

Screening for antenatal maternal depression: comparative performance of the Edinburgh Postnatal Depression Scale and Patient Health Questionnaire-9

Alberto Stefana¹, Loredana Cena², Alice Trainini², Gabriella Palumbo³, Antonella Gigantesco^{3*} and Fiorino Mirabella^{3*}

¹Dipartimento di Scienze del Sistema Nervoso e del Comportamento, Università di Pavia, Pavia, Italy

²Dipartimento di Scienze Cliniche e Sperimentali, Sezione di Neuroscienze, Osservatorio Psicologia Clinica Perinatale, Università di Brescia, Brescia, Italy

³Centro di Riferimento per le Scienze Comportamentali e la Salute Mentale, Istituto Superiore di Sanità, Rome, Italy

*These Authors equally contributed to this work and should be considered co-last Authors

Abstract

Background. Maternal antenatal depression affects 21-28% of expectants globally and negatively impacts both maternal and child health in the short and long term.

Objective. To compare the psychometric properties and clinical utility of the Edinburgh Postnatal Depression Scale (EPDS) and the Patient Health Questionnaire (PHQ-9) in pregnant individuals.

Methods. In this cross-sectional study, 953 third-trimester pregnant Italian individuals completed both the EPDS and the PHQ-9.

Results. Both scales demonstrated good internal consistency (EPDS $\omega=0.83$, PHQ-9 $\omega=0.80$) and a moderate correlation between their scores ($r=0.59$). Concordance at recommended cut-off points (≥ 14 for both) was moderate ($k=0.55$). Factor analyses indicated a bifactor solution for the EPDS (dimensions: "depression" and "anxiety") and for the PHQ-9 (dimensions: "depression", "pregnancy symptoms", "somatic"). Benchmarks for clinical change were also established.

Conclusions. The EPDS and PHQ-9 capture distinct aspects of perinatal depressive symptomatology. Clinically, these findings recommend using both scales in obstetric and gynaecologic settings to minimize false positives and negatives.

Key words

- EPDS
- PHQ-9
- antenatal depression
- anxiety
- screening
- pregnancy

INTRODUCTION

Antenatal depression is a non-psychotic unipolar depressive disorder characterized by specific feelings and thoughts about the parental role [1]. It is one of the leading complications for people during the antepartum period [1] with a worldwide prevalence estimated between 21% and 28% [2-4]. However, despite this high prevalence, antenatal depression is frequently underdiagnosed [5], with about one in five pregnant people not asked about depression during prenatal visits [6].

New-onset or pre-existing depression in pregnant people can be a significant cause of short- and long-term negative consequences on both pregnant health

and child development [7, 8], which entails, among other things, an important cost for national health care systems [9]. Studies aimed at developing and evaluating intervention programmes are consistent in highlighting the importance of early screening and prompt intervention to produce more optimal emotional health outcomes [10] and offspring health outcomes [11]. However, in Italy, where the prevalence of antenatal depression has been reported to be around 25% in 2022 [12], such programmes are not adequately integrated into clinical guidelines for appropriate practical care planning routines [13, 14]. This is partly due to the absence of a national Italian policy to screen for perina-

tal mental illness. In addition, there is limited training among healthcare providers on how to choose and use the most appropriate screening tool(s) and the cut-off point for a particular period of time.

Routine screening for perinatal depression through valid, reliable, and economical screening tools is probably the most widely accepted option [15-17]. However, no consensus has been reached on what scale can be considered the gold standard. Two of the three most frequently used screening tools are the Edinburgh Postnatal Depression Scale (EPDS) and the Patient Health Questionnaire (PHQ-9) [18, 19].

Aims of the study

The purpose of this study was to compare the psychometric properties and clinical utility of the Italian versions of EPDS and PHQ-9 among pregnant women. This study serves as an initial step toward establishing an effective screening programme for antenatal depression in Italy and possibly other Western countries.

MATERIALS AND METHODS

Study design and sample

The cross-sectional data presented here were collected as part of a larger project on screening and early intervention for maternal perinatal anxiety and depressive disorders [13]. A total of 1,159 consecutive adult pregnant people in their third trimester and of Italian nationality were asked to join the study from March 2017 to June 2018. Pregnant people who agreed to participate underwent an interview led by clinical psychologists to obtain information on current and past maternal experience with psychiatric conditions and use of psychotropic drugs. The exclusion criteria were having problems with drug or substance misuse and/or having ongoing psychotic symptoms. This study was approved by the ethics committee of the Healthcare Centre of Bologna Hospital (Comitato Etico Internazionale Bologna-Imola) (Reg. n. 77808 del 27/6/2017).

Data collection

Pregnant people who said they wanted to participate in the study signed an informed consent form. They were interviewed in a private room inside the health centre by a licensed clinical psychologist trained in evidence-based assessment techniques for perinatal mental health issues to determine their eligibility. Information on the demographic, economic, psychosocial, and reproductive characteristics of eligible participants was collected. The EPDS and PHQ-9 questionnaires were then administered.

Measures

Edinburgh Postnatal Depression Scale. The EPDS [20, 21] is the most chosen self-administered screening scale for perinatal depression [18]. The EPDS can be used to assess depression according to DSM-5 and ICD-10 criteria [22]. It assesses the frequency of each of the following depressive symptoms experienced in the previous seven days: anhedonia (two items); guilt; anxiety; panic attack; overwhelming; sleep disorders; sadness; tearfulness; and suicidal ideas. The EPDS consists of 10

items scored on a 4-point Likert scale ranging from 0 to 3. Its overall score can range from 0 to 30, and scores of 0-9, 10-11, 12-15, and ≥ 16 are commonly used as thresholds for normal, slightly increased risk of depression, increased risk of depression, and likely depression, respectively [21]. Findings from a systematic review and individual participant data meta-analysis indicate that a cut-off of ≥ 14 approximated structured clinical interview for diagnostic and statistical manual of mental disorders (SCID)-based prevalence of major depression [23].

