Journal of Affective Disorders Reports 25 (2026) 101118

Contents lists available at ScienceDirect

Journal of Affective Disorders Reports

journal homepage: www.sciencedirect.com/journal/journal-of-affective-disorders-reports

Research Paper

A comparative psychometric evaluation of the Edinburgh Postnatal Depression Scale and the Montgomery-Åsberg Depression Rating Scale-Self Report as measures of depression during pregnancy

Erik Forsell [a,*], Sandra Frööjd [a], Emelie Svanerud George [a], Mikaela Bergmark Dunkels [a,b], Marie Bendix [a,c]

[a] Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Sweden
[b] Department of Psychology, Stockholm University, Stockholm, Sweden
[c] Department of Clinical Sciences, Umeå University, Umeå, Sweden

ARTICLE INFO

Keywords:
Depressive symptoms
Pregnancy
Edinburgh Postnatal Depression Scale (EPDS)
Montgomery Åsberg Depression Rating Scale (MADRS)
Psychometry
Antenatal depression

ABSTRACT

Introduction: The Edinburgh Postnatal Depression Scale (EPDS) is the most widely used instrument in perinatal clinical practice and research based on a belief that somatic symptoms during the perinatal period impact on depression measurement validity and reliability. The Montgomery Åsberg Depression Rating Scale – Self report (MADRS-S) may have better psychometric performance as a screening and outcome instrument for depression.
Methods: Self-referred pregnant women were screened (n = 650), diagnostically assessed (n = 390), treated with internet-based cognitive behavior therapy for depression (n = 270) and followed-up diagnostically (n = 169). EPDS and MADRS-S were assessed at screening and after treatment. Item level data from all patients were used in a series of psychometric tests including confirmatory factor analysis (CFA) and item response-theory tests (IRT).
Results: Both scales showed adequate reliability. Using depression as a unidimensional latent trait in CFA, MADRS-S fitted data better than EPDS. IRT revealed that the somatic items in MADRS-S contributed less than average rather than inflating the totals. For more depressed respondents, the EPDS scores' validity decreased compared to MADRS-S.
Conclusions: MADRS-S and EPDS are reliable tools for assessing depressive symptoms in pregnant patients, but MADRS-S outperforms EPDS on several key psychometric indicators. Contrary to common belief, somatic items seem not to inflate scores, and MADRS-S retains better precision at higher symptom levels. These findings support the use of standard depression measures, such as the MADRS-S, in pregnant populations. Doing so may improve detection and treatment by integrating antenatal depression into general psychiatric care and research rather than isolating it within perinatal-specific frameworks.

1. Introduction

Depression occurs in 10–15% of pregnant women and is one of the most common pregnancy complications (Woody et al., 2017). Depression during pregnancy (antenatal depression) increases the risk for mental and physical complications in both mother and child as well as for postpartum depression (Howard and Khalifeh, 2020). More than half of pregnant women with depression are neither detected nor treated (Eakley and Lyndon, 2024). The treatment gap is partly explained by lack of identification (Faisal-Cury et al., 2021). In mild to moderate depression, cognitive behavioral therapy (CBT), including internet-delivered CBT (ICBT), is considered as a first-line treatment (Cuijpers et al., 2023; Loughnan et al., 2019; NICE, 2020; Pettman et al., 2023).

Screening is essential for identifying women at risk of depression, but a comprehensive diagnostic assessment is necessary to determine the appropriate treatment (Rondung et al., 2024). Depressive symptoms can occur across a range of mental health disorders, not just in major depressive disorder (MDD). MDD itself is a heterogeneous condition, encompassing disturbances in arousal and regulatory systems (e.g., sleep, fatigue, appetite) as well as the negative valence systems, including suicidal thoughts (Gunzler et al., 2020). Both the ICD and DSM

* Corresponding author.
E-mail address: erik.forsell@ki.se (E. Forsell).

https://doi.org/10.1016/j.jadr.2026.101118
Received 22 January 2026; Received in revised form 28 May 2026; Accepted 12 July 2026
Available online 13 July 2026
2666-9153/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

systems rely on a combination of criteria to guide clinicians during clinical assessments. Self-report scales can be valuable in supporting the diagnostic process by enhancing reliability and measuring symptom severity (Svanborg and Asberg, 2001).

A high-quality self-report instrument must be reliable, valid, sensitive, and specific. It should produce consistent results, accurately measure the intended construct, detect meaningful changes, and minimize unrelated factors. Practicality is also key, with easy administration, scoring, and interpretation. While instruments used only for screening should be characterized by high sensitivity, instruments used as outcome measures or in research to quantify severity (i.e. not just serve as an indicator that another more comprehensive assessment is needed) should be more specific and sensitive to changes, focused, and have strong psychometric properties (Svanborg and Asberg, 2001; Uher et al., 2012).

Some studies suggest that depressive symptoms during pregnancy may differ from those in the general population, but there is no strong evidence to distinguish perinatal depression from MDD outside the perinatal period (Batt et al., 2020; Nonacs and Cohen, 2003; Su et al., 2007; Viktorin et al., 2016; Yonkers et al., 2009). However, diagnosing MDD in pregnant women is challenging due to symptom overlap with common pregnancy-related issues, such as sleep disturbances from back pain, appetite changes due to nausea, and fatigue or low mood from anemia (Balserak and Lee, 2011; Benson et al., 2022; Lutterodt et al., 2019).

The Edinburgh Postnatal Depression Scale (EPDS; see Table 1) was originally developed as an easy-to-administer, unidimensional screening tool for postpartum depression, with high face validity and a reduced emphasis on somatic symptoms to minimize the risk of capturing unrelated symptoms (Cox et al., 1987; Rubertsson et al., 2011). It has since been validated as a screening instrument in pregnant women and other populations (Bergink et al., 2011; Cox et al., 1996; de Cock et al., 2011; Kozinszky and Dudas, 2015; Rubertsson et al., 2011).

Today, EPDS is the most widely used tool in perinatal clinical practice and research (Hewitt et al., 2010). A recent study by (Qiu et al., 2023) highlighted that a nine-item version, omitting the self-harm item, performs similarly and may be especially useful in research, as it reduces the need for patient follow-up. However, EPDS's widespread use has faced criticism, particularly regarding its role as a diagnostic and outcome measure. Even the instrument's originators have denounced its use as a diagnostic tool (Cox, 2017, 2019). EPDS lacks formal diagnostic criteria for depression and includes ambiguously worded questions (Matthey and Agostini, 2017; Rubertsson et al., 2011). It has been shown to identify borderline personality disorder and traits, as well as eating and obsessive-compulsive disorders, at similar rates to depression and anxiety (Judd et al., 2019; Lydsdottir et al., 2014). The use as a screening instrument for perinatal depression has been criticized as EPDS lacks a threshold value with satisfactory sensitivity without a simultaneous specificity, resulting in a high prevalence of false positives (SBU, 2021). Cut-off scores seem to depend on trimester (Su et al., 2007) and differ between populations (Matthey and Ross-Hamid, 2012).

Furthermore, it has been suggested that EPDS might be interpreted differently depending on the individual's level of distress, affecting reliability (Cunningham et al., 2015). Several studies report a two-factor structure indicating that EPDS captures both depression and anxiety (Massoudi et al., 2013; Matthey, 2008; Matthey et al., 2013; Phillips et al., 2009; Pop et al., 1992). Other studies suggest a three-factor structure; depression, anhedonia and anxiety (Coates et al., 2017). In response to these concerns, various studies have recommended modifications to improve EPDS's psychometric properties (Bina and Harrington, 2016; Zhong et al., 2015). These efforts assume that somatic symptoms are either less relevant in perinatal depression or may comprise the scale's sensitivity. Rather than replacing EPDS with more robust tools, the focus has been on refining the existing measure. Given the recurrent nature of depression, using the same instrument across perinatal and non-perinatal periods has been suggested to allow for meaningful comparisons between groups and within individuals over time (Peng et al., 2024).

