Journal of Biomedical Sciences and Health
e-ISSN: 3047-7182 | p-ISSN: 3062-6854
Available online at ejournal.unkaha.ac.id/index.php/jbsh

OPEN ACCESS

Vol. 3, No. 2, 2026 DOI: 10.34310/jbsh.v3.i2.357 Pages 148-164

Systematic Review

Oral microbiome signatures for gestational diabetes detection and prediction: a systematic review and evidence map

Fitra Adi Prayogo[1*], Mellyaning Oktaviani Sonya Kirana Sari [2]

[1] *Biomedical Science Study Program, Faculty of Nursing and Health Sciences, Karya Husada University, Semarang, Indonesia*
[2] *Biology Department, Faculty of Sciences and Mathematics, Diponegoro University*

*Correspondence: fitraadi@unkaha.ac.id

ABSTRACT

Background: Gestational diabetes mellitus (GDM) shares inflammatory and microbial pathways with periodontal disease. Whether the maternal oral microbiome could support GDM detection or prediction has not been comprehensively mapped. **Methods:** A systematic review followed PRISMA 2020 guidance across web-based, PubMed/PMC-indexed, and publisher sources plus citation chasing and a free-database supplementary search (ten primary search iterations plus two database-equivalent saturation checks, last updated 30 June 2026). Eligible studies compared pregnant women with and without GDM using a direct oral microbial measurement. Risk of bias was assessed using QUADAS-2, PROBAST, or Newcastle-Ottawa Scale/JBI tools according to design. **Results:** Of 36 records identified, 14 studies (2008-2026, six countries) met the inclusion criteria. Six used community-level 16S rRNA sequencing, seven used targeted pathogen quantification, and one combined cohort profiling with mechanistic and interventional components. Only two studies reported a discrimination model (area under the curve up to 0.89), and only two sampled the oral cavity before diagnosis; the remaining twelve measured it at or after diagnosis or within indeterminate timing windows. Most studies carried unclear or high risk of bias, particularly because confounder adjustment (including pre-pregnancy body mass index, confirmed in one open-access cohort to abolish most raw associations) and temporality were inadequately reported or methodologically problematic. **Conclusion:** Maternal oral dysbiosis is repeatedly but heterogeneously associated with GDM, and the evidence for using it as a detection or prediction tool remains preliminary, heterogeneous, and dominated by cross-sectional, post-diagnosis designs. Prospective, confounder-adjusted studies with externally validated prediction models are needed before clinical translation.

Keywords: oral microbiome; gestational diabetes mellitus; periodontal pathogens; systematic review; risk of bias

Received: 29 July 2026 | **Revised:** 19 August 2026 | **Accepted:** 29 August 2026 | **Published:** 31 August 2026

Cite this article: Prayogo, F. A., & Sari, M. O. S. K. (2026). Oral microbiome signatures for gestational diabetes detection and prediction: a systematic review and evidence map. *Journal of Biomedical Sciences and Health*, *3*(2), 148-164. https://doi.org/10.34310/jbsh.v3.i2.357

1. INTRODUCTION

Gestational diabetes mellitus (GDM) is glucose intolerance first recognized during pregnancy, and it remains one of the most common metabolic complications of gestation. GDM raises the risk of pre-eclampsia, operative delivery, fetal macrosomia, and later maternal and offspring type 2 diabetes, which is why earlier and more reliable detection has long been a research priority in obstetric and endocrine care.

Periodontal disease and diabetes are linked in both directions: hyperglycaemia favours a more pathogenic subgingival community, and periodontal inflammation can in turn worsen glycaemic control. Because pregnancy itself alters oral microbial composition through hormonal and vascular changes, GDM therefore co-occurs with two conditions already known to reshape the oral microbiome independently, which makes isolating a GDM-specific signature both scientifically interesting and methodologically difficult.

Three narrower reviews have previously touched on this literature, none of them a risk-of-bias-stratified evidence map of the maternal oral microbiome specifically. A mini-review by Corrêa and colleagues screened databases through early 2023 and synthesised nine primary studies, concluding that GDM tends to favour periodontitis-associated taxa such as *Prevotella* and *Treponema* at the expense of health-associated genera including *Streptococcus*. A separate systematic review by Camoni and colleagues addressed a related but distinct question (the oral microbiota of infants born to mothers with GDM) rather than the maternal oral microbiome itself. A third, broader narrative review by Bokoliya and colleagues surveyed both gut and oral microbiota as candidate GDM biomarkers, citing the discriminatory capacity of salivary *Streptococcus* and dental-plaque *Leptotrichia*, but without a systematic search protocol, risk-of-bias appraisal, or any oral-specific synthesis. None of the three incorporated three primary studies published in 2026 (including a large mechanistic and interventional study) or a large 2022 population-based cohort that appears to have been missed by prior searches rather than genuinely superseding them.