Patient Health Questionnaire-9. The PHQ-9 [24, 25] is a self-administered depression screening scale containing nine items corresponding to the DSM-IV criteria for depression. However, it can be used to measure depression severity also according to DSM-5 criteria. The PHQ-9 is the first choice for depressive symptoms in non psychiatric primary care settings [26]. It assesses the frequency of each of the following depressive symptoms experienced in the previous two weeks: anhedonia; depressed mood; insomnia or hypersomnia; fatigue or loss of energy; appetite disturbances; feelings of worthlessness or excessive guilt; diminished ability to think or concentrate; psychomotor agitation or retardation; and suicidal thoughts. The PHQ-9 is comprised of 9 items, each rated on a Likert scale ranging from 0 (not at all) to 3 (nearly every day). Its overall score can range from 0 to 27, and scores of 0-4, 5-9, 10-14, and ≥ 15 represent the thresholds for normal, mild depressive symptoms, moderate depressive symptoms, and moderately severe to severe depressive symptoms, respectively [24]. Findings from an individual participant data meta-analysis indicate that a cut-off of ≥ 14 most closely matched SCID major depression prevalence [27].

State Anxiety Scale. The State Anxiety Scale (SAS) of the State-Trait Anxiety Inventory-Form Y (STAI-Y) [28, 29] is a 20-item self-report measure designed to assess situational anxiety, capturing feelings experienced in the present moment. Responses are recorded on a four-point Likert scale, ranging from "not at all" to "very much so", with ten of the items being reverse-scored. The inventory has demonstrated good internal consistency, with Cronbach's alpha values ranging between 0.86 and 0.95 [28, 29]. Notably, recently, a shortened version of the STAI specifically tailored for pregnant women has been developed [30].

Statistical analyses

Descriptive statistics were carried out using means and standard deviations (SD) for continuous variables and using frequencies and percentages for categorical variables. The factor structures of both EPDS and PHQ-9 were explored as follows. Parallel analysis using the R package EFAtools v0.4.3 [31] was conducted on a poly-choric correlation matrix using mean eigenvalues and 95th percentile eigenvalues of 5,000 simulated random datasets to evaluate the number of factors that may be supported by our data. The scree plot and the eigenvalues associated with each factor were used to identify the number of meaningful factors. The sample was randomly divided into two mutually independent groups, ensuring

separate and independent samples for the exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), respectively. EFA was conducted on a matrix of inter-item polychoric correlations using the Promax rotation method. This analysis was carried out with the R package EFAtools, version 0.4.3 [31]. CFA using the R packages lavaan v0.6-12 [32] was performed to explore the factor structures of the scale. The overall fit of the models was assessed using the following criteria: a minimum threshold of 0.95 for both the comparative fit index (CFI) and the Tucker-Lewis index (TLI), and maximum thresholds of 0.06 for the root mean square error of approximation (RMSEA), and 0.08 for the standardized root mean square residual (SRMR) [33, 34]. The internal consistency of both EPDS and PHQ-9 was assessed using Cronbach's alpha, McDonald's omega total, and Pearson's product-moment correlation coefficient (r). Omega is a more effective index than alpha for assessing reliability, particularly in the case of short scales or scales with multiple dimensions [35]. Internal consistency values are categorized as follows: values between 0.70 and 0.79 are considered adequate, those between 0.80 and 0.89 are regarded as good, and values of 0.90 or higher are deemed excellent [36]. Additionally, composite reliability (CR) and average variance extracted (AVE) have to be calculated to assess convergent validity. The commonly recommended thresholds for AVE and CR are 0.50 and 0.70, respectively. However, in cases where the AVE value falls below the recommended threshold, yet the CR value is high, this scenario suggests that the convergent validity of the construct may still be considered adequate [37, 38]. The agreement between EPDS and PHQ-9 with cut-off scores ≥ 14 and ≥ 14 , respectively, was assessed using Cohen's kappa coefficient. The agreement between the two measures at different severity cut-off points (for the EPDS: 0-9, 10-11, 12-15, and ≥ 16 ; for PHQ-9: 0-4, 5-9, 10-14, and ≥ 15) was assessed using the intraclass correlation coefficient (ICC). All tests were two-tailed with the statistical significance level set at $p=0.05$. All data were coded and analysed using the Statistical Package for Social Science (SPSS) version 24 and R version 4.3.1.

RESULTS

Sample characteristics

Of the 1,159 pregnant people who met the eligibility criteria and were asked to participate in the study, 959 (83%) agreed to join. Of these participants, 953 completed EPDS and PHQ-9. Almost half (47%) of them were aged 30 to 35 years, 31% were aged 36 or older, and 22% were 29 years or younger. Regarding the level of education, 54% of the participants had high (tertiary) education, 36% had middle (secondary) education, and 10% had low (lower than secondary) education. Regarding working status, 75% were permanently employed, 9% were temporarily employed and 16% were unemployed/housewives/students/other. Regarding economic status, 48% of the participants had an average high status, 46% had a few economic problems without specific difficulties, and 6% had economic problems. Sociodemographic and reproductive information is shown in Table 1.