The Montgomery and Åsberg Depression Rating Scale (MADRS) was developed from the Comprehensive Psychopathological Rating Scale (CPRS; (Asberg et al., 1978; Montgomery and Asberg, 1979)) with the aim of measuring the severity of depressive symptoms sensitive to change in antidepressant treatment studies (Montgomery and Asberg, 1979). Both the 10-item expert-rated and 9-item self-rated versions (Svanborg and Asberg, 1994) are widely used in clinical and research settings. The Montgomery and Åsberg Depression Rating Scale - Self report (MADRS-S; see Table 1) has been validated for use with depressed patients in inpatient and outpatient settings, including via internet and telephone applications, but has not been specifically validated during pregnancy (Bondolfi et al., 2010; Fantino and Moore, 2009; Svanborg and Asberg, 1994, 2001; van Ballegooijen et al., 2016). It has demonstrated high reliability (Svanborg and Asberg, 2001) and can distinguish depression from anxiety and other mental disorders (Svanborg and Asberg, 2001). However, the relationship between scale scores and clinical depression severity remains unclear (Leucht et al., 2017). The self-rated version has been criticized for a lack of research on its cut-off values (SBU, 2012). In screening contexts, MADRS-S performs similarly to the Beck Depression Inventory (BDI; Svanborg and Ekselius, 2003), though cut-off scores vary by population (Duarte-Guerra et al., 2016), necessitating validation for different patient groups.

Factor analyses in non-perinatal populations have yielded inconsistent results: one study by Fantino and Moore (2009) found a one-dimensional structure accounting for 45% of the variance, while Thorndike et al. (2009) described a three-dimensional structure, including dysphoric apathy/retardation, psychic anxiety, and vegetative symptoms (sleep, appetite). In perinatal studies, adaptations of MADRS have been used to account for pregnancy-related changes in sleep (Kittel-Schneider et al., 2022; Wickberg and Hwang, 1996a, 1996b).

Few studies have compared EPDS with MADRS or MADRS-S and other scales during pregnancy. While EPDS showed a higher sensitivity (95%) than MADRS (91%) and BDI (68%), MADRS showed the highest specificity (95%), compared to EPDS (93%) and BDI (88%) (Harris et al., 1989). However, a more recent meta-analysis finds the overall sensitivity and specificity of EPDS to be a bit lower (sensitivity 84%, specificity 87% at cutoff $\geq$10) (Rondung et al., 2024). One study found good agreement between the Hamilton Depression Rating Scale (HAM-D), MADRS, and EPDS in identifying depression in pregnant and postpartum

Table 1
Comparison of The Edinburgh Postnatal Depression Scale (EPDS) and The Montgomery and Åsberg Depression Rating Scale – Self report (MADRS-S).

	EPDS	MADRS-S
Scale property	10 item, 4-point likert scale, 0–3 points, range 0–30	9 item, 6-point likert scale (3 undefined), 0–6 points, range 0–54
Time frame	7 days	3 days
Somatic symptoms	-	Appetite, Sleep
Suicidality	Self-harm	Suicidal stages
Remote use validation	Internet (Spek et al., 2008)	Telephone (Hedman et al., 2013), Internet (Holländare, et al., 2008)
Items	1. Ability to laugh 2. Emotional involvement 3. Self-blame 4. Anxiety 5. Panic 6. Overwhelmed 7. Sleep problems due to being unhappy 8. Mood 9. Tearfulness 10. Self-harm	1. Mood 2. Anxiety 3. Sleep 4. Appetite 5. Ability to concentrate 6. Initiative 7. Emotional involvement 8. Pessimism 9. Will to live (suicidal ideation and plans)

Notes: EPDS=Edinburgh Postnatal Depression Scale, MADRS-S= Montgomery-Åsberg Depression Rating Scale- Self report.

women (D'Agostino et al., 2024). EPDS and MADRS were significantly correlated from the second half of pregnancy through the first postnatal year. During pregnancy, 11% scored ≥10 on EPDS, while 21% scored ≥7 on MADRS. After excluding the sleep item, 13.8% of pregnant women screened positive for depression, while screening prevalence remained unchanged for fathers (Kittel-Schneider et al., 2022). A clinical trial in antenatal depression found similar remission and response rates for the EPDS, MADRS-S, and HAM-D, with MADRS-S showing slightly better sensitivity to change (Donmez et al., 2022). Another study reported moderate effect sizes for MADRS, while effect sizes for EPDS were large (Green et al., 2020). However, findings are mixed and Forsell et al. (2017) reported significant treatment effects on MADRS-S but not on EPDS in an RCT of ICBT for antenatal depression, indicating that the instruments may differ in their responsiveness to change.

Although EPDS was originally developed as a screening tool in the perinatal population, and MADRS-S as a general measure of depressive symptom severity, both instruments are today frequently used beyond their original scope. In real-world perinatal care, self-report scales are commonly applied not only for screening purposes, but also for severity assessment and symptom monitoring over the course of treatment. However, empirical evidence supporting these broader uses, particularly in pregnant populations, remains limited.

1.1. Aims

We aimed to perform a psychometric evaluation of MADRS-S and EPDS as screening and outcome measures for depression in pregnant women in a clinically relevant context where depression is already suspected. Specifically, we examined the scales' internal consistency, factor structure, and sensitivity to change, as well as their correspondence with diagnostic status where applicable. We aimed to provide evidence of their respective fit for current clinical use as aid in assessment of severity and symptom monitoring during treatment, rather than to propose new instruments or modifications of the scales. This study extends and replicates previous psychometric findings by leveraging a large, clinically characterized sample of pregnant help-seeking women.

2. Methods

2.1. Participants and procedure

The present study draws on data from a subset of participants in the DANA trial. The DANA trial investigates Internet-based Cognitive Behavioral Therapy (ICBT) for antenatal depression. Pregnant individuals between week 8 to 29 were recruited nationwide in Sweden through social media advertisements and maternal health care centers. The participants completed online questionnaires through the study's website between February 2021 and April 2024 (subgroup 1, n = 650). The study was approved by the Swedish Ethical Review Authority (Dnr 2020-06049). All patients provided written informed consent.

Those that were eligible after screening (MADRS-S score between 15–35, legal age, proficiency in Swedish,enrolled in prenatal care) was assessed with the depression segment of SCID-I adapted for the perinatal period and a structured diagnostic interview consisting of multiple modules to assess comorbidity. Those fulfilling inclusion criteria (MDD; stable medication at assessment; no ongoing psychological treatment that could interfere with the study's treatment and not clinically assessed to have highly increased suicide risk; severe psychiatric comorbidity or a medical or pregnancy related condition that was deemed a significant contraindication for participation) were included . Subgroup 2 (n = 283) completed online questionnaires at the end of treatment. Subgroup 3 (n = 190) also participated in a follow-up telephone interview conducted at the end of treatment by trained assessors and included the depression segment of SCID-I. Subgroups 2 and 3 consisted of participants progressing from subgroup 1, with decreasing sample sizes reflecting participant flow across study stages. This included attrition due to dropout, and participants who had not yet reached the relevant assessment time points, as well as participants who were excluded.*Intervention*

Participants received a pregnancy-adapted therapist-guided ICBT treatment for depression, used in regular care at Internet Psychiatry in Stockholm, based on (Forsell et al., 2017). The program includes 10 modules over 10 weeks. Modules contain reading materials, exercises, and assignments, which licenced psychologists communicate with participants through a protected messaging system.

2.2. Measures

2.2.1. MADRS-S

The Montgomery-Åsberg Depression Rating Scale Self-report version (MADRS-S) is a widely used and validated scale for measuring depression (Montgomery and Asberg, 1979; Svanborg and Asberg, 1994, 2001). It consists of 9 items on a 7-point Likert-scale with a maximum score of 54.

2.2.2. EPDS

The Edinburgh Postnatal Depression Scale (EPDS; (Cox et al., 1987) is a self-assessment questionnaire designed to measure postpartum depression and has also been validated for use within the pregnant population (Cox et al., 1987; Murray and and Cox, 1990). Each of the ten questions is rated on a four-point likert scale, with a maximum score of 30. EPDS has been validated in Sweden on the pregnant population (Rubertsson et al., 2011). In the Swedish validation study, a cutoff score of 13 indicated a likely diagnosis of depression.

