OPEN ACCESS

Scholars International Journal of Obstetrics and Gynecology

Abbreviated Key Title: Sch Int J Obstet Gynec
ISSN 2616-8235 (Print) |ISSN 2617-3492 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Review Article

Identification of Perinatal Depression in Gulf Cooperation Council (GCC) Countries: A Systematic Review

Nada A. Al-Mulla[1]*

[1]King's College London, School of Life Sciences and Medicine, London, United Kingdom

DOI: https://doi.org/10.36348/sijog.2026.v09i11.003 | **Received:** 07.09.2026 | **Accepted:** 31.10.2026 | **Published:** 04.11.2026

***Corresponding author:** Nada A. Al-Mulla
King's College London, School of Life Sciences and Medicine, London, United Kingdom

Abstract

Perinatal depression represents a significant public health concern across the Gulf Cooperation Council (GCC) region; however, psychometric validation of screening tools in these culturally distinct populations remains inadequately synthesised. Prevalence estimates vary substantially depending on screening thresholds, while cultural factors such as stigma, somatic symptom expression, and family-centred decision-making influence symptom presentation and help-seeking behaviour. This review aimed to identify and critically appraise the diagnostic accuracy of perinatal depression screening tools validated across all six GCC member states (Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the United Arab Emirates). A comprehensive search was conducted across five electronic databases (PubMed, Ovid MEDLINE, Embase, Global Health, and APA PsycINFO) without language restrictions, supplemented by manual reference screening. Eligible studies were required to validate screening tools against reference standards in GCC perinatal populations. Methodological quality was assessed using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2) framework, and findings were synthesised narratively. Only three validation studies were identified, all evaluating the Arabic Edinburgh Postnatal Depression Scale (EPDS) in the United Arab Emirates, Qatar, and Oman. Reported sensitivity and specificity were broadly acceptable; however, all studies exhibited moderate to high risk of bias, and no validation evidence was identified for Bahrain, Kuwait, or Saudi Arabia. Whilst the Arabic EPDS currently represents the most empirically supported screening tool in the region, the evidence base remains limited in scope, methodological rigour, and geographical coverage. Further research should prioritise robust validation studies in GCC countries and the development of culturally grounded screening approaches.

Keywords: Antenatal Depression; Perinatal Period; Postpartum Depression; Healthcare quality assessment; Maternal experience; Maternal Distress; Pregnancy; Depression.

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)** which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

1. INTRODUCTION

1.1 The Spectrum of Perinatal Psychiatric Disorders

Perinatal mental health encompasses psychological wellbeing during pregnancy and the period following childbirth. Although the perinatal period is widely recognised as a time of increased vulnerability to mental health difficulties, its boundaries remain inconsistently defined across diagnostic systems, clinical guidelines, and research literature. The *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) applies a "peripartum onset" specifier to disorders occurring during pregnancy or within four weeks postpartum, whereas many clinical guidelines recognise that vulnerability may extend for up to six to twelve months after birth [1]. These variations influence prevalence estimates, screening practices, referral pathways, and the identification of women requiring support [2].

Perinatal psychiatric disorders encompass a spectrum of conditions, including depression, anxiety disorders, post-traumatic stress disorder, obsessive-compulsive disorder, eating disorders, and, less commonly, postpartum psychosis [3,4]. While some disorders emerge for the first time during pregnancy or after childbirth, others may represent the continuation or recurrence of pre-existing mental health conditions. Their impact extends beyond maternal wellbeing, affecting partner mental health, family relationships, parenting capacity, and child cognitive, emotional, and

behavioural development, highlighting the importance of a family-centred perspective in perinatal care [5].

The development of perinatal mental disorders is best understood through a biopsychosocial-spiritual framework, in which biological, psychological, social, environmental, and spiritual factors interact to influence both vulnerability and resilience [6]. Common risk factors include a history of mental illness, limited social support, socioeconomic disadvantage, interpersonal violence, and unintended pregnancy [3]. However, not all emotional distress experienced during pregnancy and early parenthood reflects psychopathology. The concept of *matrescence* describes the profound psychological, relational, physical, and identity-related transitions associated with becoming a mother [7]. Although these changes are considered normative, some experiences associated with matrescence may resemble symptoms assessed by psychiatric screening instruments, creating potential challenges in distinguishing normal adaptation from clinically significant distress.

Among all perinatal mental health conditions, depression represents the most prevalent disorder and accounts for a substantial proportion of perinatal psychiatric morbidity worldwide [8]. Prevalence estimates vary considerably across settings, affecting approximately 10–13% of women in high-income countries and between 20% and 30% globally, with pooled estimates of 24.7% reported in low- and middle-income countries [9–12]. Antenatal depression is one of the strongest predictors of postpartum depression, while adverse life events, poor social support, and socioeconomic stressors consistently increase vulnerability across populations [9,13–15].

The consequences of untreated perinatal depression are considerable. Maternal outcomes include impaired functioning, chronic psychiatric illness, reduced quality of life, and maternal suicide, which remains a leading cause of maternal mortality in several high-income countries [8,16]. Perinatal depression has also been associated with adverse obstetric and child outcomes, including preterm birth, low birth weight, disrupted mother-infant bonding, and poorer cognitive and emotional development in offspring [8,17,18]. Although much of this evidence is observational and therefore susceptible to confounding by social adversity, comorbid anxiety, and obstetric complications, the overall public health burden remains substantial [19].

Given its prevalence and potential consequences, early detection and intervention have become central components of perinatal mental healthcare. Screening programmes are increasingly recommended to identify women experiencing depressive symptoms and facilitate timely assessment and treatment. However, the effectiveness of screening depends on the accuracy and cultural appropriateness of the instruments used. These considerations are particularly important in regions where cultural, linguistic, and social factors may influence symptom expression and help-seeking behaviours, such as the Gulf Cooperation Council (GCC) countries.

1.2 Perinatal Depression in the Gulf Cooperation Council Countries

The Gulf Cooperation Council (GCC), comprising Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the United Arab Emirates [20], represents a distinctive but relatively underexplored setting for perinatal mental health research. Although these countries share linguistic, religious, and cultural characteristics, they differ substantially in healthcare infrastructure, economic development, demographic composition, and social policy [20]. Rapid urbanisation, changing family structures, and the presence of large expatriate populations further shape the social environment in which pregnancy and childbirth occur [21].

Cultural and social factors play a significant role in the recognition, expression, and management of mental health difficulties during the perinatal period. Stigma surrounding mental illness may discourage disclosure of symptoms and reduce engagement with formal mental health services [21]. In some contexts, psychological distress may be understood through spiritual or supernatural frameworks, such as beliefs related to jinn possession, while religious coping strategies may be preferred to biomedical interventions [22–24]. Family-centred decision-making, including the influence of husbands and older female relatives, may further shape help-seeking behaviours and treatment choices [8]. Consequently, women may experience, interpret, and communicate symptoms of psychological distress in ways that differ from those assumed by screening instruments developed in Western populations.

Evidence increasingly suggests that perinatal depression represents a significant public health concern across the GCC. However, reported prevalence estimates vary substantially between countries and studies. In Saudi Arabia, antenatal depression prevalence has been reported at 31.9%, while postpartum depression estimates range from 32.8% to 43.4% (25–30). Studies from Oman have reported prevalence rates of 27.1% during pregnancy and 29.3% following childbirth [9,31], whereas postpartum depression prevalence in Qatar has been estimated at 17.6% (32). Such variation is influenced by differences in study design, sampling strategies, assessment timing, and methods of case identification.

Importantly, pooled evidence indicates that prevalence estimates are highly sensitive to the screening thresholds used to define probable depression. For example, use of the Edinburgh Postnatal Depression Scale (EPDS) at different cut-off scores may produce

substantially different estimates of prevalence, ranging from approximately 18% when a higher threshold is applied to as much as 45% when lower thresholds are used (33). Consequently, prevalence estimates may reflect methodological differences as much as genuine variation in population burden.

The marked heterogeneity observed across GCC studies raises important questions regarding the validity and comparability of currently available evidence. Many prevalence studies have relied exclusively on screening instruments without confirming diagnoses using recognised diagnostic reference standards. As a result, uncertainty remains regarding whether reported prevalence estimates accurately reflect the true burden of depressive disorders or are partly attributable to variation in measurement approaches. These concerns highlight the importance of examining the psychometric performance of screening instruments within GCC populations and provide a rationale for focusing not only on the epidemiology of perinatal depression but also on the tools used to identify it.

1.3 Validation of Perinatal Depression Screening Tools in GCC Countries

Accurate identification of perinatal depression relies on screening instruments that can distinguish women experiencing clinically significant depressive symptoms from those undergoing normative emotional adaptation during pregnancy and the transition to motherhood. Screening instruments are not diagnostic tools; rather, they identify individuals who may require further clinical evaluation. Their performance is therefore assessed against recognised diagnostic reference standards, typically structured psychiatric interviews such as the Structured Clinical Interview for DSM Disorders (SCID), the Mini-International Neuropsychiatric Interview (MINI), or the Composite International Diagnostic Interview (CIDI) (34).

The psychometric performance of screening instruments is commonly evaluated using measures of diagnostic accuracy, including sensitivity, specificity, positive and negative predictive values, likelihood ratios, and the area under the receiver operating characteristic (ROC) curve (35). These measures provide insight into a tool's ability to correctly classify individuals with and without depression and are essential when determining appropriate screening thresholds for clinical or public health use.

The importance of diagnostic accuracy is particularly pronounced during the perinatal period because screening errors may have significant consequences. False-negative results can delay recognition and treatment, resulting in missed opportunities for intervention during a period of heightened vulnerability. Conversely, false-positive results may contribute to unnecessary medicalisation, anxiety, stigma, and inefficient use of healthcare resources (35,36). Achieving an appropriate balance between sensitivity and specificity therefore requires careful consideration of the intended purpose of screening and the characteristics of the population in which an instrument is used.

The Edinburgh Postnatal Depression Scale (EPDS), a 10-item self-report instrument originally developed in the United Kingdom, remains the most widely used perinatal depression screening tool internationally and is commonly employed during both pregnancy and the postpartum period (37). Other frequently used instruments include the Patient Health Questionnaire-9 (PHQ-9) and the Beck Depression Inventory-II (BDI-II). Although these measures have demonstrated acceptable psychometric properties in many settings, evidence generated in one population cannot be assumed to apply universally. Screening tool performance is influenced by linguistic, cultural, social, and healthcare-system factors that shape symptom expression, help-seeking behaviours, and perceptions of mental illness. Consequently, optimal cut-off scores, sensitivity, specificity, and predictive values may vary substantially across populations, highlighting the need for local validation studies (35).

These issues are particularly relevant within GCC countries. Experiences of pregnancy, childbirth, and motherhood are shaped by cultural expectations, religious beliefs, family structures, and broader social norms. Emotional distress may be expressed through somatic symptoms, relational concerns, or spiritual frameworks that are not fully captured by conventional depression measures. Conversely, experiences associated with normal adaptation to motherhood may be interpreted as indicators of psychopathology, increasing the possibility of false-positive results. Such considerations emphasise the importance of evaluating screening instruments within the cultural contexts in which they are intended to be used.

Interpretation of the GCC evidence base is further complicated by methodological limitations. Much of the existing literature consists of cross-sectional studies employing convenience or consecutive sampling methods, limiting representativeness and reducing generalisability. Additional heterogeneity arises from variation in study populations, timing of assessment, diagnostic reference standards, and screening thresholds (30,32). Differences between diagnostic frameworks also contribute to inconsistency. For example, while the DSM-5 limits the peripartum specifier to pregnancy and the first four weeks postpartum, broader postpartum timeframes are often adopted in clinical practice and within the International Classification of Diseases (ICD), influencing case identification and prevalence estimates across studies.

