Research Article

Predictive Validity of the Postpartum Depression Predictors Inventory-Revised

Ji-Hyang Youn, RN, MSN, Ihn Sook Jeong, RN, PhD *

College of Nursing, Pusan National University, Yangsan, South Korea

A R T I C L E   I N F O

Article history:
Received 7 October 2011
Received in revised form
18 November 2011
Accepted 22 November 2011

Keywords:
depression
predictive value of tests
postpartum period

S U M M A R Y

Purpose: This study aimed to investigate the predictive validity of three versions of the Postpartum Depression Predictors Inventory-Revised (PDPI-R) in Korea.

Methods: A descriptive cross-sectional design with a self-administered questionnaire, including 43 items of the PDPI-R, using the Edinburgh Postpartum Depression Scale as the gold standard was used. Data were collected from 316 women within 6 weeks after childbirth in Busan, Korea, from August to November 2010.

Results: The postpartum depression and postpartum depressive symptom (PDS) rate was 22.5%. The area under the curve of the receiver operating characteristic curve was .882 for the prenatal version of the PDPI-R and .927 for the full version. The sensitivity and specificity were 87.3% and 85.1%, respectively, at a cutoff point of 9.5 for the full version, and 91.5% and 66.1%, respectively, at a cutoff point of 5.5 for the prenatal version. The Hosmer-Lemeshow goodness-of-fit statistics was 3.554 (p = .829) for the prenatal version and 8.305 (p = .404) for the full version; this showed a good degree of correspondence between the estimated and observed probabilities of PDS. By age, education, and socioeconomic groups, the discrimination and calibration were generally good for both the prenatal and full versions.

Conclusion: The PDPI-R showed good predictive validity among women in Korea. It is recommended that the prenatal version of the PDPI-R be used to predict PDS for pregnant women and the full version of the PDPI-R be used for women during the postpartum period.

Copyright © 2011, Korean Society of Nursing Science. Published by Elsevier. All rights reserved.

Introduction

Postpartum depression or postpartum depressive symptom (PDS), a depression or depressive symptom that develops postnatally or starts before childbirth and continues after birth, contributes greatly to postnatal maternal suicide and is a serious problem in the field of health and medicine. PDS is one of the most universal health problems worldwide. According to systematic literature reviews, the prevalence of major depressive episodes is around 23% in Korean pregnant women (Ryu, Kim, & Lee, 2010), 6.5–12.9% until the first month after childbirth, 19.2% until the third month (Gavin et al., 2005). PDS may occur anytime during the first year after childbirth (Beck & Gable, 2001). However, according to a longitudinal study, it increases gradually until the sixth week after childbirth and then begins to decrease (Gjerdingen, Frober, Chaloner, & McGover, 1993).

This prevalence, however, is highly likely to be an underestimation. It is reported that around 50% of women who have experienced PDS do not recognize negative emotions arising from it nor do they report them to health practitioners (McGill, Burrows, Holland, Langer, & Sweet, 1995). Thus, essential prevention strategies are required to identify high-risk parturient women early by assessing the risk of PDS in pregnant women prenatally using a highly predictable tool (Austin & Lumley, 2003). Various scales have been developed in order to assess the risk of PDS in pregnant women. Representative tools developed relatively include the Vulnerable Personality Style Questionnaire (Boyce, Hickey, Gilchrist, & Talley, 2001), Postpartum Depression Predictors Inventory—Revised (Beck, 2001), Brisbane Postnatal Depression Index (Webster, Pritchard, Creedy, & East, 2003), Pregnancy Risk Questionnaire (Webster et al., 2003), and Contextual Assessment of Maternity Experience (Bernazzani et al., 2005).

All the PDS prediction tools mentioned above, excluding the Vulnerable Personality Style Questionnaire include both prenatal and postnatal factors as items related to PDS; thus, they may enable minimizing the condition via adequate management of prenatal factors during pregnancy. In particular, the Postpartum Depression Predictors Inventory-Revised (PDPI-R) includes all of the factors found to be significantly related to PDS via a systematic literature review. Thus, it covers relatively more diverse aspects of PDS compared to other tools which can minimize the omission of parturient women with a high risk of PDS.

