Research article

A systematic review and meta-analysis of studies validating Edinburgh Postnatal Depression Scale in fathers

Ainul Khamisah Shafian [a], Salina Mohamed [a,b,*], Nor Jannah Nasution Raduan [a,b], Anne Yee Hway Ann [c]

[a] Department of Psychiatry, Faculty of Medicine, Universiti Teknologi, MARA, Malaysia
[b] Department of Psychiatry, Hospital Universiti Teknologi, MARA, Malaysia
[c] Department of Psychological Medicine, Universiti Malaya, Malaysia

ARTICLE INFO

ABSTRACT

The Edinburgh Postnatal Depression Scale (EPDS) is a reliable measure for detecting paternal postpartum depression. The study's purpose is to determine the appropriate cut-off scores of EPDS for fathers. Our research was conducted using PubMed, Embase, Web of Science, and Scopus. The time frame of the search was from the issuance of EPDS in 1987 until January 2021. The analysis comprised of studies that compared EPDS scores for depression from validated diagnostic interviews. For EPDS cut-off values of 7–13, a bivariate random-effects meta-analysis was used to estimate pooled sensitivity and specificity, as well as the diagnostic odds ratio. Seven studies with a total of 2393 participants were identified. The pooled sensitivity and specificity were satisfactory at cut-off values of 7–10, with significant diagnostic odds ratio. The EPDS accuracy was unaffected by the prevalence of depression, the fathers' mean age, or the translated language. The Edinburgh Postnatal Depression Scale has acceptable properties for detecting paternal postpartum depression, with cut-off scores ranging from 7 to 10.

1. Introduction

Postpartum depression is commonly perceived to be a problem limited to women with newborn babies. Research pertaining to father's mood symptoms had been conducted since 30 years ago (Quadagno et al., 1986) with the research often done in the context as the partners, in the researches designed for postpartum mothers. However, there is growing interest regarding the concept of paternal post-partum depression in this past decade even though there is no exclusive definition for paternal post-partum depression in DSM 5. Fathers also experience significant life changes after childbirth, many of which like the experiences of mothers. Perinatal depression in fathers mainly occurs later than in mothers, often developed shortly after maternal depression and even after the first year of a child's life (Goodman, 2004).

The prevalence of paternal postpartum depression was around 4%–25% globally (Kim and Swain, 2007). Detection of depression in men is relatively challenging because the signs and symptoms observed in paternal postnatal depression are somewhat different from maternal

postnatal depression. They may have milder depressive symptoms with comorbidities of anxieties and somatic symptoms. Apart from that, fathers often show externalizing symptoms, defined as depression equivalents like hostility, anger, and substance use (drugs and alcohol) (Martin et al., 2013).

Edinburgh Postnatal Depression Scale (EPDS) is the commonest instrument used for screening postnatal depression in women. The EPDS assesses symptoms of anhedonia and reactivity, anxiety, panic, coping, tearfulness, and thoughts of self-harm. The scale has been designed to avoid items that reflect difficulties that have been a regular change during the early postnatal period, for example, weight changes, sleep disturbances, and fatigue (Cox et al., 1987). It has been translated and validated in many languages other than English; however, the validation was mainly done in a population of mothers (Gibson, McKenzie-McHarg, Shakespeare, Price and Gray, 2009). Since the past decade, researchers had used EPDS to screen for paternal postnatal depression. EPDS is a 10-item self-report questionnaire about feelings experienced over the past seven days. Each item has four short statements and scored from 0 to

---

3, yielding a total range of 0–30. The cut-off scores for detecting paternal postnatal depression were varied and often lower than the cut-off scores used for mothers. Men may be considered less expressive about their feelings, thus score lower in the screening tools like EPDS, even though they experience the same level of depression (Carlberg et al., 2018).

EPDS has been specifically designed for women in peripartum period, however Matthey et al., (2001) suggested the usage of EPDS in men (Matthey et al., 2001). A recent opinion by Baldoni & Giannotti described a complex clinical presentation of paternal perinatal depression. Apart from that several screening tools had been used for screening at risk fathers, such as Edinburgh Postnatal Depression Scale, Gotland Male Depression Scale, Masculine Depression Scale and recently developed scale, Perinatal Assessment of Paternal Affectivity (PAPA) (Baldoni and Giannotti, 2020). Among the studies that used multiple screening tools, it was concluded that EPDS performed similarly to or better than the other measures are assessed. A Danish study had compared between Gotland Male Depression Scale (GMDS) and Edinburgh Postnatal Depression Scale which found that EPDS detected 5% of depression in contrast with 3.4% in GMDS (Madsen and Juhl, 2007). A review by (Darwin et al., 2021) found a lack of agreement regarding the cut-off points use in fathers. The cut-offs points ranging from ≥11 to ≥5 for depression.

While the original validation study results using this scale are promising, it is crucial to determine whether EPDS compares favorably with gold standard reference in the screening for paternal postpartum depression. This review aimed at identifying the suitable cut-off scores for the use of EPDS in the screening for depression in fathers by collating data available from EPDS validation studies that had been done before.

2. Methods

A systematic review of the literature search was conducted. The study protocol was published in the PROSPERO international register for systematic reviews (registration number CRD42021235006. We followed PRISMA guidance in reporting the review.

2.1. Data source and search strategy

The literature search was done in the following databases, PubMed, Embase, Web of Science, and Scopus. The time frame of the search was from 1987, since the year EPDS was issued until January 2021. Search terms were as follows: Edinburgh Postnatal Depression Scale/EPDS, father/paternal/dad or male partner, depress* or psychol* and validation/validity (Table 1). The search was intentionally broad to enable the identification of relevant literature across all the review areas. However, based on previous research, terms related to diagnostic accuracy (e.g., sensitivity, specificity, receiver operator characteristic curve) leads to the omission of relevant reference (P. P. Whiting et al., 2011). Thus, these terms were excluded.

