Journal of Midwifery & Women's Health www.jmwh.org
Research Article

A Brief Depression Screening Tool for Perinatal Clinical Practice: The Performance of the PHQ-2 Compared with the PHQ-9

Antonella Gigantesco[1], PsyD, Gabriella Palumbo[1], PsyD, Loredana Cena[2], PsyD, Laura Camoni[1], PsyD, Alice Trainini[2], PsyD, Alberto Stefana[3], PsyD, Fiorino Mirabella[1], PsyD

Introduction: There is ongoing interest in using brief screening instruments to identify perinatal depression in clinical practice. One ultra-brief screening instrument for depression is the Patient Health Questionnaire-2 (PHQ-2), but thus far its accuracy in perinatal clinical practice has been barely researched. In the present study, we aimed to assess the screening accuracy of the PHQ-2 against the Patient Health Questionnaire-9 (PHQ-9) in a large sample of perinatal women.

Method: A total of 1155 consecutive women attending 11 health care centers throughout Italy completed the PHQ-9 (which includes the PHQ-2) during pregnancy (27-40 weeks) or postpartum (1-13 weeks). Sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratio positive, likelihood ratio negative, and overall accuracy were calculated using cut points 3 or greater and 2 or greater.

Results: During pregnancy, PHQ-2 greater than or equal to 3 revealed low sensitivity (38.4%-44.7%) but high specificity (97.8%-99.3%). In postpartum, it revealed moderate sensitivity (56.9%-70.6%), high specificity (95.8%-99.8%), and fair overall accuracy in pregnancy (70%). The alternative threshold greater than or equal to 2 revealed very high sensitivity (pregnancy: 92.1%-95.2%; postpartum: 87.1%-95.2%), moderate specificity (pregnancy: 78.1%-83.2%; postpartum: 68.8%-81.1%) and good overall accuracy, both during pregnancy (87%) and postpartum (84%).

Discussion: The PHQ-2 provided acceptable accuracy for screening for depression compared with the PHQ-9. In perinatal screening practice, a threshold of 2 or greater should be preferred as this ensures high sensitivity, missing only approximately 6% to 8% of cases, and a false-positive rate (percentage of women classified as affected with depressive symptoms when they are not) of 19% to 25%.
J Midwifery Womens Health 2022;0:1–7

INTRODUCTION

Perinatal depression is a disabling condition that has been associated with poorer outcomes such as a reduction in a woman's ability to perform daily activities and parenting[1] and adverse consequences on the child's development.[2,3] Research studies have demonstrated that perinatal depression affects up to 20% of women,[4] with variability in prevalence across different geographical locations and populations[5] and higher prevalence among African American and Hispanic women compared with white women. Risk factors include low socioeconomic status, lack of social support, history of depression, prenatal depression or anxiety, stigma, and racial or ethnic disparity.[6] Although perinatal depression screening has increased in recent years, many women do not receive mental health treatment, with the lowest rates found for nonwhite women,[7] suggesting that race- and ethnicity-related vulnerability should be taken into consideration in identifying women at risk for perinatal depression.[8]

Screening in perinatal care settings that could be performed by health care staff including midwives and nurses has been proposed as a strategy for early detecting or launching treatment of perinatal depression because, typically, women with perinatal depression are more likely to seek care in these settings than in specialized mental health settings.[9,10]

To date, screening for depression is recommended for all women in the perinatal period by a number of organizations, including the US Preventive Services Task Force, the American College of Obstetricians and Gynecologists, the American Psychiatric Association, and the American Academy of Paediatrics.[11] Several screening instruments have been validated for use during pregnancy and the postpartum period to identify women who may benefit from further assessment. The Edinburgh Postnatal Depression Scale (EPDS)[12] is the most widely used in clinical practice and most studied instrument in perinatal populations.[13] Although the EPDS generally performs well in middle- and high-income countries,[14] a systematic review in low- and lower-middle income countries found that none of the studies achieved more than 80% on sensitivity and specificity.[15] The scale, which is the most widely translated perinatal screening instrument, consists of 10 self-reported questions including anxiety symptoms, which are a prominent feature of perinatal mental disorders, but excludes some cardinal or constitutional symptoms of depression such as anhedonia and somatic symptoms, which are captured only by indirect questions about depression.[16] The inclusion of constitutional symptoms common in pregnancy