* Correspondence to: Ihn Sook Jeong, RN, PhD, College of Nursing, Pusan National University, Beomeo-ri, Mulgeum-eup, Yangsan-si, Gyeongsangnam-do 626-870, South Korea.
E-mail address: jeongis@pusan.ac.kr (I.S. Jeong).

1976-1317/5 – see front matter Copyright © 2011, Korean Society of Nursing Science. Published by Elsevier. All rights reserved.
doi:10.1016/j.anr.2011.11.003
EPDS is used as the gold standard, it shows the highest sensitivity at a certain cutoff point. This study chose the PDPI-R as its PDS prediction tool, and evaluated its usability in Korea by testing its predictive validity.

On the other hand, in order for a PDS prediction tool to measure PDS accurately, it should satisfy three factors: discrimination, calibration, and uniformity of fit (Zimmerman et al., 1998). Discrimination is the ability to distinguish those with PDS from those without it, and is evaluated with the area under the curve (AUC) of the receiver operating characteristic (ROC) curve (Hanley & McNeil, 1983), as well as with validity indicators such as sensitivity, specificity, and predictability, and the correct classification rate. Calibration is a scale of agreement between PDS estimated by the prediction tool and the actual PDS, and is evaluated by analyzing the agreement between the number of predicted PDS women and the number of observed PDS women, standardization rate, and results of the Hosmer-Lemeshow goodness-of-fit (H-L fit) test (Hosmer & Lemeshow, 1989, pp. 82–134). Uniformity of fit means uniformity in the predictive validity of the PDS prediction tool among subgroups, and is evaluated by measuring discrimination and calibration according to the subjects' general characteristics (Zimmerman et al.).

The aim of this study was to test the predictive validity of the PDS prediction tool PDPI-R, using the EPDS as the gold standard in parturient women 3–6 weeks after childbirth and who visited gynecology clinics or women's hospitals in the Busan Metropolitan city. The specific objectives of this study were (a) to survey the participants' general characteristics and their PDS levels, (b) to measure the discrimination of PDPI-R, (c) to measure the calibration of PDPI-R, and (d) to measure the uniformity of fit of PDPI-R according to the participants' general characteristics.

**Methods**

**Study participants**

Depression occurring 4 weeks after giving birth was defined as PDS by the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV; American Psychiatric Association, 1994). Due to its higher sensitivity and specificity, the EPDS test was recommended to be administered within 6 weeks after birth (Cox, Holden, & Sagovsky, 1987). Therefore, the participants of this study were 316 women who had delivered not more than 6 weeks before and were recruited from women's hospitals, or postpartum care facilities in Busan, Korea.

**Instruments**

**PDPI-R**

The PDPI-R, a screening tool for women at risk for developing PDS, developed from the findings of an updated meta-analysis of the risk factors of PDS (Beck, 2001, 2002; Beck, Records, & Rice, 2006). The PDPI-R consisted of 13 predictors of PDS: 10 prenatal factors and 3 postnatal factors. The prenatal factors were marital status, socioeconomic status, self-esteem, prenatal depression, prenatal anxiety, unwanted/unplanned pregnancy, history of previous depression, social support from husband/mate, family and friends, marital satisfaction, and life stresses (e.g., marital changes, occupational changes, crises). Postnatal factors included child care stress, infant temperament, and maternity blues. The prenatal version of the PDPI-R, which had a scale ranging from 0 to 32, was used during pregnancy. The full version of PDPI-R including both prenatal and postnatal factors, which had a scale ranging from 0 to 39, was used after delivery. The higher the score, the more risk factors a woman had for developing PDS (Beck et al., 2006).

Although the PDPI-R was designed to be administered via an interview conducted by a clinician (Beck & Gable, 2001), after its successful use in Australian women it could be used as a self-report questionnaire (Hanna, Jarman, Savage, & Layton, 2004). One of the strengths of the PDPI-R was that it could assess both the prenatal and postnatal risk for developing PDS, while most screening instruments for developing PDS were devised to be used only during pregnancy. The risk factors identified in the tool could also be used to develop intervention programs.