2.2.3. SCID-I

Diagnosis of MDD before and after treatment was assessed with an adapted version of the depression module from the Structured Clinical Interview for DSM-IV Axis I diagnoses (SCID-I) (Spitzer et al., 1992). The adaptations made were tailored to focus on the perinatal context, with suggestions of additional follow-up questions to differentiate between symptoms of depression and normal pregnancy experiences, particularly in areas such as appetite and sleep changes. Furthermore, we modified the interview to ensure compatibility with the diagnostic criteria of the DSM-5, the current edition of the Diagnostic and Statistical Manual of Mental Disorders. Assessments were conducted by a psychology assistant, psychologist, or psychotherapy-trained midwife, under the supervision of a psychiatrist and a psychologist with expertise in perinatal psychiatry and ICBT. The initial questionnaire included demographic questions covering age, income, education, and marital status, along with reproductive information such as parity and reports of pregnancy symptoms.

2.3. Statistical analysis

2.3.1. Analytical sample

Due to varying data availability, different samples were used for different analyses. The three subgroups represent consecutive stages of participation in the study. Subgroup 1 comprised self-referred participants seeking treatment for self-identified depression, whereas subgroups 2 and 3 consisted of progressively smaller subsets who were assessed as meeting criteria for depression, with measurements obtained following treatment. Statistical analyses were performed using the largest available sample for each variable. Unless otherwise specified, analyses were based on subgroup 1 on which no inclusion or exclusion criteria have been applied.

2.3.2. Data diagnostics

Data screening for psychometric analyses was performed with Kaiser–Meyer–Olkin (KMO) test of sampling adequacy and Bartlett's test of sphericity as well as skewness and kurtosis.

2.3.3. Confirmatory factor analysis (CFA)

MADRS-S and EPDS are both intended to be used and are used as unidimensional measures of depression. Therefore, we conducted CFA, focused on assessing the fit of these scales as unidimensional measures of depression. This was done using the "lavaan" package in R (Rosseel, 2012). Fit indices were compared to general guidelines and to each other.

2.3.4. Reliability

McDonald's *Omega* was calculated using the MBESS-package in R (Kelley, 2017) and 10 000 bootstraps to get 95% CIs of the estimate in accordance with the recommendations by Dunn et al. (2014).

2.3.5. Item response theory analyses (IRT)

Item response theory (IRT) analyses were conducted in R using the *mirt* package. The Graded Response Model (GRM) was selected, as it is epistemologically the most appropriate model for MADRS-S and EPDS, since it is intended for items where each category represents increasing levels of a single latent trait. For the GRM we examined the Infotrace plots, which indicate the relative information value of each item towards the latent trait. That is, how informative each item is towards the respondent's total. Parameters were estimated using marginal maximum likelihood estimation with the expectation-maximization algorithm. Additionally, since there is limited evidence on the uniformity of distances between the different response levels on both MADRS-S and EPDS, the widely used partial credit model (PCM) was also used, as it does not assume equidistant response levels. The PCM model with ICC (intraclass correlation) curve graphs indicates the probability of each possible response on each item at various levels of the latent trait. Missing data were handled using full-information maximum likelihood estimation within the model-fitting procedure. IRT analyses were conducted separately for the two measurements which allowed examination of item and test information characteristics within each instrument while avoiding additional assumptions associated with concurrent calibration, such as strict unidimensionality across the combined item pool and invariance of the latent metric. Comparisons between instruments were therefore interpreted descriptively rather than inferentially.

2.3.6. Change analysis

Data from Pre- to Post-treatment were used to compare change both as raw scores and by z-transforming the change scores to allow for more direct comparison. Change analysis was performed on subgroup 2.

2.3.7. Correlation with gestational week

Correlation with gestational week was calculated on subgroup 2. The correlation between each item and the sum scores for both scales and gestational week reported by the patient at screening was calculated to assess potential influence from gestational week on responses.

2.3.8. Receiver operator characteristics curve (ROC) analysis for diagnosis of depression post-treatment

ROC analysis was performed on subgroup 3. Post-treatment symptom scores for the MADRS-S and the EPDS were used as indicators of a diagnosis of depression remaining after reassessment (SCID) post-treatment using ROC-analysis with the pROC-package in R (Robin et al., 2011) and DeLong's test for two correlated ROC curves. Area under the curve was assessed both using the established cutoffs for depression and by allowing the analysis to pick the optimal cutoff. In the primary ROC analyses, the optimal cut-off was determined using the default procedure in the pROC package for R, which identifies the threshold maximizing Youden's J statistic. Additional analyses were conducted using previously established cut-off values from prior literature (11 points for MADRS-S, Forsell, et al., 2020, Forsell, et al., 2025;13 for EPDS, Rubertsson et al., 2011).

3. Results

3.1. Baseline characteristics

Demographic information for participants in subgroup 1, which constitutes the primary sample for the majority of the analysis is provided in Table 2. The mean age of participants was 32.71 (SD = 4.18). The majority were in the second trimester (59%) during self-referral, and 86% were diagnosed with MDD at assessment.

3.2. Data diagnostics

KMO test of sampling adequacy was .81 and .79 for MADRS-S and EPDS, respectively.

Bartlett's test of sphericity indicated that both measures were suitable for factor analysis ($p<.001$). MADRS-S and EPDS both showed slight negative skewness (−.22 and −.24 respectively). MADRS-S was almost perfectly mesokurtic (kurtosis = −.08) and EPDS had a slight leptokurtic distribution (kurtosis = .43), again both within acceptable ranges. We concluded that data was approximately normally distributed and suitable for all planned analyses.

Table 2
Baseline sociodemographic characteristics and clinical characteristics.

Variable			Total n
Age	Mean (sd) Min - max	32.71 (4.18) 18 - 47	650
Pregnancy week at screening	Mean (sd) Min - max	17.84 (6.64) 4 - 38	650
Gestational week Divided in pregnancy trimesters	n (%)	First trimester (week 4 −12): 233 (36%) Second trimester (week 13 −27): 384 (59%) Third trimester (week 28 - 30): 33 (5%)	650
In a committed relationship	n (%)	Yes: 621 (96%) No: (4%)	650
First time parent	n (%)	Yes = 305 (47%) No = 345 (53%)	650
Highest education	n (%)	High school: 120 (18%) University: 530 (82%)	650
Employed	n (%)	Yes = 548 (84%) No = 102 (16%)	650
Recurrent depression	n (%)	Yes = 293 (78%) No = 82 (22%)	375
Diagnosed with Major depression disorder at assessment	n (%)	Yes: 336 (86%) No: 31 (8%) Unclear: 23 (6%)	390*
Reports moderate or a high level of symptoms related to the pregnancy at screening	n (%)	With sleeping difficulties: 248 (38%) With nausea and vomiting: 366 (56,3%) With pelvic pain: 106 (16%)	650

Notes: * = The smaller sample size reflects the inclusion of only participants who completed the assessment. Reductions in sample size occurred due to incomplete data, exclusion criteria, and participant withdrawal.

3.3. Reliability

For MADRS-S, Omega was estimated to be .75 [95% CI = .71-.77] and for EPDS Omega was .74 [95% CI = .69-.77], both indicating adequate reliability.

3.4. Correlation with gestational week

There was a small, but statistically significant correlation between gestational week and MADRS-S item 4 at Pre (r= −.205, p=.001) and Change score (r=.19, p=.007). There were also small but statistically significant correlations between MADRS-S item 8 and 9 at Post treatment and gestational week (r= −.14, p=.049 and r= −.202, p=.004 for items 8 and 9, respectively).

3.5. Testing unidimensionality with CFA

The CFA using a single latent trait (Depression defined as the sum of all items) for both measures are presented in Table 3. All fit indices point towards MADRS-S fitting data better, though still only with moderate or borderline acceptable fit indices. Fit indices for EPDS indicated poor model fit and did not meet any reference values for fit.

3.6. IRT analyses

The test information function (TIF) graphs (Fig. 1), which tests the information values of the full scale, show that at higher levels of depression, EPDS becomes less informative sooner and more steeply than MADRS-S measure, suggesting that for more depressed respondents, EPDS scores' validity decreases compared to MADRS-S.

The infotrace graphs, which show the information value of each item across all levels of the latent trait, are presented in Fig. 2. The comparatively higher jaggedness of the curves for EPDS (item 1, 8 & 9) compared to MADRS-S (item 1 & 7) could indicate that the EDPS has less unidistant responses.

The ICC plots for the PCM models are presented in the Supplemental materials (Supplemental Figs. 1 & 2). For MADRS-S the most uniform response classes were for items 3, 6 and 7 and for EPDS item 3 showed approximately uniform response curves. Both scales are skewed towards the 0-point responses. Most extremely for EPDS item 10, and somewhat for item 5, whereas MADRS-S was most skewed towards 0 scores for items 3 and 4.