[1]Center for Behavioural Sciences and Mental Health, National Institute of Health, Rome, Italy
[2]Department of Clinical and Experimental Sciences, Section of Neuroscience, Observatory of Perinatal Clinical Psychology, University of Brescia, Brescia, Italy
[3]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy

Correspondence
Antonella Gigantesco
Email: antonella.gigantesco@iss.it

ORCID
Antonella Gigantesco https://orcid.org/0000-0001-9262-7756

Quick Points

- Perinatal depression is a disabling condition that has been associated with poorer outcomes. Very brief screening that could be performed by midwives and nurses may be the best strategy for early detecting perinatal depression.
- One very brief screening instrument for depression is the Patient Health Questionnaire-2 (PHQ-2), but evidence of its accuracy in perinatal clinical practice is limited.
- In the present study, we assessed the screening accuracy of the PHQ-2 against the Patient Health Questionnaire-9 in a large sample of perinatal women.
- The findings showed that the PHQ-2 had high practicality and accuracy for screening for depression in primary or secondary care perinatal settings.
- The PHQ-2 may be a key first step in any perinatal depression screening and management program.

in other screening instruments, such as the Patient Health Questionnaire-9,[17,18] the Beck Depression Inventory, the Center for Epidemiologic Study Depression Scale, and the Zung Self Rating Depression Scale,[1,19] reduces their specificity for perinatal depression. With the exception of the EPDS and the Patient Health Questionnaire-9 (PHQ-9), the other mentioned instruments have from 20 to 35 questions and, therefore, require more time to complete and score. The PHQ-9 is the shortest among those instruments, with 9 items, including a suicidal ideation item, and it is used in perinatal settings.[20] The PHQ-9 has demonstrated good diagnostic operating characteristics as a screener for perinatal depression with both sensitivity and specificity greater than 0.80.[20] Moreover, the operating characteristics of the PHQ-9 have been assessed in many different samples, countries, and clinical settings for perinatal depression assessment, and this increases the generalizability of its results.[20]

Consistent with the recommendation of using a universal screening approach, there has been an increased interest in using very brief screening instruments to identify perinatal women with major depression in busy perinatal settings.[21] One such very brief screening instrument for depression is the Patient Health Questionnaire-2 (PHQ-2),[22] which consists of the first 2 items of the PHQ-9.[17] The 2 items investigate the core symptoms of depression (ie, depressed mood and anhedonia).

In settings such as primary care or some inpatient and outpatient specialty care, the PHQ-2 has been widely validated and found to be up to 87% sensitive and 77% specific using a threshold score of 2 or greater and up to 62% sensitive and 92% specific using a threshold score of 3 or greater, in studies that used fully structured interviews as reference standards.[23]

In contrast, evidence of PHQ-2 accuracy in perinatal clinical practice is limited.[24] To the best of our knowledge, only 2 studies have validated the PHQ-2 in perinatal settings comparing the PHQ-2 with a structured diagnostic interview. One of these studies showed moderate sensitivity and low specificity in pregnancy,[25] and the other relatively high postpartum sensitivity and specificity.[26] Apart from these studies, a limited number of other studies have been conducted using the EPDS as the reference standard, with mixed results, as sensitivity ranged from 19% to 93% and specificity ranged from 75% to 93%.[27–29]

Given the paucity and the inconsistency of reports, additional research in larger and different perinatal populations has been recommended to validate the PHQ-2 as part of a perinatal health care strategy to detect perinatal depression.[24]

In the present study, we aimed to (1) investigate the PHQ-2 accuracy for screening for depression in perinatal clinical practice, using the PHQ-9 as a reference standard, among a large sample of pregnant and postpartum women attending several primary or perinatal secondary care centers throughout Italy; and (2) determine the optimal threshold score of the PHQ-2 (between thresholds $\geq$2 and $\geq$3) for identifying possible cases of depression using the PHQ-9 score of 10 or greater as a proxy for a diagnosis of depression.

METHODS

Study Design and Participants

This study was a secondary analysis of a study to evaluate the prevalence of both antepartum and postpartum depression and anxiety in a sample of women in Italy.[30] A total of 1155 women were recruited (954 during pregnancy and 201 after birth) from 11 publicly funded primary or obstetrics-gynecology secondary care centers of the Observatory of Perinatal Clinical Psychology, University of Brescia, located throughout Italy. The study design was developed in mutual agreement of scientific collaboration between the University of Brescia Department of Clinical and Experimental Sciences and Observatory of Perinatal Clinical Psychology, and the Italian National Institute of Health. Information about the rationale and methodology of the larger study was detailed in the study protocol.[30] Participants were required to be able to speak and read Italian and to be pregnant or have a biological infant aged 6 months or younger. The exclusion criteria were having psychotic symptoms or substance use.

