Inflammatory biomarkers and perinatal depression: A systematic review

2.9 / 5
AVERAGE RATING

Journal Article · Biology, Medicine, Psychology
Reviewed August 3, 2026

QUALITY RATINGS

Technical Soundness 3/5
Search and selection follow PRISMA with NOS-based quality assessment, but lack of meta-analysis exploration, limited RoB integration, and overgeneralized conclusions weaken rigor.

Novelty 3/5
Extends prior PPD reviews by including more antenatal and longitudinal work, but positioning versus recent systematic reviews is not sharply articulated.

Organization & Clarity 3/5
IMRaD structure is present, but dense, partially truncated tables and some over-aggregated results reduce clarity of the synthesis.

Writing Quality 4/5
Generally well written and coherent with minor grammatical issues; arguments are understandable though some claims overstate consistency.

Statistics & Significance 2/5
Statistical findings are largely reported as p-value-based directions without standardized effect sizes, CIs, or multiple-comparison handling in the synthesis.

Reproducibility 3/5
Protocol registration and PRISMA are strengths, but incomplete search updating, limited detail on search strings, and lack of shared extraction data constrain replication.

Ethics/Privacy Compliance 2/5
The review itself adheres to ethical norms, but there is no evaluation or reporting of ethics/consent/registration status of included human studies.

ISSUES FOUND

1 critical · 4 major · 13 moderate · 15 minor · 2 polish

35 issues identified, graded 1 (polish) to 5 (critical).

- Major methodological limitations in study selection, risk-of-bias assessment, and synthesis methods (I01, I02, I03, G02, G03, G04) substantially weaken the credibility and interpretability of the review's conclusions.
- Reporting and internal consistency problems (X01, X02, I04, I09, I10, I13, I17, I21, I24, V02) undermine transparency of the evidence base and make it difficult to follow or verify the results.
- The review's discussion and conclusions overstate consistency and direction of effects for inflammatory markers relative to mixed and sometimes inverse results described earlier (I11, I12, I20, I21, I26).
- Search methods, timeframe, and use of non-database sources are incompletely reported or out of date relative to the acceptance date, threatening comprehensiveness and reproducibility (I05, I06, I07, G01, V01).
- Handling of heterogeneity in depression constructs (diagnosed depression vs symptoms vs mood vs mixed anxiety/distress) and study types (primary vs secondary analyses) is conceptually and analytically underdeveloped (I03, I08, I20, I23, I25).
- The relationship to prior postpartum depression reviews is only superficially described, limiting the reader's understanding of the review's added value and specific contributions (I15, I22).
- Procedural details on screening, selection, and inter-rater reliability beyond quality assessment are sparse, limiting confidence in the rigor of the study selection process (I14, I19).
- Presentation of key elements (PRISMA flow, captions, tables) is suboptimal and sometimes inconsistent, impeding readability and verification of the selection process and extracted data (I17, I24, V01, V02).

Executive Summary

- Major methodological limitations in study selection, risk-of-bias assessment, and synthesis methods (I01, I02, I03, G02, G03, G04) substantially weaken the credibility and interpretability of the review's conclusions.
- Reporting and internal consistency problems (X01, X02, I04, I09, I10, I13, I17, I21, I24, V02) undermine transparency of the evidence base and make it difficult to follow or verify the results.
- The review's discussion and conclusions overstate consistency and direction of effects for inflammatory markers relative to mixed and sometimes inverse results described earlier (I11, I12, I20, I21, I26).
- Search methods, timeframe, and use of non-database sources are incompletely reported or out of date relative to the acceptance date, threatening comprehensiveness and reproducibility (I05, I06, I07, G01, V01).
- Handling of heterogeneity in depression constructs (diagnosed depression vs symptoms vs mood vs mixed anxiety/distress) and study types (primary vs secondary analyses) is conceptually and analytically underdeveloped (I03, I08, I20, I23, I25).
- The relationship to prior postpartum depression reviews is only superficially described, limiting the reader's understanding of the review's added value and specific contributions (I15, I22).
- Procedural details on screening, selection, and inter-rater reliability beyond quality assessment are sparse, limiting confidence in the rigor of the study selection process (I14, I19).
- Presentation of key elements (PRISMA flow, captions, tables) is suboptimal and sometimes inconsistent, impeding readability and verification of the selection process and extracted data (I17, I24, V01, V02).
- Minor clarity issues in the aim statement and terminology (I18, I10) further reduce precision about the review's primary targets and interpretation of 'associations'.
- No formal assessment of reporting biases or certainty of evidence is provided, leaving readers without a structured appraisal of confidence in the main findings (G03, G04).