3.7. Change from pre- to post-treatment

Table 4 describes the observed change scores across scales and individual items. Change is relatively evenly distributed within the scales and the amount of change observed when normalised (z-score) is approximately the same between MADRS-S and EPDS. One item that stands out is EPDS item 10 (self-harm) which shows both extremely low values at the start and almost no change.

Table 3
Fit indices from confirmatory factor analysis of MADRS-S and EPDS.

Variable	MADRS-S	EPDS	Reference value
Chi-square	185.987	429.956	Ideally non-significant but highly depends on N. Lower is better.
P-value	0.000	0.000	>.05
CFI	0.851	0.705	≥ 0.90
TLI	0.802	0.621	≥ 0.90
RMSEA	0.095	0.132	≤ 0.08
SRMR	0.062	0.094	≤ 0.08

Notes: CFI= Comparative fit index, TLI= Tucker-Lewis index, RMSEA= Root Mean Square Error of Approximation, SRMR= Standardized Root Mean Square Residual.

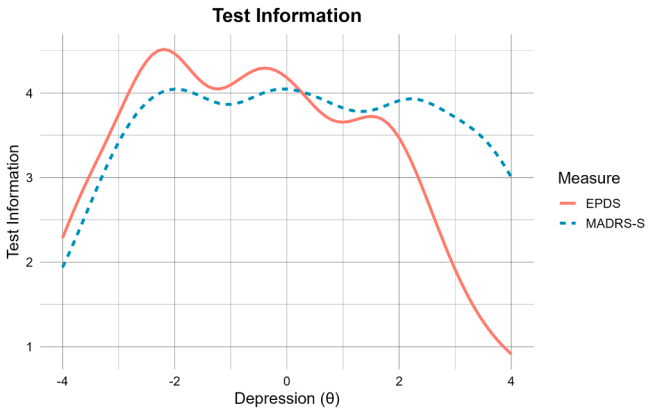

Fig. 1. The test information function (TIF) graphs.

3.8. ROC-analysis

DeLong's test for correlated ROC curves indicated that the difference in AUC for MADRS-S and EPDS in determining depression at post was not significant (AUC= .76 and .75 for MADRS-S and EPDS respectively, p=.54). Though, it is worth noting that the cutoffs selected by the ROC analysis were quite different from the ones suggested in the literature (14.5 and 10.5 for the MADRS-S and EPDS respectively). The ROC-curve graph is presented in the supplemental materials. Using the established cutoffs instead, we calculated confusion matrix statistics and found that the cutoff for MADRS-S (10 or less) had a sensitivity of .78 and a specificity of .60 and the EPDS (12 or less) had a sensitivity of .58 and a specificity of .78. Balanced accuracy was .69 for MADRS-S and .68 for EPDS.

4. Discussion

We find that both MADRS-S and EPDS are reasonable reliable psychometric questionnaires that are both sensitive to change and not highly influenced by gestational week. The confirmatory factor analyses however, show that they are far from perfectly unidimensional. Having moderate model fit for unidimensionality is not surprising in this context. MADRS-S and EPDS are designed to capture various aspects (discrete symptoms, that need not all co-occur) of a highly heterogeneous latent construct (depression). This is a reality that is far less tidy than many other measures that confirmatory factor analysis and item response theory were originally developed to assess.

Item response theory analyses showed that both scales have items that contribute more to the total than others, though interestingly, we find that the somatic items in MADRS-S (3 and 4) contribute less than average and that this seems to be driven by these items getting extra low scores (biased towards zeros) rather than inflating the totals. We also find that MADRS-S retains higher information value with higher levels of depression than EPDS.

We consistently find that MADRS-S is either as good as or outperforms EPDS, despite EPDS being specifically designed for a perinatal population. We believe this is likely due to MADRS-S being a better measure of depression, and that this aspect is more important in this population than avoiding questions that might capture somatic symptoms of pregnancy. EPDS does in fact not only avoid somatic questions, e.g. by sleep loss being conditioned to crying only. It also shifts away from depression specific questions overall and into more general psychological distress (for instance, by mixing questions of low mood with questions about anxiety or moving away from suicidality into self-harm).

Our findings are consistent with previous comparisons between EPDS and other depression specific scales. Peng et al. (2024) compared PHQ-9 to EPDS and reported both scales to be unidimensional with

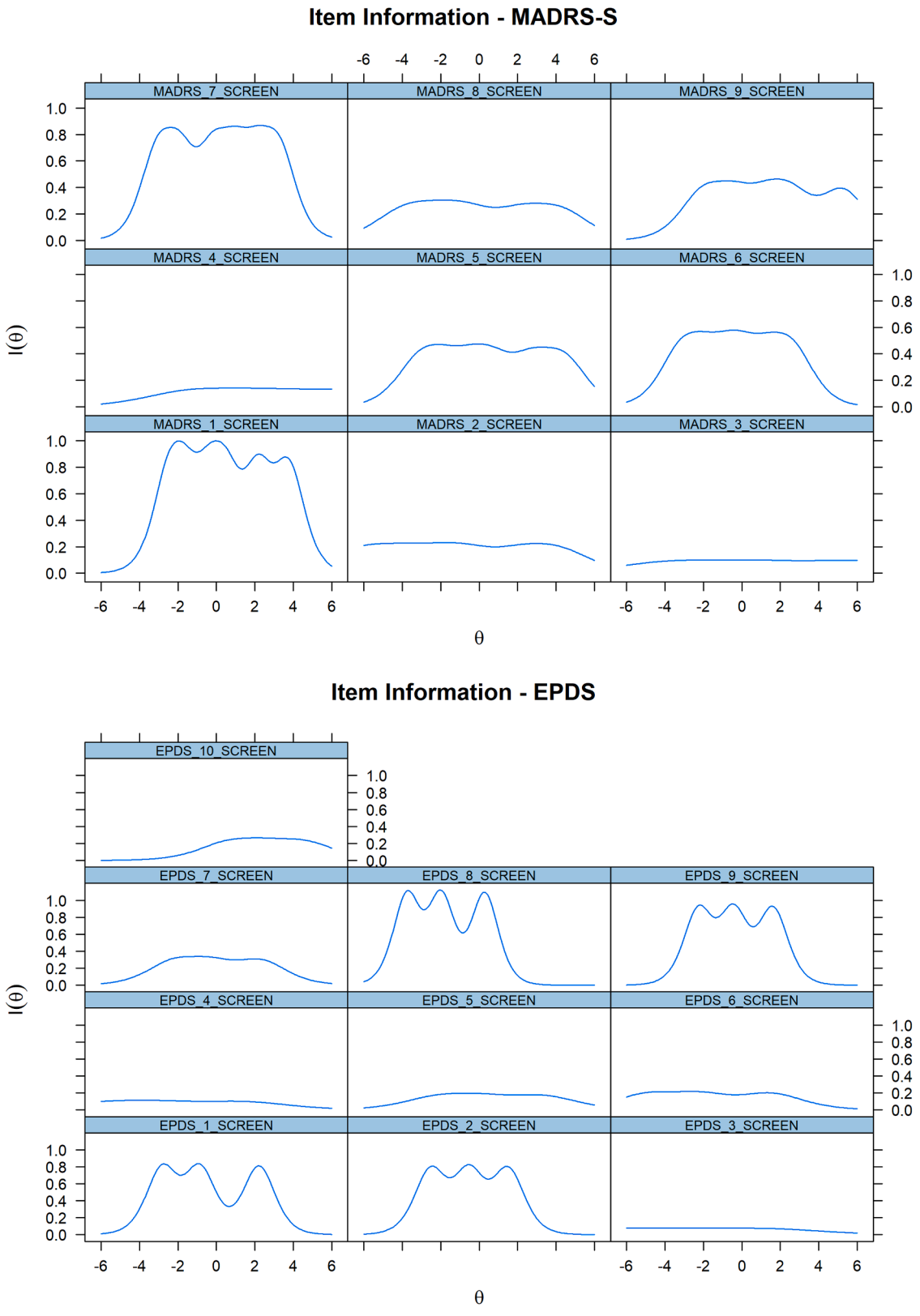

Fig. 2. Infotrace graphs.

better measurement precision at high level latent traits for the PHQ-9 (Peng et al., 2024). Similarly, Sun et al. (2019) found that women who fulfilled cardinal depressive symptoms of low-mood and anhedonia, had a much higher probability fulfilling somatic criteria (appetite, sleep, fatigue) on the PHQ-9 indicating that these items follow the general depressive level of the respondent rather than inflating it.

