all Japanese women, we should be sensitive to cultural or linguistic differences. Cross-cultural studies should be performed colleting populations from multiple countries. This was, however, beyond the scope of the current study, and awaits future international studies.

5. Conclusions

Our study showed that the core symptoms of AND were four non-somatic symptoms, particularly 'depressed mood' and 'loss of interest'. It also demonstrated that AND did not exist alone, but was accompanied by nausea and vomiting in the first trimester. Hence, we propose a new category among pregnant women: emesis–depression complex.

References

1. Yin, X.; Sun, N.; Jiang, N.; Xu, X.; Gan, Y.; Zhang, J.; Qiu, L.; Yang, C.; Shi, X.; Chang, J.; et al. Prevalence and associated factors of antenatal depression: Symptomatic reviews and meta-analyses. *Clin. Psychol. Rev.* **2021**, *83*, 101932. [CrossRef] [PubMed]
2. Feighner, J.P.; Robins, E.; Guze, S.; Woodruff, R.A.; Winokur, G.; Minoz, R. Diagnostic criteria for use in psychiatric research. *Arch. Gen. Psychiatry* **1972**, *26*, 57–63. [CrossRef] [PubMed]
3. Spitzer, R.L.; Endicott, J.; Robins, E. *Research Diagnostic Criteria (RDC) for a Selected Group of Functional Disorders*, 3rd ed.; New York State Psychiatric Institute: New York, NY, USA, 1978.
4. Endicott, J. Measurement of depression in patients with cancer. *Cancer* **1984**, *53*, 2243–2248. [CrossRef] [PubMed]
5. Akechi, T.; Ietsugu, T.; Sukigara, M.; Okamura, H.; Nakano, T.; Akizuki, N.; Okamura, M.; Shimizu, K.; Okuyama, T.; Furukawa, T.A.; et al. Symptoms indicator of severity of depression in cancer patients: A comparison of the DSM-IV criteria with alternative diagnostic criteria. *Gen. Hosp. Psychiatry* **2009**, *31*, 225–232. [CrossRef] [PubMed]
6. von Ammon Cavanaugh, S. Depression in the medically ill: Critical issues in diagnostic assessment. *Psychosomatics* **1983**, *36*, 48–59. [CrossRef]
7. Zigmond, A.S.; Snaith, R.P. The hospital anxiety and depression scale. *Acta Psychiatr.* **1983**, *67*, 361–370. [CrossRef]
8. Beck, A.; Ward, C.; Mendelson, M.; Mock, J.; Erbaugh, J. An inventory for measuring depression. *Arch. Gen. Psychiatry* **1961**, *4*, 561–571. [CrossRef]

9.　Einarson, T.R.; Piwko, C.; Koren, G. Quantifying the global rates of nausea and vomiting of pregnancy: A meta-analysis. *J. Popul. Ther. Clin. Pharmacol.* **2013**, *20*, e171–e183.

10.　Matthey, S.; Ross-Hamid, C. The validity of DSM symptoms for depression and anxiety disorders during pregnancy. *J. Affect. Disord.* **2011**, *133*, 546–552. [CrossRef]

11.　Aksoy, H.; Aksoy, U.; Karadağ, Ō.I.; Hacimusalar, Y.; Açmaz, G.; Aykut, G.; Çağli, F.; Yücel, B.; Aydin, T.; Babayiğit, A. Depression levels in patients with hyperemesis gravidarum: A prospective case-control study. *SpringerPlus* **2015**, *4*, 34. [CrossRef]

12.　Fell, D.B.; Dodds, L.; Joseph, K.S.; Allen, V.M.; Butler, B. Risk factors for hyperemesis gravidarum requiring hospital admission during pregnancy. *Obstet. Gynecol.* **2006**, *107*, 277–284. [CrossRef] [PubMed]