For this study, one of the researchers translated the instrument into Korean after getting permission from the developer (C. T. Beck), and a nursing researcher bilingual in Korean and English translated it into English. Coincidence between translation and back-translation was very high because the items of PDPI-R were short and simple. Next, we submitted the translated version of PDPI-R to a nursing professor and three physicians specializing in maternal care and familiar with the PDS lexical and content validation. Nursing professor and three physicians showed almost 100% of content validity index and asked us to make minor modifications such as typo correction or smooth expression without change of meaning. We also conducted a pilot test with five women who were within 6 weeks after delivery for the feasibility of the translated version of PDPI-R. The participants of pilot test asked to divide family support into the husband's and wife's sides of the family, in keeping with the Korean culture. We revised the instrument accordingly. Therefore, the prenatal version of the PDPI-R had 38 items with 10 risk factors from 0 to 3, and the final version of the PDPI-R had 45 items with 13 risk factors ranging from 0 to 45.

**EPDS**

In this study, the level of PDS was assessed using the EPDS (Cox et al., 1987), which was one of the most widely used instruments for approximating PDS (Boyd, Le, & Somberg, 2005). The EPDS consisted of 10 items for assessing a woman's experience during the previous 7 days, and not just how she felt on the day of the survey. Each item was measured on a 4-point scale with a range of 0–3 (resulting range, 0–30); the higher the score, the higher the level of PDS. The EPDS had been validated for use either in pregnancy (Milgrom, Ericksen, Negri, & Gemmill, 2005) or postpartum. In Korea, the EPDS had been translated and validated with Korean women (Han, Kim, & Park, 2004; Kim, Hur, Kim, Oh, & Shin, 2008). Han et al. (2004) showed a 100% sensitivity, a 78.4% specificity, and a 45.8% correct classification rate at a 9.5 cutoff point for both minor and major PDS, and a 100% sensitivity, a 90.5% specificity, and a 75.0% correct classification rate at a 12.5 cutoff point for major PDS only. Kim et al. (2005) determined the optimal cutoff point of EPDS for 236 pregnant women in Korea, resulting in a 79.7% sensitivity, an 87.1% specificity, and a .818 AUC of the ROC curve at 9.5 (9/10), the best cutoff point. According to the previous findings of Korean studies, we used 9.5 (9/10) as the cutoff point for the indication of probable PDS and divided the participants into a postpartum group (≥ 9.5) and a nonpostpartum group (<9.5). We used the Korean version of EPDS validated by Kim et al. (2005), and the internal consistency measured with Cronbach's alpha was .851 in this study.

**Data collection**

Data were collected from August to November 2010 after receiving an ethics approval from the Institutional Review Board, which is affiliated with the school to which the researchers belong. At first, we made a list of the hospitals specializing in women's care, and postpartum care facilities located in the Busan area, and contacted the directors of the nursing departments of each of these institutions to get permission for the study. Finally, we selected 3
institutions with a consideration of the number of delivery and outpatient visitors within a month among the 10 institutions that agreed to join this study. One of these institutions had 200–250 deliveries per month, the other had 100–150 deliveries per month, and another facility gave postpartum care to an average of 40–80 women.

With permission from the directors of the nursing departments, one researcher visited the two hospitals to distribute and collect the questionnaires, and mailed the questionnaires, with return envelopes, to the director of the postpartum care facility. The director of the nursing department of the care facility was responsible for distributing and retrieving the questionnaires.

Data analysis

Collected data were coded and analyzed using SPSS version 18.0 for Windows (SPSS Inc., Chicago, IL, USA). All statistical tests were performed a .05 level of significance for two-tailed tests.

The general characteristics of the participants were analyzed with means and standard deviations for continuous variables, and frequencies and percentages for categorical variables. The means and standard deviations of the PDPI-R and EPDS scores were obtained in total and by characteristics. The probabilities of PDS expected by the PDPI-R were compared with those observed by the EPDS (EPDS≥9.5). The relative risks for PDS by general characteristics were presented with odds ratios (ORs) and 95% confidence intervals (CIs).

The discriminative power of the PDPI-R was measured with the AUC of the ROC which was a measure of the overall discriminatory power of the prognostic model in distinguishing those who had PDS from those who did not (Hanley & McNeil, 1983). Next, we calculated the sensitivity (true positive), specificity (true negative), positive predictive value (1-false positive), and negative predictive value (1-false negative) at certain cutoff points with the ROC curve.