Table 1

Socio-demographics and reproductive characteristics of the sample

	n (%)
Age	
18-29	212 (22.1)
30-35	454 (47.4)
>35	292 (30.5)
Marital status	
Married or cohabiting	882 (92.6)
Single, separated, or divorced/owidowed	70 (7.4)
Educational level	
University	509 (53.5)
Secondary	343 (36.0)
Primary or illiterate	100 (10.5)
Working status	
Permanent employee	705 (74.5)
Temporary employee	90 (9.5)
Student, homemaker, or unemployed	151 (16.0)
Economic status	
Average high status	454 (47.9)
A few problems without specific difficulties	435 (45.9)
Same or many problems	58 (6.2)

Parallel analysis

The number of factors suggested by the parallel analyses with Hull's method, Principal component analysis (PCA), and EFA was as follows: one, one, and five for the EPDS; and one, two, and three for the PHQ-9. Furthermore, an examination of the scree plot evidenced one or two factors for both scales. Given the number of items, it is not plausible to have more than three factors in the EPDS.

Exploratory factor analysis

Regarding EPDS, EFAs comparing the two models indicated by parallel analyses (i.e., one- and three-factor models) were performed (see Table 2). Eigenvalues and percentage of cumulative variance were as follows: 4.49 (44.9%) for the one-factor solution; 3.27 (32.7%) and 1.95 (52.2%) for the two-factor solution. We labeled these two factors as "depression" and "anxiety".

Regarding PHQ-9, EFAs comparing the three models suggested by parallel analyses (i.e., one- two- and three-factor models) were performed (see Table 3). Eigenvalues and percentage of cumulative variance were as follows: 3.77 (42.0%) for the one-factor solution; 2.68 (29.8%), 1.43 (45.7%), and 1.34 (60.5%) for the three-factor solution (item 1 loaded on both the first and the second factors). We labeled these three factors as "depression", "pregnancy symptoms" and "somatic". EFA could not be estimated for the two-factor model because no solutions were achieved across which averaging was possible for this model.

Table 2

Loadings and percentage of cumulative variance for the Edinburgh Postnatal Depression Scale (EPDS) and the Patient Health Questionnaire (PHQ-9)

		1-factor model	2-factor models		
Item content abbreviated		F1	F1	F2	
EPDS	1. Laugh	0.69	0.75	-0.01	
	2. Enjoyment	0.68	0.81	-0.09	
	3. Self-blame	0.61	0.25	0.46	
	4. Anxious	0.57	-0.09	0.83	
	5. Scared	0.61	0.05	0.71	
	6. Hard to cope	0.56	0.23	0.42	
	7. Hard to sleep	0.70	0.59	0.17	
	8. Sad	0.83	0.78	0.11	
	9. Crying	0.78	0.69	0.16	
	10. Self-harm	0.60	0.46	0.19	
Cumulative variance/%		44.9	32.7	52.2	
		1-factor model	3-factor models		
Item content abbreviated		F1	F1	F2	F3
PHQ-9	1. Anhedonia	0.75	0.44	0.36	0.14
	2. Depressed mood	0.68	0.72	0.17	-0.10
	3. Sleeping difficulties	0.33	-0.22	0.92	-0.01
	4. Fatigue.	0.57	0.23	0.56	0.02
	5. Appetite changes	0.59	0.21	0.25	0.30
	6. Feeling of worthlessness	0.74	0.80	-0.06	0.07
	7. Concentrations difficulties	0.64	0.20	0.10	0.49
	8. Psychomotor agitation	0.59	-0.05	-0.04	0.89
	9. Suicide ideation	0.81	1.01	-0.20	0.08
Cumulative variance/%		42.0	29.8	45.7	60.5

F1: factor 1; F2: factor 2.

Bold fonts show loadings of >0.30.

The table reports average loadings from 72 exploratory factor analyses, conducted using the mean method without any trimming (trim=0). These analyses were performed by the R package EFATools [31] and varied across various factor extraction and rotation methods: initial communalities, criterion type, number of factors for Promax rotation, rotation method type, and type of Varimax rotation.

Table 3

Confirmatory factor analysis indices of the factor models of the Edinburgh Postnatal Depression Scale (EPDS) and the Patient Health Questionnaire (PHQ-9)

Factor solution	χ^2 value	df	CFI	TLI	RMSEA	SRMR		
EPDS	One-factor model	222.62	35	0.84	0.80	0.11	0.07	
	Two-factor model	122.86	34	0.93	0.90	0.08	0.04	
	Bi-factor model	66.50	25	0.97	0.94	0.06	0.03	
	EPDS-4A	14.61	2	0.98	0.94	0.08	0.03	
PHQ-9	One-factor model	167.21	27	0.81	0.74	0.11	0.07	
	Three-factor model (item 1 on Factor 1)	89.81	24	0.91	0.87	0.08	0.05	
	Bi-factor model (item 1 on Factor 1)	Computation of modification indices for the bifactor model was not feasible						
	Three-factor model (item 1 on Factor 2)	114.93	24	0.88	0.81	0.09	0.06	
Bi-factor model (item 1 on Factor 2)	32.59	18	0.98	0.96	0.04	0.03		

The items' scale assignments are those indicated in Table 2 using bold fonts. CFI: comparative fit index; df: degree of freedom; TLI: Tucker-Lewis index, RMSEA: maximum thresholds of 0.06 for the root mean square error of approximation; SRMR: standardized root mean square residual; χ^2 : chi-squared.

Confirmatory factor analysis

A series of CFAs were conducted to test the solutions indicated by EFAs, including bifactor models. *Table 3* presents the fit indices for each factor model. For both EPDS and PHQ-9, bifactor models demonstrated the best fit.