13.　Kjeldgaard, H.K.; Eberhard-Gran, M.; Benth, J.Š.; Nordeng, H.; Vikanes, V.Å. History of depression and risk of hyperemesis gravidarum: A population-based cohort study. *Arch. Women's Ment. Health* **2017**, *20*, 397–404. [CrossRef] [PubMed]

14.　Kjeldgaard, H.K.; Vikanes, Å.; Benth, J.Š.; Junge, C.; Garthus-Niegel, S.; Eberhard-Gran, M. The association between the degree of nausea in pregnancy and subsequent posttraumatic stress. *Arch. Women's Ment. Health* **2019**, *22*, 493–501. [CrossRef] [PubMed]

15.　Meltzer-Brody, S.; Maegbaek, M.L.; Medland, S.E.; Miller, W.C.; Sullivan, P.; Munk-Olsen, T. Obstetrical, pregnancy and socio-economic predictors for new-onset severe postpartum psychiatric disorders in primiparous women. *Psychol. Med.* **2017**, *47*, 1427–1441. [CrossRef] [PubMed]

16.　Pirimoglu, Z.M.; Guzelmeric, K.; Alpay, B.; Balcik, O.; Unal, O.; Turan, M.C. Psychological factors of hyperemesis gravidarum by using the SCL-90-R questionnaire. *Clin. Exp. Obstet. Gynecol.* **2009**, *37*, 56–59.

17.　Seng, J.S.; Schrot, J.A.; van de Ven, C.; Liberzon, I. Service use data analysis or pre-pregnancy psychiatric and somatic diagnosis in women with hyperemesis gravidarum. *J. Psychosom. Obstet. Gynaecol.* **2007**, *28*, 209–217. [CrossRef]

18.　Hizli, D.; Kamalak, Z.; Kosus, A.; Kosus, N.; Akkurt, G. Hyperemesis gravidarum and depression in pregnancy: Is there an association? *J. Psychosom. Obstet. Gynaecol.* **2012**, *33*, 171–175. [CrossRef]

19.　Kitamura, T.; Shima, S.; Sugawara, M.; Toda, M. Psychological and social correlates of the onset of affective disorders among pregnant women. *Psychol. Med.* **1993**, *23*, 967–975. [CrossRef]

20.　McCormack, D.; Scott-Heyes, G.; McCusker, C.G. The impact of hyperemesis gravidarum on maternal mental health and maternal-fetal attachment. *J. Psychosom. Obstet. Gynaecol.* **2011**, *32*, 79–87. [CrossRef]

21.　Mitchell-Jones, N.; Gallos, I.; Farren, J.; Tobias, A.; Bottomley, C.; Bourne, T. Psychological morbidity associated with hyperemesis gravidarum: A systematic review and meta-analysis. *BJOG Int. J. Obstet. Gynaecol.* **2016**, *124*, 20–30. [CrossRef]

22.　Kaufman, L.; Rousseeuw, P. *Finding Groups in Data: An Introduction to Cluster Analysis*; Wiley: New York, NY, USA, 1990.

23.　Lord, F.M. *Applications of Item Response Theory to Practical Testing Problems*; Routledge: Abingdon, UK, 1980.

24.　Nugent, W.R. Understanding DIF and DTF: Description, methods, and implications for social work research. *J. Soc. Soc. Work Res.* **2017**, *8*, 305–334. [CrossRef]

25.　Samejima, F. Estimation of reliability coefficients using the test information function and its modifications. *Appl. Psychol. Meas.* **1994**, *18*, 229–244. [CrossRef]

26.　Wakamatsu, M.; Minatani, M.; Hada, A.; Kitamura, T. 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 J. Depress.* **2021**, *10*, 121–137. [CrossRef]

27.　Hada, A.; Minatani, M.; Wakamatsu, M.; Koren, G.; Kitamura, T. The pregnancy-unique quantification of emesis and nausea (PUQE-24): Configural, measurement, and structural invariance between nulliparas and multiparas and across two measurement time points. *Healthcare* **2021**, *9*, 1553. [CrossRef] [PubMed]