Additionally, we searched the reference list of included studies and relevant reviews. Records were imported into referencing software (Endnote version X9), and duplicates were removed. The study was conducted from early January 2021 until May 2021 and a repeated search was done to avoid missing any additional new article by the end of the study.

Table 1. Complete search strategy from PUBMED.

#	Searches	Results
1.	Psychol* OR depress*	2,277,323
2.	Father* OR male partner* OR dad* OR paternal*	174536
3.	EPDS* OR Edinburgh Postnatal Depression Scale*	3448
4.	Validation* OR validity*	447722
5.	#1 AND #2	58531
6.	#1 AND #2 AND #3	231
7.	#1 AND #2 AND #3 AND #4	12

2.2. Study selection

Studies were selected following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guideline. We selected studies that reported the validity of EPDS for screening of postpartum depression in fathers. In addition, the studies had to provide sufficient data to allow us to calculate contingency tables.

2.2.1. Population

The population of interest was fathers in the antepartum or postpartum period up to 1 year of postpartum.

2.2.2. Intervention

Administration of the Edinburgh Postnatal Depression Scale to screen for depression was performed. Any mode of administration was included, e.g., face-to-face, telephone interview, or postage.

2.2.3. Reference test

The validity and accuracy of the EPDS were compared with a gold standard, defined as a structured or semi-structured clinical interview to diagnose depression. Structured interviews conform to a standardized set of questions with uniform sequence of questioning and systematic rating of the relevant response whilst semi-structured interview has predetermined questions for symptoms specified, however the interviewer has the flexibility to modify or augment the standard inquiries with individualized probes to get more accurate rating for specific symptoms. Examples of structured clinical interview include Composite International Diagnostic Interview (CIDI), Clinical Interview Schedule-Revised (CIS-R) and Diagnostic Interview Schedule (DIS) and semi-structured interview includes Structured Clinical Interview for DSM (SCID) and Schedule for Affective Disorder and Schizophrenia (SADS) (Mueller and Segal, 2015).

2.2.4. Exclusion criteria

Studies were excluded when EPDS-3A (that scores for anxiety), EPDS-partner, or EPDS-13 item scale is used. In addition, studies were excluded if the population was not in the antepartum or postpartum period. Literature search, commentaries, or qualitative studies is also excluded.

2.3. Data extraction

Studies were selected following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guideline. The extraction of the data and the citations were assessed for inclusion based on the title and abstract by the first author (AK) and second author (SM). Suppose sufficient information was not available from the abstract to determine the inclusion or exclusion, the citation was included in the next step of the search, in which a full-text copy of the citation was obtained.

Full-text copies were examined against inclusion and exclusion criteria by the first author (AK) and second author (SM), and those meeting criteria were included in the review. Any discrepancies were resolved by the third author (NJ).

2.4. Data quality

The quality of the selected studies was assessed using a grading system based on Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) that assesses the study's risk of bias and applicability based on four domains: Patient selection, Index test, Reference Standard, and Flow and Timing (Table 2). All questions were retained in each domain. Additional questions incorporated for the translated version of EPDS and translated versions of the reference interview to ensure the correct translation procedure being used. The assessment of data quality was performed by the first author (AK).

Table 2. Risk of bias assessment using QUADAS 2.

Study	Patient selection:consecutive or random sampling	Patient selection: case control design avoided	Patient selection: inappropriate exclusions avoided	Patient selection: Overall risk of bias	Index test:EPDS interpreted blind to reference test	Index test: threshold prespecified or reported multiple cutoffs	Index test: appropriate translation	Index test: if translated, psychometric properties reported	Index test overall risk of bias
Areias, Portugal	Yes	Yes	Yes	Low	No	No	Yes	No	High
Matthey, Australia	Yes	Yes	?	Low	Yes	Yes	NA	NA	Low
Edmondson, UK	Yes	Yes	Yes	Low	Yes	Yes	NA	NA	Low
Lai. BP,Hong Kong	Yes	Yes	Yes	Low	Yes	Yes	Yes	Yes	Low
Massoudi, Sweden	Yes	Yes	Yes	Low	Yes	Yes	Yes	Yes	Low
Tran, Vietnam	Yes	Yes	Yes	Low	Yes	Yes	Yes	Yes	Low
Shaheen,Saudi Arabia	Yes	Yes	Yes	Low	Yes	Yes	Yes	No	Low

Study	Reference test: Correctly classifies target condition	Reference test: interpreted blind to EPDS	Reference test: appropriate translation	Reference test: if translated,psychometric properties reported	Reference test: overall risk of bias	Flow/timing: interval of two weeks or less	Flow/timing: all participants received the same reference test	Flow/timing:all participants included in the analysis	Flow/timing: Overall risk of bias
Areias, Portugal	Yes	No	NA	NA	High	Yes	Yes	Yes	Low
Matthey, Australia	Yes	No	NA	NA	Low	Yes	Yes	Yes	Low
Edmondson, UK	Yes	No	NA	NA	High	No	No	No	High
Lai. BP,Hong Kong	Yes	Yes	Yes	?	Low	Yes	No	Yes	Low
Massoudi, Sweden	Yes	Yes	?	?	Low	Yes	No	Yes	Low
Tran, Vietnam	Yes	Yes	Yes	NA	Low	Yes	Yes	Yes	Low
Shaheen,Saudi Arabia	Yes	Yes	?	?	Low	Yes	No	Yes	Low

2.5. Data analysis and synthesis

For each cut-off score, 2 × 2 tables are constructed that computed the sensitivity, specificity, positive and negative predictive values. A bivariate meta-analysis is performed to obtain pooled estimates of specificity and sensitivity and their associated 95% confidence intervals (CIs).