Kammerer et al. (2009) evaluated the validity of the diagnosis depressive episode according to DSM-IV during pregnancy and found that only the appetite item was unable to discriminate depressed from healthy pregnant women. The authors also reported that somatic

Table 4
Sum and item scores and change after treatment.

Variable	Mean Pre	Sd Pre	Mean Post	Sd Post	Mean Change	Sd Change	Mean Change-Z
MADRS_sum	21.48	6.45	12.43	7.16	−8.64	6.56	−4.67
MADRS_1	2.2	1.21	1.08	1.02	−1.1	1.34	−2.73
MADRS_2	2.96	1.14	1.76	1.23	−1.19	1.2	−3.65
MADRS_3	2.45	1.4	1.97	1.35	−0.49	1.52	−2.1
MADRS_4	1.45	1.31	0.7	0.92	−0.62	1.1	−1.58
MADRS_5	2.65	1.2	1.7	1.17	−0.88	1.26	−2.93
MADRS_6	2.91	1.2	1.69	1.18	−1.17	1.3	−3.4
MADRS_7	2.47	1.01	1.36	1	−1.1	1.18	−3.55
MADRS_8	2.96	1.34	1.48	1.2	−1.45	1.35	−3.28
MADRS_9	1.42	0.91	0.69	0.81	−0.65	0.85	−2.28
EPDS_sum	15.22	4.01	10.27	4.95	−4.91	4.91	−5.02
EPDS_1	1.69	0.58	1.08	0.69	−0.62	0.69	−3.99
EPDS_2	1.55	0.7	0.8	0.78	−0.76	0.89	−3.32
EPDS_3	2.04	0.78	1.68	0.78	−0.35	0.88	−3.06
EPDS_4	1.92	0.75	1.36	0.85	−0.56	0.88	−3.29
EPDS_5	1.11	0.82	0.69	0.79	−0.4	0.88	−1.83
EPDS_6	1.9	0.63	1.48	0.73	−0.4	0.78	−3.64
EPDS_7	1.31	0.79	0.76	0.78	−0.56	0.87	−2.37
EPDS_8	2.1	0.63	1.45	0.69	−0.66	0.79	−4.4
EPDS_9	1.38	0.71	0.83	0.7	−0.54	0.91	−2.7
EPDS_10	0.23	0.51	0.14	0.41	−0.07	0.48	−0.59

Notes: EPDS=Edinburgh Postnatal Depression Scale, MADRS= Montgomery-Åsberg Depression Rating Scale- Self report.

symptoms were significantly different between pregnant and postpartum women. Yonkers et al. (2009) found a significant difference between depressed and non-depressed pregnant women regarding degree and quantity of somatic symptoms. Thus, it does not seem reasonable to exempt somatic symptoms from psychological instruments and instruments that include somatic questions should be able to detect depression even during pregnancy (Castro et al., 2015).

Thus, our own, as well as previous findings are all contrary to the generally held belief amongst researchers and clinicians that a depression rating from a pregnant woman is likely an overestimation if it has been done with a scale that contains somatic items.

Regarding having no explicit suicidal ideation item, but instead a broad self-harm item in the EPDS our results indicate that the self-harm item in the EPDS might not be capturing suicide risk at all. This is indicated by the ICC-IRT plots showing that this item is scored 0 more often than not, even at quite high total scores in the sample, and far beyond how the suicidality item on MADRS-S is scored. In this sample, respondents were much more likely to endorse suicidal ideation than self-harm.

There is a dearth of evidence about the psychometric properties of the MADRS-S measure in general and specifically administered online. Therefore, it is hard to determine if the moderate performance of the MADRS-S in this pregnant sample is any better or worse than in a non-pregnant sample in a similar context. Future research is needed to assess the psychometric qualities of the MADRS-S in general as well as when administered to pregnant and non-pregnant women.

For whatever reason, pregnancy seems to be regarded as an exclusionary state for depression due to the somatic symptoms and changes that occur during this period. There are many other states and comorbidities where this is not the case. Somatic symptoms such as poor sleep or low appetite being caused, or exacerbated, by chronic pain, age, diseases, injury, or self-harm would likely not be viewed as exclusionary when assessing depression in any other population but rather considered a complicating comorbidity. Why then is “comorbid” pregnancy and depression viewed differently? This creates the risk that neither primary care, psychiatry nor maternity care provide adequate care for this already vulnerable population.

4.1. Clinical recommendations

More research is needed, but our results indicate that self-report measures developed for the general population serve their intended purpose also during pregnancy. In the present sample, MADRS-S And EPDS demonstrated broadly similar diagnostic performance, with MADRS-S demonstrating comparable or slightly better performance across several psychometric indices. While these differences were generally modest, they were consistent across analyses and suggest that MADRS_S performs at least as well as pregnancy specific measures in a clinical context.

Importantly, these findings should not be interpreted as indicating a clear diagnostic superiority of one instrument over the other. Rather, they support the conclusion that pregnancy per se does not necessitate fundamentally different approaches to depression assessment. From a clinical and implementation perspective, adapting assessment exclusively for pregnancy may in some cases complicate routine care and reduce comparability with depression assessments conducted at other life stages. In this respect, the use of established general-population instruments such as MADRS-S appears reasonable when depression is already suspected or established. Furthermore, the suicide item of the MADRS-S aids detection of perinatal patients at risk for suicidal behavior. Based on MADRS-S, suicide detection methods have even been implemented in remote therapy settings with good results (Hedman et al., 2014).

The use of instruments with potentially increased performance in regard to patient safety and psychometric properties that are familiar to professionals outside the perinatal realm might have the additional benefit of normalizing depression during pregnancy. This could reduce the risk of underdiagnosis by framing antenatal depression as a common disorder that can be handled in primary care or psychiatry settings. In contrast, relying exclusively on perinatal-specific measures like the EPDS may unintentionally reinforce the notion that depression in pregnancy requires separate, special assessment, potentially hindering depression detection and treatment.

We realize that the EPDS is internationally widely established and a central part of perinatal care chains. Thus, change may not be easily achieved. Our and others’ research highlights though the need of more nuanced knowledge on the application and results of an EPDS assessment: A) Midwives feeling comfortable using the EPDS must be aware that they are screening for more general psychological distress rather than depression. B) Both clinicians and researchers need to be aware that a positive screening result is not the same as a diagnosis of depression. Proper diagnostics are crucial. C) Clinicians and researchers interested in treatment outcome assessment should use scales with better psychometric performance. Somatic pregnancy symptoms are not

a hindrance. E) Suicidality is part of depression and a threat that needs proper attention – but the EPDS is not able to detect this.

5. Limitations

One limitation is that the current study only looks at patients who have self-selected into a psychiatric treatment setting specifically for antenatal depression. This means that all respondents do themselves believe that they are depressed, are help-seeking and positive to ICBT. In our data, most of the participants did receive a diagnosis of depression at the start of the study. Consequently, the study was not designed to evaluate MADRS-S or EPDS as general population screening instruments. Rather, the focus was on their performance as screening and outcome measures in a context where there is already a clinical decision to assess depression in individuals with a high likelihood of depressive disorders. This focus may further explain the limited emphasis on cut-off optimization and diagnostic classification relative to other psychometric aspects examined in the study.

A related limitation is the lack of non-depressed but pregnant controls and non-pregnant but depressed controls which limits the resolution of our findings. Future studies should include participants from the pregnant population in general as well as depressed and non-pregnant individuals.

6. Conclusion

This study suggests that both the MADRS-S and EPDS are reliable tools for assessing depressive symptoms in pregnant patients, but that the MADRS-S outperforms the EPDS on several key psychometric indicators. Contrary to common belief, somatic items did not inflate scores, and MADRS-S retained better precision at higher symptom levels. These findings support the use of standard depression measures, such as the MADRS-S, in pregnant populations. Doing so may improve detection and treatment by integrating antenatal depression into general psychiatric care rather than isolating it within perinatal-specific frameworks.