28.　Hada, A.; Minatani, M.; Wakamatsu, M.; Kitamura, T. Disability during early pregnancy: Using the Sheehan Disability Scale during the first trimester in Japan. *Healthcare* **2022**, *10*, 2514. [CrossRef] [PubMed]

29.　Yamada, F.; Kataoka, Y.; Minatani, M.; Hada, A.; Wakamatsu, M.; Kitamura, T. The NVP QOL Questionnaire: Psychometric properties of the self-report measure of health-related quality of life for nausea and vomiting during pregnancy. *PCN Rep.* **2022**, *1*, e21. [CrossRef]

30.　Spitzer, R.L.; Kronke, K.; Williams, J.B.W. Patient Health Questionnaire Study Group Validity and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA* **1999**, *282*, 1737–1744. [CrossRef]

31.　Inagaki, M.; Ohtsuki, T.; Yonemoto, N.; Kawashima, Y.; Saitoh, A.; Oikawa, Y.; Kurosawa, M.; Muramatsu, K.; Furukawa, T.A.; Yamada, M. 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. *Gen. Hosp. Psychiatry* **2013**, *35*, 592–597. [CrossRef]

32.　Muramatsu, K.; Kamijima, K. Puraimarikea shinnryou to utubyou sukuri-ningu tsuru: Patient Health Questionnaire-9 nihongoban 'Kokoroto Karadano Shitsumonhyou' (Primary care and depression screening tool: The Japanese version of the Patient health Questionnaire-9 'Questionnaire of Mind and Body'). *Shindan Chiryou* **2009**, *97*, 1465–1473. (In Japanese)

33.　Ebrahimi, N.; Maltepe, C.; Garcia-Bournissen, F.; Koren, G. Nausea and vomiting of pregnancy: Using the 24-hour pregnancy-unique quantification of emesis (PUQE-24) scale. *JOGC* **2009**, *31*, 803–807. [CrossRef]

34.　Magee, L.A.; Chandra, K.; Mazzotta, P.; Stewart, D.; Koren, G.; Guyatt, G.H. Development of a health-related quality of the instrument for nausea and vomiting of pregnancy. *Am. J. Obstet. Gynecol.* **2002**, *186*, S232–S238. [CrossRef] [PubMed]

35.　Sheehan, D.V. *The Anxiety Disease*; Scribner: New York, NY, USA, 1983.

36.　Bastien, C.H.; Vallières, A.; Morin, C.M. Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Med.* **2001**, *2*, 297–307. [CrossRef] [PubMed]

37. Morin, C.M.; Belleville, G.; Bélanger, L.; Ivers, H. The Insomnia Severity Index: Psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep* **2011**, *34*, 601–608. [CrossRef] [PubMed]

38. Shinohara, E.; Minatani, M.; Wakamatsu, M.; Hada, A.; Kitamura, T. The Insomnia Severity Index: Factor structure and measurement and structural invariance across perinatal measurement time points (manuscript preparation). *Healthcare* **2023**, *11*, 1194. [CrossRef]

39. Zwaig, M.H.; Campbell, G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin. Chem.* **1993**, *39*, 561–577. [CrossRef]

40. Samejima, F. Estimation of latent ability using a response pattern of graded scores 1. *ETS Res. Bull. Ser.* **1968**, i-169. [CrossRef]

41. Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Psychom. Monogr. Suppl.* **1969**, *34 Pt 2*, 100. [CrossRef]

42. Reeve, B.B.; Hays, R.D.; Bjorner, J.B.; Cook, K.F.; Crane, P.K.; Teresi, J.A.; Thissen, D.; Revicki, D.A.; Weiss, D.J.; Hambleton, R.K.; et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med. Care* **2007**, *45* (Suppl. S1), S22–S31. [CrossRef]

43. The Patient-Reported Outcome Measurement Information System; PROMIS® Instrument Development and Validation Scientific Standards Version 2.0, (Revised May 2013). 2013. Available online: https://www.mcgill.ca/can-pro-network/files/can-pro-network/promisstandards_vers2.0_final.pdf (accessed on 18 May 2023).