The H-L test statistic in the logistic regression model was used to evaluate the calibration of the PDPI-R (Hosmer & Lemeshow, 1989, pp. 82–134). Participants were divided into approximately 10 groups of roughly the same size based on the percentiles of the estimated probabilities. This goodness-of-fit statistic has a chi-square distribution and the discrepancies between the observed and expected number of observations in these groups were summarized by the Pearson chi-square statistic and p value (Hosmer & Lemeshow). Percentage agreement and kappa between the observed and expected number of PDSS were obtained. Standardized mortality ratios (SMR) were obtained by dividing the observed number by the expected number of PDSS for each category, and 95% CI were calculated.

For evaluating the uniformity of fit in different subgroups, we stratified patients by age group (<30 yr vs. ≥30 yr), education level (high school and below vs. college and above), and perceived socioeconomic status (low vs. middle or high level). The discrimination and calibration of the PDPI-R were reanalyzed in each subgroup.

Results

General characteristics of the participants

The average age of the participants was 31.5 years old, 77.2% graduated from college and above, and 85.1% reported that they were of middle or high level perceived socioeconomic status (Table 1).

Distribution of the PDPI-R and EPDS scores

The distribution of the PDPI-R and EPDS scores and their relationship with the general characteristics were summarized in Table 1. The average PDPI-R score was 6.3 in the prenatal version and 8.3 in the full version. The average PDPI-R score of the full version was related to age (p = .040), education level (p = .004), perceived socioeconomic status (p < .001), and weeks after delivery (p = .027), while the average PDPI-R score of the full version was

Table 1

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>n (%)</th>
<th>PDPI-R (M ± SD)</th>
<th>EPDS (M ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Prenatal version</td>
<td>Full version</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.3 ± 5.0</td>
<td>8.3 ± 5.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.4 ± 4.6</td>
<td>22.5</td>
</tr>
<tr>
<td>Age (yr)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;30 (M ± SD = 31.5 ± 3.6)</td>
<td>95 (30.1)</td>
<td>7.3 ± 6.1</td>
<td>9.3 ± 6.9</td>
</tr>
<tr>
<td></td>
<td>221 (69.9)</td>
<td>5.8 ± 4.3</td>
<td>7.9 ± 4.6</td>
</tr>
<tr>
<td>Education level</td>
<td>p</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤High school</td>
<td>0.04</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>&gt;College f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤High school</td>
<td>72 (22.8)</td>
<td>8.0 ± 6.0</td>
<td>10.4 ± 6.6</td>
</tr>
<tr>
<td>&gt;College f</td>
<td>244 (77.2)</td>
<td>5.8 ± 4.5</td>
<td>7.7 ± 4.9</td>
</tr>
<tr>
<td>Perceived SES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low level</td>
<td>47 (14.9)</td>
<td>9.7 ± 6.0</td>
<td>12.2 ± 6.6</td>
</tr>
<tr>
<td>High or middle level</td>
<td>269 (85.1)</td>
<td>5.7 ± 4.5</td>
<td>7.7 ± 4.9</td>
</tr>
<tr>
<td>Time after delivery (wk)</td>
<td>p</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>83 (26.3)</td>
<td>7.4 ± 5.0</td>
<td>8.8 ± 5.4</td>
</tr>
<tr>
<td>4</td>
<td>73 (23.1)</td>
<td>6.7 ± 4.7</td>
<td>9.2 ± 5.4</td>
</tr>
<tr>
<td>6</td>
<td>92 (29.1)</td>
<td>5.6 ± 5.0</td>
<td>7.9 ± 5.3</td>
</tr>
<tr>
<td>Tukey test</td>
<td>3 wk &gt; 6 wk (p = .047)</td>
<td>4 wk &gt; 6 wk (p = .22)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Tukey method was used for multiple comparisons in “Time after delivery”. SES = socioeconomic status.
related to education level \((p = .002)\) and perceived socioeconomic status \((p < .001)\).