Prioritized Issues

I01

MAJOR · 4/5

Section	Methods
Location	Unknown
Evidence (≤25w)	"Different versions of the Newcastle-Ottawa Scale (NOS) were used to assess the methodological quality of selected studies"
Problem	Risk-of-bias assessment is limited to NOS total scores; no domain-level judgments or description of prevalent biases, and results are only briefly summarized.
Standard Violated	Systematic review standards (e.g., PRISMA) expect transparent, domain-wise risk-of-bias assessment and integration into interpretation, not just summary scores.
Impact	Without explicit domains (selection, comparability, outcome) and their influence on conclusions, readers cannot gauge how bias patterns affect overall findings.
Confidence	High

I02

MAJOR · 4/5

Section	Discussion
Location	Unknown
Evidence (≤25w)	"a combined quantitative synthesis of all eligible studies was not feasible owing to the large variability in inflammatory markers assessed"
Problem	Justification for not doing any meta-analysis is asserted but not supported by any formal assessment of heterogeneity or identification of analyzable subsets.
Standard Violated	Evidence-synthesis best practice requires exploring meta-analysis for subsets with comparable exposures/outcomes and justifying non-use with quantitative heterogeneity metrics or protocol-based rules.
Impact	Potentially meta-analyzable markers (e.g., CRP, IL-6, TNF-α with EPDS/CES-D) are only narratively summarized, limiting precision and risking subjective weighting of evidence.
Confidence	Medium

I03

MAJOR · 4/5

Section	Methods
Location	Unknown
Evidence (≤25w)	"Depression assessed using ICD/DSM diagnostic criteria either through diagnostic interview or expert opinion. Alternatively, depressive symptomatology assessed using standardized psychometric instruments"
Problem	Eligibility criteria allow heterogeneous constructs (diagnosed depression, depressive symptoms, "depressive mood") without clear boundaries or handling of mixed anxiety/distress scales.
Standard Violated	Clinical systematic reviews should define target condition precisely and specify how different instruments and constructs map to that definition.
Impact	Pooling studies capturing general distress or mood with studies of diagnosed depression may blur associations and compromise construct validity of depression-related biomarkers.
Confidence	High

I26

MODERATE · 3/5

Section	Conclusion
Location	Unknown
Evidence (≤25w)	"the findings of the present work provide evidence of increased proinflammatory markers among women with depressive symptoms."
Problem	Conclusion suggests generalized increase in proinflammatory markers, despite Results emphasizing mixed, marker- and period-specific, and sometimes inverse associations.
Standard Violated	Conclusions should be appropriately qualified to reflect heterogeneity and uncertainty instead of broad generalizations.
Impact	Risk of overgeneralization may misinform clinicians or policymakers regarding biomarker screening or mechanistic certainty.
Confidence	High

I13

MODERATE · 3/5

Section	Results
Location	Unknown
Evidence (≤25w)	"various biomarkers were used for the investigations of association with depressive symptoms"
Problem	Across the Results, there is no systematic reporting of effect sizes, confidence intervals, or adjustment for multiple comparisons when summarizing individual studies' statistical findings.
Standard Violated	SAMPL and good reporting practice expect quantitative synthesis to highlight effect size and precision, not only p-values or direction.
Impact	Without standardized statistical parameters, readers cannot compare strength of associations across studies or judge clinical relevance.
Confidence	Medium

I12

MODERATE · 3/5

Section	Discussion
Location	Unknown
Evidence (≤25w)	"associations were somewhat different when focusing at pregnancy compared to the delivery time-point and postpartum, and mainly referred to increased levels of IL-6, IL-8, CRP and TNF-α among depressed."
Problem	Claim of "mainly" increased IL-6, IL-8, CRP, TNF-α downplays substantial null and inverse associations explicitly described for these markers in earlier sections.
Standard Violated	Synthesis should balance supporting and conflicting evidence and avoid emphasis that misrepresents heterogeneous findings.
Impact	Risk of misguiding meta-analytic or translational efforts towards a narrow biomarker set while true effects may be modest or context-specific.
Confidence	High

I11

MODERATE · 3/5

Section	Discussion
Location	Unknown
Evidence (≤25w)	"the most consistent finding of the present study was the significant association between elevated CRP levels and depressive symptoms during pregnancy."
Problem	This statement overstates consistency given earlier acknowledgment that only 1 of 3 CRP studies antenatally and 3 postpartum longitudinal studies showed associations, with inverse associations postpartum.
Standard Violated	Conclusions should accurately reflect the balance of positive, null, and inconsistent findings reported in the Results section.
Impact	Readers may infer stronger, more uniform evidence for CRP as a biomarker of perinatal depression than is warranted, affecting clinical and research decisions.
Confidence	High