For the EPDS, the bifactor model comprising a general factor and two specific factors – depression (items 1, 2, 7, 8, 9, and 10) and anxiety (items 3, 4, 5, and 6) – yielded the following fit indices: $\chi^2(df=25)=66.50$, CFI=0.97, TLI=0.94, RMSEA=0.06 (90% CI=0.04, 0.08), and SRMR=0.03. *Supplementary Figure 1a (available online)* provides a graphical representation of this model. Additionally, we tested the EPDS-4A model, which includes the EPDS-3A – consisting of consisting of items 3, 4 and 5 [39, 40] – plus item 6. This addition was suggested by our data and corroborated by the only other study investigating the factorial structure of the Italian version of the EPDS in a perinatal population [41]. The EPDS-4A model demonstrated a good fit: $\chi^2(df=2)=14.61$, CFI=0.98, TLI=0.94, RMSEA=0.08 (90% CI=0.05, 0.12), and SRMR=0.03. *Supplementary Figure 1b (available online)* displays a graphical representation of this model. For the PHQ-9, a bifactor model with a general factor and three specific factors – depression (items 2, 6, and 9), pregnancy symptoms (items 1, 3 and 4), and somatic (items 5, 7 and 8) – showed the following fit indices: $\chi^2(df=18)=32.59$, CFI=0.98, TLI=0.96, RMSEA=0.04 (90% CI=0.02, 0.06), and SRMR=0.03. *Supplementary Figure 1c (available online)* provides a graphical representation of this model.

Reliability

For the EPDS, the Cronbach's alpha value of the EPDS was 0.81, and the McDonald's omega total was 0.83. The CR was 0.81, while the AVE was 0.31. The correlations between the items and the total scores ranged from 0.20 (item 10) to 0.83 (item 8), with an average inter-item correlation of $r=0.29$. For the PHQ-9, the Cronbach's alpha was 0.76, and the McDonald's omega total was 0.80. The CR was 0.75, and the AVE was 0.31. The correlations between the items and the total scores for the PHQ-9 ranged from 0.33 (item 9) to 0.63 (item 1), with an average inter-item correlation of $r=0.24$.

Correlation

Pearson's correlation coefficient between EPDS and PHQ-9 was $r=0.59$ ($p < 0.001$). The correlation coefficients between, on the one hand, the SAS and, on the other hand, the EPDS-4A, the EPDS, and the PHQ-9 were respectively $r=0.46$ ($p < 0.001$), $r=0.55$ ($p < 0.001$), and $r=0.48$ ($p < 0.001$). We also evaluated the correlation between the EPDS-3A and the SAS: and $r=0.43$ ($p < 0.001$).

Severity ratings

The mean EPDS score was 4.7 (SD=3.9), while the mean PHQ-9 score was 4.3 (SD=3.0). When the scales were divided into four cut-off threshold groups (that is, 0-9, 10-11, 12-15, and ≥ 16 for the EPDS; 0-4, 5-9, 10-14, and ≥ 15 for the PHQ-9), the EPDS identified

4% of pregnant people ($n=37$) as having a possible depressive disorder (score from 12 to 15), while PHQ-9 classified 6% of subjects ($n=56$) as from moderately to severely depressed (score from 10 to 14). Additionally, 2% of pregnant people ($n=19$) were identified as likely depressed (score ≥ 16), while the PHQ-9 scale classified 1% of subjects ($n=9$) as moderately severely or severely depressed (score ≥ 15). The ICC was 0.46 (0.32-0.56), indicating poor reliability [42]. *Table 4* shows the distribution of study participants according to their severity of depressive symptoms.

Agreement at different cut-off scores for major depression

When applying EPDS and PHQ-9 cut-off thresholds (which is ≥ 14 for both scales) to estimate major depression prevalence, 95% ($n=902$) of pregnant people were concordantly classified. More specifically, 4% ($n=35$) persons were classified as depressed on both the EPDS and PHQ-9 scales, while 91% ($n=867$) expectant people were classified as not depressed on both scales (*Table 5*). The % agreement between the two scales was 95%, $k=0.55$, indicating moderate agreement.

Critical change benchmarks

Table 6 reports the following benchmarks for assessing clinical changes: critical changes at the 90% and 95% confidence intervals, minimally important difference, and minimum change for reliable change [43]. These four benchmarks are essential tools for clinicians, helping to assess whether alterations in a patient's scores are substantial beyond mere measurement error and have clinical relevance. The critical change values at both the 90% and 95% confidence levels signify the least amount of score change necessary to confidently assert that the observed change is not a result of chance or measurement inaccuracies. The minimally important difference denotes the smallest score variation perceived by patients as advantageous, crucial to assessing the impact

Table 4
Depression severity based on the Edinburgh Postnatal Depression Scale (EPDS) and the Patient Health Questionnaire (PHQ-9) ($n=1,153$)

	Depression severity	n	%
EPDS	0-9: Normal	850	89.2
	10-11 Slightly increased risk of depression	47	4.9
	12-15: Increased risk of depression	37	3.9
	≥ 16 : Likely depression	19	2.0
	Total score ($M \pm SD$)	953	4.7 \pm 3.9
PHQ-9	0-4: Normal	589	61.8
	5-9: Mild depressive symptoms	299	31.4
	10-14: Moderate depressive symptoms	56	5.9
	≥ 15 : Moderately severe to severe depressive symptoms	9	0.9
	Total score ($M \pm SD$)	953	4.3 \pm 3.0

Table 5

Comparison of the Edinburgh Postnatal Depression Scale (EPDS) and the Patient Health Questionnaire (PHQ-9) in identifying probable major depression

	PHQ-9 depression	PHQ-9 no depression	Total
EPDS depression	35 (3.7)	21 (2.2)	56 (5.9)
EPDS no depression	30 (3.1)	867 (91.0)	897 (94.1)
Total	65 (6.8)	888 (93.2)	953 (100%)

Table 6

Critical change benchmarks

	EPDS	PHQ-9
90% CC	2.64	2.21
95% CC	3.15	2.63
MID	1.95	1.5
MCRC	4.46	3.72

CC: critical change; EPDS: Edinburgh Postnatal Depression Scale; MID: minimal important difference; MCRC: minimum change for a reliable change; PHQ-9: Patient Health Questionnaire.

of treatment. Meanwhile, the minimum change for a reliable change indicates the extent of change needed to be deemed statistically robust, confirming that the observed variation is not attributable to random fluctuation. Using these benchmarks, clinicians can effectively track patient progress and judiciously assess the impact of therapeutic interventions, determining when a change in scores is significant enough to warrant modifications to the treatment strategy.