This review therefore asks a more specific and more clinically oriented question than either predecessor: which maternal oral microbial taxa, community structures, or multivariable signatures are associated with, discriminate, or prospectively predict GDM, and how much confidence does the underlying study design warrant for each claim? We distinguish three tiers of evidence throughout: plain association, cross-sectional discrimination (a model separating already-diagnosed cases from controls), and prospective prediction (a model or exposure measured before diagnosis). We make this distinction because these tiers imply very different levels of readiness for any future clinical application, and because conflating them is, in our reading of the existing narrative reviews, precisely where the literature's optimism has exceeded what the evidence supports.

2. METHODS

2.1 Protocol and Reporting

This review followed the PRISMA 2020 statement for reporting systematic reviews (Page et al., 2021). No prospective protocol was registered; this is declared as a limitation in the Discussion.

2.2 Eligibility Criteria

Studies were eligible if they enrolled pregnant human participants and directly measured the oral microbial community or specific oral pathogens (by culture, targeted PCR/qPCR, microarray, 16S rRNA sequencing, shotgun metagenomics (Quince et al., 2017), or another direct microbial assay) from saliva, supragingival plaque, subgingival plaque, gingival crevicular fluid, tongue or oral swab, or a related oral specimen. Eligible comparators were pregnant women without GDM; factorial studies were eligible when GDM-versus-non-GDM data could be extracted separately from other stratifying factors such as periodontitis status. Eligible designs were prospective cohort, case-control, cross-sectional comparative, nested case-control, and longitudinal observational studies.

We excluded narrative and systematic reviews, studies restricted to neonatal or infant oral microbiota, studies of gut-only or placental-only microbiota without a separable oral

outcome, studies in participants with pre-existing type 1 or type 2 diabetes rather than gestation-onset diabetes, studies without a usable non-GDM comparator, and studies that assessed oral-fluid biomarkers, extracellular vesicles, or clinical periodontal parameters without any direct microbial community or pathogen measurement.

2.3 Information Sources and Search Strategy

Because this review was not conducted through a commercial bibliographic database export, we treat the search as a transparent, reproducible replication rather than a Scopus- or Web of Science-style database count, and we report exact queries and dates for that reason. Twelve search iterations were run between 28 May and 30 June 2026: an initial web/PubMed-indexed search combining oral microbiome and gestational diabetes terms; a PMC-focused search adding periodontal pathogen and specimen-type terms; a publisher and DOI search across BMJ, Springer Nature, Elsevier, Taylor & Francis, Wiley, and Acta Odontologica Scandinavica; a backward citation search from an existing mini-review; five further supplementary searches that specifically targeted the 2022–2026 publication gap and non-PubMed/PMC sources, together with a targeted backward/forward citation search once a large population-based study (Liu et al., 2022) was identified as a likely omission from the initial four searches; a free-database supplementary search of OpenAlex- and Lens.org-indexed content (direct OpenAlex API queries were attempted but blocked by persistent rate-limiting, so this search was instead run through general web search restricted to these platforms' indexed content); and two additional database-equivalent searches on 30 June 2026 using PubMed/MEDLINE (three independent queries: "oral microbiome gestational diabetes", "periodontal bacteria gestational diabetes pregnancy", and "salivary microbiota GDM pregnancy 16S", yielding 109 unique records after deduplication) and Scite-indexed literature ("oral microbiome gestational diabetes mellitus", 30 records screened against full metadata). These two database-equivalent iterations served as post hoc saturation checks against the studies already identified through the preceding ten iterations and are reported separately from the primary PRISMA study-selection flow.

These two final iterations returned no additional primary studies meeting inclusion criteria beyond the 14 already identified, providing support for search saturation across both manually searched and database-indexed sources.

Table S1 (Search Log) reproduces the exact query strings, sources, dates, and yield of every search.