I10

MODERATE · 3/5

Section	Results
Location	Unknown
Evidence (≤25w)	"Most of the studies (11/14) found an association between inflammatory biomarkers and depression status."
Problem	"Association" is used broadly without distinguishing directionality, magnitude, statistical robustness, or whether results persisted after adjustment for confounders.
Standard Violated	Evidence synthesis should differentiate positive, negative, null, and adjusted vs unadjusted associations, not aggregate them into a binary "association found" metric.
Impact	Over-simplification could overstate evidence supporting specific biomarkers as related to perinatal depression.
Confidence	High

I04

MODERATE · 3/5

Section	Results
Location	Table 1 / Table 2
Evidence (≤25w)	"In general, independently of the study design, studies with high quality (higher or equal to 8 points) are scarce."
Problem	Interpretation of NOS scores is qualitative ("high quality") without specifying thresholds a priori or how this influenced weighting of evidence or sensitivity analyses.
Standard Violated	Risk-of-bias tools should be used to classify and, where possible, stratify or sensitize analyses; arbitrary post hoc categories should be explicitly justified.
Impact	Readers cannot discern whether conclusions are driven by lower-quality observational studies or robust ones, limiting confidence in inferred biomarker-depression links.
Confidence	High

I08

MODERATE · 3/5

Section	Methods
Location	Unknown
Evidence (≤25w)	"The following data was extracted from each selected study: country of origin; study design; number of subjects; socio-economic status/ethnicity; obstetric information"
Problem	Data extraction items listed do not include critical risk-of-bias covariates such as adjustment factors, handling of confounders, or assay quality parameters.
Standard Violated	Systematic reviews of observational biomarkers should predefine extraction of confounding control and methodological details relevant to bias.
Impact	Inadequate characterization of confounding/control hampers nuanced interpretation of whether observed associations are likely causal or confounded.
Confidence	Medium

I07

MODERATE · 3/5

Section	Methods
Location	Unknown
Evidence (≤25w)	"Case reports, experimental studies, reviews, qualitative analysis, meta-analysis, gray literature or replicated data were excluded."
Problem	Non-database sources beyond gray literature (e.g., trial registries, conference proceedings, reference list hand-searching) are not systematically addressed.
Standard Violated	Comprehensive systematic reviews should describe all information sources and supplementary search methods to minimize publication bias.
Impact	Potential omission of unpublished or non-indexed studies may bias evidence towards positive, published findings on inflammatory markers.
Confidence	Medium

I06

MODERATE · 3/5

Section	Methods
Location	Unknown
Evidence (≤25w)	"On February 23rd, 2022, a new search was conducted using the same search formula"
Problem	Last search update is February 2022 while acceptance is May 2024; no rationale or search for 2022–2024 literature is described.
Standard Violated	Systematic reviews are expected to have up-to-date searches close to submission/publication or to explain limitations of older searches.
Impact	Key studies on perinatal inflammation and depression from 2022–2024 may be missing, affecting completeness and contemporary relevance of conclusions.
Confidence	High

I05

MODERATE · 3/5

Section	Methods
Location	S1 Table
Evidence (≤25w)	"for search terms see S1 Table"
Problem	The main text omits full search strings and provides limited detail on filters and handling of non-database sources, relying on a supplement not described in detail.
Standard Violated	PRISMA requires transparent reporting of full electronic search strategies for at least one database to permit replication.
Impact	Without the explicit strategy in the article text, replicability of study identification and assessment of search adequacy are constrained.
Confidence	High

I25

MINOR · 2/5

Section	Results
Location	S2 Table
Evidence (≤25w)	"27 were studies in which the association between depression and inflammatory markers wasn't the principal aim but... presented them in the S2 Table."
Problem	Secondary-analysis studies are included qualitatively but not clearly distinguished in the narrative synthesis from primary-aim studies, nor is their impact on conclusions delineated.
Standard Violated	Systematic reviews should stratify primary vs secondary analyses and indicate how each group contributes to overall inferences.
Impact	Mixing primary and secondary analyses may overrepresent exploratory findings and obscure strength of evidence focused on perinatal depression.
Confidence	Medium