The average EPDS score was 6.4 (21.3 based on 100 points), and was related to education level \((p < .001)\), perceived socioeconomic status \((p < .001)\), and weeks after delivery \((p = .019)\). In total, 22.5% of the participants showed a 9.5 or higher score, and 10.8% received 12.5 or higher on the EPDS. PDS was significantly higher for participants under 30 years of age \((OR = 1.80, 95\% CI = 1.21–2.70)\), with a high school and below education level \((OR = 1.73, 95\% CI = 1.14–2.62)\), and with a lower perceived socioeconomic status \((OR = 1.54, 95\% CI = 1.17–2.02)\) than that of their counterparts. PDS was significantly higher at 3 or 4 weeks after delivery than at 5 or 6 weeks after delivery (Table 1).

**Discussion**

The discrimination findings of the prenatal version and full version of the PDPI-R were shown in Table 2 and Fig. 1. The AUC of the ROC curve for the prenatal version of the PDPI-R was .882 (95\% CI = .838–.925), while the sensitivity, specificity, positive predictability, negative predictability, and correct classification rate were 91.5%, 66.1%, 43.9%, 96.4%, and 71.8%, respectively, at a 5.5 (5/6) cutoff point.

The AUC of the ROC curve for the full version of the PDPI-R was .927 (95\% CI = .893–.961), while the sensitivity, specificity, positive predictability, negative predictability, and correct classification rate were 90.1%, 82.0%, 59.2%, 96.6%, and 83.8%, respectively, at a 9.5 (9/10) cutoff point.

**Calibration**

The calibration findings of the prenatal and full versions of the PDPI-R were shown in Table 3.

The value (chi-square) of the HL fit statistic was 3.554 \((p = .829)\) of the prenatal version of the PDPI-R and 8.305 \((p = .404)\) for the full version of the PDPI-R. Because the null hypothesis of the HL fit test was that the model fits well, these results mean that the expected number of PDSs from this PDPI-R equation and the observed full version of the PDPI-R. Considering the number of observed and expected PDSs, the prenatal version of the PDPI-R underestimated PDS in participants over 30 years of age, and the full version of the PDPI-R underestimated PDS in participants less than 30 years of age (Table 4).

**Uniformity of fit**

In participants stratified according to age group, education level, and socioeconomic status, discrimination was generally good in all subgroups with a range from .863 to .947 in the AUCs of the ROC curve, while the chi-square of the H-L fit statistic showed a good fit for all subgroups. Most percentage agreement was over 80%, except among those participants less than 30 years old, for the prenatal version of the PDPI-R. The value of the H-L fit statistic was 3.554 \((p = .829)\) of the prenatal version of the PDPI-R and 8.305 \((p = .404)\) for the full version of the PDPI-R. Because the null hypothesis of the HL fit test was that the model fits well, these results mean that the expected number of PDSs from this PDPI-R equation and the observed full version of the PDPI-R. The AUC of the ROC curve for the full version of the PDPI-R was .927 (95\% CI = .893–.961), while the sensitivity, specificity, positive predictability, negative predictability, and correct classification rate were 90.1%, 82.0%, 59.2%, 96.6%, and 83.8%, respectively, at a 9.5 (9/10) cutoff point.

**Calibration**

The calibration findings of the prenatal and full versions of the PDPI-R were shown in Table 3.

The value (chi-square) of the HL fit statistic was 3.554 \((p = .829)\) of the prenatal version of the PDPI-R and 8.305 \((p = .404)\) for the full version of the PDPI-R. Because the null hypothesis of the HL fit test was that the model fits well, these results mean that the expected number of PDSs from this PDPI-R equation and the observed number of PDSs from the EPDS were similar for each category.

The percent agreement between the expected and observed number of PDSs was 85.1% for the prenatal version of the PDPI-R, and 88.0% for the full version of the PDPI-R.

The SMR, which was obtained by dividing the observed number by the expected number of PDSs for each category of the PDPI-R score, was .96 (95\% CI = 0.79–1.12) for the version of the PDPI-R and .99 (95\% CI = 0.72–1.25) for the full version of the PDPI-R.