Although international organisations such as the World Health Organization advocate routine

screening and psychosocial support as components of comprehensive maternal healthcare (11), effective implementation requires more than the selection of a screening instrument. Culturally appropriate screening processes, trained healthcare professionals, clear referral pathways, access to diagnostic assessment, and availability of treatment services are all necessary to ensure that screening leads to meaningful clinical outcomes. Within GCC countries, implementation may be further challenged by stigma surrounding mental illness, variability in service provision, and uncertainty regarding the cultural validity of available instruments (22).

Despite the increasing use of perinatal depression screening tools throughout the GCC, evidence relating to their psychometric validation remains fragmented. Many studies have implemented screening instruments without formally evaluating their performance against an appropriate diagnostic reference standard. To date, no published review has comprehensively synthesised evidence concerning the validation of perinatal depression screening tools across all six GCC countries. Addressing this gap is essential for determining whether commonly used instruments are appropriately calibrated for women in the region and for informing the development of effective, culturally responsive screening programmes.

1.4 Aims and Objectives

The foregoing review of the literature establishes that perinatal depression is a significant and consequential public health problem across the GCC region; that the existing evidence base is characterised by substantial methodological heterogeneity and limited use of diagnostic gold standards; and that no published review has synthesised the psychometric validation evidence for perinatal depression screening tools across the GCC as a whole. The present review directly addresses this gap. Grounded in the diagnostic accuracy framework and attentive to the sociocultural specificities of the GCC context, the review presents a critical appraisal of perinatal depression screening tools used across the region.

In addressing this gap, the present work makes an original contribution to the evidence base in two respects. First, it provides the first systematic synthesis of psychometric validation evidence for perinatal depression screening tools across the GCC as a whole, generating regionally grounded estimates of diagnostic accuracy that can directly inform clinical practice and screening policy. Second, by integrating the diagnostic accuracy evidence with a culturally sensitive analytical lens, it moves beyond the technical appraisal of instruments to address the broader question of how screening systems can be designed to be both psychometrically valid and culturally appropriate. Building on these findings, the review outlines the initial strategies needed to advocate for a comprehensive screening programme capable of capturing the diverse human experiences that shape pregnancy, birth, and the transition to motherhood across the region's multiple cultural contexts. Such a programme would enable a more context-sensitive understanding of local expressions of matrescence and the nuances of perinatal psychopathology, ensuring that assessment and care are better aligned with the lived realities of women and families in the GCC.

The specific objectives are to:

Primary Objective:

1. Systematically identify which screening tools for perinatal depression have been validated for use among women during the perinatal period across all six GCC countries (Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the United Arab Emirates) (20).

Secondary Objectives

1. To critically evaluate the psychometric properties of the identified screening tools.
2. To identify and synthesise prevalence estimates of perinatal depression reported in the included studies.
3. To identify and synthesise evidence on cultural, contextual, and systemic barriers to the effective implementation of screening programmes in the GCC.

2. METHODOLOGY

2.1 Research Design

A systematic review design was adopted to synthesise evidence on the validation of perinatal depression screening tools within the GCC region. This approach provides a structured and rigorous framework for identifying, evaluating, and integrating primary research, and is particularly suited to addressing the identified evidence gap given the anticipated variability in screening instruments, settings, and population characteristics across studies.

Systematic reviews represent a robust method for synthesising evidence on diagnostic test accuracy, using transparent and reproducible processes to identify, appraise, and integrate all available studies. They also require specialised methodological approaches distinct from intervention reviews, including the application of dedicated quality assessment tools such as QUADAS-2 (38) and reporting guidelines such as PRISMA-DTA (39). Furthermore, given the expected heterogeneity across GCC countries, a systematic narrative synthesis is considered more appropriate than an individual participant data meta-analysis, which would require access to raw data from primary studies. Importantly, this methodology allows for critical appraisal of study quality, ensuring that conclusions are interpreted in light of the strength and limitations of the available evidence.

2.2 Methodological Approach

This review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews

and Meta-Analyses (PRISMA) 2020 guidelines, with application of the PRISMA extension for diagnostic test accuracy (PRISMA-DTA) where appropriate (39). Adherence to these frameworks ensured a systematic and transparent process for study identification, selection, and reporting.

The methodological approach encompasses all stages of the review process, including the development of eligibility criteria, search strategy, study selection, data extraction, quality appraisal, data synthesis, ethical considerations, and identification of methodological limitations. The use of predefined criteria reduces the risk of selection bias and enhances the reliability, transparency, and reproducibility of the findings.

This study was registered with PROSPERO (CRD420261373029) to promote methodological transparency and enable independent scrutiny of the planned procedures, thereby strengthening the credibility of the review.

2.3 Eligibility Criteria

2.3.1 Introduction to Eligibility Criteria

The review was restricted to studies involving pregnant and postpartum women residing in Gulf Cooperation Council (GCC) countries during the perinatal period, defined as the antenatal phase from conception to delivery and the postnatal phase from delivery up to 12 months postpartum. Eligible studies were required to evaluate a screening or case-identification instrument for perinatal depression and compare its performance against a reference standard used for validation purposes. Screening instruments could be self-administered or interviewer-administered and included established, locally developed, or culturally adapted measures of perinatal depression.

No restrictions were placed on the type of reference standard employed, provided that it served as the comparator against which the screening instrument's performance was assessed. Reference standards therefore varied across studies and included structured clinical diagnostic interviews as well as other validated psychological measures selected by the original investigators for validation purposes. Variations in the methodological robustness of these reference standards were considered during quality appraisal and taken into account when interpreting the findings.

Studies were excluded if they evaluated instruments designed primarily to measure constructs other than perinatal depression or failed to include a comparison against a reference standard. No restrictions were placed on study design, language, or publication date in order to maximise retrieval sensitivity. Both prospective and retrospective studies were eligible, including cross-sectional and longitudinal designs, provided that participants completed both the screening instrument and the reference standard assessment (Table 1).

2.3.2 Table 1: PICO (Population, Intervention, Comparison, Outcome) Framework

2.4 Information Sources and Search Strategy

2.4.1 Databases Searched

A comprehensive electronic search was conducted across five databases: PubMed, Ovid MEDLINE, Embase, Global Health, and APA PsycINFO. This multi-database approach was selected to ensure comprehensive coverage of peer-reviewed biomedical and psychological literature relevant to the review objectives.

2.4.2 Search Strategy Development

The search strategy was developed in collaboration with a specialist information scientist and structured around four conceptual domains: (1) the perinatal period (population), (2) depression and common mental disorders (condition), (3) screening, diagnostic assessment, and validation (intervention), and (4) the GCC region (setting). Controlled vocabulary terms (e.g., MeSH headings) were combined with free-text keywords, and Boolean operators (AND, OR) were used to maximise retrieval. The strategy was adapted for each database as required, with no restrictions applied to language or publication date. The Ovid MEDLINE search strategy is presented as an illustrative example of the approach used across all databases (see Table 2). Reference lists of all included studies were also screened manually to identify additional eligible studies not captured through electronic searches. All search strategies, dates, and record counts were documented. Complete search strategies for all databases are reported in Appendix B.

Table 1. Predefined eligibility criteria organised by PICO components

Component	Inclusion Criteria	Exclusion Criteria
Population	• Pregnant and postpartum women during the perinatal period (antenatal: conception to delivery; postnatal: delivery to 12 months postpartum) • Women residing in any of the six GCC countries: Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, UAE • All ages, parities, and sociodemographic backgrounds	• Women outside the defined perinatal period • Studies conducted outside GCC countries • GCC nationals residing outside the GCC region
Intervention	• Studies evaluating screening or case-identification tools for perinatal depression (e.g., EPDS, PHQ-9, BDI, locally developed or culturally adapted instruments) • Self-administered or interviewer administered tools • Tool development studies that include formal validation against a diagnostic reference standard • Studies reporting sufficient data to assess diagnostic accuracy or validation performance	• Tools primarily measuring quality of life, maternal-infant bonding, or other non-depression constructs • Studies assessing screening tools for conditions other than perinatal depression • Studies without a gold-standard diagnostic comparison • Studies focused exclusively on treatment without a screening or case identification component
Comparator	• Studies employing any type of reference standard, including structured clinical diagnostic interviews or other validated psychological measures selected by the original investigators for validation purposes. • Studies where interviewer training is unclear are included but flagged in quality assessment	• Studies lacking any form of reference standard comparison
Outcomes	• Primary: AUC, sensitivity, specificity, PPV, NPV, positive/negative likelihood ratios, Cronbach's alpha, optimal cut-off thresholds • Secondary: Prevalence from diagnostic interviews; test-retest reliability; inter-rater reliability; construct validity; screening acceptability; cultural and contextual implementation barriers • Outcomes measured during pregnancy or up to 12 months postpartum	• Outcomes assessed beyond 12 months postpartum • Studies reporting screening scores only, without psychometric or diagnostic accuracy data

Table 2: Ovid MEDLINE Search Strategy (Illustrative Example)

Line	Search String
1.	prenatal care/
2.	Pregnancy/
3.	Maternal Health/
4.	(antenatal OR ante-natal OR antepartum OR ante-partum OR birth OR childbirth OR maternal OR mother OR natal OR perinatal OR peri-natal OR peripartum OR peri-partum OR postnatal OR post-natal OR postpartum OR post-partum OR pregnan* OR prenatal OR pre-natal OR puerperal OR trimester OR women).ti,ab
5.	1 OR 2 OR 3 OR 4
6.	mental disorders/ OR exp mood disorders/
7.	Depression/
8.	6 OR 7
9.	5 AND 8
10.	Depression, Postpartum/
11.	(antenatal or ante-natal or antepartum or ante-partum or birth or childbirth or maternal or mother or natal or perinatal or peri-natal or peripartum or peri-partum or postnatal or post-natal or postpartum or post-partum or pregnan* or prenatal or pre-natal or puerperal or trimester or women) adj5 (adjustment or affective or common mental disorder* or CMD or depression or depress* or distress* or mood or psychiatric or psychological).ti,ab.
12.	9 OR 10 OR 11
13.	(assessment OR checklist OR diagnosis OR diagnostic OR identification OR identify OR index OR instrument OR interview OR inventory OR questionnaire OR reliable OR reliability OR scale OR score OR scoring OR screen OR screening OR sensitive OR sensitivity OR specific OR specificity OR standard OR standardised OR standardized OR survey OR test OR tool OR validation OR validity).ti,ab
14.	("Gulf Cooperation Council" OR GCC OR Bahrain OR Bahraini OR Kuwait OR Kuwaiti OR Oman OR Omani OR Qatar OR Qatari OR Saudi Arabia OR Saudi OR "United Arab Emirates" OR UAE OR Emirati OR Gulf OR "Arab Gulf" OR "Persian Gulf" OR "Middle East" OR "Arab countries" OR Arab*).ti,ab
15.	12 AND 13 AND 14

2.5 Study Selection Process

Titles and abstracts retrieved from the electronic searches were independently screened by two reviewers against the predefined eligibility criteria. Records deemed potentially eligible or unclear by either reviewer progressed to full-text review. Full-text articles were then independently assessed for final inclusion, and reasons for exclusion were documented systematically.