I24

MINOR · 2/5

Section	Figures & Tables
Location	Fig 1
Evidence (≤25w)	"Fig 1. Flow diagram of the search selection for the included studies."
Problem	PRISMA flow is referenced but not fully described in text; numbers are summarized but reasons for full-text exclusions (the "44 studies" in P25) are not itemized by category.
Standard Violated	PRISMA recommends reporting reasons for exclusion at full-text stage in the flow diagram or text.
Impact	Lack of categorized exclusion reasons reduces transparency about potential selection bias (e.g., language, outcome, design).
Confidence	High

I23

MINOR · 2/5

Section	General
Location	Unknown
Evidence (≤25w)	"All relevant data are within the manuscript and its Supporting Information files."
Problem	"All relevant data" claim is not backed by clear indication that extracted datasets (e.g., per-study effect estimates) or coding sheets are available, beyond summary tables.
Standard Violated	Data availability statements should specify what datasets are shared and in what format to enable secondary analyses.
Impact	Others cannot readily reuse detailed extraction for re-analysis or alternative syntheses, limiting transparency and reproducibility.
Confidence	Medium

I22

MINOR · 2/5

Section	Discussion
Location	Unknown
Evidence (≤25w)	"in line with the main conclusion of a recent systematic review, stating that… [32]."
Problem	The relationship to review [32] is briefly asserted but not critically compared (e.g., differences in inclusion period, markers, or analytic approach).
Standard Violated	When citing overlapping systematic reviews, authors should provide a concise comparative analysis, not only a single-point agreement.
Impact	Unclear how this review advances or differs from [32], which may affect decisions about which synthesis to rely on for practice or future research planning.
Confidence	Medium

I21

MINOR · 2/5

Section	Results
Location	Unknown
Evidence (≤25w)	"three of them showed lower TNF-α in PPD group… another study displays a higher TNF-α at 8–12 gestational weeks in depressed subjects"
Problem	Opposing directions of TNF-α associations are described but not explored quantitatively (e.g., no consideration of context, timing, or assay differences) or flagged as potential effect modification.
Standard Violated	Systematic reviews should not only catalogue inconsistencies but attempt structured exploration (e.g., by timepoint, fluid) when contradictions arise.
Impact	Readers are left without guidance on how to interpret conflicting TNF-α results; potential moderators remain speculative rather than systematically assessed.
Confidence	Medium

I20

MINOR · 2/5

Section	Discussion
Location	Unknown
Evidence (≤25w)	"definition of exposure and outcome among inflammation and depression is not always straight-forward."
Problem	Although key limitations are acknowledged, the discussion does not systematically map which specific markers, fluids, or depression measures are most affected by definitional ambiguity.
Standard Violated	Limitations sections should, where possible, specify which evidence segments are most compromised rather than only general statements.
Impact	Readers cannot easily prioritize which biomarker–depression associations are most reliable versus most methodologically fragile.
Confidence	Medium

I19

MINOR · 2/5

Section	Methods
Location	Unknown
Evidence (≤25w)	"The full text of qualifying articles was then assessed against the same standard by different pairs of authors"
Problem	No explicit statement that screening and selection followed predefined protocol decisions regarding handling of disagreements beyond "discussion"; no escalation rules beyond informal group consensus are specified.
Standard Violated	Transparent systematic review methods should detail adjudication procedures for conflicts (e.g., third reviewer, majority vote).
Impact	Ambiguous adjudication processes may permit subjective bias in inclusion/exclusion, especially for borderline cases.
Confidence	Medium

I17

MINOR · 2/5

Section	Results
Location	Table 1
Evidence (≤25w)	"Table 1. Cross-sectional and case-control studies on inflammatory biomarkers and depression."
Problem	Tables 1 and 2 are extremely dense, with truncated entries and interleaved narrative text, making it difficult to follow columns (e.g., markers, results, quality score) and verify extraction.
Standard Violated	Data tables in systematic reviews should be clearly structured, readable, and fully aligned with column headings for transparency.
Impact	Poor legibility may obscure key methodological differences across studies and hinder external verification or re-analysis.
Confidence	High

I16

MINOR · 2/5

Section	Ethics/General
Location	Unknown
Evidence (≤25w)	"we excluded from this systematic review results of inflammatory markers related to quantification of immune cells as well as physical properties"
Problem	Ethical/IRB approval, consent procedures, and trial registration for included studies are not described or synthesized anywhere in the manuscript.
Standard Violated	For human-subject systematic reviews, authors should comment on ethics/consent status of primary studies or at least note whether this information was collected.
Impact	Readers cannot assess whether included evidence adheres to contemporary ethics and privacy standards, particularly relevant in perinatal populations.
Confidence	High