2.4 Study Selection

Records from all searches were pooled and deduplicated first by DOI, then by exact title, then by first author/year/journal. Two screening stages were applied independently by two reviewers: title/abstract screening against the eligibility criteria above, followed by full-text or extended-abstract assessment of records that passed the first stage. Disagreements were resolved by discussion until consensus was reached. Reasons for exclusion at each stage were recorded for every excluded record (Table S2, Excluded and Borderline).

2.5 Data Extraction

For each included study, we extracted first author, year, journal, country, study design, population and sample size, the GDM-versus-comparator definition used, oral specimen type, microbial method, timing of oral sampling relative to GDM diagnosis, the main result relevant to this review, and any reported discrimination or prediction metric (area under the receiver operating characteristic curve, sensitivity, or specificity). Because several 2026 records were extracted from abstracts rather than full text, every extracted field carries an explicit confidence rating (High, Moderate, or Moderate-High) reflecting how much of the entry was directly stated versus inferred, and fields still pending full-text confirmation are flagged as such rather than estimated.

2.6 Risk-of-Bias Assessment

Because the fourteen included studies span diagnostic-accuracy, prediction-model, cohort, and cross-sectional designs, no single tool was applicable to all of them. We applied QUADAS-2 (Whiting et al., 2011) to the one study reporting a diagnostic-accuracy-style discrimination statistic without a combined multivariable model; PROBAST (Wolff et al., 2019) to the one study reporting a multivariable

machine-learning prediction model; the Newcastle-Ottawa Scale (Wells et al., n.d.) to cohort and case-control studies; and the JBI Critical Appraisal Checklist for Analytical Cross-Sectional Studies to the remaining cross-sectional or ambiguously labelled comparative studies, noting explicitly where a study's stated design label did not match how it was actually analysed.

One 2026 study (Gao et al.) was found on closer reading to combine three separable components: a large observational cohort, mechanistic animal and cell experiments, and a registered randomized controlled trial of a topical intervention. It was therefore left as not yet classifiable pending full-text extraction of each component separately, rather than being forced into a single ill-fitting tool. To harmonise domain-level judgements across these four instruments into a single traffic-light summary, we mapped each tool's items to five common domains (selection/representativeness, exposure measurement, confounding/comparability, outcome/reference standard, and temporality/flow-and-timing) as follows: for NOS-assessed studies, the Selection and Comparability stars mapped to Selection and Confounding domains respectively, the Exposure/Outcome stars to Exposure Measurement and Outcome domains; for JBI-assessed studies, items addressing sampling, condition measurement, and confound identification mapped analogously. An overall judgement of "high risk" was assigned when any consequential domain was rated high; "unclear" when no domain was rated high but at least one consequential domain lacked sufficient reporting; and "low risk" only when all domains were rated low, which no study achieved in this review.

Because most full texts were behind institutional paywalls or blocked by automated access challenges, the majority of individual risk-of-bias domains are reported as unclear rather than rated low or high; we treat an unclear rating as an explicit statement that the abstract did not report the relevant information, not as a proxy for low risk. Accordingly, the domain-level ratings for studies assessed without full text should be read as provisional risk-of-bias assessments based on accessible reporting, not as definitive methodological appraisals. Full text was successfully obtained for one study (Crusell et al., 2020, open access) and partially obtained for a second (Gao et al., 2026, early-access metadata only); one further study's recruitment setting was cross-checked against an independent secondary systematic review (Dasanayake et al., 2008, as re-abstracted by Abariga & Whitcomb, 2016) that had itself extracted full-text details. Overall judgements combined domain-level ratings conservatively: a study was never rated low risk overall, while any consequential domain remained unclear.

2.7 Data Synthesis

Given wide heterogeneity in oral specimen type, microbial method, 16S variable region, bioinformatic pipeline, GDM diagnostic criteria, and confounder handling, a meta-analysis was not attempted. Findings are instead synthesised narratively and organised into an evidence map that separates plain association evidence, cross-sectional discrimination evidence, and prospective, prediagnostic evidence, cross-referenced against each study's risk-of-bias rating.

3. RESULTS

3.1 Study Selection

The primary search (ten iterations, 28 May – 30 June 2026, including a final free-database supplementary search of OpenAlex-, Lens.org-, and Europe PMC-indexed content) captured 36 candidate records. After removing seven duplicates, 29 unique records were screened by title and abstract; ten were excluded at this stage (five review articles, two studies in neonatal rather than maternal oral microbiota, one study in pre-existing rather than gestational diabetes, one gut-only study, and one study addressing an outcome other than GDM).