**Uniformity of fit**

In participants stratified according to age group, education level, and socioeconomic status, discrimination was generally good in all subgroups with a range from .863 to .947 in the AUCs of the ROC curve, while the chi-square of the H-L fit statistic showed a good fit for all subgroups. Most percentage agreement was over 80%, except among those participants less than 30 years old, for the prenatal version of the PDPI-R. The value of the H-L fit statistic was 3.554 \((p = .829)\) of the prenatal version of the PDPI-R and 8.305 \((p = .404)\) for the full version of the PDPI-R. Because the null hypothesis of the HL fit test was that the model fits well, these results mean that the expected number of PDSs from this PDPI-R equation and the observed number of PDSs from the EPDS were similar for each category.

The percent agreement between the expected and observed number of PDSs was 85.1% for the prenatal version of the PDPI-R, and 88.0% for the full version of the PDPI-R.

The SMR, which was obtained by dividing the observed number by the expected number of PDSs for each category of the PDPI-R score, was .96 (95\% CI = 0.79–1.12) for the version of the PDPI-R and .99 (95\% CI = 0.72–1.25) for the full version of the PDPI-R.

**Discussion**

This study tested the predictive validity of the PDS prediction tool PDPI-R, using the EPDS as the gold standard, with 316 parturient women 3–6 weeks after childbirth.

According to the results of measuring with the EPDS, 22.5% of the participants received a score of 9.5 or higher, while 10.8% received 12.5 or higher. In their systematic literature review of theses published between 1980 and 2004, Gavin et al. (2005) reported that the prevalence of major depressive episodes until 3 months after childbirth was 19.2%; a study in Jordan reported that 22% of women at 6–8 weeks postpartum received a score of 13 or higher (Mohammad, Gamble, & Creedy, 2010); and a study in the United States reported that 13% of women at 4–8 weeks postpartum received a score of 10 or higher (Horowitz, Murphy, Gregory, & Wojcik, 2011). While all of these studies measured PDS using the EPDS, they are not comparable with one another.

![Fig. 1. Receiver operating characteristic curve: Postpartum Depression Predictors Inventory-Revised (PDPI-R) on the Edinburgh Postnatal Depression Scale (EPDS). Note. AUC = area under the curve.](image-url)
because they measured EPDS at different period after childbirth and used different cutoff points. Cox et al. (1987), who developed the tool, suggested 12.5 as a cutoff point, but studies in Western Europe used various cutoff points ranging from 9.5 to 12.5. In contrast, a study by Ryu et al. (2010) showed that 23% of women during the postpartum period were regarded as having PDS, similar to our finding at the same cutoff point. Considering that Korean studies such as those by Han et al. (2004) and Kim et al. (2008) suggested 9/10 as the cutoff point while Japanese (Yoshida et al., 1997) and Chinese ones (Lee et al., 1998) suggested 10.5 and 9.5, respectively, it is adequate to use 9/10 as the standard cutoff in future research in Korea.

In order to test predictive validity in this study, we evaluated the discrimination, calibration, and uniformity of fit calibration according to the general characteristics. Discrimination is the ability to distinguish those with PDS from those without, and it is generally measured with the AUC of the ROC curve (Hanley & McNeil, 1983). The AUC is evaluated to be “noninformative (AUC = .5), less accurate (.5 < AUC ≤ .7), moderately accurate (.7 < AUC ≤ .9), highly accurate (.9 < AUC < 1.0), or perfect (AUC = 1)” (Greiner, Pfeiffer, & Smith, 2000). In our study, the full and prenatal versions of the PDPI-R had ROC AUCs of .927 and .882, respectively, showing relatively substantial discrimination.