I15

MINOR · 2/5

Section	Discussion
Location	Unknown
Evidence (≤25w)	"Our systematic review endeavors to thoroughly investigate the relationship between inflammation and depression throughout the perinatal period."
Problem	Positioning relative to prior PPD-focused systematic reviews (e.g., [30], [32]) remains high-level and does not clearly specify what is novel beyond including more recent antenatal-focused studies.
Standard Violated	Academic standards expect explicit differentiation of new contributions versus existing reviews, especially when multiple systematic reviews exist in the area.
Impact	Readers may overestimate the incremental value of this review or have difficulty understanding its unique contribution to the literature.
Confidence	Medium

I14

MINOR · 2/5

Section	Methods
Location	Unknown
Evidence (≤25w)	"All the inter-rater agreements between authors were verified prior to resolving disagreements."
Problem	For quality assessment, only percent agreement and kappa are given; no similar metrics for title/abstract or full-text screening are reported.
Standard Violated	Systematic reviews should report reliability of screening stages as well as risk-of-bias assessments to support rigor of selection.
Impact	Unclear screening reliability raises possibility of inconsistent inclusion/exclusion of borderline studies.
Confidence	Medium

I09

MINOR · 2/5

Section	Results
Location	Unknown
Evidence (≤25w)	"31 studies involved repeated assessment time points and longitudinal analyses both during pregnancy (4 studies) and from pregnancy to postpartum (27 studies)."
Problem	Counts of longitudinal studies sometimes conflict with later descriptions (e.g., "five" postpartum longitudinal in P32 vs "five" again in P118), with no integrated overview table of designs/time windows.
Standard Violated	Results should provide internally consistent, clearly tabulated accounting of study designs and follow-up periods.
Impact	Ambiguity about which studies contribute to each longitudinal category complicates interpretation of temporal and predictive findings.
Confidence	Medium

I18

POLISH · 1/5

Section	Introduction
Location	Unknown
Evidence (≤25w)	"we aim to delve into longitudinal studies to not only investigate cross-sectional between inflammation and perinatal depression but also to elucidate possible bidirectional effects"
Problem	Aim statement contains a grammatical omission ("investigate cross-sectional between"), reducing clarity about whether cross-sectional or longitudinal associations are primary targets.
Standard Violated	Research aims should be clearly and grammatically articulated to avoid ambiguity about study scope.
Impact	Minor confusion for readers on primary analytic focus (cross-sectional vs longitudinal).
Confidence	High

Consistency Audit

I11

Type	claim_scope
Evidence A	"the most consistent finding... was the significant association between elevated CRP levels and depressive symptoms during pregnancy." (P127)
Evidence B	"CRP– 1 in 3 studies that have assessed this biomarker" (P108)
Inconsistency	Discussion claims CRP–depression association is "most consistent" whereas Results acknowledge only 1 of 3 antenatal CRP studies show such association, with postpartum inverse relations as well.

I12

Type	claim_scope
Evidence A	"associations were somewhat different... and mainly referred to increased levels of IL-6, IL-8, CRP and TNF-α among depressed." (P125)
Evidence B	"IL-6 was assessed in 10... but only in two was found a significant positive association" (P108)
Inconsistency	The narrative that associations "mainly" involve increased IL-6, IL-8, CRP, TNF-α overstates consistency compared with detailed mixed findings in the Results.

I26

Type	claim_scope
Evidence A	"provide evidence of increased proinflammatory markers among women with depressive symptoms." (P136)
Evidence B	"not many prominent and consistent associations... there is a little clarity regarding a consistent immune profile" (P125)
Inconsistency	Conclusion generalizes to "increased proinflammatory markers" despite earlier recognition of limited, inconsistent immune profiles across markers and timepoints.

Guideline Compliance

Audited against PRISMA (2020); study type: systematic review.

Item	Status	Evidence / Note
PRISMA-1 — Identification as a systematic review	Reported	"Inflammatory biomarkers and perinatal depression: A systematic review" — Title explicitly labels the article as a systematic review.
PRISMA-5 — Eligibility criteria	Reported	"Inclusion criteria: • Written in English or French or Spanish or Portuguese • Observational studies • Depression assessed using ICD/DSM... Exclusion criteria: • Case reports or experimental studies..." — Inclusion and exclusion criteria are listed and the focus group (observational designs, perinatal period) is described.
PRISMA-6 — Information sources and dates	Reported	"An initial article search was conducted on July 3rd, 2020, through three electronic databases (Pubmed, Web of Science and PsychInfo)... On February 23rd, 2022, a new search was conducted" — All databases and the dates of each search are given.
PRISMA-7 — Search strategy	Partial	"An initial article search was conducted... (for search terms see S1 Table)." — They refer to full strategies in S1 Table but do not provide them in the manuscript text; filters and exact strings are not reported here.
PRISMA-8 — Selection process	Reported	"Two authors (ASF and MM) screened the titles and abstracts... independently... Any discrepancies were resolved first through discussion amongst the pair, and if a consensus could still not be reached, by conferring with other group members." — Number of reviewers, independence, and disagreement resolution for selection are described.
PRISMA-9 — Data collection process	Reported	"The following data was extracted from each selected study... the data extraction were conducted independently by two authors." — Data items and that extraction was done independently by two reviewers are described.