Funding sources

Söderström Königska Foundation (SLS-941378, SLS-941098, SLS-969514, SLS-982400, SLS-982143, SLS-995658), FORTE (2021-00867), Bror Gadelius Memory Foundation (129900957728, 129900529626) KI Foundations, Allmänna BB Minnesfond (FoUI-1012658), Thuring Foundation (2020-00627, 2021-00692, 2022-00746, 2023-021). No sponsor took any part in the design, execution, analysis, interpretation or publishing decisions of the study.

This research uses data from the DANA trial - Internet CBT for Antenatal Depression (trial registration: ClinicalTrials.gov Identifier NCT05533138).

CRediT authorship contribution statement

Erik Forsell: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sandra Fröjd:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Emelie Svanerud George:** Writing – review & editing, Investigation, Data curation. **Mikaela Bergmark Dunkels:** Writing – review & editing, Methodology, Conceptualization. **Marie Bendix:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

None.

Acknowledgements

The authors would like to thank the funding bodies and the psychiatric clinic as well as the clinicians, master's students and research staff for enabling this research to happen. We would especially like to thank all the patients who participated.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jadr.2026.101118.

References

Asberg, M., Montgomery, S.A., Perris, C., Schalling, D., Sedvall, G., 1978. A comprehensive psychopathological rating scale. Acta Psychiatr. Scand. Suppl. 271, 5–27. https://doi.org/10.1111/j.1600-0447.1978.tb02357.x.

Balserak, B.I., Lee, K., 2011. Sleep disturbances and sleep-related disorders in pregnancy. In: Kryger, M.H., Roth, T., Dement, W.C. (Eds.), Principles and Practice of Sleep Medicine, 5th ed. Elsevier Health Sciences, pp. 1572–1586. https://doi.org/10.1016/B978-1-4160-6645-3.00138-9.

Batt, M.M., Duffy, K.A., Novick, A.M., Metcalf, C.A., Epperson, C.N., 2020. Is postpartum depression different from depression occurring outside of the perinatal period? A review of the evidence. Focus (Am. Psychiatr. Publ.) 18 (2), 106–119. https://doi.org/10.1176/appi.focus.20190045.

Benson, A.E., Shatzel, J.J., Ryan, K.S., Hedges, M.A., Martens, K., Aslan, J.E., Lo, J.O., 2022. The incidence, complications, and treatment of iron deficiency in pregnancy. Eur. J. Haematol. 109 (6), 633–642. https://doi.org/10.1111/ejh.13870.

Bergink, V., Kooistra, L., Lambregtse-van den Berg, M.P., Wijnen, H., Bunevicius, R., van Baar, A., Pop, V., 2011. Validation of the Edinburgh Depression Scale during pregnancy. J. Psychosom. Res. 70 (4), 385–389. https://doi.org/10.1016/j.jpsychores.2010.07.008.

Bina, R., Harrington, D., 2016. The Edinburgh Postnatal Depression scale: screening tool for postpartum anxiety as well? Findings from a confirmatory factor analysis of the hebrew version. Matern. Child Health J. 20 (4), 904–914. https://doi.org/10.1007/s10995-015-1879-7.

Bondolfi, G., Jermann, F., Rouget, B.W., Gex-Fabry, M., McQuillan, A., Dupont-Willemin, A., Aubry, J.M., Nguyen, C., 2010. Self- and clinician-rated Montgomery-Asberg Depression Rating Scale: evaluation in clinical practice. J. Affect. Disord. 121 (3), 268–272. https://doi.org/10.1016/j.jad.2009.06.037.

Castro, E.C.T., Martins Brancaglion, M.Y., Nogueira Cardoso, M., Bergo Protzner, A., Duarte Garcia, F., Nicolato, R., Lopes, P.A.R.A., Vitor Leite, H., Corrêa, H., 2015. What is the best tool for screening antenatal depression? J. Affect. Disord. 178, 12–17. https://doi.org/10.1016/j.jad.2015.02.003.

Coates, R., Ayers, S., de Visser, R., 2017. Factor structure of the Edinburgh Postnatal Depression Scale in a population-based sample. Psychol. Assess. 29 (8), 1016–1027. https://doi.org/10.1037/pas0000397.

Cox, J., 2017. Use and misuse of the Edinburgh Postnatal Depression Scale (EPDS): a ten point 'survival analysis. Arch. Womens Ment. Health 20 (6), 789–790. https://doi.org/10.1007/s00737-017-0789-7.

Cox, J., 2019. Thirty years with the Edinburgh Postnatal Depression Scale: voices from the past and recommendations for the future. Br. J. Psychiatry 214 (3), 127–129. https://doi.org/10.1192/bjp.2018.245.

Cox, J.L., Chapman, G., Murray, D., Jones, P., 1996. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. J. Affect. Disord. 39 (3), 185–189. https://doi.org/10.1016/0165-0327(96)00008-0.

Cox, J.L., Holden, J.M., Sagovsky, R., 1987. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression scale. Br. J. Psychiatry 150, 782–786. https://doi.org/10.1192/bjp.150.6.782.

Cuijpers, P., Franco, P., Ciharova, M., Miguel, C., Segre, L., Quero, S., Karyotaki, E., 2023. Psychological treatment of perinatal depression: a meta-analysis. Psychol. Med. 53 (6), 2596–2608. https://doi.org/10.1017/s0033291721004529.

Cunningham, N.K., Brown, P.M., Page, A.C., 2015. Does the Edinburgh Postnatal Depression Scale measure the same constructs across time? Arch. Womens Ment. Health 18 (6), 793–804. https://doi.org/10.1007/s00737-014-0485-9.

D'Agostino, A., Garbazza, C., Malpetti, D., Azzimonti, L., Mangili, F., Stein, H.C., Del Giudice, R., Cicolin, A., Cirignotta, F., Manconi, M., LOS Group, 2024. Optimal risk and diagnosis assessment strategies in perinatal depression: a machine learning approach from the life-ON study cohort. Psychiatry Res. 332, 115687. https://doi.org/10.1016/j.psychres.2023.115687.

de Cock, E.S., Emons, W.H., Nefs, G., Pop, V.J., Pouwer, F., 2011. Dimensionality and scale properties of the Edinburgh Depression Scale (EDS) in patients with type 2 diabetes mellitus: the DiaDDzoB study. BMC Psychiatry 11, 141. https://doi.org/10.1186/1471-244x-11-141.

Donmez, M., Yorguner, N., Kora, K., Topcuoglu, V., 2022. Efficacy of bright light therapy in perinatal depression: a randomized, double-blind, placebo-controlled study. J. Psychiatr. Res. 149, 315–322. https://doi.org/10.1016/j.jpsychires.2022.02.027.

Duarte-Guerra, L.S., Gorenstein, C., Paiva-Medeiros, P.F., Santo, M.A., Lotufo Neto, F., Wang, Y.P., 2016. Clinical utility of the Montgomery-Åsberg Depression Rating Scale for the detection of depression among bariatric surgery candidates. BMC Psychiatry 16, 119. https://doi.org/10.1186/s12888-016-0823-8.