The cutoff point of the PDPI-R was estimated through ROC analysis. Because it is very important for a PDS prediction tool to identify parturient women with a high risk of PDS early, it is desirable to raise sensitivity as high as possible while maintaining a level of specificity so that all pregnant women suspected to have the risk of PDS can be detected. Sensitivity is the proportion of subjects who had PDS in the EPDS that was expected correctly by model of the PDPI-R score; specificity is the proportion of subjects who did not have PDS in the EPDS that was expected correctly by model of the PDPI-R score. In terms of this criterion, the optimal cutoff point of the PDPI-R based on the results of this study were 5.5 for the prenatal version and 9.5 for the full version. In a study by Beck et al. (2006), who developed the inventory, the prenatal version of the PDPI-R explained 67% of the variance of PDS and its sensitivity, specificity, and ROC AUC was 76%, 54%, and .673 at a cutoff point of 10.5 using the EPDS as the gold standard PDS, respectively (Beck et al., 2006). We cannot explain the exact reason of the big difference in cutting score between the two studies. However, the study of Beck et al. (2006) showed “less accuracy” in AUC and lower level of specificity than that of our study, so we recommend 5.5 as a cutoff point of prenatal version of PDPI-R and further studies to confirm the cutoff point of PDPI-R. In the prospective study by Oppo et al. (2009), in which interviews using the DSM-IV were used as the gold standard, the prenatal version showed a sensitivity of 76% and a specificity of 71% at a cutoff of 3.5, while the full version showed a sensitivity and specificity of 82% and 75%, respectively, at a cutoff point of 5.5, demonstrating large differences in its cutoff points from the results of Beck et al. (2006). The possible reason for the differences in studies is the tool used for the measurement of PDS. Our study and the Beck et al. (2006) study determined PDS by measuring the EPDS, but Oppo et al. (2009) determined whether a subject belonged to the category of DSM-IV by consulting a psychiatrist or trained psychological counselor if the score of the EPDS was 13 or higher. Beck et al. (2006) also determined whether a subject belonged to the category of DSM-IV by consulting a psychiatric nurse for subjects with a high EPDS score. In other words, these previous studies used different criteria for the final diagnosis of PDS.

Because it is very important for it is very important for a PDS prediction tool to identify parturient women with a high risk of PDS early, it is desirable to raise sensitivity as high as possible while maintaining a level of specificity so that all pregnant women suspected to have the risk of PDS can be detected. Sensitivity is the proportion of subjects who had PDS in the EPDS that was expected correctly by model of the PDPI-R score; specificity is the proportion of subjects who did not have PDS in the EPDS that was expected correctly by model of the PDPI-R score. In terms of this criterion, the optimal cutoff point of the PDPI-R based on the results of this study were 5.5 for the prenatal version and 9.5 for the full version. In a study by Beck et al. (2006), who developed the inventory, the prenatal version of the PDPI-R explained 67% of the variance of PDS and its sensitivity, specificity, and ROC AUC was 76%, 54%, and .673 at a cutoff point of 10.5 using the EPDS as the gold standard PDS, respectively (Beck et al., 2006). We cannot explain the exact reason of the big difference in cutting score between the two studies. However, the study of Beck et al. (2006) showed “less accuracy” in AUC and lower level of specificity than that of our study, so we recommend 5.5 as a cutoff point of prenatal version of PDPI-R and further studies to confirm the cutoff point of PDPI-R. In the prospective study by Oppo et al. (2009), in which interviews using the DSM-IV were used as the gold standard, the prenatal version showed a sensitivity of 76% and a specificity of 71% at a cutoff of 3.5, while the full version showed a sensitivity and specificity of 82% and 75%, respectively, at a cutoff point of 5.5, demonstrating large differences in its cutoff points from the results of Beck et al. (2006). The possible reason for the differences in studies is the tool used for the measurement of PDS. Our study and the Beck et al. (2006) study determined PDS by measuring the EPDS, but Oppo et al. (2009) determined whether a subject belonged to the category of DSM-IV by consulting a psychiatrist or trained psychological counselor if the score of the EPDS was 13 or higher. Beck et al. (2006) also determined whether a subject belonged to the category of DSM-IV by consulting a psychiatric nurse for subjects with a high EPDS score. In other words, these previous studies used different criteria for the final diagnosis of PDS. Calibration is a measure of agreement between the number of PDS women predicted by a PDS prediction tool and the number of actual PDS women. In this study, all three versions of the PDPI-R showed high calibration. In the results of the H-L fit test using a logistic regression model, all three versions were fit for the model.

### Table 3

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Hosner-Lemeshow goodness-of-fit statistics</th>
<th>Observed (O) and expected (E) PDS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>χ²</td>
<td>p</td>
</tr>
<tr>
<td>PDPI-R prenatal version</td>
<td>3.554</td>
<td>.829</td>
</tr>
<tr>
<td>PDPI-R full version</td>
<td>8.305</td>
<td>.404</td>
</tr>
</tbody>
</table>

Note. O/E – observed/expected; PDPI-R – Postpartum Depression Predictors Inventory-Revised.