Item	Status	Evidence / Note
PRISMA-11 — Risk of bias assessment	Reported	"Different versions of the Newcastle-Ottawa Scale (NOS) were used to assess the methodological quality of selected studies... All the inter-rater agreements between authors were verified" — Tool (NOS), scope (case-control, cohort, cross-sectional), and assessor pairing with reliability are reported.
PRISMA-13a — Synthesis methods	Partial	"Due to this diversity, results are presented in two steps. Firstly, cross-sectional, and case-control studies... Secondly, longitudinal studies are presented" — They state how studies are grouped for narrative syntheses but do not describe detailed synthesis methods (e.g. how effects were combined or compared).
PRISMA-13d — Meta-analysis model and heterogeneity	Not applicable	No meta-analysis was performed; only qualitative synthesis is described, so no meta-analytic model or heterogeneity assessment applies.
PRISMA-14 — Reporting bias assessment	Missing	No methods are described for assessing risk of bias due to missing results or publication bias.
PRISMA-16a — Study selection flow	Reported	"A total of 3527 relevant references were initially identified... 806 duplicated references were removed... 83 studies were included in the review. A flow diagram of the search selection for the included studies is presented in Fig 1." — Numbers at each selection stage and a flow diagram are reported; reasons for full-text exclusions are summarized as meeting exclusion criteria.
PRISMA-17 — Characteristics of included studies	Reported	"As can be seen in Table 1, the quality of the fourteen studies... Regarding the longitudinal studies, the quality of the four studies... Table 1. Cross-sectional and case-control studies on inflammatory biomarkers and depression." — Tables 1–2 list each included study with citation and key characteristics.
PRISMA-18 — Risk of bias in included studies	Reported	"Different versions of the Newcastle-Ottawa Scale (NOS) were used... As can be seen in Table 1, the quality of the fourteen studies..." — Risk-of-bias/quality scores per study are summarized in tables as described.
PRISMA-21 — Certainty of evidence	Missing	No formal assessment of certainty or confidence (e.g. GRADE or equivalent) for bodies of evidence is reported.
PRISMA-24a — Registration and protocol	Reported	"This study followed the recommendations outlined in the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) guidelines [34,35] and has been registered on... PROSPERO (Registration ID: CRD42020210080). Protocol details are available at https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=210080." — Registration database, ID, and protocol access URL are provided.
PRISMA-25 — Support and funding	Reported	"Funding: AS-F was supported by FCT and the Portuguese Ministry of Science, Technology and Higher Education... The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript." — Funding sources and funders' role (no role) are clearly reported.

Action Plan

I02, G02

Action	Develop and report a structured synthesis plan that includes: (a) a priori criteria for when meta-analysis is attempted (e.g., same marker, comparable time window, similar depression construct, compatible effect measures); (b) formal assessment of statistical and clinical heterogeneity (e.g., I^2, τ^2, visual forest plots, stratified summaries) for any candidate subsets; and (c) explicit justification, grounded in these assessments, for when a meta-analysis is not feasible, with detailed description of the narrative synthesis methods used instead (how effects are grouped, weighted, and compared).
Effort (S/M/L)	L
Priority (P0–P2)	P0
Success Criteria	Methods section contains a dedicated “Synthesis methods” subsection that specifies quantitative and narrative synthesis procedures; at least one formal heterogeneity assessment is presented for a candidate analyzable subset, or transparent, criteria-based justification is given for the absence of any meta-analysis. Reviewers can trace how individual study results were combined or compared for each marker/time window.