Summary receiver operating characteristic curves is constructed, with each data in this space represents a separate study. The heterogeneity between studies is assessed using the I^2 statistics for the pooled diagnostic odds ratio (OR), which describes the percentage of total variation across studies caused by heterogeneity rather than chance. We considered the I^2 value of 25% to below, 50% to be moderate, and 75% to be
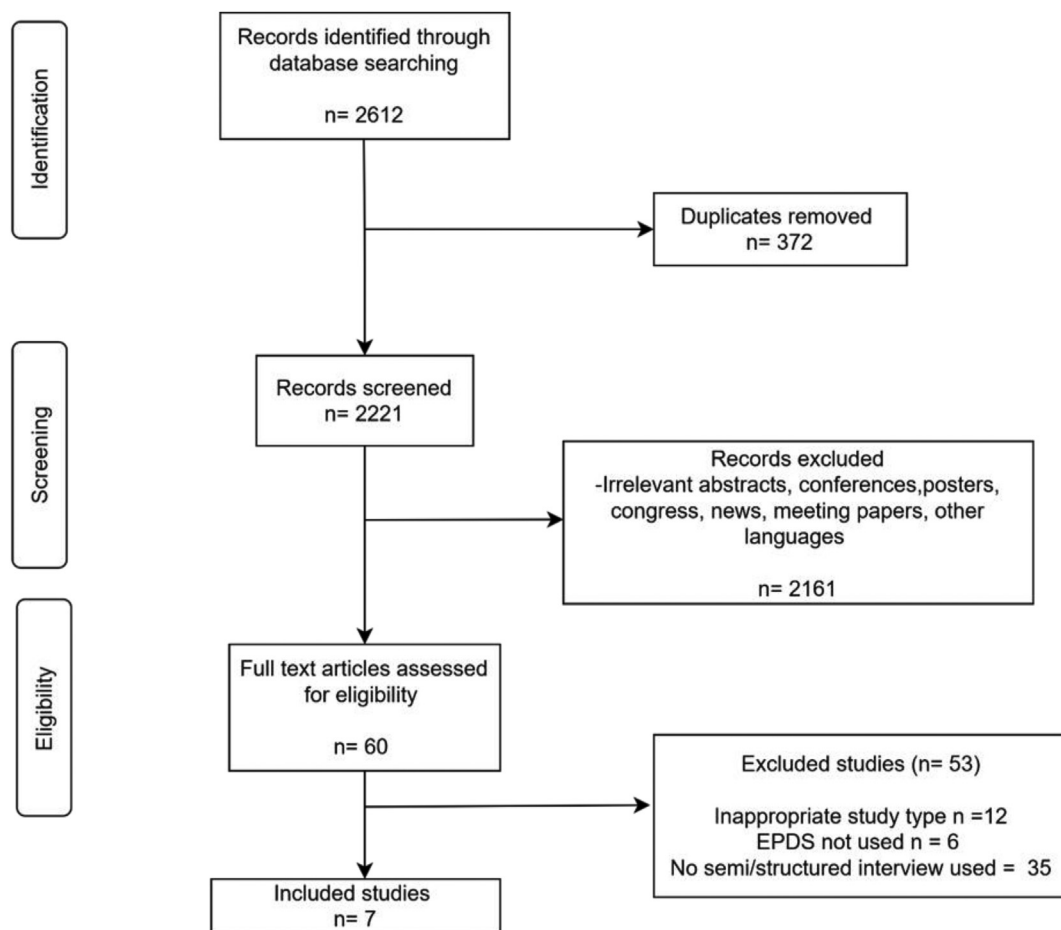

Figure 1. Flow of the study through the review process.

high. The cause of heterogeneity is explored if there is significant between-study heterogeneity. We identified outside the 95% confidence curve by visually inspecting the summary receiver operating characteristic curve plots.

A meta-regression analysis of the logit diagnostic OR is performed to identify sources of heterogeneity. We investigated heterogeneity resulting from the characteristic of the sample or study design by exploring the effects of potential predictive variables. Publication and small study bias are examined using a funnel plot. Statistical analysis of data was conducted using MetaDiSc version 1.4 (Zamora et al., 2006), and the publication bias is analyzed using Comprehensive Meta-Analysis (CMA) software.

3. Results

A total of 2221 citations were screened for eligibility after duplicates had been removed. The full text was obtained for 60 citations, of which 53 were excluded. The exclusion was as follows; six studies did not use Edinburgh Postnatal Depression Scale as the screening tool, 35 studies did not use semi/structured interviews as a diagnostic reference standard, 12 citations were of inappropriate study types. Among the twelve citations with inappropriate study type, two citations include the general population; seven citations were literature search, commentaries, and qualitative study. In addition, one study assessed the depressive symptoms at three days postpartum. The flow of the review process is reported in Figure 1.

3.1. Study characteristics

A total of 7 studies were eligible for inclusion in the review. The characteristics of these studies are reported in Table 3. The size of the study samples varied between 42 to 882, with the total number of participants are 2393. All the studies were drawn from community samples (recruitment is done in maternal-child services like the antenatal and well-baby clinic and health visiting services). Two out of the eight studies use the EPDS in the antepartum population (Tran et al., 2012) (Areias et al., 1996). The mean age of the study participants varied from 26.2 to 35 years old. Five of the studies used languages other than English, including Vietnamese, Chinese, Arabic, and Portuguese. 6 out of 7 studies used the self-reporting method. In contrast, one study by (Tran et al., 2012) used structured interviews because the self-report questionnaire is not a norm and unfamiliar for the Population in Vietnam. According to the semi-structured/structured interview, the percentage of participants that met the diagnostic criteria of depression according to the semi-structured/structured interview ranged from 2.9% (Matthey et al., 2001) to 23.8% (Areias et al., 1996). The timing of administration of questionnaires also varies, ranging from 24 weeks of gestation up to 12 months of the postpartum period.