### Table 4

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Hosner-Lemeshow goodness-of-fit statistics</th>
<th>Observed and expected PDS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>χ²</td>
<td>p</td>
</tr>
<tr>
<td>PDPI-R prenatal version</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (yr) &lt;30</td>
<td>.869 (.797–.941)</td>
<td>4.458 .814</td>
</tr>
<tr>
<td>≥30</td>
<td>.890 (.833–.947)</td>
<td>5.809 .562</td>
</tr>
<tr>
<td>Education ≤High school</td>
<td>.892 (.821–.963)</td>
<td>5.817 .561</td>
</tr>
<tr>
<td>≥College</td>
<td>.872 (.579–.760)</td>
<td>3.208 .921</td>
</tr>
<tr>
<td>Perceived SES Low level</td>
<td>.901 (.818–.980)</td>
<td>2.663 .914</td>
</tr>
<tr>
<td>High or middle level</td>
<td>.863 (.807–.919)</td>
<td>4.911 .767</td>
</tr>
<tr>
<td>PDPI-R full version</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years) &lt;30</td>
<td>.926 (.873–.979)</td>
<td>3.595 .825</td>
</tr>
<tr>
<td>≥30</td>
<td>.928 (.885–.972)</td>
<td>9.487 .303</td>
</tr>
<tr>
<td>Education ≤High school</td>
<td>.947 (.897–.996)</td>
<td>6.976 .539</td>
</tr>
<tr>
<td>≥College</td>
<td>.916 (.871–.960)</td>
<td>3.911 .865</td>
</tr>
<tr>
<td>Perceived SES Low level</td>
<td>.922 (.844–.999)</td>
<td>3.552 .830</td>
</tr>
<tr>
<td>High or middle level</td>
<td>.918 (.875–.961)</td>
<td>5.528 .596</td>
</tr>
</tbody>
</table>

Note. AUC – Area under the curve; ROC – receiver operating characteristic; CI – confidence interval; O/E – observed/expected; PDS – postpartum depression or postpartum depressive symptom; SES – socioeconomic status.
That is, PDS predicted by the PDPI-R showed high agreement with actual PDS evaluated with the EPDS. In addition, the kappa value obtained by correcting accidental agreement in the calibration between predicted and observed PDS was a substantial .63 for all versions. In general, a kappa value of .21 or higher is considered fair, .41—.60 moderate, and .61—.80 substantial (Viera & Garrett, 2005).

In order to test the uniformity of the predictive validity of the PDS prediction tool according to subgroups, we assessed the discrimination calibration according to the participants’ general characteristics, such as age, education level, and socioeconomic level. The full version of the PDPI-R showed a substantial uniformity of discrimination among the subgroups according to the three general characteristics. The uniformity of calibration also showed a satisfactory level for all of the characteristics. The prenatal version of the PDPI-R showed a somewhat lower AUC of the ROC curve than the full version did, but it was still over .85, showing that its discrimination was acceptable. Its calibration was also satisfactory.

This study is meaningful in that it suggested the optimal cutoff point in using the PDPI-R and the clinical utility of the tool, but it has a number of limitations to be considered while interpreting the results. First, because the study used self-reported responses, the data may not be accurate. Second, because it surveyed both prenatal and postnatal factors only postpartum, there may be a recall bias.

Conclusion

The PDPI-R showed good predictive validity among women in Korea. Therefore, we recommend that the prenatal version of the PDPI-R be used to predict PDS for pregnant women and the full version of the PDPI-R be used for women during the postpartum period. The recommended cutoff points were 5.5 (5/6) and 9.5 (9/10) for prenatal version of the PDPI-R and full version of the PDPI-R, respectively. We also recommend further study to verify these cutoff points of two instruments in a representative sample in Korea because our study was performed in the limited sample in one area of Korea.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

We would like to thank the directors of the three study institutions as well as the participants for their time, and Dr. C.T. Beck for giving his permission to use the PDPI-R questionnaire for this study.

References