I01, I04, G03, G04

Action	Upgrade risk-of-bias and certainty assessment by (a) performing domain-level risk-of-bias judgments (e.g., selection, measurement, confounding) for each included study rather than relying solely on NOS totals; (b) summarizing the predominant bias concerns per marker and per time period; (c) conducting at least basic sensitivity analyses or stratified summaries by risk-of-bias level; and (d) implementing a structured certainty-of-evidence framework (e.g., GRADE or adapted approach) for each main marker–period–depression construct combination, with explicit criteria for downgrading or upgrading.
Effort (S/M/L)	L
Priority (P0–P2)	P0
Success Criteria	Results include a table or figure with domain-level risk-of-bias ratings for each study and narrative/structured summaries of key bias domains by marker; at least one sensitivity or stratified analysis by risk-of-bias is presented; Discussion or a dedicated section presents certainty ratings (e.g., high/moderate/low/very low) for each primary outcome, with reasons for each rating clearly stated.

X01, X02, V02, I09, I24, V01

Action	Reconcile and clearly document all study flow and counting inconsistencies by (a) re-checking all numbers for records identified, screened, excluded, and included; (b) ensuring counts of longitudinal, cross-sectional, and repeated-measures studies are internally consistent across text, tables, and figures; (c) revising the PRISMA diagram and adding a detailed caption; and (d) itemizing full-text exclusion reasons by category for the 44 excluded studies in the text or a supplementary table.
Effort (S/M/L)	M
Priority (P0–P2)	P0
Success Criteria	All numerical counts (records identified, screened, excluded, included; numbers of longitudinal/postpartum/antenatal studies) are internally consistent across abstract, Methods, Results, PRISMA figure, and tables; PRISMA flowchart has a descriptive caption and a text paragraph that reproduces the key numbers; a table or narrative itemizes main reasons for full-text exclusions with counts that add up to the total excluded.

I03, I08, I20

Action	Clarify and analytically handle heterogeneity in depression-related constructs and key covariates by (a) explicitly defining categories (e.g., diagnosed MDD, validated depressive symptom scales, broad distress/mixed anxiety-depression, unspecified "depressive mood"); (b) revising eligibility criteria to indicate which constructs are included and how mixed scales are handled; (c) updating data extraction to capture depression measure type, cut-offs, and handling of comorbid anxiety; and (d) structuring Results and Discussion to summarize findings separately by these categories and to specify which markers/fluids are most affected by this definitional ambiguity.
Effort (S/M/L)	L
Priority (P0–P2)	P0
Success Criteria	Eligibility criteria explicitly differentiate depression constructs and specify inclusion/exclusion rules for mixed scales; extraction tables have columns for depression measure, diagnostic vs symptom-based, and comorbidity handling; Results subsections or summary tables present marker associations stratified by depression construct; Discussion explicitly identifies markers/fluids where definitional heterogeneity most limits interpretability.

I05, G01

Action	Fully document the search strategy in the main text or an accessible supplement by (a) providing complete database-specific search strings, including Boolean operators, truncations, and limits; (b) detailing any filters (language, date, study type) applied; and (c) clearly describing how non-database records (if any) were identified and managed, with explicit linkage to the supplementary material.
Effort (S/M/L)	M
Priority (P0–P2)	P1
Success Criteria	A PRISMA-compliant "Search strategy" subsection (and/or an appendix) lists full search strings for each database and details all filters and limits; the manuscript text explicitly refers to the supplementary file where the verbatim strategies are presented; another reader could replicate the search without additional assumptions.

I06, I07

Action	Update and expand the search to cover the period from February 2022 to at least near acceptance and clarify non-database sources by (a) re-running all database searches to a current cut-off date; (b) systematically searching additional sources such as trial registries, conference abstracts, and reference lists with stated procedures; and (c) screening and integrating any newly identified studies, with documentation of their influence on the synthesis and conclusions.
Effort (S/M/L)	L
Priority (P0–P2)	P0
Success Criteria	Search methods specify a final search date that is within 6–12 months of acceptance; at least one non-database source category (e.g., trial registry, reference list hand-search) is addressed with clear methods; if new studies are found, they appear in updated tables and are incorporated into results and discussion, with a brief statement describing whether and how they modified conclusions.

I10, I13, I21

Action	Standardize and enrich reporting of individual-study results by (a) consistently stating direction and magnitude of associations (including effect estimates and 95% CIs where reported); (b) indicating whether associations are crude or adjusted and for which confounders; (c) noting any correction for multiple comparisons used in the primary studies; and (d) explicitly flagging and discussing contexts that may explain opposite directions of association (e.g., timing, biological matrix, assay methodology, sample characteristics) for key markers such as TNF-α.
Effort (S/M/L)	M
Priority (P0–P2)	P1
Success Criteria	Results text and summary tables report, for the majority of included studies, the type of effect measure, direction, approximate magnitude, CI or p-value, and adjustment status; statements about “associations” routinely specify positive/negative, crude/adjusted; at least one dedicated paragraph or table addresses effect modification/context for conflicting TNF-α (and similar) findings.