Data were available from six out of seven included studies to explore the accuracy of the EPDS questionnaire. Study done in Vietnam by (Tran et al., 2012) provided raw data up until the cut off scores of 5/6, hence limiting the analysis of further cut off scores. Even though study by (Areias et al., 1996) did not mention the appropriate cut-off scores for EPDS used in fathers, however there were adequate data that can be integrated in the review.

3.2. Diagnostic instruments

Three studies using Structured Clinical Interview for DSM-IV (SCID) (Edmondson et al., 2010; Lai et al., 2010; Tran et al., 2012), one study used Primary Care Evaluation of Mental Disorders (PRIME-MD) (Massoudi et al., 2013), Schedule for Affective Disorders (SADS) (Areias et al., 1996) and Diagnostic Interview Schedule (DIS) (Matthey et al., 2001)

and a study underwent interview by a psychologist using Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM 5) (Shaheen et al., 2019).

3.3. Diagnostic criteria

All the studies included measures of depression compared to the reference test; three of the studies also assess generalized anxiety disorder, panic disorder, or its comorbidities (Massoudi et al., 2013; Matthey et al., 2001; Tran et al., 2012). One of the studies used "distress caseness" in detecting depressive and anxiety disorders, as the criteria for the diagnosis were fulfilled but not the duration of the illness (Matthey et al., 2001). There were other screening tools used in four of the studies. A study on Vietnamese fathers was also validating the 12-item General Health Questionnaire (GHQ-12) and Zung's Self-rated Anxiety Scale (Zung SAS) (Tran et al., 2012). A study based on the Swedish father used the Hospital Anxiety and Depression Scale (HAD-A) anxiety component to assess the anxiety subscale (Massoudi et al., 2013).

3.4. Quality assessment

The methodological quality of the included studies was assessed using the QUADAS-2 tool. (P. F. P. F. Whiting et al., 2011). The quality assessment results are presented in Table 4, Figure 2. The studies included in the review were of mixed methodological quality. Four out of the seven studies were judged to have a low risk of bias in all the domains. All the studies used consecutive recruitment for their patient selection. Two of the studies were judged to have a high risk of bias in the reference test domain, as the assessor was not blinded, and there was no mention of the translation and the psychometric assessment of the translation (Areias et al., 1996; Edmondson et al., 2010). In the flow/time domain, one study was judged to have a high risk of bias as there was an interval of 4.8 weeks between the administration of the index and the reference test. All the studies performed well in the applicability domain; all were judged to have a low risk of bias.

3.5. Sensitivity and specificity

Sensitivity and specificity varied largely between studies for each of the EPDS cut-off scores, the sensitivity ranged from 40% (Areias et al., 1996) to 100% (Edmondson et al., 2010) and specificity of 58.1%–93% to detect paternal postnatal depression.

The pooled sensitivity value decrease with increasing cut-off points, from 0.87(95% CI: 0.78–0.93) at cut-off points of 7, and the lowest pooled sensitivity at cut-off points of 13, 0.37 (95% CI: 0.28–0.47). The pooled specificity increases as the cut-off scores increase, ranging from 0.713 (95% CI: 0.68–0.74) to 0.96 (95% CI: 0.95–0.97), as shown in Table 5. Four studies available and included in the calculation of pooled sensitivity at the cut-off scores of >7, and 5 studies are eligible at the cut-off scores of >13.

3.6. Likelihood ratio

Likelihood ratios describe how many times more likely a person with the disease is to receive a particular test result than a person without the disease, with the advantages over sensitivity and specificity because they are less likely to change with the prevalence of the disorder. As the cut-off points of EPDS increased, the positive likelihood ratio also increased, ranging from 3.31 at the cut-off scores of 7–13.16 at the cut-off scores of 13. It has been suggested that a positive likelihood ratio of more than 10 provides convincing diagnostic evidence, while the value of more than 5 provides strong evidence in clinical practice.

Table 3. Characteristics of included studies.

Authors	Country	Age (mean and range)	Sample	Time of administration	Setting	%meeting criteria for depression	Other screening tools used	Diagnostic criteria	EPDS cut off points reported	Method of administration	Language	Reference Test
Areias,1996	Portugal	26.2 (20–37)	42	Six months antenatal, 3mo and 12mo postnatal	Maternity clinic and home	23.80%		Depression (major and minor)	Not mention	Self-report	Portuguese	SADS
Matthey, 2001	Australia	29.1 (range not reported)	208	6–7 weeks postpartum	Parent-hood class	2.90%	CES-D	Distress caseness for depression and anxiety/depression only/anxiety	>9	Self-report	English	DIS
Edmondson, 2010	United Kingdom	35 (range not reported)	189	14 weeks postpartum	Maternity hospital	10%		>10 MDD, >8 casenesses of depression and GAD	>10	Postal-self report	English	SCID (DSM IV)
Lai,2010	Hong Kong	33.4 (18–59)	551	8 weeks postpartum	Postnatal ward	7.23%	BDI, PHQ-9	Depression	>10	Self-report	Chinese	SCID-NP
Tran,2012	Vietnam	31 (20–49)	231	28 weeks ANC to 4–6 weeks postpartum	Commune health centers	7.40%	Zung SAS, GHQ-12		>4	Individual structured interview	Vietnamese	SCID
Massoudi, 2013	Sweden	33 (20–51)	882	12 weeks postpartum	Child health services	6.10%	HAD-A	Depression, depression and anxiety, anxiety only	>9	Self-report	Swedish	PRIME MD
Shaheen, 2019	Saudi Arabia	34.97 (range not reported)	290	up to 6 months postpartum	Postnatal ward and birth registration center	16.60%		Depression (major and minor)	>8	Self-report	Arabic	DSM5

Abbreviations: SADS (Schedule for Affective Disorders and Schizophrenia, CES-D (Centre for Epidemiologic Studies Depression Scale), DIS (Diagnostic Interview Schedule), SCID (Structured Clinical Interview for DSM Disorders), PRIME MD (Primary Care Evaluation of Mental Disorders), DSM (Diagnostic and Statistic Manual of Mental Disorders), BDI (Beck Depression Inventory), PHQ-9(Patient Health Questionnaire-9), HAD-A (Hospital and Anxiety Depression Scale).