Data Collection and Questionnaires

All recruited women were required to complete the Italian version of the PHQ-9 questionnaire[18] once during the prenatal or postpartum period, depending on the characteristics of each health care center. The PHQ-9 is a 9-item self-report questionnaire designed to screen for depression in primary care and other specialty care settings.[23] It contains items about symptoms of depression such as anhedonia and low mood, in

Table 1. Definitions of Parameters for Evaluation of Screening Tests

Term	Definition
Sensitivity	The proportion of people with a condition who are correctly identified by a test as indeed having that condition.
Specificity	The proportion of people without a condition who are correctly identified by a test as indeed not having the condition.
Positive predictive value	The probability that people with a positive test result indeed do have the condition of interest.
Negative predictive value	The probability that people with a negative test result indeed do not have the condition of interest.
Likelihood ratio positive	Likelihood ratio for positive test result tells how much more likely the positive test result occurs in participants with the condition compared with those without the condition of interest and is usually >1 because it is more likely that the positive test result will occur in participants with the condition than in participants without the condition.
Likelihood ratio negative	Likelihood ratio negative tells us how much less likely the negative test result occurs in participants with the condition than in those without the condition and is usually <1 because it is less likely that a negative test result occurs in participants with than in participants without the condition of interest.

conjunction with problems concerning physical activity, appetite, concentration, energy, self-esteem, sleep, and suicidal ideation. Participants answer each item on a 4-point Likert scale (0 to 3), which indicates the frequency they have experienced the symptom during the past 2 weeks. The total score ranges from 0 to 27, with higher scores indicating more severe depressive symptoms. Participants also completed a demographic questionnaire that included age (years), marital status (married or cohabitating; single, separated, divorced, or widowed), educational level (elementary or middle school; secondary school; university), working status (student, homemaker, or unemployed; temporary employee; permanent employee), economic status (several problems; a few problems without specific difficulties; average to high status), and children living at the time of the current pregnancy/birth (yes or no).

Data Analysis

All statistical analyses were conducted using SPSS created for Windows, version 26.0. The PHQ-9 was considered the reference standard. The PHQ-2 was analyzed using total score at a threshold of 3 or more, as is usually recommended to define the result as positive.[22,31] In addition, the PHQ-2 was also analyzed using a total score of 2 or more. The PHQ-9 total score was transformed into a binary variable to indicate at least probable minor depression using the threshold score of 10 or more during pregnancy and postpartum, as recommended by the literature.[20] Prevalence of depression risk according to the PHQ-9 and positive PHQ-2 responses are presented as frequencies and percentages with 95% CIs based on the SE. The screening performance of the PHQ-2 was assessed using sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), likelihood ratio positive (LR+), and likelihood ratio negative (LR-)[32] (see Table 1). As a single summary descriptor of PHQ-2 accuracy, the balanced accuracy was also calculated [(sensitivity + specificity)/2].

In addition, the adequacy of the antenatal and postnatal sample sizes was evaluated post hoc, based on the prevalence of depression using the PHQ-9, the sensitivity and specificity of the PHQ-2 using a threshold of 2 or more and 3 or more, 80% power, and $\alpha = .05$.[33]

Ethical Consideration

The research was assessed and approved by the ethics committee of the Healthcare Centre of Bologna (registration number 77805, dated June 27, 2017). All women received information orally and in writing about the study's content and implications. If they were willing to participate, they signed the informed consent form.

RESULTS

Both groups of pregnant and postpartum participants were primarily in their thirties. Overall, the majority of them were married or lived with their partner and were well educated. Furthermore, the majority were employed in paid work, and only a few had serious economic difficulties (Table 2). All the women were of Italian origin and residing in Italy.

The percentage of antenatal and postnatal women with at least probable minor depression determined by the PHQ-9 was 6.8% (95% CI, 5.2%-8.4%) and 12.4% (95% CI, 7.9%-17.1%), respectively. These values were used to evaluate the adequacy of the antenatal and postnatal sample sizes. The estimated minimum sample size required ranged from between 236 (postnatal threshold ≥2) and 1328 (antenatal threshold ≥3), depending on the measured sensitivity and specificity of the PHQ-2.