Disagreements at either stage were resolved through discussion until consensus was reached, with a third reviewer available if needed. Duplicate records were identified and removed using reference management software. Screening decisions were managed using Rayyan (40), a systematic review software. The study selection process is presented using a PRISMA 2020 flow diagram (39).

2.6 Data Extraction

A standardised data extraction form was developed and piloted on a subset of included studies prior to full extraction. Data were independently extracted by two reviewers, with any discrepancies resolved through discussion until consensus was achieved. Extracted data included study characteristics (author(s), year, country, design, setting, and sample size), participant characteristics (age, antenatal or postnatal status, gestational age or weeks postpartum, recruitment method, and eligibility criteria), screening tool characteristics (instrument name, language, mode and timing of administration, and cut-off scores), and details of the reference standard used for validation. Where reported, the timing of administration relative to screening was also extracted. Outcome data included measures of diagnostic and psychometric performance where available and consistently reported across studies.

2.7 Quality Assessment and Risk of Bias

Methodological quality and risk of bias were assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool (39). QUADAS-2 was selected because it is specifically designed for evaluating diagnostic accuracy studies and provides a structured and transparent approach to assessing risk of bias across four domains: patient selection, index test, reference standard, and flow and timing (39). In this review, QUADAS-2 was used primarily to assess risk of bias; although the tool also includes applicability domains, these were not formally analysed because evaluating applicability across different clinical and cultural settings was beyond the scope of the review. The

adapted QUADAS-2 framework used in this review is provided in Appendix A.

Quality assessment was conducted independently by two reviewers, with disagreements resolved through discussion and consensus. Studies were not excluded on the basis of methodological quality. Instead, quality assessments were used to inform interpretation of the findings, particularly where variations in study methods may have influenced reported diagnostic performance. Results are presented using summary tables and traffic-light plots to illustrate risk of bias across the included studies.

2.8 Data Synthesis

Diagnostic accuracy data, including sensitivity, specificity, positive predictive value, and negative predictive value, were extracted for each reported cut-off score. Given the small number of eligible studies and substantial heterogeneity in populations, reference standards, assessment timing, and cut-off thresholds, findings were synthesised narratively rather than through meta-analysis. Table 3 summarises the included studies by screening instrument, population, perinatal period, and country, and presents key study characteristics and psychometric properties. The synthesis focused on identifying patterns in diagnostic performance and optimal cut-off thresholds while considering variations in study methodology and risk of bias.

3. RESULTS

3.1 Study Selection

The systematic search identified 3,510 records across all databases. After the removal of 1,858 duplicates, 1,652 unique records remained for title and abstract screening. Of these, nine full-text articles were retrieved for detailed assessment. Following full-text review, three studies met all inclusion criteria and were included in the review. The PRISMA flow diagram (Figure 1) provides a detailed overview of the study selection process, including reasons for exclusion at each stage. The primary reasons for exclusion at the full-text stage were the absence of a reference standard diagnostic assessment and lack of focus on perinatal depression.

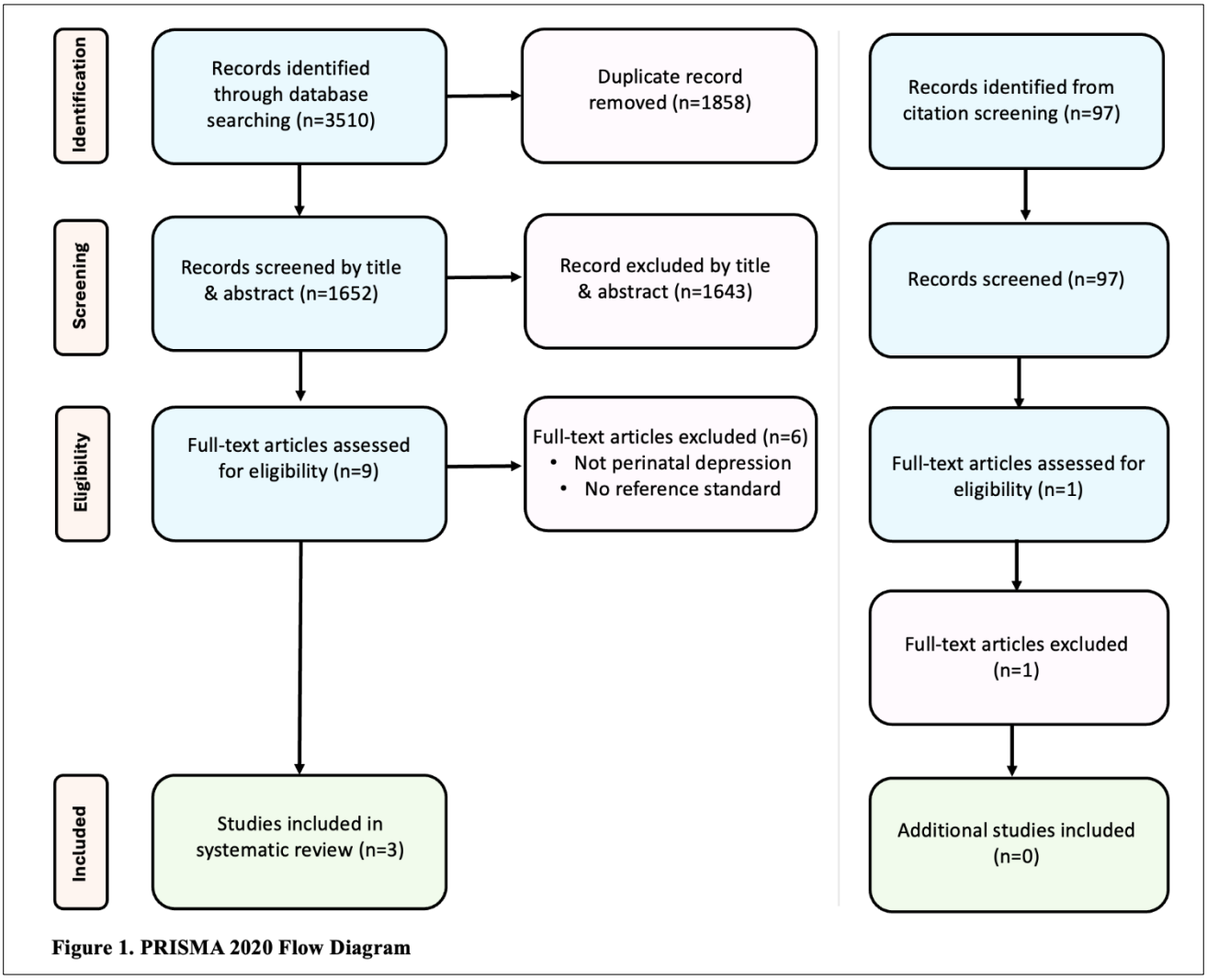


Figure 1. PRISMA 2020 Flow Diagram

3.2 Study Characteristics

Three studies met the inclusion criteria for this review. All studies were conducted in the GCC and evaluated the psychometric properties of the Arabic version of the EPDS for screening perinatal depression. The characteristics of the included studies are presented in Table 3.

Table 3. Characteristics of Included Studies

Study	Setting	Study design	Sample Size	Condition	Screening tool	Comparator	Language	Optimal cut-off	Sensitivity	Specificity	Prevalence	Other
Ghubash (1997)	Urban (New Dubai Hospital, UAE)	Prospective	95	Postnatal depression	EPDS	PSE	Arabic	EPDS: 10	0.91	0.84	EPDS: (cut-off 10): 26%	At cut-off 12: sens 0.73; spec 0.90
Naja (2019)	Urban (nine primary health care centres, Qatar)	Cross-sectional	128	Antenatal depression	EPDS	MINI	Arabic	EPDS: ≥13	EPDS: 0.87	EPDS: 0.90	MINI: 24%	EPDS: AUC 0.951 (SE=0.02; 95% CI=0.91 to 0.99); PPV 0.75
					BDI-II			BDI-II: ≥19	BDI-II: 0.96	BDI-II:0.73		BDI-II: AUC 0.912 (SE=0.025; 95% CI=0.86 to 0.96); PPV 0.54
Al-Abri (2025)	Mixed urban (one tertiary hospital + five primary health centres, Oman)	Cross-sectional	352	Antenatal depression	EPDS	PHQ-9	Arabic	EPDS: 11 overall; trimester-specific: 10 (first), 11 (second), 12 (third)	0.863 overall; 1.00 (first trimester); 0.862 (second trimester); 0.786 (third trimester)	0.714 overall; 0.632 (first trimester); 0.699 (second trimester); 0.810 (third trimester)	EPDS: (cut-off 12): 33%	AUC = 0.85; PPV = 45.7%; NPV = 94.9%

3.3 Narrative Summary of Included Studies

Three studies comprising 575 participants were included in this review. All three studies were conducted in the GCC: one in the United Arab Emirates (41), one in Qatar (42), and one in Oman (33). Study designs included one prospective study (41) and two cross-sectional studies (33,42). Sample sizes ranged from 95 to 352 participants.

All three studies evaluated the EPDS for screening perinatal depression using a diagnostic reference standard (33,41,42). One study focused exclusively on postnatal depression (41), whilst the remaining two examined antenatal depression (33,42). Diagnostic reference standards varied across studies: the PSE (Present State Examination) with Catego classification (41), the Mini International Neuropsychiatric Interview (42), and the Patient Health Questionnaire-9 (33). All studies administered the EPDS in Arabic.

Optimal cut-off scores ranged from 10 to 13 across studies. Ghubash *et al.,* (1997) reported two cut-offs (10 and 12), with cut-off 10 demonstrating higher sensitivity (0.91) and cut-off 12 demonstrating higher specificity (0.90) (41). Naja *et al.,* (2019) identified an optimal cut-off of ≥13 (sensitivity 0.87, specificity 0.90) (Naja *et al.,* 2019) (42), whilst Al-Abri *et al.,* (2025) reported an overall optimal cut-off of 11 (sensitivity 0.863, specificity 0.714) and proposed trimester-specific cut-offs (10 for first trimester, 11 for second trimester, 12 for third trimester) (33). The prevalence of perinatal depression ranged from 24% to 33%, depending on the diagnostic instrument employed and the timing within the perinatal period.

Timing of screening varied across studies. Ghubash *et al.,* (1997) administered the EPDS at one week postpartum with diagnostic assessment at eight weeks postpartum (41). The two antenatal studies recruited participants across all trimesters, with the majority of participants in the second and third trimesters (Naja *et al.,*, 2019: 48% second trimester, 30% third trimester (42); Al-Abri *et al.,*, 2025: 45.2% at ≥28 weeks gestation (33)).

3.4 Study by Study Methodological Appraisal

3.4.1 Ghubash (1997)

Ghubash *et al.,* (1997) demonstrates several methodological strengths that support aspects of internal validity, although important limitations substantially affected the overall quality of the study (41). The use of a consecutive sampling strategy reduced the likelihood of selection bias and enhanced representativeness within the clinical setting. The prospective design avoided the inflated diagnostic accuracy estimates commonly observed in case-control studies. All participants underwent both the index test (the Edinburgh Postnatal Depression Scale) and a validated diagnostic reference standard (the Present State Examination with Catego classification). This approach ensured consistency in verification and minimised the risk of partial or

differential verification bias. The translation of the Edinburgh Postnatal Depression Scale followed rigorous forward and backward procedures, and the reported internal consistency indicated acceptable reliability within this population.