Table 4. Methodological quality of included studies.

The average negative likelihood ratio ranged from 0.143 to 0.62, from the cut-off scores of 7–13. Thus, a negative likelihood ratio of less than 0.1 provide convincing evidence, and below 0.2 may provide strong diagnostic evidence.

3.7. Diagnostic odds ratio

The diagnostic odds ratios were significant, ranging from 15.83 to 24.71 throughout all cut-off values. The diagnostic odds ratio (DOR) is a measure of test performance that combines the strength of sensitivity and specificity, but it is independent of the prevalence rate. The DOR is found to be statistically significant ($p < 0.05$) between the cut-off scores of 7–10.

In view of high heterogeneity, fixed effect meta-regression was performed to estimate the relationship between the diagnostic odds ratio and the covariates. Accuracy of EPDS did not vary according to the prevalence of depression ($p = 0.47$), the mean age of the fathers ($p = 0.88$), translation of EPDS ($p = 0.92$), and country of the original research (English speaking countries vs others) $p = 0.69$.

3.8. Publication bias

Typical of small studies, it is crucial to detect publication bias. This is caused by the absence of information due to the non-publication of missing studies or reporting only a select number of the publications on the Edinburgh Postnatal Depression Scale in fathers. The funnel plot in Figure 3 shows no evidence of publication bias, given that Egger's test for a regression intercept gave a p-value of 0.670, indicating no evidence of publication bias.

4. Discussion

This systematic review intended to address the growing interest in screening for a father's mental health during the perinatal period. This study aims to determine the accuracy and validity of the Edinburgh Postnatal Depression Scale in paternal postpartum depression when compared to structured/semi-structured interviews. This review identifies seven articles that met the inclusion criteria. The samples include fathers in the perinatal period up to 1 year of postpartum using EPDS and semi-structured/structured interviews to diagnose paternal depression. There were significant differences in terms of the language used for EPDS, as 5 out of 7 studies were using a translated version of various languages. The reference test, the timing of testing, and the mean age of the population sampled were also different from study to study. The differences highlighted above contributed to moderate to high heterogeneity in the study (Patrick Bossuyt, 2013).

Heterogeneity in test accuracy review is common since measures of test accuracy are not a fixed property of a test, and the study itself varies with the population, settings, characteristics, and conduct of the test that include skills and experience of assessors and practitioners (Takwoingi et al., 2015). Attempts were made to reduce the heterogeneity by excluding outlying studies, but most values were still within the moderate to high range.

4.1. Accuracy of the EPDS as a screening tool for paternal postpartum depression

Our results support the findings that EPDS has acceptable diagnostic properties for detecting paternal postpartum depression despite clinical heterogeneity in the included studies. There were no significant differences in pooled sensitivity and specificity for cut-off scores between 7 to 10. The pooled sensitivity ranging from 0.71(95% CI: 0.62–0.79) to 0.87 (95%CI 0.78–0.93), pooled specificity ranged from 0.71(95% CI: 0.68–0.74) – 0.86 (95% CI: 0.83–0.87). The diagnostic odds ratio ranged from 16.59-24.71 ($p < 0.05$) at the cut-off score of 7–10.

If the EPDS is to be used, the clinicians might consider this cut-off score, which appears to have increased sensitivity without compromising specificity value. For example, if the prevalence of paternal postpartum depression is 10%, for every 1000 patients seen, the EPDS would identify 80 out of 100 patients with paternal postpartum depression if the sensitivity is 0.80. A total of 198 out of 1000 patients will have false-positive results (or incorrectly identified as having paternal postpartum depression), given the specificity is 0.78.

4.2. Strengths

This is the first review determining the accuracy and attempts to determine the cut-offs point in screening for paternal postpartum depression. The review used a comprehensive search strategy that included a study of EPDS published in 1987 and January 2021 and includes the studies that compare with the gold standard reference. The overall methodological quality of the included studies has low risk of bias.
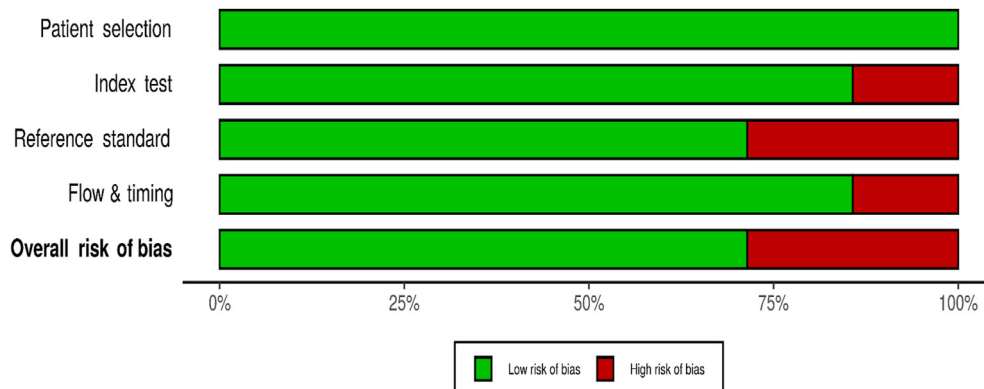

Figure 2. Methodological quality of included studies.