Table 2. Sociodemographic Characteristics of Participants Screened with PHQ-9

Characteristic	Antenatal (from 27 to 40 wk) (N = 954)	Postnatal (from 1 to 13 wk) (N = 201)
Age		
18-29	209 (21.9)	30 (14.9)
30-35	453 (47.5)	79 (39.3)
>35	292 (30.6)	92 (45.8)
Marital status		
Married or cohabitating	878 (92.6)	183 (91.5)
Single, separated, divorced or widowed	70 (7.4)	17 (8.5)
Educational level		
Elementary or middle school	100 (10.5)	26 (13.0)
High school	341 (36.0)	83 (41.5)
University	507 (53.5)	91 (45.5)
Economic status[a]		
Several problems	58 (6.2)	16 (8.1)
A few problems without specific difficulties	433 (45.9)	99 (50.0)
Average to high status	452 (47.9)	83 (41.9)
Working status		
Student, homemaker, or unemployed	151 (16.0)	39 (19.6)
Temporary employee	89 (9.4)	20 (10.1)
Permanent employee	702 (74.6)	140 (70.4)
Children living at the time of this pregnancy/birth		
No	798 (83.6)	137 (68.2)
Yes	157 (16.4)	64 (31.8)

Abbreviation: PHQ-9, Patient Health Questionnaire 9.
[a] Several problems: having debts, difficulty or inability to pay daily expenses and rent; a few problems without specific difficulties: relatively modest standard of living but without particular difficulties; average high status: home owned, possibility of taking holidays or travelling for pleasure.

Table 3. Performance of the PHQ-2 at Different Thresholds Using the PHQ-9 as the Reference

	Antenatal (from 27 to 40 wk) (N = 954)		Postnatal (from 1 to 13 wk) (N = 201)	
PHQ-2	**Threshold ≥3 (n = 38)**	**Threshold ≥2 (n = 232)**	**Threshold ≥3 (n = 18)**	**Threshold ≥2 (n = 66)**
Sensitivity (95% CI)	41.5 (38.4-44.7)	93.8 (92.1-95.2)	64.0 (56.9-70.6)	92.0 (87.1-95.2)
Specificity (95% CI)	98.8 (97.8-99.3)	80.8 (78.1-83.2)	98.9 (95.8-99.8)	75.4 (68.8-81.1)
PPV (95% CI)	71.0 (68.0-73.9)	26.3 (23.5-29.2)	88.9 (83.5-92.7)	34.8 (28.4-41.9)
NPV (95% CI)	95.9 (94.3-97.0)	99.4 (98.7-99.8)	95.1 (90.8-97.5)	98.5 (95.3-99.6)
LR+ (95% CI)	34.6 (18.0-66.5)	4.9 (4.2-5.7)	58.2 (14.2-238.3)	3.7 (2.8-5.0)
LR- (95% CI)	0.59 (0.48-0.73)	0.1 (0.0-0.2)	0.36 (0.22-0.61)	0.1 (0.0-0.4)

Abbreviations: PHQ, Patient Health Questionnaire-2; PPV, positive predictive value; NPV, negative predictive value; LR+, likelihood ratio for a positive result; LR-, likelihood ratio for a negative result.

The positive rate determined using the PHQ-2 score threshold of 3 or greater was 4.0% (95% CI, 2.8%-5.2%) during pregnancy and 8.9% (95% CI, 5%-12.8%) postpartum. Using the lower threshold of 2 or greater, the positive rate was 24.3% (CI, 21.7%-27.2%) during pregnancy and 32.8% (CI, 26.6%-40%) postpartum.

The rate of probable minor depression was consistently higher during the postpartum period than during pregnancy whether the PHQ-9 or PHQ-2 was used.

The performance of the PHQ-2 in screening for depression risk using the PHQ-9 score as the reference during pregnancy and postpartum is presented in Table 3.

Using the threshold of 3 or more, the PHQ-2 had low sensitivity, and, as consequence, the percentage of affected women with a false-negative result (women classified by the PHQ-2 as not affected when they were actually affected with depressive symptoms) was substantial during pregnancy (58.5%) and considerable postpartum (36%). Looking at the PPVs, the probability that a woman with a positive result indeed had depression was high in pregnancy (71%) and postpartum (88.9%). The NPVs were very high both during pregnancy and postpartum (95.9% and 95.1%, respectively) and indicated the high percentages of women with a negative test result who were free from depressive symptoms as indicated

by the PHQ-9. The LRs+ were all bigger than 10 but suffered from very wide and inconclusive CIs. The LRs- were moderately higher than desirable (ie, <0.20).