DISCUSSION

Both EPDS and PHQ-9 have been shown to have good internal consistency and homogeneity [44] when administered to a sample of Italian third-trimester pregnant people. The item-total correlations were acceptable for the vast majority of items on both scales and aligned with the results on the internal consistency of the scales. The two scales have a moderate positive correlation [45]. EFA suggested a two-factor model (named “depression” and “anxiety” factors) for EPDS and a three-factor model (named “depression”, “somatic”, and “pregnancy symptoms” factors) for PHQ-9. CFA showed that for both scales, the bifactor model was the best fit. This model suggests the presence of a general factor assessing antepartum depression, which accounts for the shared variance across all items. Additionally, it identifies group factors (two for the EPDS and three for the PHQ-9) that capture the common variance within specific item clusters, beyond the influence of the general factor [46].

When the scales’ recommended cut-off score (i.e., ≥ 14 for both EPDS and PHQ-9) was used, the EPDS and the PHQ-9 identified comparable proportions of subjects considered as clinically depressed: 6% and 7%, respectively. However, EPDS identified a higher proportion of subjects as normal/not depressed (89% for EPDS vs 62% for PHQ-9). Vice versa, PHQ-9 identified a higher proportion of subjects as affected by subdiagnostic symptoms of depression (5% for EPDS vs 31%

for PHQ-9). Taking into account the different severity cut-off scores, the concordance between the scales was poor [42] in our sample. Overall, these results confirmed that EPDS and PHQ-9 are similar tools, but measure different aspects of antenatal depressive symptoms.

The variation observed between these two scales aligned with results from a previous study with pregnant Peruvians [47] that suggested as a possible explanation the fact that PHQ-9 but not EPDS includes items addressing somatic symptoms. This might be important because some people can emphasize somatic complaints rather than reporting feelings of sadness [48]. However, in our case, this explanation is unlikely since there is evidence suggesting that white women with depression or depressive symptoms report fewer somatic symptoms than Hispanic/Latina women [49]. Furthermore, there exists the possibility of an overlap between (pathological) symptoms of depression and (normal) somatic complaints of pregnancy, which can lead to an overdiagnosis of depressive disorders in such a population [50, 51]. In fact, it has been observed that appetite increase and increase in energy (e.g., agitation) are uninformative with regard to a major depressive disorder diagnosis in pregnant women [52]. Therefore, somatic symptoms, which may be caused by normal physiological changes associated with pregnancy can increase the false-positive rate of depression during the antenatal period; this is the reason for the absence of any somatic symptom items on the EPDS. However, it is possible that the elimination of somatic symptoms (e.g., sleep disturbances, fatigue, psychomotor retardation) from the depression scale might result in the loss of clinically useful information, such as a specific pattern of symptoms or indicators of depression severity.

A further possible explanation for the variation between the scales is that, while EPDS was developed using items drawn from three scales for anxiety and depression (that is, the Irritability, Depression and Anxiety Scale [53], the Hospital Anxiety and Depression Scale [54], and the Anxiety and Depression Scale [55]), the PHQ-9 was developed specifically to identify depressive disorders based on DSM-IV [56] criteria. The presence of an EPDS anxiety subscale has been largely demonstrated [57-60] and a positive correlation between the results of the EPDS anxiety subscale and those of scales specifically developed to measure anxiety [61-63].

However, consistent with a previous study [61], we found that the EPDS total score in our sample was more highly correlated with the SAS than both the EPDS-4A and the EPDS-3A. Even the correlation was higher between PHQ-9 and SAS, than between the lat-

ter and the EPDS anxiety subscale score. A possible explanation is that while the ideal tool for assessing anxiety should measure both the negative affect (which is the clinical characteristic shared by anxiety and depression) and the physiological arousal (which is a typical symptom of anxiety), the EPDS-4A does not contain items specifically related to hyperarousal. A further possible explanation rests on the fact that anxiety is a multidimensional construct and can be generalized or focused on specific aspects/situations [64]. Therefore, a four-item anxiety subscale is unlikely to be able to accurately and reliably measure perinatal anxiety. Lastly, given that items 3, 4, and 5 are the only EPDS items that contain a subjective negative judgment about feelings, the measurement of anxiety could be less accurate in people with low self-esteem.

Study limitations

This study has some limitations that are worth mentioning. First, we evaluated depressive symptomatology with self-report instruments without supplementing that assessment with a diagnostic interview to actually make a diagnosis of depression according to the criteria of DSM-5-TR [48]. Thus, validity cannot be established for either scale; a criterion validity comparison between them cannot be tested. Second, the cross-sectional nature of the data prevents the possibility of evaluating whether and how the performance of EPDS and PHQ-9 changes during the perinatal period. Future work should include longitudinal studies with at least two time-point assessments to evaluate the predictive validity of both the instruments, which was not analyzed here.

CONCLUSIONS

The current study offers new evidence on screening tools for antenatal depression. Our results show that EPDS and PHQ-9 have satisfactory internal consistency and identify similar proportions of antenatal depression. However, the observed differences indicate that their ability to screen for depression during pregnancy is not identical because they partially focus on

different symptoms of depression. EPDS and PHQ-9 capture partially distinct features of depressive symptomatology: anxiety symptoms and somatic symptoms, respectively. These findings suggest that when using these scales for clinical purposes with people at risk of antenatal depression, they should be used in combination – rather than substituted – to reduce depression both false-positive and false-negative results.