The remaining 19 reports were assessed in full; five were excluded (one lacking a non-GDM comparator, one using Mendelian randomization without direct maternal oral sampling, and three lacking any direct oral microbial index test: one used oral-fluid protein biomarkers, one used clinical periodontal and inflammatory parameters only, and one characterised extracellular vesicles rather than microbial community or pathogen load). Fourteen studies met all eligibility criteria and were included in the evidence map (Figure 1). Two additional database-equivalent iterations

conducted on 30 June 2026 as post hoc saturation checks — PubMed/MEDLINE (three independent queries, 109 unique records after deduplication) and Scite-indexed literature (30 records with full metadata screening) — identified no additional primary study beyond the 14 already included. Because these iterations served to test completeness rather than to discover new candidates, they are reported separately from the PRISMA study-selection flow above. This null result across both manually searched and database-indexed sources is treated as supportive of, though not proof of, search saturation for this specific, narrow topic.

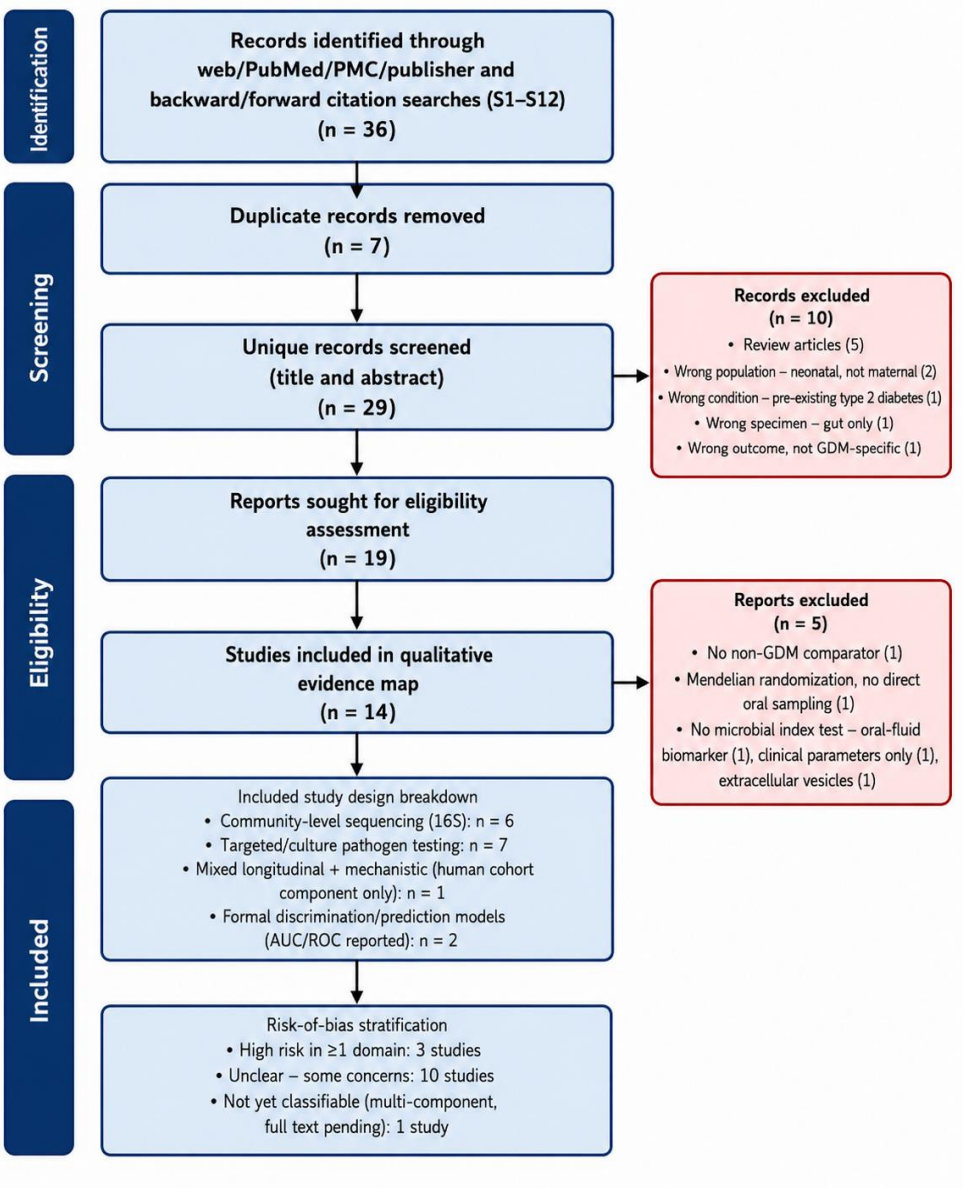

Figure 1. PRISMA 2020 flow diagram of study identification, screening, and inclusion.