44. Rimehaug, S.A.; Kaat, A.J.; Nordvik, J.E.; Klokkerud, M.; Robinson, H.S. Psychometric properties of the PROMIS-57 questionnaire, Norwegian version. *Qual. Life Res.* **2022**, *31*, 269–280. [CrossRef]

45. Mokken, R.A. *Theory and Procedure of Scale Analysis: With Applications in Political Research*; De Gruyter Mouton: Berlin, Germany; New York, NY, USA, 2011. [CrossRef]

46. Revelle, W. *Psych: Procedures for Psychological, Psychometric, and Personality Research*, R package version 2.2.9; Northwestern University: Evanston, IL, USA, 2022.

47. Rosseel, Y. Lavaan: An R package for structural equation modeling. *J. Stat. Softw.* **2012**, *48*, 1–36. [CrossRef]

48. Van der Ark, L.A. Mokken scale analysis in R. *J. Stat. Softw.* **2007**, *20*, 1–19. [CrossRef]

49. Rizopoulos, D. ltm: An R package for latent variable modelling and item response theory analyses. *J. Stat. Softw.* **2006**, *17*, 1–25. [CrossRef]

50. Leu, S.H.; Chou, J.Y.; Lee, P.C.; Cheng, H.C.; Shao, W.C.; Hsien, W.L.; Huang, C.L.; Chen, V.C.H. Validity and reliability of the Chinese version of the Sheehan Disability Scale (SDS-C). *Asia-Pac. Psychiatry* **2015**, *7*, 215–222. [CrossRef] [PubMed]

51. Luciano, J.V.; Bertsch, J.; Salvador-Carulla, L.; Tomás, J.M.; Fernández, A.; Pinto-Meza, A.; Haro, J.M.; Palao, D.J.; Serrano-Blanco, A. Factor structure, internal consistency, and construct validity of the Sheehan Disability Scale in a Spanish primary care sample. *J. Eval. Clin. Pract.* **2010**, *16*, 895–901. [CrossRef] [PubMed]

52. Koot, M.H.; Boelig, R.C.; van't Hooft, J.; Limpens, J.; Rosebom, T.J.; Painter, R.C.; Grooten, I.J. Variation in hyperemesis gravidarum definition and outcome reporting in randomised clinical trials: A systematic review. *BJOG* **2018**, *125*, 1514–1521. [CrossRef] [PubMed]

53. Huey, N.S.; Guan, N.C.; Gill, J.S.; Hui, K.O.; Sulaiman, A.H.; Kangasundram, S. Core symptoms of major depressive disorder among palliative care patients. *Int. J. Environ.* **2018**, *15*, 1758. [CrossRef] [PubMed]

54. Kitamura, T.; Shima, S.; Sugawara, M.; Toda, M.A. Clinical and psychosocial correlates of antenatal depression: A review. *Psychother. Psychosom.* **1996**, *65*, 117–123. [CrossRef]

55. Salomonsson, B. *Psychodynamic Interventions in Pregnancy and Infancy: Clinical and Theoretical Perspectives*; Routledge: Abingdon, UK, 2018.

56. Evans, K.; Spildy, H.; Morrell, J.C. Non-pharmacological interventions to reduce the symptoms of mild to moderate anxiety in pregnant women: A systematic review and narrative synthesis of women's views on the acceptability of and satisfaction with interventions. *Arch. Women's Ment. Health* **2020**, *23*, 11–28. [CrossRef] [PubMed]

57. Angst, J.; Dobler-Mikola, A. Do the diagnostic criteria determine the sex ratio in depression? *J. Affect.* **1984**, *7*, 189–198. [CrossRef]

58. Angst, J.; Dobler-Mikola, A. The Zurich study III. Diagnosis of depression. *Eur. Arch. Psychiatry Neurol. Sci.* **1984**, *234*, 30–37. [CrossRef]