Finally, the current study further highlights the need to continue exploring the psychometric properties of both EPDS and PHQ-9 to assess maternal depression, with the general aim of particularly improving the quality of assessment of antenatal mental health. It remains crucial to establish which symptoms can be considered reliable and valid indicators of antenatal depression. Further validations of both EPDS and PHQ-9 in other countries using larger sample sizes are recommended to support the advancement of research and clinical guidelines for the appropriate screening of maternal mental health.

Authors' contributions

AS and FM developed the outline of this manuscript, performed the statistical analysis, and contributed to the writing; LC, AT, and GP searched the literature, performed quality analysis and contributed to writing; AG contributed to the final version of the manuscript and supervised the entire study. All Authors read and approved the final manuscript.

The first Author (AS) worked on this article with funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement n. 101030608.

Conflict of interest statement

The Authors have no conflicts of interest to declare. All co-Authors have seen and agree with the contents of the manuscript and there is no financial interest to report.

Received on 13 September 2023.

Accepted on 10 January 2024.

REFERENCES

- Howard LM, Khalifeh H. Perinatal mental health: a review of progress and challenges. *World Psychiatry*. 2020;19(3):313-27. doi: <https://doi.org/10.1002/wps.20769>
- Al-Abri, K, Edge D, Armitage CJ. Prevalence and correlates of perinatal depression. *Soc Psychiatry Psychiatr Epidemiol*. 2023;58:1581-90. doi: <https://doi.org/10.1007/s00127-022-02386-9>
- Yin, X, Sun, N, Jiang N, Xu X, Gan Y, Zhang J, et al. Prevalence and associated factors of antenatal depression: Systematic reviews and meta-analyses. *Clin Psychol Rev*. 2021;83:101932.
- Cena L, Mirabella F, Palumbo G, Gigantesco A, Trainini A, Stefana A. Prevalence of maternal antenatal and postnatal depression and their association with sociodemographic and socioeconomic factors: A multicentre study in Italy. *Journal of Affective Disorders*. 2021;279:217-21. doi: <https://doi.org/10.1016/j.jad.2020.09.136>
- Faisal-Cury A, Rodrigues DMO, Matijasevich A. Are pregnant women at higher risk of depression underdiagnosis? *J Affective Disorders*. 2021;283:192-7. doi: <https://doi.org/10.1016/j.jad.2021.01.057>
- Bauman BL, Ko JY, Cox S, D'Angelo Mph DV, Warner L, Folger S, Tevendale HD, Coy KC, Harrison L, Barfield WD. Vital signs: postpartum depressive symptoms and provider discussions about perinatal depression – United States, 2018. *Morbidity Mortality Weekly Report*. 2020;69(19):575-81. doi: <https://doi.org/10.15585/mmwr.mm6919a2>
- Dadi AF, Miller ER, Bisetegn TA, Mwanri L. Global burden of antenatal depression and its association with adverse birth outcomes: an umbrella review. *BMC Public Health*. 20220(1):173. doi: <https://doi.org/10.1186/s12889-020-8293-9>