However, several methodological weaknesses introduce a high risk of bias (Table 4). The restriction of the sample to a highly selected group of local women limits external validity and reduces the applicability of findings to more diverse populations. The small sample size reduces statistical power and increases the likelihood of imprecise estimates of sensitivity and specificity. In addition, the evaluation of multiple cut off points introduces a risk of overestimation of diagnostic accuracy due to data driven optimization (35,36). The absence of clearly reported blinding procedures raises concerns about potential information bias. Most critically, the substantial time interval between administration of the index test and the reference standard represents a major limitation. Changes in clinical status over time may have resulted in misclassification, thereby biasing estimates of diagnostic performance.

Overall, this study is judged to be at high risk of bias, with moderate to high concerns regarding applicability, particularly in relation to the representativeness of the study population and the temporal integrity of the diagnostic assessment.

3.4.2 Naja *et al.*, (2019)

Naja *et al.,* (2019) present a study with several methodological strengths alongside notable limitations that influence its overall quality (42). A key strength is the use of the Mini International Neuropsychiatric Interview as the reference standard, which is a structured and validated tool for diagnosing major depressive disorder (34). The study population comprises pregnant women across different trimesters attending primary healthcare settings, which aligns with the intended population for antenatal depression screening. All participants received both index tests, the Edinburgh Postnatal Depression Scale and Beck Depression Inventory II, as well as the same reference standard, thereby mising the risk of differential verification bias. Recruitment across nine primary healthcare centres enhances sample heterogeneity and supports external validity. The study also provides preliminary evidence on the diagnostic accuracy of Arabic versions of commonly used screening tools in an understudied antenatal population.

Despite these strengths, several methodological limitations result in a high risk of bias (Table 4). The use of data driven cut off scores rather than prespecified thresholds introduces a significant risk of overfitting and likely inflates estimates of diagnostic accuracy (35,36). This approach limits reproducibility and generalisability in external populations. Furthermore, key methodological details are insufficiently reported. There is no clear description of sampling methods, blinding procedures, or the time interval between index test administration and the reference standard. As a result, the risk of bias remains unclear across multiple domains. Although the Mini International Neuropsychiatric Interview is acceptable, the absence of the Structured Clinical Interview for diagnostic classification represents a limitation. The relatively small sample size further reduces the precision of diagnostic accuracy estimates.

In summary, applicability concerns are relatively low due to the appropriate population and setting, but the overall risk of bias is high. The reported sensitivity and specificity are likely optimistic, and the proposed cut off values should be interpreted as preliminary and requiring validation in larger and more methodologically robust studies.

3.4.3 Al-Abri (2025)

The Omani validation study demonstrates several methodological strengths that support internal consistency, although important limitations affect overall validity (33). A key strength is the concurrent administration of both the index test and comparator instrument, which eliminates temporal bias and ensures stability of symptom reporting across assessments. Complete participant retention, with all 352 participants included in the analysis, removes the risk of attrition bias. The use of culturally validated Arabic instruments strengthens the relevance of the findings to the target population. Recruitment across both tertiary and primary healthcare settings improves representativeness within the healthcare system and captures a broader clinical population.

However, several limitations introduce a high risk of bias (Table 4). The use of convenience sampling, combined with the exclusion of irregular clinic attenders, creates substantial selection bias and may restrict the clinical spectrum. This could lead to underestimation of depression prevalence and biased estimates of diagnostic accuracy. The use of data driven threshold optimisation further limits generalisability, as such methods tend to inflate performance metrics and may not replicate in external validation studies (35,36). The absence of clearly reported blinding procedures raises concerns about interpretation bias, although the self-administered format offers partial mitigation.

The most important limitation relates to the choice of reference standard. The use of the PHQ-9 rather than a structured diagnostic interview introduces classification uncertainty. As both instruments assess overlapping symptom domains, this creates circularity and may artificially increase agreement without establishing true diagnostic accuracy.

Overall, this study is judged to be at high risk of bias, particularly in patient selection and reference standard domains. While applicability to the Omani healthcare context is strong, the methodological limitations require cautious interpretation, as the findings are likely to overestimate diagnostic performance (Table 4).

Table 4. Risk of bias (quality) of studies included in review as assessed using QADAS-2

Study	Patient Selection	Index Test	Reference Standard	Flow and Timing
Ghubash (1997)	HIGH	HIGH	UNCLEAR	HIGH
Naja (2019)	UNCLEAR	HIGH	UNCLEAR	UNCLEAR
Al-Abri (2025)	HIGH	HIGH	UNCLEAR	LOW

Low risk of bias
Unclear risk of bias
High risk of bias

3.5 Cross Study Methodological Synthesis

Across the included studies, several consistent methodological concerns emerge. Whilst most studies used appropriate screening tools in relevant populations, the widespread reliance on data-driven thresholds represents a major source of bias and is likely to inflate diagnostic accuracy estimates (35,36). Limitations in reporting, particularly regarding blinding procedures and timing between index tests and reference standards, introduce further uncertainty in risk of bias assessment.

Variation in the choice of reference standards also affects interpretability, with some studies relying on screening tools rather than structured diagnostic interviews, thereby reducing diagnostic certainty. In addition, issues related to selection bias and restricted population sampling reduce the generalisability of findings across different clinical and cultural contexts. Taken together, the overall body of evidence is characterised by moderate to high risk of bias. These limitations have important implications for the synthesis and interpretation of diagnostic accuracy estimates and support the need for cautious weighting of individual studies in meta-analysis.

4. DISCUSSION

4.1 Overview of Key Findings

This review evaluated the diagnostic accuracy, psychometric properties, and implementation considerations of perinatal depression screening tools across Gulf Cooperation Council countries. Three validation studies were identified, conducted in the United Arab Emirates, Qatar, and Oman, with a combined sample of 575 participants spanning antenatal and postnatal populations (33,41,42). All studies assessed the Arabic version of the EPDS against differing reference standards, reporting optimal cut-off scores ranging from 10 to 13, with sensitivity estimates between 0.863 and 0.91 and specificity estimates between 0.714 and 0.90. These findings indicate broadly acceptable diagnostic performance across the included settings. Methodological appraisal using QUADAS-2 (38) revealed a consistent pattern of moderate to high risk of bias across all included studies, attributable primarily to data-driven threshold optimisation, inadequate reporting of blinding procedures, variation in reference standards, and selection bias. No validation studies were identified for Kuwait, Bahrain, or Saudi Arabia, representing a substantial evidence gap for the rest of GCC member states.

Overall, these findings confirm that whilst the Arabic EPDS demonstrates acceptable psychometric properties in selected GCC populations, the evidence base remains limited in scope, methodological rigour, and geographical coverage. The absence of studies employing gold-standard structured diagnostic interviews, combined with the widespread use of cut-off optimisation, suggests that reported accuracy estimates are likely optimistic and may not generalise to routine clinical practice. Equally significant is the absence of any systematic examination of cultural acceptability, implementation barriers, or the influence of somatic symptom expression on screening performance, omissions that carry particular weight given the sociocultural specificities of the GCC context.

4.2 Interpretation of Findings

4.2.1 Diagnostic Accuracy and Psychometric Performance

The sensitivity and specificity values reported across the three included studies appear broadly comparable to international benchmarks for the EPDS (32,43). However, this apparent consistency requires cautious interpretation in light of the methodological limitations identified through QUADAS-2 appraisal. A key concern is the reliance on data-driven threshold optimisation across all studies, an approach associated with overfitting: cut-off scores selected to maximise performance within the validation sample reflect sample-

specific optimisation rather than accuracy that can be confidently generalised to independent populations.

Variation in optimal cut-off thresholds across studies, ranging from 10 to 13, further complicates interpretation (33,41,42). These differences may reflect genuine variation in symptom presentation across gestational stages, particularly given the proposal of trimester-specific thresholds by Al-Abri and colleagues (33). Alternatively, they may arise from heterogeneity in study populations and reference standards. The identification of a higher antenatal threshold of 13 by Naja and colleagues is particularly informative in this context (42). The improved specificity associated with this cut-off may be clinically advantageous in settings where over-referral carries social consequences related to stigma and imposes additional demands on limited specialist services. In contrast, the lower threshold of 10 reported by Ghubash and colleagues in the postnatal period prioritises sensitivity, reflecting the heightened clinical risk associated with missed cases during early postpartum stages (41). Together, these findings underscore that optimal cut-off values are inherently context-dependent and cannot be assumed to transfer across populations without empirical validation.

Interpretation is further complicated by the variation in reference standards employed across the included studies. Ghubash and colleagues used the Present State Examination with Catego classification (41), Naja and colleagues used the Mini International Neuropsychiatric Interview (42), and Al-Abri and colleagues used the Patient Health Questionnaire-9 (PHQ-9) as a comparator instrument (33). The use of the PHQ-9 as a reference standard is methodologically problematic because both the PHQ-9 and EPDS assess overlapping symptom domains. Their apparent agreement may therefore reflect shared item content rather than accurate case identification. This circularity inflates reported diagnostic performance and represents a major source of bias, substantially limiting the comparability of findings across studies.

4.2.2 Comparison with International Studies

When situated within the broader international validation literature, the diagnostic performance of the EPDS in GCC studies appears broadly comparable, though important methodological differences constrain direct comparison. A comprehensive individual participant data meta-analysis by Levis and colleagues, encompassing 58 studies and 15,557 participants from diverse settings, identified an optimal cut-off of ≥11 for detecting major depression in perinatal women, with sensitivity of 0.81 (95% CI 0.75–0.87) and specificity of 0.88 (95% CI 0.85–0.91) when semi-structured interviews served as the reference standard (32). At the commonly employed threshold of ≥13, sensitivity decreased to 0.66 (95% CI 0.58–0.74) whilst specificity increased to 0.95 (95% CI 0.92–0.96), illustrating the inherent trade-off between case detection and false-positive rates (32). These findings suggest that the optimal cut-off may be lower than the traditionally recommended threshold of ≥13, though the appropriate choice remains dependent on clinical context and the relative costs of false negatives versus false positives.

A more recent systematic review and meta-analysis by Park and Kim, synthesising data from 17 studies involving 2,902 women across 12 countries, reported pooled sensitivity of 0.79 (95% CI 0.74–0.84) and specificity of 0.88 (95% CI 0.82–0.92), with a summary receiver operating characteristic area under the curve (AUC) of 0.89 (43). The authors recommended a cut-off of 13, consistent with the optimal range of 9–13 identified in previous reviews, whilst acknowledging considerable variability across individual studies (43). Subgroup analyses demonstrated that diagnostic performance remained robust across pregnancy (sensitivity 0.81, specificity 0.87, AUC 0.90) and the postpartum period (sensitivity 0.79, specificity 0.92, AUC 0.90), though performance was somewhat lower in mixed perinatal samples (sensitivity 0.72, specificity 0.83, AUC 0.74) (43), supporting the EPDS's utility as a screening instrument across both antenatal and postnatal settings.

Individual validation studies from culturally diverse contexts illuminate the extent to which performance varies with population characteristics and methodological choices. In Turkey, Aydin and colleagues validated the EPDS among 341 postpartum women attending primary healthcare clinics, identifying an optimal cut-off of 12.5 for combined major and minor depression, with sensitivity of 75.5%, specificity of 71.5%, and an AUC of 0.72 (44). These values are notably lower than those reported in many international studies, leading the authors to characterise the scale's validity as respectable but not excellent. The authors attributed these findings partly to translation quality, suggesting that verbatim literal translation may have impaired comprehension, and partly to population characteristics, including educational level and familiarity with survey instruments (44). The self-administered format was also identified as a potential contributor to increased false negatives, highlighting the importance of administration mode in screening accuracy (44).