Using the threshold of 2 or more, the PHQ-2 was highly sensitive (only ~7% false-negative, both in pregnancy and postpartum) with acceptable specificity (80.8% in pregnancy and 75.4% postpartum). In contrast, the PPVs were quite low, especially in pregnancy (26.3%). The LRs+ were moderately high (4.9 during pregnancy and 3.7 postnatal). The LRs- were strongly low (0.1).

At the threshold of 3 or greater, the PHQ-2 demonstrated fair accuracy during pregnancy (70%) and fairly good accuracy postpartum (81%). At the lower threshold of 2 or greater, the PHQ-2 demonstrated good accuracy both during pregnancy (87%) and postpartum (84%).

DISCUSSION

One of the main findings of the present study was that at the threshold of 3 or greater the PHQ-2 was poorly sensitive for identifying perinatal women at risk for depression; therefore, many cases of probable depression remained undetected, especially during pregnancy. In contrast, we found the PHQ-2 to be highly specific, suggesting a very low risk of false positives and response burden.

Looking at the PPVs and NPVs, our findings indicated a good performance of the PHQ-2 at this threshold. In contrast, the LRs+ suffered from very wide and inconclusive CIs, and the LRs- were higher than desirable; that is, the negative result may likely occur in participants with the condition more than in those without the condition.

Using the threshold score of 2 or more, the PHQ-2 was highly sensitive. Although the specificity values were lower than the corresponding values at the threshold of 3 or greater, they were still acceptable, and the PHQ-2 ultimately demonstrated good balanced accuracy for screening during both pregnancy and postpartum.

Looking at the PPVs, our findings indicate that they were very low, especially in pregnancy. However, it should be noted that PPV is partly dependent on the prevalence of the condition in the population being tested, and the prevalence of depressive disorders in our sample was low, especially in pregnancy. Under a scenario in which the prevalence of depression is higher, as for example, in general clinical settings where a substantial proportion of patients who have a chronic physical health problem or likely comorbid psychopathological or stress symptoms, the PPV will be accordingly higher.[34]

The LRs+ were moderately indicative of a presence of depression. However, they were better than the median values reported for other depression case-finding instruments.[35] In contrast, the LRs- were strongly indicative of an absence of depression. In this situation, the clinician in presence of a negative PHQ-2 result may feel strongly confident that depression could be ruled out.

Differently from predictive values, LRs are largely independent of the setting in which a screening test is used, as they are quite stable with changes in prevalence of the disease in the population. As a consequence, they are probably the best way to evaluate the strength of a screening test.

When using the threshold of 3 or more, the PHQ-2 had higher LRs+ than when using the threshold of 2 or more. This indicates that the shift in odds favoring the condition of depression will be relatively larger when using the threshold of 3 or more. However, when using the threshold of 3 or more, the PHQ-2 demonstrated higher LRs- than when using the threshold of 2 or more. As is known, the LR- value indicates the change in odds favoring the condition given a negative test result. Because a negative test result is supposed to reduce the odds that a condition is present, it is desirable for a test to have a low LR- value (ie, <0.20). A small LR- indicates a test that is useful for ruling out a condition when the result is negative. Our results demonstrated that a negative result using the threshold of 3 or more was less indicative of an absence of depression compared with a negative result using the threshold of 2 or more.

In other words, a PHQ-2 score of 3 or more indicates greater likelihood of being affected with depression but also gives a high chance of misclassifying the women affected as nonaffected with depression. The lower threshold of 2 or more will yield higher sensitivity and acceptable specificity, and ultimately, the PHQ-2 will correctly classify most of the affected women as affected but also give some chance (from 19% to 25%) of misclassifying the nonaffected women as affected.

The ability to compare our findings was hampered by a lack of comparable perinatal studies. Our findings are partially similar to those of a study of Smith et al,[25] which reported moderate PHQ-2 sensitivity (77%) but low specificity (59%) in a cohort of pregnant women, using a PHQ-2 cutoff point of 3 or more. Another postnatal study[26] showed lower sensitivity (75%) and relatively higher specificity (88%) than those observed in the present study at a cutoff point of 2 or more. However, it should be noted that these studies are not actually comparable with our study because they assessed the PHQ-2 diagnostic accuracy against a structured interview.