3.2 Study Characteristics

The 14 included studies were published between 2008 and 2026 and were conducted in six countries: China (six studies, plus one further study whose country was inferred rather than explicitly stated but is likely also China), Turkey (two studies), India (two studies), and single studies from the United States, Brazil, and Denmark (Table 1). Sample sizes ranged

from 52 total participants in the smallest study (a simple case-control design; Ganiger et al., 2019) to 3,523 in the largest population-based cohort (Liu et al., 2022). Six studies used community-level 16S rRNA sequencing, seven used targeted culture or PCR-based pathogen quantification, and one combined longitudinal human cohort profiling with mechanistic animal/cell experiments and a registered randomized controlled trial. Oral specimens included saliva (six studies), subgingival plaque (six studies, several combined with supragingival plaque), and gingival crevicular fluid (one study).

Table 1. Characteristics of the 14 included studies.

No.	Author, Year	Country	Design	N	Specimen	Method	Sampling timing
1	Dasanayake, 2008	USA	Prospective cohort	265 (22 GDM)	Subgingival plaque	Targeted PCR	~7 wk before diagnosis
2	Gogeneni, 2015	Turkey	Case-control	117 (59 GDM / 58 non-GDM)	Subgingival plaque	Targeted quantification	After diagnosis
3	Wang, 2018	China	Comparative cohort*	149 GDM / 197 non-GDM	Saliva	16S rRNA sequencing	At delivery
4	Cortez, 2019	Brazil	Cross-sectional	26 GDM / 42 non-GDM	Saliva	16S rRNA sequencing	After diagnosis
5	Yao, 2019	China	Matched case-control	59 GDM / 59 matched	Supra+subgingival plaque	Culture-based	2nd trimester/after
6	Ganiger, 2019	India	Case-control	26 GDM / 26 controls	Subgingival plaque	Targeted qPCR	After diagnosis
7	Crusell, 2020	Denmark	Longitudinal cohort	212 (50 GDM)†	Saliva	16S rRNA sequencing	Late pregnancy + postpartum
8	Xu, 2020	China	Case-control	30 GDM / 31 controls	Saliva	16S rRNA sequencing	3rd trimester, after diagnosis
9	Li, 2021	China	Case-control + ML	111 (44 GDM)	Saliva + plaque	16S + SVM/random forest	Around/after diagnosis
10	Zhang, 2021	China	Factorial cross-sectional	69 (4 groups)	Saliva + plaque (both sites)	16S V4 amplicon	2nd trimester, 20–28 wk
11	Cömert, 2026	Turkey	Comparative observational	101 GDM / 98 non-GDM	Subgingival plaque	Targeted assessment	16–36 wk gestation
12	Bandi, 2026	India	Cross-sectional	80 (40 GDM / 40 non-GDM)	Subgingival plaque	7 pathogens, targeted	During pregnancy
13	Gao, 2026	China	Cohort + mechanistic + RCT	>2,500 / 534 profiled / 40 RCT	Oral (site TBC)	Community profiling	Longitudinal
14	Liu, 2022	China (likely)‡	Prospective population cohort	3,523	Gingival crevicular fluid	Targeted (*P. gingivalis*)	12–16 wk, before 24–28 wk OGTT

Wang (2018) is labelled a comparative cohort in its own report, but oral sampling occurred at a single time point (delivery) after GDM status was already known; it is treated here as functionally cross-sectional for the oral-microbiome outcome. †Corrected during data

extraction: 50 GDM, 160 normoglycaemic, and 2 of uncertain status. ‡Inferred from the co-authorship network (a related Chinese-language companion publication); not explicitly stated in the retrieved abstract. §An independent secondary systematic review (Abariga & Whitcomb, 2016) reports n = 262 for this cohort; the primary full text (Dasanayake et al., 2008, PMC2561333) states n = 265. This review follows the primary source.