In contrast, a validation study conducted in Kerman, Iran, involving 200 postnatal women and using DSM-IV Axis I criteria as the reference standard, demonstrated substantially higher accuracy (45). An optimal cut-off of 12/13 for major depression yielded a sensitivity of 95.3% and a specificity of 87.9%, with an AUC of 98%. For combined major and minor depression, the optimal cut-off was 11/12, with a sensitivity of 81.2% and a specificity of 91.6% (AUC 93%) (45). The authors noted consistency with findings from previous international studies, suggesting cross-cultural stability in the EPDS psychometric properties

when appropriately translated and administered. However, they acknowledged that verbal administration for illiterate participants introduced potential response bias, although this approach was considered necessary to ensure generalisability to the broader population (45).

A systematic review and meta-analysis specific to the Indian context by Fellmeth and colleagues, encompassing seven studies and 1,003 women, found that at the threshold of ≥13, the EPDS demonstrated pooled sensitivity of 88.9% (95% CI 77.4–94.9) and specificity of 93.4% (95% CI 81.5–97.8) (46). When optimal thresholds from individual studies (ranging from ≥9 to ≥13) were applied, pooled sensitivity increased to 94.4% (95% CI 81.7–98.4) with specificity of 90.8% (95% CI 83.7–95.0) (46), suggesting that the EPDS performs particularly well in Indian settings. Notably, this review identified no validated alternative screening tools for perinatal common mental disorders in India, highlighting the EPDS as the sole psychometrically validated instrument in this context (46), a finding that resonates with the current review's observation of a similarly narrow evidence base within the GCC.

Across these international studies, methodological heterogeneity emerges as a consistent and important source of variation in reported EPDS performance. Reference standards varied considerably, with studies employing the Structured Clinical Interview for DSM Disorders (SCID), the MINI, DSM-IV criteria, and ICD criteria, each carrying differing requirements for structure and interviewer training (32,43,45,46). Levis and colleagues noted that 34 of their included studies used the SCID, 18 used the MINI, and four used fully structured interviews such as the Composite International Diagnostic Interview, with the MINI analysed separately owing to its distinct design characteristics (32). The choice of reference standard directly influences sensitivity and specificity estimates, as semi-structured interviews permit greater clinical judgement whilst fully structured interviews prioritise standardisation (32). Levis and colleagues further reported that their sensitivity estimates were 6–13% lower and specificity estimates 4–7% higher than those from a previous meta-analysis conducted 13 years earlier, attributing these differences to the larger sample and the inclusion of data across all cut-off values (32).

Cultural and linguistic factors emerge with equal consistency as determinants of EPDS validity. The Turkish validation study's concerns about translation quality (44) are consistent with the broader observation that the EPDS, originally developed in English-speaking, Westernised societies, may require careful cultural adaptation to maintain psychometric integrity in non-Western contexts. The Iranian study, despite its strong psychometric outcomes, acknowledged the challenge of verbal administration for illiterate participants (45), underscoring the need to balance standardisation with accessibility in diverse populations. Park and Kim similarly noted that cut-off variability across their included studies likely reflected both methodological differences and genuine cultural variation in symptom expression and reporting (43). These considerations are directly relevant to the GCC context, where somatic presentations, stigma, and linguistic diversity are well documented barriers to screening validity, yet these factors were not empirically examined in the included GCC studies.

Limitations identified across the international literature provide important context for interpreting the present findings. Levis and colleagues noted moderate heterogeneity and insufficient data for subgroup analyses based on cultural factors such as country or language, with limited representation of low- and middle-income countries (32). Park and Kim acknowledged that small study numbers in some subgroups precluded identification of sources of heterogeneity, and that the exclusion of women under 18 or with high-risk pregnancies limited generalizability (43). Fellmeth and colleagues emphasised that the absence of a validated alternative screening tool in India highlights a significant gap in the evidence base, a gap that parallels the situation in the GCC (46). Taken together, these findings suggest that while the EPDS demonstrates consistent utility across diverse cultural settings, its performance remains strongly influenced by methodological and contextual factors. Apparent agreement with international benchmarks does not exclude the presence of important local differences.

In summary, the findings of the present study are consistent with international evidence supporting the EPDS as a useful screening tool for perinatal depression across diverse cultural settings. However, performance is consistently influenced by methodological factors such as the choice of reference standard, translation quality, mode of administration, and population characteristics. The strong performance observed in Indian settings (46), alongside the more modest values reported in Turkey (44), highlights the importance of local validation and culturally appropriate cut-off selection, a conclusion that is equally relevant to the GCC context.

4.2.3 Patterns and Evidence Gaps

Prevalence estimates of perinatal depression across the included studies ranged from 24% to 33% (33,41,42). This variation is consistent with the wider GCC literature but raises important methodological concerns. Prevalence estimates are highly sensitive to differences in measurement methods and study design and cannot be meaningfully compared across settings without established psychometric equivalence of the instruments used. In the absence of such validation, observed differences in prevalence may reflect variation in measurement properties rather than true differences in population burden.

A key finding is the complete absence of eligible validation studies from Kuwait, Bahrain, and Saudi Arabia. This is particularly notable for Saudi Arabia, which has the largest population in the region and a substantial body of perinatal mental health research. Although potentially relevant studies were identified, none met the criterion of validation against a reference standard, suggesting that much of the existing literature relies on screening tools without formal psychometric validation.

The concentration of validation evidence in only three GCC member states further constrains generalisability. The United Arab Emirates (UAE), Qatar, and Oman differ in population structure, healthcare systems, and socioeconomic context, and these differences are especially pronounced relative to the unrepresented countries. In particular, the high proportion of expatriate populations in Qatar and the UAE may influence the composition of antenatal samples, potentially limiting the applicability of findings to national populations or to less demographically diverse settings.

4.2.4 Positive Predictive Value, Clinical Utility, and Threshold Validity

Variation in positive predictive value (PPV) across GCC validation studies has important implications for clinical utility. The Omani study reported a PPV of 45.7% at a cut-off of 11(33), while the Qatari study reported a PPV of 75% at a cut-off of ≥13 (42). These estimates should be interpreted in relation to the relatively high prevalence observed in the validation samples (24%-33%), since PPV is directly dependent on prevalence. Evidence from international validation studies indicates that PPV is likely to decrease in lower-prevalence settings even when sensitivity and specificity remain comparable; accordingly, the reported values may not directly reflect performance in routine clinical contexts.

The clinical implications are notable. A PPV of 45.7% indicates that a substantial proportion of individuals who screen positive may not meet diagnostic criteria upon further assessment, which may increase referral demands on specialist services (33). The higher PPV observed at a cut-off of ≥13 in the Qatari study suggests improved specificity at higher thresholds, although this is likely to be accompanied by reduced sensitivity (42). Notably, none of the included studies evaluated how screening thresholds might be optimised in relation to service capacity, referral pathways, or implementation context. This gap limits the translation of psychometric findings into practical guidance.

The proposal of trimester-specific cut-offs in the Omani study represents a novel contribution, though one that requires cautious interpretation (33). Lower thresholds were associated with higher sensitivity and lower specificity in early pregnancy, whilst higher thresholds in later trimesters showed the reverse pattern. Although this may reflect genuine variation in symptom presentation across gestation, the thresholds were derived through data-driven optimisation within subgroups of the same sample. Given the smaller subgroup sizes, these estimates may be unstable, and the reported 100% sensitivity in the first trimester is unlikely to generalise without independent validation. Evidence from other validation studies has generally supported the use of a single cut-off across pregnancy, and the feasibility of implementing trimester-specific thresholds in routine care is further constrained by the need for simple, standardised protocols. In the absence of replication in independent GCC samples, trimester-specific cut-offs should be regarded as preliminary, and further research is needed to determine whether gestational stage meaningfully influences screening performance.

4.2.5 Cultural and Contextual Barriers to Screening Validity and Implementation

The available evidence highlights substantial cultural, contextual, and systemic barriers that directly affect the validity, acceptability, and overall effectiveness of perinatal depression screening within GCC populations. Although these barriers were not empirically examined within the included studies, their consistent identification in the wider literature enables a critical synthesis of the factors constraining implementation.

At the cultural level, divergent illness conceptualisations represent a primary challenge. Perinatal distress is frequently interpreted through non-biomedical frameworks encompassing spiritual attribution, social adversity, and religious meaning (23), and these explanatory models shape both how women interpret screening items and whether questions are perceived as relevant. Standardised instruments consequently risk limited acceptability, as responses may reflect culturally mediated interpretations of distress rather than the psychopathological constructs the instruments are designed to measure.

Somatic symptom expression further complicates screening accuracy. In GCC populations, psychological distress is often communicated through somatic complaints, including pain, fatigue, and gastrointestinal disturbance. Screening tools such as the EPDS, which minimise somatic content to avoid overlap with perinatal physiology, may therefore fail to capture clinically significant presentations (37). In addition, culturally specific idioms of distress provide alternative linguistic frameworks through which symptoms are articulated (47). The absence of integration of such idioms into screening instruments reduces sensitivity to locally meaningful expressions, increasing the risk of under-identification and limiting the effectiveness of screening programmes.

Stigma represents a pervasive systemic barrier. Concerns regarding confidentiality, fear of social judgement, and perceived consequences for family reputation contribute to deliberate under-reporting of symptoms (24), directly undermining the validity of self-report measures and reducing participation rates. Healthcare delivery structures introduce further constraints: brief, checklist-based tools restrict opportunities for clarification and contextualisation, particularly for individuals whose experiences do not align with predefined symptom categories, reducing both acceptability and the reliability of responses.

Collectivist social organisation within GCC societies additionally shapes help-seeking patterns and decision-making. Perinatal mental health is often embedded within family systems, with partners and extended family members playing a central role in recognising symptoms and determining whether care is sought (24). Screening approaches that focus exclusively on the individual do not capture these relational dynamics, limiting their ecological validity and reducing their effectiveness in identifying cases mediated by family influence.

Taken together, these barriers indicate that effective screening implementation in the GCC extends well beyond linguistic translation. Cultural congruence, symptom representation, social acceptability, and healthcare structure interact to influence how screening tools are understood, whether they are engaged with, and how accurately they identify clinically relevant distress. Without the incorporation of culturally grounded frameworks and system-level considerations, screening programmes remain at risk of reduced acceptability, incomplete uptake, and diminished effectiveness.

4.3 Alignment with Study Context

The findings of this review align closely with the study objectives and conceptual framework. The review identified three eligible studies that validated screening tools for perinatal depression among women in GCC countries. All studies evaluated the Arabic version of the Edinburgh Postnatal Depression Scale (EPDS), and no validated alternative instruments were identified. This suggests that perinatal depression screening in the GCC is highly dependent on a single instrument, reflecting a limited and standardised evidence base.

The review also evaluated the psychometric performance of the Arabic EPDS through QUADAS-2 appraisal and narrative synthesis. Across the included studies, the EPDS demonstrated generally acceptable sensitivity and specificity. Additional indicators, including area under the receiver operating characteristic curve, predictive values, likelihood ratios, internal consistency, and optimal cut-off thresholds, were reported inconsistently, limiting meaningful comparison across studies. Furthermore, the presence of moderate to high risk of bias reduced confidence in the reported estimates and restricted the generalisability of the findings.

Evidence relating to cultural, contextual, and systemic barriers to screening implementation was limited. None of the included studies explicitly examined cultural validity, participant acceptability, or factors influencing screening uptake. Consequently, these issues could only be considered through contextual interpretation. This absence represents an important finding in itself, highlighting a significant gap in understanding the real-world implementation of perinatal depression screening within GCC populations.