The strengths of the present study include the use of a large sample and several perinatal clinical centers throughout Italy. The main limitation consists of the PHQ-9 as a reference standard instead of a structured diagnostic interview. Although this limitation suggests some caution in interpreting our results, this study is the first in Italy to provide evidence of validity against a gold standard screening instrument for the PHQ-2 in primary and secondary care perinatal settings and corroborated the notion that the PHQ-2 is an acceptable ultra-brief tool for screening for depression.[23]

The use of a 2-item instrument to measure the severity of depressive symptoms implies some reduction in psychometric reliability, given the known correlation between reliability and number of items, and in the breadth of the assessment, as the instrument does not provide information about the presence of all 9 symptoms that underlie the diagnostic criteria for a major depressive episode. To partially mitigate this limitation, a recent individual participant data meta-analysis of 44 studies involving 10,627 participants reported only small differences in sensitivity and specificity between the PHQ-2 and the PHQ-9,[23] which suggests that the psychometric performance of the PHQ-2 is only marginally lower than that of the full PHQ-9.

Implications for Practice

Screening for depression in perinatal settings can be done in stages whereby various instruments are combined.[29] Stages may involve a 2-step process in which PHQ-2 is used to identify potential cases and, for women who screen positive, a second, diagnostic instrument or psychiatric examination administered by a specialist is used to confirm the diagnosis according to *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* criteria.

The evidence suggests that case identification by use of screening tools may be good and cost-effective clinical practice in primary and secondary perinatal care settings if these settings provide interdepartmental collaboration with mental health services for consultation, referral and prompt treatment.[36] If those settings have an integrated health professional, such as a licensed social worker or other mental health trained professional, that team member can provide immediate triage for a positive screen, offer support, facilitate mental health referral, and coordinate follow-up.

As opposed to referral to a mental health service, patients and clinicians have recently described a benefit to integrating mental health care into perinatal services through the presence of embedded mental health providers who could conduct psychotherapy and monitor patients.[7] This integrated approach increased convenience and avoided the stigma associated with a mental health referral, which affects especially racial and ethnic minorities and women with low income.[1,7,8]

There is evidence supporting the health benefits of programs in which other staff (eg, nurses, midwives, or trained therapists) provide part of the depression care, and it is documented that integrating mental health services into primary care may be the most viable way of closing treatment gap for mental health in low-resource settings.[37]

Taken together, this discussion underscores the importance of having services to ensure accurate diagnosis and treatment of depression in women, whether by specialists in mental health or by primary care or perinatal clinicians.

CONCLUSION

The PHQ-2 is an available and acceptable first step in depression screening and management program in primary or secondary care perinatal settings. Depression screening alone is insufficient and should be followed by evaluation by a qualified health care provider to ensure correct diagnosis. Using a threshold score of 2 or more provided high accuracy for screening as shown by high sensitivity (missing only 7% of cases) and acceptable specificity. Using a threshold score of 3 or more provided lower sensitivity, which was unacceptable for clinical screening of depression. Therefore, a threshold of 2 or more should be recommended for perinatal depression screening in clinical practice because the current recommended PHQ-2 threshold of 3 or more could lead to underdetection.

CONFLICT OF INTEREST

The authors have no conflicts of interest to disclose.