Dunn, T.J., Baguley, T., Brunsden, V., 2014. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br. J. Psychol. 105 (3), 399–412. https://doi.org/10.1111/bjop.12046.
Eakley, R., Lyndon, A., 2024. Disparities in screening and treatment patterns for depression and anxiety during pregnancy: an integrative review. J. Midwifery Womens Health 69 (6), 847–862. https://doi.org/10.1111/jmwh.13679.
Faisal-Cury, A., Rodrigues, D.M.O., Matijasevich, A., 2021. Are pregnant women at higher risk of depression underdiagnosis? J. Affect. Disord. 283, 192–197. https://doi.org/10.1016/j.jad.2021.01.057.
Fantino, B., Moore, N., 2009. The self-reported Montgomery-Asberg Depression Rating Scale is a useful evaluative tool in major depressive disorder. BMC Psychiatry 9, 26. https://doi.org/10.1186/1471-244x-9-26.
Forsell, E., Bendix, M., Holländare, F., von Schultz, B.S., Nasiell, J., Blomdahl-Wetterholm, M., Eriksson, C., Kvarned, S., Soderbergder Linden, J.L.V., Söderberg, E., Jokinen, J., Wide, K., Kaldo, V., 2017. Internet delivered Cognitive behavior therapy for antenatal depression: a randomised controlled trial. J. Affect. Disord. 221, 56–64. https://doi.org/10.1016/j.jad.2017.06.013.
Forsell, E., Isacsson, N., Blom, K., Jernelöv, S., Ben Abdesslem, F., Lindefors, N., Boman, M., Kaldo, V., 2020. Predicting treatment failure in regular care internet-delivered cognitive behavior therapy for depression and anxiety using only weekly symptom measures. J. Consult. Clin. Psychol. 88 (4), 311–321. https://doi.org/10.1037/ccp0000462.
Forsell, E., Mattsson, S., Hentati Isacsson, N., Kaldo, V., 2025. Accuracy of therapists' predictions of outcome in internet-delivered cognitive behavior therapy for depression and anxiety in routine psychiatric care. J. Consult. Clin. Psychol. 93 (3), 176–190. https://doi.org/10.1037/ccp0000943.
Green, S.M., Donegan, E., McCabe, R.E., Streiner, D.L., Agako, A., Frey, B.N., 2020. Cognitive behavioral therapy for perinatal anxiety: a randomized controlled trial. Aust. NZ J. Psychiatry 54 (4), 423–432. https://doi.org/10.1177/0004867419898528.
Gunzler, D., Sehgal, A.R., Kauffman, K., Davey, C.H., Dolata, J., Figueroa, M., Huml, A., Pencak, J., Sajatovic, M., 2020. Identify depressive phenotypes by applying RDOC domains to the PHQ-9. Psychiatry Res. 286, 112872. https://doi.org/10.1016/j.psychres.2020.112872.
Harris, B., Huckle, P., Thomas, R., Johns, S., Fung, H., 1989. The use of rating scales to identify post-natal depression. Br. J. Psychiatry 154, 813–817. https://doi.org/10.1192/bjp.154.6.813.
Hedman, E., Ljótsson, B., Kaldo, V., Hesser, H., El Alaoui, S., Kraepelien, M., Andersson, E., Rück, C., Svanborg, C., Andersson, G., Lindefors, N., 2014. Effectiveness of internet-based cognitive behaviour therapy for depression in routine psychiatric care. J. Affect. Disord. 155, 49–58. https://doi.org/10.1016/j.jad.2013.10.023.
Hewitt, C.E., Gilbody, S.M., Mann, R., Brealey, S., 2010. Instruments to identify post-natal depression: which methods have been the most extensively validated, in what setting and in which language? Int. J. Psychiatry Clin. Pract. 14 (1), 72–76. https://doi.org/10.3109/13651500903198020.
Howard, L.M., Khalifeh, H., 2020. Perinatal mental health: a review of progress and challenges. World Psychiatry 19 (3), 313–327. https://doi.org/10.1002/wps.20769.
Judd, F., Lorimer, S., Thomson, R.H., Hay, A., 2019. Screening for depression with the Edinburgh Postnatal Depression Scale and finding borderline personality disorder. Aust. NZ J. Psychiatry 53 (5), 424–432. https://doi.org/10.1177/0004867418804067.
Kammerer, M., Marks, M.N., Pinard, C., Taylor, A., von Castelberg, B., Kunzli, H., Glover, V., 2009. Symptoms associated with the DSM IV diagnosis of depression in pregnancy and post partum. Arch. Womens Ment. Health 12 (3), 135–141. https://doi.org/10.1007/s00737-009-0062-9.
Kelley, K. (2017). MBESS (Version 4.9.3) [Computer software and manual]. In.
Kittel-Schneider, S., Davidova, P., Kalok, M., Essel, C., Ahmed, F.B., Kingeter, Y., Matentzoglu, M., Leutritz, A.L., Kersken, K., Koreny, C., Weber, H., Kollert, L., McNeill, R.V., Reif, A., Bahlmann, F., Trautmann-Villalba, P., 2022. A pilot study of multilevel analysis of BDNF in paternal and maternal perinatal depression. Arch. Womens Ment. Health 25 (1), 237–249. https://doi.org/10.1007/s00737-021-01197-2.
Kozinszky, Z., Dudas, R.B., 2015. Validation studies of the Edinburgh Postnatal Depression Scale for the antenatal period. J. Affect. Disord. 176, 95–105. https://doi.org/10.1016/j.jad.2015.01.044.
Leucht, S., Fennema, H., Engel, R.R., Kaspers-Janssen, M., Lepping, P., Szegedi, A., 2017. What does the MADRS mean? Equipercentile linking with the CGI using a company database of mirtazapine studies. J. Affect. Disord. 210, 287–293. https://doi.org/10.1016/j.jad.2016.12.041.
Loughnan, S.A., Joubert, A.E., Grierson, A., Andrews, G., Newby, J.M., 2019. Internet-delivered psychological interventions for clinical anxiety and depression in perinatal women: a systematic review and meta-analysis. Arch. Womens Ment. Health 22 (6), 737–750. https://doi.org/10.1007/s00737-019-00961-9.
Lutterodt, M.C., Kähler, P., Kragstrup, J., Nicolaisdottir, D.R., Siersma, V., Ertmann, R.K., 2019. Examining to what extent pregnancy-related physical symptoms worry women in the first trimester of pregnancy: a cross-sectional study in general practice. BJGP Open 3 (4), bjgpopen19X101674. https://doi.org/10.3399/bjgpopen19X101674.
Lydsdottir, L.B., Howard, L.M., Olafsdottir, H., Thome, M., Tyrfingsson, P., Sigurdsson, J. F., 2014. The mental health characteristics of pregnant women with depressive symptoms identified by the Edinburgh Postnatal Depression Scale. J. Clin. Psychiatry 75 (4), 393–398. https://doi.org/10.4088/JCP.13m08646.
Massoudi, P., Hwang, C.P., Wickberg, B., 2013. How well does the Edinburgh Postnatal Depression Scale identify depression and anxiety in fathers? A validation study in a population based Swedish sample. J. Affect. Disord. 149 (1–3), 67–74. https://doi.org/10.1016/j.jad.2013.01.005.
Matthey, S., 2008. Using the Edinburgh Postnatal Depression Scale to screen for anxiety disorders. Depress. Anxiety 25 (11), 926–931. https://doi.org/10.1002/da.20415.
Matthey, S., Agostini, F., 2017. Using the Edinburgh Postnatal Depression Scale for women and men-some cautionary thoughts. Arch. Womens Ment. Health 20 (2), 345–354. https://doi.org/10.1007/s00737-016-0710-9.
Matthey, S., Fisher, J., Rowe, H., 2013. Using the Edinburgh postnatal depression scale to screen for anxiety disorders: conceptual and methodological considerations. J. Affect. Disord. 146 (2), 224–230. https://doi.org/10.1016/j.jad.2012.09.009.
Matthey, S., Ross-Hamid, C., 2012. Repeat testing on the Edinburgh Depression Scale and the HADS-A in pregnancy: differentiating between transient and enduring distress. J. Affect. Disord. 141 (2–3), 213–221. https://doi.org/10.1016/j.jad.2012.02.037.
Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. Br. J. Psychiatry 134, 382–389. https://doi.org/10.1192/bjp.134.4.382.
Murray, D., Cox, J.L., 1990. Screening for depression during pregnancy with the edinburgh depression scale (EDDS). J. Reprod. Infant Psychol. 8 (2), 99–107. https://doi.org/10.1080/02646839008403615.
NICE, 2020. Antenatal and Postnatal Mental Health: Clinical Management and Service Guidance. NICE-National Institute for Health and Care Excellence. Retrieved from. www.nice.org.uk/guidance/cg192.
Nonacs, R., Cohen, L.S., 2003. Assessment and treatment of depression during pregnancy: an update. Psychiatr. Clin. N. Am. 26 (3), 547–562. https://doi.org/10.1016/s0193-953x(03)00046-7.
Peng, Y., Wu, Y., Huang, Y., Zhou, N., Fang, W., Xi, J., 2024. Unidimensional or multidimensional? Revisiting the psychometrics of PHQ-9 and EPDS using bifactor model and item response theory in 2939 Chinese perinatal women. J. Affect. Disord. 367, 686–695. https://doi.org/10.1016/j.jad.2024.07.148.
Pettman, D., O'Mahen, H., Blomberg, O., Svanberg, A.S., von Essen, L., Woodford, J., 2023. Effectiveness of cognitive behavioural therapy-based interventions for maternal perinatal depression: a systematic review and meta-analysis. BMC Psychiatry 23 (1), 208. https://doi.org/10.1186/s12888-023-04547-9.
Phillips, J., Charles, M., Sharpe, L., Matthey, S., 2009. Validation of the subscales of the Edinburgh Postnatal Depression Scale in a sample of women with unsettled infants. J. Affect. Disord. 118 (1–3), 101–112. https://doi.org/10.1016/j.jad.2009.02.004.
Pop, V.J., Komproe, I.H., van Son, M.J., 1992. Characteristics of the Edinburgh post natal depression scale in the Netherlands. J. Affect. Disord. 26 (2), 105–110. https://doi.org/10.1016/0165-0327(92)90041-4.
Qiu, X., Wu, Y., Sun, Y., Levis, B., Tian, J., Boruff, J.T., Cuijpers, P., Ioannidis, J.P.A., Markham, S., Ziegelstein, R.C., Vigod, S.N., Benedetti, A., Thombs, B.D., DEPRESsion Screening Data (DEPRESSD) EPDS Group, 2023. Individual participant data meta-analysis to compare EPDS accuracy to detect major depression with and without the self-harm item. Sci. Rep. 13 (1), 4026. https://doi.org/10.1038/s41598-023-29114-w.
Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., Müller, M., (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. (1471-2105 (Electronic)).
Rondung, E., Massoudi, P., Nieminen, K., Wickberg, B., Peira, N., Silverstein, R., Moberg, K., Lundqvist, M., Grundberg, Å., Hultcrantz, M., 2024. Identification of depression and anxiety during pregnancy: a systematic review and meta-analysis of test accuracy. Acta Obstet. Gynecol. Scand. 103 (3), 423–436. https://doi.org/10.1111/aogs.14734.
Rosseel, Y., 2012. lavaan: an R package for structural equation modeling. J. Stat. Softw. 48 (2), 1–36. https://doi.org/10.18637/jss.v048.i02.
Rubertsson, C., Börjesson, K., Berglund, A., Josefsson, A., Sydsjö, G., 2011. The Swedish validation of Edinburgh Postnatal Depression Scale (EPDS) during pregnancy. Nord. J. Psychiatry 65 (6), 414–418. https://doi.org/10.3109/08039488.2011.590606.
SBU. (2012). Diagnostik och uppföljning av förstämningssyndrom: en systematisk litteraturöversikt. Stockholm Retrieved from https://www.sbu.se/contentassets/56adfbafa780477b9bc44d1abcdbb65e/Forstamningssyndrom_2012.pdf.
SBU. (2021). Förlossningsrädsla, depression och ångest under graviditet: en Systematisk översikt och Utvärdering Av medicinska, hälsoekonomiska, Sociala och etiska aspekter. Stockholm Retrieved from https://www.sbu.se/322.
Spitzer, R.L., Williams, J.B., Gibbon, M., First, M.B., 1992. The Structured Clinical Interview for DSM-III-R (SCID). I: history, rationale, and description. Arch. Gen. Psychiatry 49 (8), 624. https://doi.org/10.1001/archpsyc.1992.01820080032005.
Su, K.P., Chiu, T.H., Huang, C.L., Ho, M., Lee, C.C., Wu, P.L., Lin, C.Y., Liau, C.H., Liao, C. C., Chiu, W.C., Pariante, C.M., 2007. Different cutoff points for different trimesters? The use of Edinburgh Postnatal Depression Scale and Beck Depression Inventory to screen for depression in pregnant Taiwanese women. Gen. Hosp. Psychiatry 29 (5), 436–441. https://doi.org/10.1016/j.genhosppsych.2007.05.005.
Sun, J.W., Cao, D.F., Li, J.H., Zhang, X., Wang, Y., Bai, H.Y., Lin, P.Z., Zhang, H.H., Cao, F.L., 2019. Profiles and characteristics of clinical subtypes of perinatal depressive symptoms: a latent class analysis. J. Adv. Nurs. 75 (11), 2753–2765. https://doi.org/10.1111/jan.14136.
Svanborg, P., Asberg, M., 1994. A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating Scale. Acta Psychiatr. Scand. 89 (1), 21–28. https://doi.org/10.1111/j.1600-0447.1994.tb01480.x.
Svanborg, P., Asberg, M., 2001. A comparison between the Beck Depression Inventory (BDI) and the self-rating version of the Montgomery Asberg Depression Rating Scale (MADRS). J. Affect. Disord. 64 (2–3), 203–216. https://doi.org/10.1016/s0165-0327(00)00242-1.
Svanborg, P., Ekselius, L., 2003. Self-assessment of DSM-IV criteria for major depression in psychiatric out- and inpatients. Nord. J. Psychiatry 57 (4), 291–296. https://doi.org/10.1080/08039480307281.
Thorndike, F.P., Carlbring, P., Smyth, F.L., Magee, J.C., Gonder-Frederick, L., Ost, L.G., Ritterband, L.M., 2009. Web-based measurement: effect of completing single or