Table 5. Pooled estimates of sensitivity, specificity, positive and negative likelihood ratio, diagnostic odds ratio, and heterogeneity by cut off-scores.

Cut off score	Number of studies	Sensitivity(95%CI)	Specificity (95% CI)	Positive likelihood ratio(95%CI)	Negative Likelihood Ratio (95%CI)	Diagnostic Odds Ratio (95%ci)	P-value	Heterogeneity I^2
7	4	0.87 (0.78–0.93)	0.713 (0.68–0.74)	3.31 (2.01–5.45)	0.143 (0.036–0.571)	24.71 (4.42–138.11)	0.036	64.9 %
8	6	0.80 (0.72–0.87)	0.78 (0.75–0.80)	4.489 (2.69–7.49)	0.271 (0.129–0.570)	19.612 (9.19–41.85)	0.192	32.6 %
9	6	0.79 (0.71–0.86)	0.80 (0.78–0.83)	4.67 (2.85–7.67)	0.265 (0.122–0.572)	19.491 (7.52–50.53)	0.025	60.9 %
10	6	0.71 (0.62–0.79)	0.86 (0.83–0.87)	5.31 (3.12–9.04)	0.370 (0.211–0.652)	16.59 (6.099–45.14)	0.0041	70.9 %
11	6	0.57 (0.47–0.66)	0.92 (0.90–0.93)	7.24 (3.65–14.34)	0.514 (0.337–0.783)	15.83 (6.11–41.00)	0.0052	70.0 %
12	6	0.43 (0.34–0.53)	0.95 (0.94–0.97)	11.32 (6.67–19.20)	0.61 (0.39–0.94)	20.95 (8.44–51.99)	0.0715	50.7 %
13	5	0.37 (0.28–0.47)	0.96 (0.95–0.97)	13.16 (7.58–22.84)	0.62 (0.42–0.92)	22.41 (8.794–57.097)	0.0001	55.9 %

4.3. Limitations of the review

This review has several limitations. First, the study selection was performed by one author, which might introduce bias in selecting the studies or errors in extraction of data. Second, we were unable to explain a large amount of heterogeneity between studies fully. Hence caution should be used in interpreting the results. Apart from that, pooled sensitivity and specificity in this review are possibly inflated due to varying cut-off points reported in original studies. There are possibilities that cut-off score not reported had low sensitivity and specificity, which may cause potential bias.

4.4. Implications for further research

Men have the tendency to delay seeking treatment when they become ill. This has been largely attributed to the concept of "traditional masculine behaviour" which is common among white middle class men. The delay in help seeking behaviour contributed to varying prevalence of paternal postpartum depression (Galdas et al., 2005). Two studies from Asian region, reported large difference in cut off scores, in which study done by (Tran et al., 2012) suggest cut off score of >4 or more while Lai et al. (2010) which conducted the study in Hong Kong suggest a cut off score of >9. These large differences can be explained by the differences in sociocultural factors. Culturally, it was described in the original report from Vietnam that in general, the population is less expressive of negative emotions, hence may reduce the likelihood that items in the questionnaire pertaining to negative feelings would be recognized. Another contributing factors for differences in cut-off scores include the socio-demographic differences in which one study is conducted for the rural population while the other was conducted in an urban setting.

Study by Pakaluk and Price (2020), suggested that father and mother had distinct role but is complementary to each other in raising a child, in which father was found to be more protective towards the family while mothers take the nurturing role. The differences in this role, might need to be consider in the validation of the scale (Pakaluk and Price, 2020).

Further validation studies are needed to establish the validity of EPDS for identifying paternal postpartum depression in the future. Researchers should report the validation for all cut-off points to ensure the completion of data. In addition, the full reporting of cut-offs points may help ensure a more accurate estimation of diagnostic accuracy in future meta-analyses.

Apart from the validation studies, the suitable time frame for detection of paternal postpartum depression still falls in grey areas as paternal depression is closely correlated with maternal depression, and usually developed later than mothers (Garfield et al., 2014; Nath et al., 2016).

A growing effort in developing and validating screening tool specific for detection of depression in fathers during the perinatal period such as Perinatal Assessment of Paternal Affectivity (PAPA) has been developed, it is of importance to compare EPDS and this scales and determine the validity and cost effectiveness of using these tools to screen the fathers (Baldoni, S, Agostini, Schimmenti and Caretti, 2016).