The review only partially addressed the objective of collating prevalence estimates derived from diagnostic reference standards and evaluating additional psychometric properties. All included studies reported prevalence estimates; however, the prevalence data presented in this review were extracted and reported as presented in the original publications. Consequently, prevalence estimates were derived either from screening tool thresholds or from diagnostic reference standards, depending on the methodology of the individual study, and this distinction was explicitly reported within the narrative synthesis. Additional psychometric properties, including test-retest reliability, inter-rater reliability, and construct validity, were not consistently reported and therefore could not be systematically evaluated. These limitations reduce the comparability of prevalence estimates across studies and restrict the ability to draw firm conclusions regarding the broader psychometric performance of the EPDS in GCC populations.

Finally, the review addressed the objective of developing a culturally informed screening framework for the GCC by synthesising the available evidence into a series of contextually relevant recommendations. These recommendations encompass screening timing, adaptation of screening instruments, and the establishment of tiered referral pathways. The proposed framework also emphasises the importance of narrative and context-sensitive approaches to complement standardised screening measures and better capture the sociocultural complexities of GCC populations.

4.4 Strengths and Limitations

4.4.1 Strengths of the Review

This review makes several methodological contributions that enhance the credibility and utility of its findings. A comprehensive search strategy was conducted across five major bibliographic databases, namely PubMed, Ovid MEDLINE, Embase, Global Health, and APA PsycINFO. To maximise retrieval sensitivity and identify potentially relevant studies not captured through database searching, the reference lists of all eligible studies were also manually screened. No language or date restrictions were applied, which

minimised selection bias and enabled a more inclusive identification of the available evidence.

Prospective registration with PROSPERO prior to data extraction strengthens methodological transparency and reduces the risk of selective reporting. The review protocol is publicly available through the PROSPERO registration record, and adherence to PRISMA 2020 and PRISMA-DTA reporting standards further facilitates independent scrutiny of the review process. The application of QUADAS 2 for quality assessment provides a standardised and validated framework for evaluating risk of bias and applicability in diagnostic accuracy studies, enabling systematic and replicable appraisal of study quality. Finally, the inclusion of all six GCC member states, rather than a single national context, supports broader regional inference and enables the identification of geographical gaps in the existing literature.

4.4.2 Limitations and Their Implications

A narrative synthesis was conducted which, although methodologically appropriate given the available evidence, is inherently more susceptible to subjective interpretation than statistical synthesis. The findings should therefore be understood as qualitative assessments of patterns across studies rather than as precise pooled estimates of diagnostic accuracy.

The small number of included studies (n = 3) further constrains the robustness of the review. With such a limited evidence base, individual studies exert a disproportionate influence on overall conclusions, reducing both statistical power and inferential reliability. Restriction to peer-reviewed published studies introduces the potential for publication bias, as studies reporting favourable or statistically significant findings are more likely to be published. Grey literature was excluded owing to concerns about replicability and methodological rigour, as such sources frequently lack the standardised reporting necessary for robust critical appraisal. The review is also dependent on the completeness and accuracy of data reported within the primary studies; the findings should consequently be regarded as preliminary, pending confirmation from future methodologically rigorous investigations.

A further limitation relates to comparisons with international validation evidence. Although aggregate diagnostic accuracy estimates appear broadly consistent with international benchmarks, these comparisons are constrained by differences in population context and study design. The underrepresentation of Middle Eastern populations in international meta-analyses, combined with methodological heterogeneity within the GCC studies, limits comparability. Apparent concordance with international evidence may therefore obscure important contextual differences and should be interpreted with appropriate caution.

4.5 Implications

4.5.1 Clinical Practice

The findings have several potential implications for clinical practice in GCC perinatal settings. The Arabic version of the EPDS appears to be the most widely supported screening tool currently available in the region, although its performance may vary depending on contextual factors. Optimal thresholds are likely to differ according to gestational stage and population characteristics and should be considered in relation to local prevalence and available clinical resources.

The cultural and contextual factors identified in this review may also influence how screening is implemented. Approaches to administration may benefit from sensitivity to privacy, stigma, and communication style, alongside framing screening as a routine component of maternal care. Clinician awareness of variation in symptom presentation and of broader social and family contexts may also support more effective identification and engagement. It is important that screening is accompanied by appropriate referral pathways and follow up care, as limited support structures may reduce the overall effectiveness of screening initiatives.

The available evidence suggests that repeated assessment across the perinatal period may support improved identification of depressive symptoms as they emerge or change over time. While specific models of implementation vary, screening at multiple points during pregnancy and the postnatal period may offer a more comprehensive approach to detection. Finally, a four-tier referral framework is proposed to provide a proportionate and structured response to identified individuals (Figure 2). Universal screening and psychoeducation at the primary care level would constitute the first tier. Brief psychological intervention and enhanced monitoring for individuals identified in physician led practice as requiring further support would form the second tier. Specialist community perinatal mental health services, including comprehensive psychiatric assessment and evidence based psychological therapies, would form the third tier for moderate to severe or complex presentations. Inpatient and crisis services would constitute the fourth tier, with the development of Mother Baby Units representing the preferred model of inpatient care. This framework reflects a continuum of care that integrates universal approaches with more intensive support where required. The development of appropriately resourced care pathways remains an important consideration for service planning within the region.

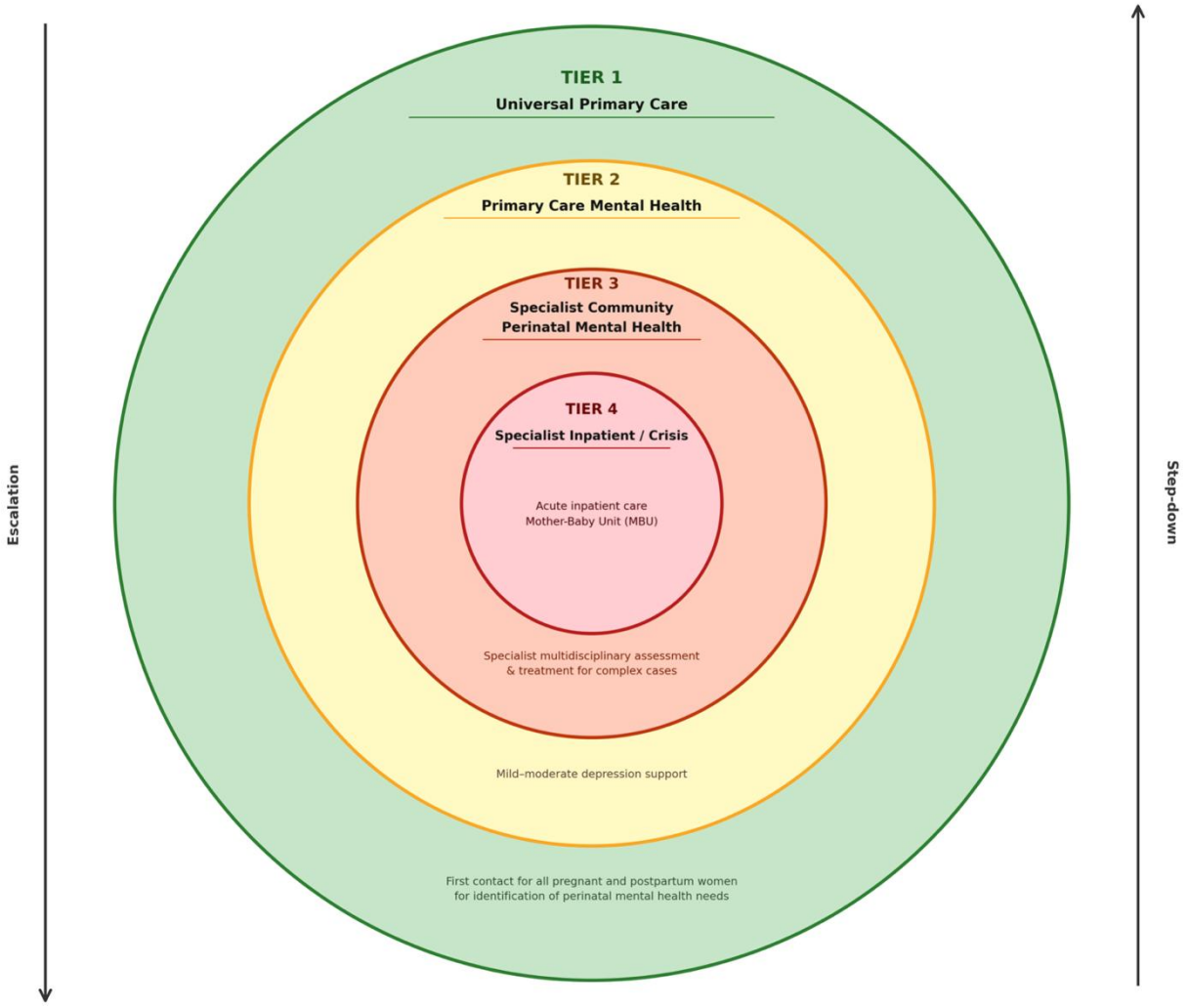


Figure 2. Proposed four-tier perinatal mental health referral pathway adapted for GCC healthcare settings

4.5.2 Future Research Priorities

The findings identify several priority areas for future research. Most urgently, methodologically robust validation studies are needed in Kuwait, Bahrain, and Saudi Arabia, which remain entirely unrepresented in the current evidence base. Future validation studies should employ pre-specified cut-off thresholds, gold-standard structured diagnostic interviews such as the SCID-5 or MINI, adequate sample sizes, and rigorous blinding procedures, in order to generate unbiased and generalisable estimates of diagnostic accuracy. Studies should also include stratified analyses by gestational period, nationality, language, parity, and socioeconomic status, to establish the generalisability of psychometric parameters across the diverse populations accessing GCC antenatal services.

The evidence also supports investment in the development of a GCC-specific screening tool that addresses the cultural validity limitations of the Arabic EPDS. Such a tool should be developed through a systematic multi-phase process beginning with qualitative exploration of how perinatal depression is conceptualised, experienced, and expressed by women of diverse cultural backgrounds in GCC healthcare settings. This qualitative phase would generate emic perspectives on perinatal distress, identifying locally relevant symptoms, idioms of distress, and help-seeking patterns, with explicit attention to the roles of stigma, family dynamics, spiritual beliefs, and somatic symptom expression. Candidate items generated from this phase, drawing on both existing validated instruments and novel culturally grounded items, should undergo rigorous psychometric testing in large, representative, multi-site samples spanning all six GCC member states. Complementary narrative and qualitative methods, including patient journaling and semi-structured interviews conducted by culturally matched interviewers, are necessary for capturing presentations of distress that standardised questionnaires systematically

overlook, and their feasibility and validity in GCC perinatal populations require dedicated investigation.

5. CONCLUSION

This review provides the first comprehensive synthesis of perinatal depression screening tools across the Gulf Cooperation Council. Only three validation studies were identified, all evaluating the Arabic EPDS in UAE, Qatari, and Omani populations. Whilst these studies report broadly acceptable performance, this finding is substantially undermined by consistent moderate to high risk of bias and the complete absence of validation evidence from three of the six GCC member states. The Arabic EPDS thus emerges as the most empirically supported screening option currently available in the region, yet the evidence base remains too limited and methodologically fragile to support definitive clinical guidance or region-wide policy implementation.