I11, I12, I26

Action	Align interpretive statements in the Discussion and Conclusion with the mixed nature of the findings by (a) revising language that suggests generalized or predominantly increased inflammatory markers; (b) explicitly acknowledging the proportion of null and inverse findings for IL-6, IL-8, CRP, TNF-α and other key markers; and (c) reframing conclusions to be marker-, period-, and context-specific rather than global, indicating where evidence is consistent vs inconsistent.
Effort (S/M/L)	S
Priority (P0–P2)	P0
Success Criteria	Discussion and Conclusion sections contain no unqualified statements implying a generalized proinflammatory state; instead, they include marker-specific summaries that mention the number/fraction of studies supporting positive, null, and inverse associations for each major marker; claims of “mainly increased” markers are either removed or qualified with explicit reference to the mixed evidence.

I14, I19

Action	Strengthen reporting of study selection procedures by (a) specifying how many reviewers independently screened titles/abstracts and full texts; (b) providing inter-rater agreement metrics (e.g., percent agreement, kappa) for screening stages as well as quality assessment; and (c) clarifying predefined rules for resolving disagreements, including any escalation beyond informal discussion (e.g., third reviewer arbitration).
Effort (S/M/L)	S
Priority (P0–P2)	P2
Success Criteria	Methods state the number of reviewers per screening stage, describe independence, and report at least one agreement statistic at title/abstract and full-text levels; a clear protocol for resolving disagreements (discussion plus named third reviewer or equivalent) is described; these processes are consistent across the manuscript.

l15, l22

Action	Provide a more detailed and critical comparison to key prior reviews (e.g., [30], [32]) by (a) tabulating or narratively outlining differences in inclusion periods, populations (antenatal vs postpartum), inflammatory markers covered, and analytic approaches; (b) specifying how the current review's broader or more recent scope modifies or extends prior conclusions; and (c) highlighting any genuinely novel insights that arise specifically from the additional antenatal or newer studies.
Effort (S/M/L)	S
Priority (P0–P2)	P2
Success Criteria	Discussion includes a clearly labeled subsection contrasting this review with at least the two main prior PPD-focused reviews; a table or structured paragraph delineates differences in time frame, inclusion criteria, markers, and methods; specific novel contributions (e.g., antenatal trajectory patterns, marker-period combinations not previously synthesized) are explicitly stated.

l17, l23, l25

Action	Improve transparency and usability of data presentation by (a) redesigning Tables 1 and 2 to avoid truncated entries and interleaved narrative text, ensuring clear columns for study design, population, period, marker(s), depression construct, effect size/direction, adjustment, and risk-of-bias; (b) clearly labeling which studies are secondary analyses versus primary-aim studies and indicating in the synthesis how they are considered; and (c) clarifying what underlying data are available (e.g., extraction sheets, coded variables) and where, or adjusting claims about "all relevant data" accordingly.
Effort (S/M/L)	M
Priority (P0–P2)	P1
Success Criteria	Revised tables have readable, non-truncated cells and discrete columns that allow readers to reconstruct key features and findings of each study; secondary analyses are explicitly marked (e.g., with a column or superscript) and their influence explicitly discussed in Results or Discussion; the manuscript specifies whether and where detailed extracted datasets or coding sheets are available (repository, supplementary file) or tones down data availability claims if they are not shared.

l18

Action	Clarify the aim statement in the Introduction to unambiguously specify whether both cross-sectional and longitudinal associations are primary targets and to correct the grammatical omission.
Effort (S/M/L)	S
Priority (P0–P2)	P2
Success Criteria	The Introduction contains a single, grammatically correct aim sentence that explicitly states the focus on cross-sectional, longitudinal, or both types of associations between inflammatory markers and perinatal depression, without ambiguity.

G03, G04

Action	Explicitly address risks of reporting bias and overall certainty by (a) describing any methods used to assess publication or reporting bias (e.g., consideration of small-study effects, trial registry comparisons, qualitative assessment of selective outcome reporting) even if formal funnel plots are not feasible; and (b) integrating these considerations into the certainty-of-evidence judgments for the main outcomes.
Effort (S/M/L)	M
Priority (P0–P2)	P1
Success Criteria	Methods include a subsection describing how potential publication/reporting bias was assessed; Results or Discussion explicitly comment on likely direction and magnitude of reporting biases; certainty-of-evidence judgments for each key marker incorporate an assessment of suspected reporting bias (e.g., GRADE domain for publication bias addressed with clear justification).