4.5. Implication of practice

Parental depression has been suggested to weaken a parent's ability to regulate their offspring's emotions and affect temperamental development (Hanington et al., 2010). It is also associated with an increased risk of behavioural problems in children (Ramchandani, Stein, Evans and O'Connor, 2005). Hence, it is essential to screen and detect a father's affective symptoms as the impact of psychological distress in men is significant to the whole family's well-being (Mangialavori et al., 2021). Screening for perinatal depression in mothers had been routinely taking place in the maternal-child health care system (Fisher and Garfield, 2016). Early detection and intervention may reduce paternal depression (Walsh et al., 2020). In many countries, perinatal period is the time of frequent medical contact for both mother and child. Maternal and child services usually includes regular follow up and administration of routine
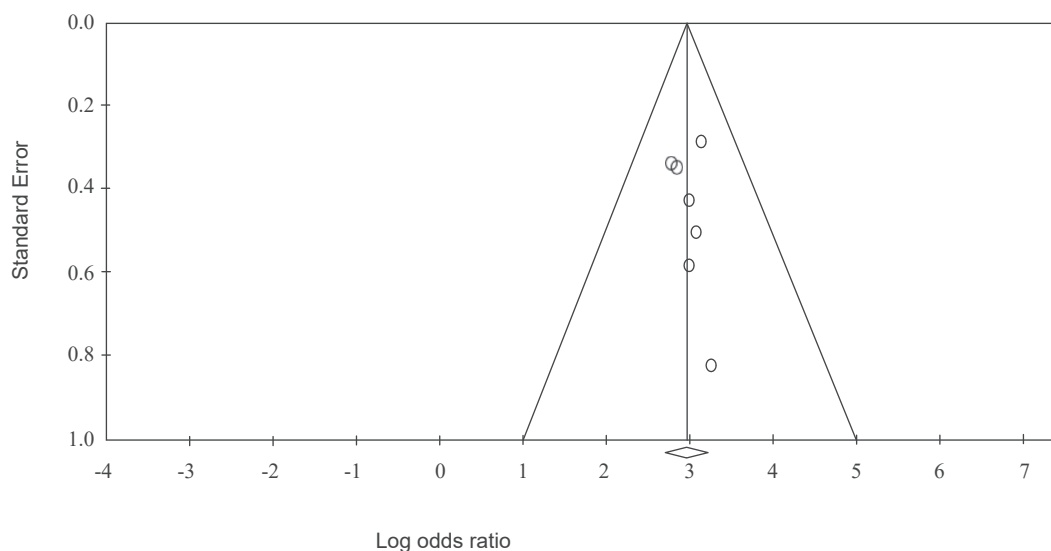
Figure 3. Funnel Plot of Standard Error by Log odds ratio.

immunisation for the babies under five years old. This well-child visits had been suggested to be one of the opportunities to screen fathers as more fathers had been attending the centres, increasing involvement in their child's health care and learning valuable information about their child's health and development (Garfield and Isacco, 2006).

Healthcare cost was also found to be higher in paternal postpartum depression. A study was done in 2011 (Edoka et al., 2011) found a statistically significant mean cost difference of £95.37 in community care services between fathers with and without depression.

EPDS was developed to screen for postnatal depression in mothers; however, growing studies used EPDS to screen paternal postpartum depression. All studies included in the review suggest using a lower cut-off score of EPDS for paternal depression compared to cut-off scores in the maternal population. However, the result shall be interpreted with caution, as multiple factors may influence paternal postpartum depression in different countries such as culture, socio-economic status, education, the perception of the role of a father (Shwalb, 2014). Furthermore, the time when EPDS was used and how male in different societies may have different presentations when it comes to depression and this may influence the EPDS score.

Based on the review, using the EPDS will generate a substantial proportion of false positives, which is costly for the service provider, as a further assessment is needed. However, most screening tools had the same difficulties, and it is designed to indicate the possibility of an illness. EPDS is free in terms of its availability and easily administered self-report questionnaire, and it remains a potential tool in expanding and improving paternal mental health.

5. Conclusion

The current review demonstrated the validity of the EPDS for screening depression in fathers. Clinicians may wish to consider using lower cut-off scores ranging from 7 to 10 of EPDS screening for paternal postpartum depression as it confers the optimum balance between sensitivity and specificity. However, due to the small number of studies included, it requires further validation of larger population samples. The future review may consider differentiating the presence of anxiety, depression, or distress during the transition to parenthood in fathers. It is crucial to include fathers in perinatal mental health care as it will prevent future problems in fathers and the negative impact that could happen to the children and family.

Declarations

Author contribution statement

Ainul Khamisah Shafian: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Salina Mohamed: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Nor Jannah Nasution Raduan: Conceived and designed the experiments; Performed the experiments; Wrote the paper.

Anne Yee Hway Ann: Analyzed and interpreted the data.

Data availability statement

Data included in article/supplementary material/referenced in article.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Areias, M.E., Kumar, R., Barros, H., Figueiredo, E., 1996. Comparative incidence of depression in women and men, during pregnancy and after childbirth. Validation of the Edinburgh Postnatal Depression Scale in Portuguese mothers. Br. J. Psychiatr. 169 (1), 30–35.

Baldoni, F., Giannotti, M., 2020. Perinatal distress in fathers: toward a gender-based screening of paternal perinatal depressive and affective disorders. Front. Psychol. 11 (1892).

Baldoni, F., S, M., Agostini, F., Schimmenti, A., Caretti, V., 2016. Perinatal Assessment of Paternal Affectivity (PAPA): preliminary report on a new screening tool. Infant Ment. Health J. 37, 132–133.

Carlberg, M., Edhborg, M., Lindberg, L., 2018. Paternal perinatal depression assessed by the Edinburgh postnatal depression scale and the Gotland male depression scale: prevalence and possible risk factors. Am. J. Men's Health 12 (4), 720–729.

Cox, J.L., Holden, J.M., Sagovsky, R., 1987. Detection of postnatal depression. Development of the 10-item Edinburgh postnatal depression scale. Br. J. Psychiatr. 150, 782–786.

Darwin, Z., Domoney, J., Iles, J., Bristow, F., Siew, J., Sethna, V., 2021. Assessing the mental health of fathers, other Co-parents, and partners in the perinatal period: mixed methods evidence synthesis. Front. Psychiatr. 11, 585479-585479.

Edmondson, O.J., Psychogiou, L., Vlachos, H., Netsi, E., Ramchandani, P.G., 2010. Depression in fathers in the postnatal period: assessment of the Edinburgh Postnatal Depression Scale as a screening measure. J. Affect. Disord. 125 (1-3), 365–368.

Edoka, I.P., Petrou, S., Ramchandani, P.G., 2011. Healthcare costs of paternal depression in the postnatal period. J. Affect. Disord. 133 (1-2), 356–360.

Fisher, S.D., Garfield, C., 2016. Opportunities to detect and manage perinatal depression in men. Am. Fam. Phys. 93 (10), 824–825.