REFERENCES

1. Dagher RK, Bruckheim HE, Colpe LJ, Edwards E, White DB. Perinatal depression: challenges and opportunities. *J Womens Health (Larchmt)*. 2021;30(2):154-159. https://doi.org/10.1089/jwh.2020.8862.
2. Aktar E, Qu J, Lawrence PJ Tollenaar MS, Elzinga BM, Bögels SM. Fetal and infant outcomes in the offspring of parents with perinatal mental disorders: earliest influences. *Front Psychiatry*. 2019;10:391. https://doi.org/10.3389/fpsyt.2019.00391.
3. Erickson N, Julian M, Muzik M. Perinatal depression, PTSD, and trauma: impact on mother- infant attachment and interventions to mitigate the transmission of risk. *Int Rev Psychiatry*. 2019;31(3):245-263. https://doi.org/10.1080/09540261.2018.1563529.
4. Woody CA, Ferrari AJ, Siskind DJ, Whiteford HA, Harris MG. A systematic review and meta-regression of the prevalence and incidence of perinatal depression. *J Affect Disord*. 2017;219:86-92. https://doi.org/10.1016/j.jad.2017.05.003.
5. Heck JL. Postpartum depression in American Indian/Alaska Native women: a scoping review. *MCN Am J Matern Child Nurs*. 2021;46(1):6-13. https://doi.org/10.1097/NMC.0000000000000671.
6. Heck JL, Wilson JS, Parker JG. "It took away the joy:" First American mothers' experiences with postpartum depression. *MCN Am J Matern Child Nurs*. 2022;47(1):13-18. https://doi.org/10.1097/NMC.0000000000000776.
7. Iturralde E, Hsiao CA, Nkemere L, et al. Engagement in perinatal depression treatment: a qualitative study of barriers across and within racial/ethnic groups. *BMC Pregnancy Childbirth*. 2021;21(1):512. https://doi.org/10.1186/s12884-021-03969-1.
8. Mukherjee S, Trepka MJ, Pierre-Victor D, Bahelah R, Avent T. Racial/ethnic disparities in antenatal depression in the United States: a systematic review. *Matern Child Health J*. 2016;20(9):1780-1797. https://doi.org/10.1007/s10995-016-1989-x.
9. Sambrook Smith M, Lawrence V, Sadler E, Easter A. Barriers to accessing mental health services for women with perinatal mental illness: systematic review and meta-synthesis of qualitative studies in the UK. *BMJ Open*. 2019;9(1):e024803. https://doi.org/10.1136/bmjopen-2018-024803.
10. Byatt N, Xu W, Levin LL, More Simas TA. Perinatal depression care pathway for obstetric settings. *Int Rev Psychiatry*. 2019;31(3):210-228. https://doi.org/10.1080/09540261.2018.1534725.
11. American College of Obstetricians and Gynecologists. ACOG Committee Opinion no. 757: screening for perinatal depression. *Obstet Gynecol*. 2018;132(5):e208-e212. https://doi.org/10.1097/AOG.0000000000002927.
12. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh postnatal depression scale. *Br J Psychiatry*. 1987; 150:782-786. https://doi.org/10.1192/bjp.150.6.782.
13. Chorwe-Sungani G, Chipps J. Validity and utility of instruments for screening of depression in women attending antenatal clinics in Blantyre district in Malawi. *S Afr Fam Pract*. 2018;60(4):114-120. https://doi.org/10.1080/20786190.2018.1432136.
14. Ali GC, Ryan G, DeSilva MJ. Validated screening tools for common mental disorders in low- and middle-income countries: a systematic review. *PLoS One* 2016;11(6):e0156939. https://doi.org/10.1371/journal.pone.0156939.
15. Shrestha SD, Pradhan R, Tran TD, Gualano RC, Fisher JRW. Reliability and validity of the Edinburgh Postnatal Depression Scale (EPDS) for detecting perinatal common mental disorders (PCMDs) among women in low-and lower-middle-income countries: a systematic review. *BMC Pregnancy Childbirth*. 2016;16:72. https://doi.org/10.1186/s12884-016-0859-2.
16. Naja S, Al Kubaisi N, Singh R, Abdalla H, Bougmiza I. Screening for antenatal depression and its determinants among pregnant women in Qatar: revisiting the biopsychosocial model. *BMC Pregnancy Childbirth*. 2021;21(1):330. https://doi.org/10.1186/s12884-021-03793-7.