Importantly, the contribution of this review extends beyond technical appraisal. By situating diagnostic accuracy evidence within a culturally informed analytical framework that reflects the sociocultural realities of GCC populations, including somatic expressions of distress, collectivist family dynamics, stigma, and linguistic diversity, the review directly addresses the fundamental question of whether existing screening tools are fit for purpose. The absence of such considerations in the primary studies represents a critical limitation in the current evidence base. The findings therefore make clear that psychometric validation alone is insufficient. Effective identification of perinatal depression in the GCC requires the integration of cultural validity, implementation feasibility, and healthcare system capacity within a coherent screening strategy.

The originality of this study lies in both its scope and its implications. It provides the first region wide synthesis of psychometric validation evidence for perinatal depression screening, generating conclusions that carry direct relevance for clinical practice and policy development. At the same time, it exposes the gaps that currently constrain equitable identification of affected women across the region. These findings point to a clear and urgent research and policy agenda. This includes the need for rigorously designed validation studies in under represented GCC member states, systematic investigation of cultural validity and implementation barriers, and the development of screening instruments grounded in local conceptualisations of perinatal distress. Equally, the establishment of tiered referral pathways and specialised services such as Mother Baby Units is essential to ensure that identification is matched by appropriate care.

Without these coordinated advances, screening efforts risk remaining fragmented, culturally misaligned, and clinically insufficient. With them, there is a credible pathway towards a regionally responsive and equitable screening system that can meaningfully improve the detection and treatment of perinatal depression. The implications are substantial, extending beyond individual symptom recognition to the protection of maternal wellbeing, optimisation of infant developmental trajectories, and strengthening of family health across the Gulf Cooperation Council.

6. Conflict of Interest

The author declares no commercial or financial relationships that could be construed as a potential conflict of interest.

7. Data Availability Statement

No primary data were generated or analysed in this study. All data were derived from previously published sources.

8. Ethical Considerations

This review is a secondary synthesis of published data and does not involve the collection of primary data or direct contact with research participants. Accordingly, formal ethical approval was not required. All data used in this review were derived from previously published studies that had independently obtained ethical approval from their respective institutional review boards or ethics committees. The review was conducted in accordance with the principles of research integrity and transparency, as evidenced by the inclusion of a predefined protocol registered with PROSPERO (CRD420261373029), ensuring adherence to PRISMA 2020 reporting standards.

9. Author Contributions

N.A. conceived the study, conducted the literature review, analysed and synthesised the findings, and wrote the manuscript. The work was undertaken as part of the MSc Women & Children's Health programme at King's College London and was completed under the academic guidance of the Department of Women and Children's Health. N.A. is accountable for all aspects of the work and takes responsibility for the integrity and accuracy of the research.

10. Funding: This study received no financial support.

11. Acknowledgments

The author gratefully acknowledges King's College London, the Faculty of Life Sciences and Medicine, the Institute of Psychiatry, Psychology and Neuroscience, the School of Life Course and Population Health Sciences, and the Department of Women and Children's Health for providing the academic environment and support necessary for this work. Special thanks are extended to Dr Sam Burton (Supervisor) and Professor Abigail Easter (Co-Supervisor) for their invaluable guidance, expertise, and encouragement throughout the research process.

The author also wishes to express sincere gratitude to Dr Paul Taylor and Dr Ashley Boyle, Module Leads for the Research Project in Women's and Children's Health, for their academic leadership and support during the completion of the research project. Appreciation is also extended to Professor Rachel Tribe, Programme Lead for the Women & Children's Health MSc, for her leadership and oversight of the programme, and to Dr Kim Jonas for her contributions to the academic learning experience. The author further thanks John Prudhoe (Academic Skills Tutor) and Pablo Paganotto (Senior Library Assistant) for their assistance with academic writing, literature searching, and manuscript preparation.

This research project was awarded the Miriam Menkin Prize for Outstanding Research (MSc Women's & Children's Health), in recognition of its academic excellence and contribution to the field of maternal and child health research.

Artificial intelligence (AI) tools were used solely to assist with vocabulary refinement and grammatical accuracy during manuscript preparation. Reference management was conducted using Zotero, and citations were formatted in accordance with the Vancouver referencing style.

All claims expressed in this manuscript are solely those of the author and do not necessarily represent the views of the affiliated institutions, supervisors, programme leaders, module leads, supporting staff, or any other individuals acknowledged herein.

REFERENCES

1. Putnam KT, Wilcox M, Robertson-Blackmore E, Sharkey K, Bergink V, Munk-Olsen T, *et al.,* Clinical phenotypes of perinatal depression and time of symptom onset: analysis of data from an international consortium. The Lancet Psychiatry. 2017 Jun 1;4(6):477–85. doi:10.1016/S2215-0366(17)30136-0 PubMed PMID: 28476427.
2. Harrison S, Alderdice F. Challenges of defining and measuring perinatal anxiety. Journal of Reproductive and Infant Psychology. 2020 Jan 1;38(1):1–2. doi:10.1080/02646838.2020.1703526 PubMed PMID: 31964192.
3. Howard LM, Molyneaux E, Dennis CL, Rochat T, Stein A, Milgrom J. Non-psychotic mental disorders in the perinatal period. The Lancet. 2014 Nov 15;384(9956):1775–88. doi:10.1016/S0140-6736(14)61276-9 PubMed PMID: 25455248.
4. Jones I, Chandra PS, Dazzan P, Howard LM. Bipolar disorder, affective psychosis, and schizophrenia in pregnancy and the post-partum period. Lancet. 2014 Nov 15;384(9956):1789–99. doi:10.1016/S0140-6736(14)61278-2 PubMed PMID: 25455249.
5. Reinsperger I, Paul JL, Zechmeister-Koss I. International best-practice models for perinatal and infant mental health care – a scoping review. Front Psychiatry. 2025 Jun 17;16. doi:10.3389/fpsyt.2025.1536145
6. Blount AJ, Adams CR, Anderson-Berry AL, Hanson C, Schneider K, Pendyala G. Biopsychosocial Factors during the Perinatal Period: Risks, Preventative Factors, and Implications for Healthcare Professionals. International Journal of Environmental Research and Public Health. 2021 Jan;18(15):8206. doi:10.3390/ijerph18158206
7. Athan AM. A critical need for the concept of matrescence in perinatal psychiatry. Front Psychiatry. 2024 Jun 10;15. doi:10.3389/fpsyt.2024.1364845
8. Meltzer-Brody S, Howard LM, Bergink V, Vigod S, Jones I, Munk-Olsen T, *et al.,* Postpartum psychiatric disorders. Nat Rev Dis Primers. 2018 Apr 26;4(1):18022. doi:10.1038/nrdp.2018.22
9. Al-abri K, Edge D, Armitage CJ. Prevalence and correlates of perinatal depression. Soc Psychiatry Psychiatr Epidemiol. 2023 Nov 1;58(11):1581–90. doi:10.1007/s00127-022-02386-9
10. Pan T, Zeng Y, Chai X, Wen Z, Tan X, Sun M. Global Prevalence of Perinatal Depression and Its Determinants Among Rural Women: A Systematic Review and Meta-Analysis. Depress Anxiety. 2024 Sep 20;2024:1882604. doi:10.1155/2024/1882604 PubMed PMID: 40226692; PubMed Central PMCID: PMC11919136.
11. World Health Organization [Internet]. [cited 2026 May 26]. Perinatal mental health. Available from: https://www.who.int/teams/mental-health-and-substance-use/promotion-prevention/perinatal-mental-health
12. Roddy Mitchell A, Gordon H, Lindquist A, Walker SP, Homer CSE, Middleton A, *et al.,* Prevalence of Perinatal Depression in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. JAMA Psychiatry. 2023 May 1;80(5):425–31. doi:10.1001/jamapsychiatry.2023.0069 PubMed PMID: 36884232; PubMed Central PMCID: PMC9996459.
13. Hutchens BF, Kearney J. Risk Factors for Postpartum Depression: An Umbrella Review. Journal of Midwifery & Women's Health. 2020;65(1):96–108. doi:10.1111/jmwh.13067
14. Xu M, Luo Y, Huang Y, Liu Y, Ding L. Risk factors for postpartum depression: an umbrella review. Front Public Health. 2026 Jan 22;13. doi:10.3389/fpubh.2025.1714668
15. Gelaye B, Rondon M, Araya R, Williams MA. Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. Lancet Psychiatry. 2016 Oct;3(10):973–82. doi:10.1016/S2215-0366(16)30284-X PubMed PMID: 27650773; PubMed Central PMCID: PMC5155709.

16. Sharma V, Mazmanian D, Palagini L, Bramante A. Postpartum psychosis: Revisiting the phenomenology, nosology, and treatment. Journal of Affective Disorders Reports. 2022 Dec 1;10:100378. doi:10.1016/j.jadr.2022.100378
17. Dadi AF, Akalu TY, Wolde HF, Baraki AG. Effect of perinatal depression on birth and infant health outcomes: a systematic review and meta-analysis of observational studies from Africa. Arch Public Health. 2022 Jan 20;80:34. doi:10.1186/s13690-022-00792-8 PubMed PMID: 35057865; PubMed Central PMCID: PMC8772173.
18. Stojanov J, Stanković M, Žikić O, Antonijević J. Postpartum psychiatric disorders: Review of the research history, classification, epidemiological data, etiological factors and clinical presentations. AFMN. 2018 Dec 1;36(3):167–76. doi:10.5937/afmnai1903167s
19. Grote NK, Bridge JA, Gavin AR, Melville JL, Iyengar S, Katon WJ. A meta-analysis of depression during pregnancy and the risk of preterm birth, low birth weight, and intrauterine growth restriction. Arch Gen Psychiatry. 2010 Oct;67(10):1012–24. doi:10.1001/archgenpsychiatry.2010.111 PubMed PMID: 20921117; PubMed Central PMCID: PMC3025772.
20. Gulf Cooperation Council. In: Wikipedia [Internet]. 2026 [cited 2026 Jul 14]. Available from: https://en.wikipedia.org/w/index.php?title=Gulf_Cooperation_Council&oldid=1362715475
21. OKASHA A, KARAM E, OKASHA T. Mental health services in the Arab world. World Psychiatry. 2012 Feb;11(1):52–4. doi:10.1016/j.wpsyc.2012.01.008 PubMed PMID: 22295010; PubMed Central PMCID: PMC3266748.
22. Sayed MA. Mental Health Services in the United Arab Emirates: Challenges and Opportunities. Int J Emerg Ment Health. 2015;17(3). doi:10.4172/1522-4821.1000263
23. Dardas LA, Simmons LA. The stigma of mental illness in Arab families: a concept analysis. J Psychiatr Ment Health Nurs. 2015 Nov;22(9):668–79. doi:10.1111/jpm.12237 PubMed PMID: 26118332.
24. Ghubash R, Eapen V. Postpartum mental illness: perspectives from an Arabian Gulf population. Psychol Rep. 2009 Aug;105(1):127–36. doi:10.2466/PR0.105.1.127-136 PubMed PMID: 19810439.
25. Doraiswamy S, Jithesh A, Chaabane S, Abraham A, Chaabna K, Cheema S. Perinatal Mental Illness in the Middle East and North Africa Region—A Systematic Overview. International Journal of Environmental Research and Public Health. 2020 Jan;17(15):5487. doi:10.3390/ijerph17155487
26. Alhomaid TA, Alburaidi O, Garallah N, Alghofaili S, Alhayyani RM, Alkhalifah B, *et al.,* Prevalence of Depression Among Urban and Rural Residents of Saudi Arabia Compared With Other Gulf Cooperation Council Countries: A Systematic Review. Cureus. 2025 Aug;17(8):e90321. doi:10.7759/cureus.90321 PubMed PMID: 40964597; PubMed Central PMCID: PMC12440573.
27. Al-Hejji Z, Al-Khudhair M, Al-Musaileem M, Al-Eithan M. Prevalence and associated risk factors of antenatal depression among women attending antenatal clinics in primary health care centers in the Ministry of Health in Al-Ahsa City, Saudi Arabia. J Family Med Prim Care. 2019 Dec;8(12):3900–7. doi:10.4103/jfmpc.jfmpc_724_19 PubMed PMID: 31879633; PubMed Central PMCID: PMC6924245.
28. Alshahrani SASAS, Al-Saleh NF, Siddiqui AF, Khalil SN, Alsaleem MA, AlHefdhi HA, *et al.,* Burden and Risk Factors of Postpartum Depression in Southwest Saudi Arabia. J Clin Med. 2023 May 13;12(10):3444. doi:10.3390/jcm12103444 PubMed PMID: 37240551; PubMed Central PMCID: PMC10219287.
29. Alzahrani J, Al-Ghamdi S, Aldossari K, Al-Ajmi M, Al-Ajmi D, Alanazi F, *et al.,* Postpartum Depression Prevalence and Associated Factors: An Observational Study in Saudi Arabia. Medicina (Kaunas). 2022 Nov 4;58(11):1595. doi:10.3390/medicina58111595 PubMed PMID: 36363552; PubMed Central PMCID: PMC9696672.
30. Zarroug M, Altaf MF, Shaikh S, Tidjani A, Bashir O, Ayash MI, *et al.,* The Prevalence and Risk Factors of Postpartum Depression Among Saudi Arabian Women-A Systematic Review and Meta-Analysis. Healthcare (Basel). 2025 Aug 18;13(16):2040. doi:10.3390/healthcare13162040 PubMed PMID: 40868656; PubMed Central PMCID: PMC12385902.
31. Burgut FT, Bener A, Ghuloum S, Sheikh J. A study of postpartum depression and maternal risk factors in Qatar. J Psychosom Obstet Gynaecol. 2013 Jun;34(2):90–7. doi:10.3109/0167482X.2013.786036 PubMed PMID: 23701432.
32. Levis B, Negeri Z, Sun Y, Benedetti A, Thombs BD, He C, *et al.,* Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. BMJ. 2020 Nov 11;371:m4022. doi:10.1136/bmj.m4022 PubMed PMID: 33177069; PubMed Central PMCID: PMC7656313.
33. Al-Abri K, Sumri HA, Al-Azri M, Yesodharan DK, Al Kindi R. Validation of the arabic version of the edinburgh postnatal depression scale for screening antenatal depression in Oman. BMC Pregnancy Childbirth. 2025 Jul 2;25(1):677. doi:10.1186/s12884-025-07793-9 PubMed PMID: 40604594; PubMed Central PMCID: PMC12220313.
34. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, *et al.,* The Mini-International Neuropsychiatric Interview (M.I.N.I.): the