Galdas, P.M., Cheater, F., Marshall, P., 2005. Men and health help-seeking behaviour: literature review. J. Adv. Nurs. 49 (6), 616–623.

Garfield, C.F., Duncan, G., Rutsohn, J., McDade, T.W., Adam, E.K., Coley, R.L., Chase-Lansdale, P.L., 2014. A longitudinal study of paternal mental health during transition to fatherhood as young adults. Pediatrics 133 (5), 836–843.

Garfield, C.F., Isacco, A., 2006. Fathers and the well-child visit. Pediatrics 117 (4), e637–645.

Gibson, J., McKenzie-McHarg, K., Shakespeare, J., Price, J., Gray, R., 2009. A systematic review of studies validating the Edinburgh Postnatal Depression Scale in antepartum and postpartum women. Acta Psychiatr. Scand. 119 (5), 350–364.

Goodman, J.H., 2004. Paternal postpartum depression, its relationship to maternal postpartum depression, and implications for family health. J. Adv. Nurs. 45 (1), 26–35.

Hanington, L., Ramchandani, P., Stein, A., 2010. Parental depression and child temperament: assessing child to parent effects in a longitudinal population study. Infant Behav. Dev. 33 (1), 88–95.

Kim, P., Swain, J.E., 2007. Sad dads: paternal postpartum depression. Psychiatry (Edgmont (Pa.: Township)) 4 (2), 35–47. Retrieved from. https://pubmed.ncbi.nlm.nih.gov/20805898. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2922346/.

Lai, B.P., Tang, A.K., Lee, D.T., Yip, A.S., Chung, T.K., 2010. Detecting postnatal depression in Chinese men: a comparison of three instruments. Psychiatr. Res. 180 (2-3), 80–85.

Madsen, S.A., Juhl, T., 2007. Paternal depression in the postnatal period assessed with traditional and male depression scales. J. Men's Health Gend. 4 (1), 26–31.

Mangialavori, S., Giannotti, M., Cacioppo, M., Spelzini, F., Baldoni, F., 2021. Screening for early signs of paternal perinatal affective disorder in expectant fathers: a cluster analysis approach. J. Personalized Med. 11 (1), 1–15.

Martin, L.A., Neighbors, H.W., Griffith, D.M., 2013. The experience of symptoms of depression in men vs women: analysis of the national comorbidity survey replication. JAMA Psychiatr. 70 (10), 1100–1106.

Massoudi, P., Hwang, C.P., Wickberg, B., 2013. How well does the Edinburgh Postnatal Depression Scale identify depression and anxiety in fathers? A validation study in a population based Swedish sample. J. Affect. Disord. 149 (1-3), 67–74.

Matthey, S., Barnett, B., Kavanagh, D.J., Howie, P., 2001. Validation of the Edinburgh Postnatal Depression Scale for men, and comparison of item endorsement with their partners. J. Affect. Disord. 64 (2-3), 175–184.

Mueller, A., Segal, D., 2015. Structured versus semistructured versus unstructured interviews. Encyclop. Clin. Psychol. 1–7.

Nath, S., Psychogiou, L., Kuyken, W., Ford, T., Ryan, E., Russell, G., 2016. The prevalence of depressive symptoms among fathers and associated risk factors during the first seven years of their child's life: findings from the Millennium Cohort Study. BMC Publ. Health 16 (1), 509.

Pakaluk, C.R., Price, J., 2020. Are mothers and fathers interchangeable caregivers? Marriage Fam. Rev. 56 (8), 784–793.

Patrick Bossuyt, C., Deeks, Jon, Hyde, Chris, Leeflang, Mariska, Scholten, Rob, 2013. Chapter 11: interpreting result and drawing conclusions. In: Cochrane Handbook for Systematic Review of Diagnostic Test Accuracy Version 0.9. The Cochrane Collaboration.

Quadagno, D.M., Dixon, L.A., Denney, N.W., Buck, H.W., 1986. Postpartum moods in men and women. Am. J. Obstet. Gynecol. 154 (5), 1018–1023.

Ramchandani, P., Stein, A., Evans, J., O'Connor, T.G., 2005. Paternal depression in the postnatal period and child development: a prospective population study. Lancet 365 (9478), 2201–2205.

Shaheen, N.A., AlAtiq, Y., Thomas, A., Alanazi, H.A., AlZahrani, Z.E., Younis, S.A.R., Hussein, M.A., 2019. Paternal postnatal depression among fathers of newborn in Saudi Arabia. Am. J. Men's Health 13 (1), 1557988319831219.

Shwalb, D.W., Shwalb, B.J., 2014. Fatherhood in Brazil, Bangladesh, Russia, Japan, and Australia. Online Read. Psychol. Cult. 6 (3).

Takwoingi, Y., Riley, R.D., Deeks, J.J., 2015. Meta-analysis of diagnostic accuracy studies in mental health. Evid. Base Ment. Health 18 (4), 103–109.

Tran, T.D., Tran, T., Fisher, J., 2012. Validation of three psychometric instruments for screening for perinatal common mental disorders in men in the north of Vietnam. J. Affect. Disord. 136 (1-2), 104–109.

Walsh, T.B., Davis, R.N., Garfield, C., 2020. A call to action: screening fathers for perinatal depression. Pediatrics 145 (1), e20191193.

Whiting, P., Westwood, M., Beynon, R., Burke, M., Sterne, J.A.C., Glanville, J., 2011. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. J. Clin. Epidemiol. 64 (6), 602–607.

Whiting, P.F., Rutjes, A.W., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., Bossuyt, P.M., 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 155 (8), 529–536.

Zamora, J., Abraira, V., Muriel, A., Khan, K., Coomarasamy, A., 2006. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med. Res. Methodol. 6 (1), 31.