17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.
18. Mazzotti E, Fassone G, Picardi A, et al. The Patient Health Questionnaire (PHQ) for the screening of psychiatric disorders: a validation study versus the Structured Clinical Interview for DSM-IV axis I (SCID-I). *J Psychopathol*. 2003;9:235-242.
19. Sambrook Smith M, Cairns L, Pullen LSW, Opondo C, Fellmeth G, Alderdice F. Validated tools to identify common mental disorders in the perinatal period: A systematic review of systematic reviews. *J Affect Disord*. 2022; 298(Pt A):634-643. https://doi.org/10.1016/j.jad.2021.11.011.
20. Wang L, Kroenke K, Stump TE, Monahan PO. Screening for perinatal depression with the patient health questionnaire depression scale (PHQ-9): a systematic review and meta-analysis. *Gen Hosp Psychiatry*. 2021;68:74-82. https://doi.org/10.1016/j.genhosppsych.2020.12.007.
21. Howard LM, Ryan EG, Trevillion K, et al. Accuracy of the Whooley questions and the Edinburgh Postnatal Depression Scale in identifying depression and other mental disorders in early pregnancy. *Br J Psychiatry*. 2018;212(1):50-56. https://doi.org/10.1192/bjp.2017.9.
22. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care*. 2003;41(11):1284-1292. https://doi.org/10.1097/01.MLR.0000093487.78664.3C.
23. Levis B, Sun Y, He C, et al. Accuracy of the PHQ-2 alone and in combination with the PHQ-9 for screening to detect major depression: systematic review and meta-analysis. *JAMA* 2020;323(22):2290-2300. https://doi.org/10.1001/jama.2020.6504.
24. Manea L, Gilbody S, Hewitt C, et al. Identifying depression with the PHQ-2: a diagnostic meta-analysis. *J Affect Disord*. 2016;203:382-395. https://doi.org/10.1016/j.jad.2016.06.003.
25. Smith MV, Gotman N, Lin H, Yonkers KA. Do the PHQ-8 and the PHQ-2 accurately screen for depressive disorders in a sample of pregnant women? *Gen Hosp Psychiatry*. 2010;32(5):544-548. https://doi.org/10.1016/j.genhosppsych.2010.04.011.
26. Gjerdingen D, Crow S, McGovern P, Miner M, Center B. Postpartum depression screening at well-child visits: validity of a 2-question screen and the PHQ-9. *Ann Fam Med*. 2009;7(1):63-70. https://doi.org/10.1370/afm.933.
27. Cutler CB, Legano LA, Dreyer BP, et al. Screening for maternal depression in a low education population using a two item questionnaire. *Arch Women's Ment Health*. 2007;10(6):277-283. https://doi.org/10.1007/s00737-007-0202-z.
28. Bennett IM, Coco A, Coyne JC, et al. Efficiency of a two-item pre-screen to reduce the burden of depression screening in pregnancy and postpartum: an IMPLICIT network study. *J Am Board Fam Med*. 2008;21(4):317-325. https://doi.org/10.3122/jabfm.2008.04.080048.
29. Slavin V, Creedy DK, Gamble J. Comparison of screening accuracy of the Patient Health Questionnaire-2 using two case-identification methods during pregnancy and postpartum. *BMC Pregnancy Childbirth*. 2020;20(1):211. https://doi.org/10.1186/s12884-020-02891-2.
30. Cena L, Palumbo G, Mirabella F et al. Perspectives on early screening and prompt intervention to identify and treat maternal perinatal mental health. Protocol for a prospective multicenter study in Italy. *Front Psychol*. 2020;11:365. https://doi.org/10.3389/fpsyg.2020.00365
31. Nijagal MA, Wissig S, Stowell C, et al. Standardized outcome measures for pregnancy and childbirth, an ICHOM proposal. *BMC Health Serv Res*. 2018;18(1):953. https://doi.org/10.1186/s12913-018-3732-3.
32. Sackett DL, Haynes RB, Tugwell P, Guyatt GH. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. 2nd ed. Lippincott Williams & Wilkins; 1991.
33. Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med*. 1996;3(9):895-900. https://doi.org/10.1111/j.1553-2712.1996.tb03538.x.
34. Meader N, Mitchell AJ, Chew-Graham C, et al. Case identification of depression in patients with chronic physical health problems: a diagnostic accuracy meta-analysis of 113 studies. *Br J Gen Pract*. 2011:61(593):e808-820. https://doi.org/10.3399/bjgp11/613151.
35. Williams JW Jr, Noël PH, Cordes JA, Ramirez G, Pignone M. Is this patient clinically depressed?. *JAMA*. 2002;287(9):1160-1170. https://doi.org/10.1001/jama.287.9.1160.
36. Howard LM, Khalifeh H. Perinatal mental health: a review of progress and challenges. *World Psychiatry*. 2020;19(3):313-327. https://doi.org/10.1002/wps.20769.
37. Chorwe-Sungani G, Mwagomba M, Kulisewa K, Chirwa E, Jere D, Chipps J. Protocol for assessing feasibility, acceptability and fidelity of screening for antenatal depression (FAFSAD) by midwives in Blantyre District, Malawi. *Pilot Feasibility Stud*. 2021;7(1):32. https://doi.org/10.1186/s40814-021-00775-6.