development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry. 1998;59 Suppl 20:22-33;quiz 34-57. PubMed PMID: 9881538.
35. Leeflang MMG, Moons KGM, Reitsma JB, Zwinderman AH. Bias in Sensitivity and Specificity Caused by Data-Driven Selection of Optimal Cutoff Values: Mechanisms, Magnitude, and Solutions. Clin Chem. 2008 Apr 1;54(4):729–37. doi:10.1373/clinchem.2007.096032
36. Leeflang MMG, Bossuyt PMM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. J Clin Epidemiol. 2009 Jan;62(1):5–12. doi:10.1016/j.jclinepi.2008.04.007 PubMed PMID: 18778913.
37. Cox JL, Holden JM, Sagovsky R. Detection of Postnatal Depression: Development of the 10-item Edinburgh Postnatal Depression Scale. The British Journal of Psychiatry. 1987 Jun;150(6):782–6. doi:10.1192/bjp.150.6.782
38. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al.,* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529–36. doi:10.7326/0003-4819-155-8-201110180-00009 PubMed PMID: 22007046.
39. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al.,* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews [Internet]. 2021 Mar 29. doi:10.1136/bmj.n71
40. AI-Powered Systematic Review & Literature Review Platform [Internet]. [cited 2026 Jun 10]. Available from: https://www.rayyan.com/
41. Ghubash R, Abou-Saleh MT, Daradkeh TK. The validity of the Arabic Edinburgh Postnatal Depression Scale. Soc Psychiatry Psychiatr Epidemiol. 1997 Nov;32(8):474–6. doi:10.1007/BF00789142 PubMed PMID: 9409163.
42. Naja S, Al-Kubaisi N, Chehab M, Al-Dahshan A, Abuhashem N, Bougmiza I. Psychometric properties of the Arabic version of EPDS and BDI-II as a screening tool for antenatal depression: evidence from Qatar. BMJ Open. 2019 Sep 13;9(9):e030365. doi:10.1136/bmjopen-2019-030365 PubMed PMID: 31519679; PubMed Central PMCID: PMC6747648.
43. Park SH, Kim JI. Predictive validity of the Edinburgh postnatal depression scale and other tools for screening depression in pregnant and postpartum women: a systematic review and meta-analysis. Arch Gynecol Obstet. 2023 May 1;307(5):1331–45. doi:10.1007/s00404-022-06525-0
44. Aydin N, Inandi T, Yigit A, Hodoglugil NNS. Validation of the Turkish version of the Edinburgh Postnatal Depression Scale among women within their first postpartum year. Soc Psychiatry Psychiatr Epidemiol. 2004 Jun;39(6):483–6. doi:10.1007/s00127-004-0770-4 PubMed PMID: 15205733.
45. Mazhari S, Nakhaee N. Validation of the Edinburgh Postnatal Depression Scale in an Iranian sample. Arch Womens Ment Health. 2007;10(6):293–7. doi:10.1007/s00737-007-0204-x PubMed PMID: 18058062.
46. Fellmeth G, Harrison S, Opondo C, Nair M, Kurinczuk JJ, Alderdice F. Validated screening tools to identify common mental disorders in perinatal and postpartum women in India: a systematic review and meta-analysis. BMC Psychiatry. 2021 Apr 20;21(1):200. doi:10.1186/s12888-021-03190-6
47. Keyes CLM, Ryff CD. Somatization and mental health: A comparative study of the idiom of distress hypothesis. Social Science & Medicine. 2003 Nov 1;57(10):1833–45. doi:10.1016/S0277-9536(03)00017-0

Appendix A
QUADAS-2 Empty Worksheet
Quality Assessment of Diagnostic Accuracy Studies – 2

DOMAIN 1: Patient Selection
Signalling Questions

#	Signalling Question	Response (Yes / No / Unclear)	Justification / Evidence from Study
1.1	Was a consecutive or random sample of patients enrolled?		
1.2	Was a case-control design avoided?		
1.3	Did the study avoid inappropriate exclusions?		

Risk of Bias Judgement

Judgement	Select (✓)
Low — All signalling questions answered 'Yes'	
High — Any signalling question answered 'No'	
Unclear — Insufficient information to judge	

DOMAIN 2: Index Test
Domain 2: Index Test

#	Signalling Question	Response (Yes / No / Unclear)	Justification / Evidence from Study
2.1	Were the index test results interpreted without knowledge of the results of the reference standard?		
2.2	If a threshold was used, was it pre-specified?		

Risk of Bias Judgement

Judgement	Select (✓)
Low — All signalling questions answered 'Yes'	
High — Any signalling question answered 'No'	
Unclear — Insufficient information to judge	

DOMAIN 3: Reference Standard
Signalling Questions

#	Signalling Question	Response (Yes / No / Unclear)	Justification / Evidence from Study
3.1	Is the reference standard likely to correctly classify the target condition?		
3.2	Were the reference standard results interpreted without knowledge of the results of the index test?		

Risk of Bias Judgement

Judgement	Select (✓)
Low — All signalling questions answered 'Yes'	
High — Any signalling question answered 'No'	
Unclear — Insufficient information to judge	

DOMAIN 4: Flow and Timing
Signalling Questions

#	Signalling Question	Response (Yes / No / Unclear)	Justification / Evidence from Study
4.1	Was there an appropriate interval between the index test(s) and the reference standard?		
4.2	Did all patients receive a reference standard?		
4.3	Did all patients receive the same reference standard?		
4.4	Were all patients included in the analysis?		

Risk of Bias Judgement

Judgement	Select (✓)
Low — All signalling questions answered 'Yes'	
High — Any signalling question answered 'No'	
Unclear — Insufficient information to judge	

Overall Quality Rating

Rating	Select (✓)
High Quality — Low risk of bias across all domains; low applicability concerns	
Moderate Quality — Unclear risk in one or more domains; low applicability concerns	
Low Quality — High risk of bias in one or more domains	

Reference: Whiting PF, Rutjes AWS, Westwood ME, *et al.,* QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine* 2011;155(8):529–536. doi:10.7326/0003-4819-155-8-201110180-00009

Appendix B. Database Search Strategies
PubMed
Combined Search Strategy
(Pregnancy[MeSH] OR Perinatal Care[MeSH] OR Postpartum Period[MeSH] OR pregnancy OR prenatal OR antenatal OR perinatal OR postpartum) AND (Depression[MeSH] OR Depressive Disorder[MeSH]

OR Mood Disorders[MeSH] OR Anxiety Disorders[MeSH] OR Stress Disorders, Post-Traumatic[MeSH] OR depression OR anxiety OR "perinatal mental health" OR PTSD) AND (Mass Screening[MeSH] OR Diagnosis[MeSH] OR Psychometrics[MeSH] OR screening OR assessment OR validation OR EPDS OR PHQ-9 OR BDI) AND (Middle East[MeSH] OR Arab Countries[MeSH] OR "Gulf Cooperation Council" OR GCC OR Saudi Arabia OR "United Arab Emirates" OR UAE OR Qatar OR Oman OR Kuwait OR Bahrain)

Ovid MEDLINE

Search Line	Search Query
1	prenatal care/
2	Pregnancy/
3	Maternal Health/
4	Pregnancy-related keywords
5	1 OR 2 OR 3 OR 4
6	mental disorders/ OR exp mood disorders/
7	Depression/
8	6 OR 7
9	5 AND 8
10	Depression, Postpartum/
11	Perinatal mental health keywords
12	9 OR 10 OR 11
13	Screening and validation terms
14	GCC/Middle East terms
15	12 AND 13 AND 14

Embase Classic + Embase

Search Line	Search Query
1	exp prenatal care/
2	exp pregnancy/
3	maternal health care/
4	Pregnancy-related keywords
5	1 OR 2 OR 3 OR 4
6	exp mental disorder/ OR exp mood disorder/
7	exp depression/
8	6 OR 7
9	5 AND 8
10	Perinatal mental health keywords
11	9 OR 10
12	Screening and validation terms
13	GCC/Middle East terms
14	11 AND 12 AND 13

Global Health and APA PsycInfo

Search Line	Search Query
1	exp Pregnancy/
2	exp Prenatal Care/
3	maternal health care/
4	Pregnancy-related keywords
5	1 OR 2 OR 3 OR 4
6	exp Mood Disorders/ OR exp Mental Disorders/
7	exp Depression/
8	6 OR 7
9	5 AND 8
10	exp Postpartum Depression/
11	Perinatal mental health keywords
12	9 OR 10 OR 11
13	Screening and validation terms
14	GCC/Middle East terms
15	12 AND 13 AND 14