multiple items per webpage. Comput. Hum. Behav. 25 (2), 393–401. https://doi.org/10.1016/j.chb.2008.05.006.
Uher, R., Perlis, R.H., Placentino, A., Dernovšek, M.Z., Henigsberg, N., Mors, O., Maier, W., McGuffin, P., Farmer, A., 2012. Self-report and clinician-rated measures of depression severity: can one replace the other? Depress. Anxiety 29 (12), 1043–1049. https://doi.org/10.1002/da.21993.
van Ballegooijen, W., Riper, H., Cuijpers, P., van Oppen, P., Smit, J.H., 2016. Validation of online psychometric instruments for common mental health disorders: a systematic review. BMC Psychiatry 16, 45. https://doi.org/10.1186/s12888-016-0735-7.
Viktorin, A., Meltzer-Brody, S., Kuja-Halkola, R., Sullivan, P.F., Landén, M., Lichtenstein, P., Magnusson, P.K., 2016. Heritability of Perinatal Depression and Genetic Overlap With Nonperinatal Depression. Am. J. Psychiatry 173 (2), 158–165. https://doi.org/10.1176/appi.ajp.2015.15010085.
Wickberg, B., Hwang, C.P., 1996a. Counselling of postnatal depression: a controlled study on a population based Swedish sample. J. Affect. Disord. 39 (3), 209–216. https://doi.org/10.1016/0165-0327(96)00034-1.
Wickberg, B., Hwang, C.P., 1996b. The Edinburgh Postnatal Depression Scale: validation on a Swedish community sample. Acta Psychiatr. Scand. 94 (3), 181–184. https://doi.org/10.1111/j.1600-0447.1996.tb09845.x.
Woody, C.A., Ferrari, A.J., Siskind, D.J., Whiteford, H.A., Harris, M.G., 2017. A systematic review and meta-regression of the prevalence and incidence of perinatal depression. J. Affect. Disord. 219, 86–92. https://doi.org/10.1016/j.jad.2017.05.003.
Yonkers, K.A., Smith, M.V., Gotman, N., Belanger, K., 2009. Typical somatic symptoms of pregnancy and their impact on a diagnosis of major depressive disorder. Gen. Hosp. Psychiatry 31 (4), 327–333. https://doi.org/10.1016/j.genhosppsych.2009.03.005.
Zhong, Q.Y., Gelaye, B., Rondon, M.B., Sánchez, S.E., Simon, G.E., Henderson, D.C., Barrios, Y.V., Sánchez, P.M., Williams, M.A., 2015. Using the patient health questionnaire (PHQ-9) and the Edinburgh Postnatal Depression Scale (EPDS) to assess suicidal ideation among pregnant women in Lima, Peru. Arch. Womens Ment. Health 18 (6), 783–792. https://doi.org/10.1007/s00737-014-0481-0.

Glossary

CFA: Confirmatory Factor Analysis
DSM: Diagnostic and Statistical Manual of Mental Disorders
EPDS: Edinburgh postnatal depression scale
GRM: Graded response model
ICBT: internet cognitive behavioral therapy
ICC: intraclass correlation coefficient
IRT: Item response theory
KMO: Kaiser–Meyer–Olkin test
MADRS: Montgomery Åsberg Depression Rating Scale
MADRS-S: self-rating version of the Montgomery Åsberg Depression Rating Scale
MDD: Major depressive disorder
PCM: Partial Credit model
PHQ-9: Patient Health Questionnaire for depression
ROC: Receiver operating characteristics