8. Jahan N, Went TR, Sultan W, Sapkota A, Khurshid H, Qureshi IA, Alfonso M. Untreated depression during pregnancy and its effect on pregnancy outcomes: A systematic review. *Cureus*. 2021;13(8):e17251. doi: 10.7759/cureus.17251
9. Knapp M, Wong, G. Economics and mental health: the current scenario. *World Psychiatry*. 2020;19(1):3-14. doi: <https://doi.org/10.1002/wps.20692>
10. Reilly N, Kingston D, Loxton D, Talcevska K, Austin, MP. A narrative review of studies addressing the clinical effectiveness of perinatal depression screening programs. *Women and Birth*. 2020;33(1):51-9. doi: <https://doi.org/10.1016/j.wombi.2019.03.004>
11. Smith A, Twynstra J, Seabrook JA. Antenatal depression and offspring health outcomes. *Obstetric Medicine*. 2020;13(2):55-61. doi: <https://doi.org/10.1177/1753495X1984301>
12. Camoni L, Mirabella F, Gigantesco A, Brescianini S, Ferri M, Palumbo G, Calamandrei G, on Behalf of the Perinatal Mental Health Network. The Impact of the COVID-19 pandemic on women's perinatal mental health: Preliminary data on the risk of perinatal depression/anxiety from a national survey in Italy. *Int J Environ Res Public Health*. 2022;19(22):14822. doi: <https://doi.org/10.3390/ijerph192214822>
13. Cena L, Palumbo G, Mirabella F, Gigantesco A, Stefana A, Trainini A, Tralli N, Imbasciati A. 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. doi: <https://doi.org/10.3389/fpsyg.2020.00365>
14. Cena L, Rota M, Calza S, Massardi B, Trainini A, Stefana A. Estimating the impact of the COVID-19 pandemic on maternal and perinatal health care services in Italy: Results of a self-administered survey. *Front Public Health*. 2021;9:701638. doi: <https://doi.org/10.3389/fpubh.2021.701638>
15. Accortt EE, Wong MS. It is time for routine screening for perinatal mood and anxiety disorders in obstetrics and gynecology settings. *Obstetrical & Gynecological Survey*. 2017;72(9):553-68. doi: <https://doi.org/10.1097/OGX.0000000000000477>
16. Gigantesco A, Palumbo G, Cena L, Camoni L, Trainini A, Stefana A, Mirabella F. A brief depression screening tool for perinatal clinical practice: The performance of the PHQ-2 compared with the PHQ-9. *Journal of Midwifery & Women's Health*. 2022;67(5):586-92. doi: <https://doi.org/10.1111/jmwh.13389>
17. Cena L, Biban P, Janos J, Lavelli M, Langfus J, Tsai A, Youngstrom EA, Stefana A. The collateral impact of COVID-19 emergency on neonatal intensive care units and family-centered care: challenges and opportunities. *Front Psychol*. 2021;12:630594. doi: <https://doi.org/10.3389/fpsyg.2021.630594>
18. Smith MS, 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 Dis*. 2022;298:634-43. doi: <https://doi.org/10.1016/j.jad.2021.11.011>
19. Stefana A, Langfus JA, Palumbo G, Cena L, Trainini A, Gigantesco A, Mirabella F. Comparing the factor structures and reliabilities of the EPDS and the PHQ-9 for screening antepartum and postpartum depression: a multigroup confirmatory factor analysis. *Archives of Women's Mental Health*. 2023;26(5):659-68. doi: <https://doi.org/10.1007/s00737-023-01337-w>
20. Benvenuti P, Ferrara M, Niccolai C, Valoriani V, Cox JL. The Edinburgh Postnatal Depression Scale: validation for an Italian sample. *J Affect Dis*. 1999;53(2):137-41. doi: [https://doi.org/10.1016/s0165-0327\(98\)00102-5](https://doi.org/10.1016/s0165-0327(98)00102-5)
21. Hewitt C, Gilbody S, Brealey S, Paulden M, Palmer S, Mann R, Green J, Morrell J, Barkham M, Light K, Richards D. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. *Health Technol Assessment*. 2009;13(36):1-145,147-230. doi: 10.3310/hta13360
22. Smith-Nielsen J, Matthey S, Lange T, Væver MS. Validation of the Edinburgh Postnatal Depression Scale against both DSM-5 and ICD-10 diagnostic criteria for depression. *BMC Psychiatry*. 2018;18(1):393. doi: 10.1186/s12888-018-1965-7
23. Lyubenova A, Neupane D, Levis B, Wu Y, Sun Y, He C, Krishnan A, Bhandari PM, Negeri Z, Imran M, Rice DB, Azar M, Chiovitti MJ, Saadat N, Riehm KE, Boruff JT, Ioannidis JPA, Cuijpers P, Gilbody S, Kloda LA, et al. Depression prevalence based on the Edinburgh Postnatal Depression Scale compared to structured clinical interview for DSM Disorders classification: Systematic review and individual participant data meta-analysis. *Int J Methods in Psychiatr Res*. 2021;30(1):e1860. doi: 10.1002/mpr.1860
24. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Int Med*. 2001;16(9):606-13. doi: 10.1046/j.1525-1497.2001.016009606.x
25. Mazzotti E, Fassone G, Picardi A, Sagoni E, Ramieri L, Lega I, Pasquini P. 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). *Italian J Psychopath*. 2003;9:235-42.
26. Thase ME. Recommendations for screening for depression in adults. *JAMA*. 2016;315(4):349-50. doi: <https://doi.org/10.1001/jama.2015.18406>
27. Levis B, Benedetti A, Ioannidis JPA, Sun Y, Negeri Z, He C, Wu Y, Krishnan A, Bhandari PM, Neupane, D, Imran M, Rice DB, Riehm KE, Saadat N, Azar M, Boruff J, Cuijpers P, Gilbody S, Kloda LA, McMillan D, et al. Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: individual participant data meta-analysis. *J Clin Epidemiol*. 2020;122:115-28. e1. doi: 10.1016/j.jclinepi.2020.02.002
28. Spielberger CD. *Manual for the State-Trait Inventory STAI (Form Y)*. Palo Alto, CA: Consulting Psychologists Press; 1983.
29. Spielberger CD, Pedrabissi L, Santinello M. *Inventario per l'ansia di stato e di tratto: Nuova versione italiana dello STAI, forma Y: manuale*. Firenze: Organizzazioni speciali; 1989.
30. Stefana A, Cena L, Trainini A, Palumbo G, Mirabella F, Gigantesco A. Development of a short version of the Spielberger state and trait anxiety inventory for pregnant women. *PsyArXiv*. 2023. doi: <https://doi.org/10.31234/osf.io/vtgbj>
31. Steiner MD, Grieder S. EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools. *J Open Source Software*. 2020;5(53):2521. doi: <https://doi.org/10.21105/joss.02521>
32. Rosseel Y. Lavaan: An R package for structural equation modeling. *J Statistical Software*. 2011;48:1-36.
33. Hoyle RH. *Handbook of structural equation modelling*. 2nd ed. New York, NY: The Guilford Press; 2023.
34. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999;6(1):1-55. doi: <https://doi.org/10.1080/10705519909540118>

35. Revelle W, Zinbarg RE. Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*. 2009;74(1):145-54. doi: <https://doi.org/10.1007/s11336-008-9102-z>
36. Youngstrom EA, Van Meter A, Frazier TW, Hunsley J, Prinstein MJ, Ong ML, Youngstrom JK. Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*. 2017;24(4):331-63. doi: <https://doi.org/10.1111/cpsp.12207>
37. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *J Marketing Res*. 1981;18:39-50. doi: <https://doi.org/10.2307/3151312>
38. Hair JF, Black WC, Babin BJ, Anderson RE. *Multivariate data analysis*. 8th ed. Canage; 2019.
39. Matthey, S. Using the Edinburgh Postnatal Depression Scale to screen for anxiety disorders. *Depression and anxiety*. 2008;25(11):926-31. doi: <https://doi.org/10.1002/da.20415>
40. Pan Y, Xu J. Can EPDS and EPDS-3A be used to replace GAD-7 to screen the anxiety of pregnant women during pregnancy examination? *Int J Gynaecol Obstet*. 2024;164(3):902-11. doi: 10.1002/ijgo.15053
41. Della Vedova AM, Loscalzo Y, Giannini M, Matthey S. An exploratory and confirmatory factor analysis study of the EPDS in postnatal Italian-speaking women. *J Reprod Infant Psychol*. 2022;40(2):168-80. doi: 10.1080/02646838.2020.1822993
42. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropractic Med*. 2016;15(2):155-63. doi: <https://doi.org/10.1016/j.jcm.2016.02.012>
43. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991;59(1):12-9. doi: 10.1037//0022-006x.59.1.12
44. Streiner DL, Norman GR, Cairney J. *Health measurement scales: A practical guide to their development and use*, 5th ed. Oxford: Oxford University Press; 2015.
45. Ratner B. The correlation coefficient: Its values range between+ 1/- 1, or do they? *Journal of Targeting, Measurement and Analysis for Marketing*. 2009;17(2):139-42. doi: <https://doi.org/10.1057/jt.2009.5>
46. Reise SP. The rediscovery of bifactor measurement models. *Multivariate Behav Res*. 2012;47:667-96. doi: <https://doi.org/10.1080/00273171.2012.715555>
47. Zhong Q, Gelaye B, Rondon M, Sánchez SE, Garcí, PJ, Sánchez E, Barrios YV, Simon G E, Henderson, DC, Cripe SM, Williams MA. Comparative performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for screening antepartum depression. *J Affect Dis*. 2014;162:1-7. doi: <https://doi-org.proxy.unibs.it/10.1016/j.jad.2014.03.028>
48. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*, 5th rev ed. 2022. doi: <https://doi.org/10.1176/appi.books.9780890425787>
49. Lara-Cinisomo S, Akinbode TD, Wood J. A systematic review of somatic symptoms in women with depression or depressive symptoms: Do race or ethnicity matter? *Journal of Women's Health*. 2020;29(10):1273-82. doi: <https://doi.org/10.1089/jwh.2019.7975>
50. Klein MH, Essex MJ. Pregnant or depressed? The effect of overlap between symptoms of depression and somatic complaints of pregnancy on rates of major depression in the second trimester. *Depression*. 1994;2(6):308-14. doi: <https://doi.org/10.1002/depr.3050020606>
51. Sugawara M, Sakamoto S, Kitamura T, Toda MA, Shima S. Structure of depressive symptoms in pregnancy and the postpartum period. *J Affect Dis*. 1999;54(1-2):161-9. doi: [https://doi.org/10.1016/s0165-0327\(98\)00187-6](https://doi.org/10.1016/s0165-0327(98)00187-6)
52. Yonkers KA, Smith MV, Gotman N, Belanger K. Typical somatic symptoms of pregnancy and their impact on a diagnosis of major depressive disorder. *General Hospital Psychiatry*. 2009;31(4):327-33. doi: <https://doi.org/10.1016/j.genhosppsych.2009.03.005>
53. Snaith RP, Constantopoulos AA, Jardine MY, McGuffin P. A clinical scale for the self-assessment of irritability. *British J Psychiatry*. 1978;132:164-71. doi: <https://doi.org/10.1192/bjp.132.2.164>
54. Zigmond AS, Snaith RP. Hospital Anxiety and Depression Scale (HADS). *APA PsycTests*. 1983. doi: <https://doi.org/10.1037/t03589-000>
55. Bedford A, Fiulds GA, Sheffield BF. A new personal disturbance scale (DSSI/sAD). *British J Social Clin Psychol*. 1976;15(4):387-94. doi: <https://doi.org/10.1111/j.2044-8260.1976.tb00050.x>
56. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*, 4th ed. Washington, DC: APA; 1994.
57. Bina R, Harrington D. The Edinburgh Postnatal Depression Scale: screening tool for postpartum anxiety as well? Findings from a confirmatory factor analysis of the Hebrew version. *Maternal Child Health J*. 2016;20(4):904-14. doi: <https://doi.org/10.1007/s10995-015-1879-7>
58. Loyal D, Sutter AL, Rasclé N. Screening beyond postpartum depression: Occluded anxiety component in the EPDS (EPDS-3A) in French mothers. *Maternal Child Health J*. 2020;24(3):369-77. doi: <https://doi.org/10.1007/s10995-020-02885-8>
59. Matthey S, Fisher J, Rowe H. Using the Edinburgh postnatal depression scale to screen for anxiety disorders: Conceptual and methodological considerations. *J Affect Dis*. 2013;146(2):224-30. doi: <https://doi.org/10.1016/j.jad.2012.09.009>
60. Cena L, Gigantesco A, Mirabella F, Palumbo G, Camoni L, Trainini A, Stefana A. Prevalence of comorbid anxiety and depressive symptomatology in the third trimester of pregnancy: Analysing its association with sociodemographic, obstetric, and mental health features. *J Affect Dis*. 2021;295:1398-406. doi: <https://doi.org/10.1016/j.jad.2021.09.015>
61. Brouwers EP, van Baar AL & Pop VJ. Does the Edinburgh Postnatal Depression Scale measure anxiety? *J Psychosomatic Res*. 2001;51(5):659-63. doi: [https://doi.org/10.1016/s0022-3999\(01\)00245-8](https://doi.org/10.1016/s0022-3999(01)00245-8)
62. Fairbrother N, Corbyn B, Thordarson DS, Ma A, Surm D. Screening for perinatal anxiety disorders: Room to grow. *J Affect Dis*. 2019;250:363-70. doi: <https://doi.org/10.1016/j.jad.2019.03.052>
63. Tuohy A, McVey C. Subscales measuring symptoms of non-specific depression, anhedonia, and anxiety in the Edinburgh Postnatal Depression Scale. *British J Clin Psychol*. 2008;47(2):153-69. doi: <https://doi.org/10.1111/j.2044-8260.2008.tb00463.x>
64. National Institute of Mental Health, NIH. *Anxiety disorders. Transforming the understanding and treatment of mental illnesses*. Bethesda, MD: NIMH; 2018. Available from: <https://www.nimh.nih.gov/health/topics/anxiety-disorders/index.shtml>