3.3 Timing of Oral Sampling Relative to Diagnosis

A cross-cutting pattern emerged that bears directly on this review's title question of detection versus prediction. Eight of the 14 studies explicitly state that oral sampling occurred after GDM classification or diagnosis (Gogeneni, 2015; Wang, 2018; Cortez, 2019; Yao, 2019; Ganiger, 2019; Crusell, 2020; Xu, 2020; Li, 2021). A further four studies used sampling windows that overlap with, or are not clearly distinguished from, the standard 24-28 week diagnostic window (Zhang, 2021, 20-28 weeks; Cömert, 2026, 16-36 weeks; Bandi, 2026, "during pregnancy"; Gao, 2026, "longitudinal across pregnancy").

Only two studies – Dasanayake (2008), sampling approximately seven weeks before diagnosis, and Liu (2022), sampling at 12-16 weeks against a 24-28 week oral glucose tolerance test – clearly measured the oral cavity before GDM was known. This means that for the majority of this evidence base, reverse causation (hyperglycaemia altering the oral microbiome, rather than the microbiome preceding and possibly contributing to GDM) cannot be excluded, and that most reported associations describe discrimination of already-diagnosed cases rather than prediction of future onset.

3.4 Narrative Synthesis by Evidence Tier

3.4.1 Community-level sequencing studies

Six studies profiled the broader oral community by 16S rRNA sequencing (Wang, 2018; Cortez, 2019; Crusell, 2020; Xu, 2020; Li, 2021; Zhang, 2021).

Findings were directionally consistent but individually modest: Wang and colleagues reported shifts in *Proteobacteria*, *Firmicutes*, *Neisseria*, and *Leptotrichia* relative abundance between GDM and non-GDM mothers at delivery; Crusell and colleagues, in the only study with open full-text access obtained for this review, found only two differentially abundant operational taxonomic units in the third trimester and eight postpartum, and reported that most raw associations were abolished after adjusting for pre-pregnancy body mass index; Cortez and colleagues found oral-specific differences to be limited or non-significant, an informative negative result rarely emphasised in prior narrative reviews;

Xu and colleagues reported lower oral diversity and a moderate discrimination signal for *Leptotrichia* (area under the curve approximately 0.70, not externally validated); Li and colleagues reported the strongest discrimination in this review, with a machine-learning model combining *Lautropia*, *Neisseria*, and *Streptococcus* features reaching an area under the curve up to 0.83 for microbial features alone and up to 0.89 when combined with clinical variables; and Zhang and colleagues, using a four-group factorial design that separated periodontitis from GDM status, found that GDM alone increased oral Shannon diversity, while periodontitis was independently associated with plaque community structure; because these two outcomes (diversity and community structure) were not measured on a directly comparable scale, this design demonstrates that both factors matter independently rather than establishing which has the larger effect.

3.4.2 Targeted and culture-based pathogen studies

Seven studies quantified specific periodontal pathogens rather than the whole community (Dasanayake, 2008; Gogeneni, 2015; Yao, 2019; Ganiger, 2019; Cömert, 2026; Bandi, 2026; Liu, 2022). The most consistent target across these studies was *Porphyromonas gingivalis*: Ganiger and colleagues found higher *P. gingivalis* and *Prevotella intermedia* burdens in subgingival plaque of women with GDM; Cömert and colleagues, in a 2026 study of subgingival pathobionts at the deepest periodontal pockets, reported a mixed and partly counter-directional signal: *P. gingivalis* did not differ between groups, while *P. intermedia* and *Treponema denticola* were paradoxically lower in the GDM group, even as elevated *T. forsythia* independently raised GDM risk in regression

analysis; and Bandi and colleagues, restricting both comparison arms to women with periodontitis (thereby holding periodontitis constant as a design feature rather than a confounder), still found increased periodontal pathogenic burden associated with GDM, with clinical and glycaemic parameters as independent determinants.

Dasanayake and colleagues, the earliest and one of only two prediagnostic studies in this review, did not find plaque pathogen measures to differ clearly between women who later developed GDM and those who did not, with other inflammatory and clinical factors instead predicting GDM in that cohort; this result is notable because it complicates a straightforward "more pathogens, more GDM" narrative. Liu and colleagues, the largest study in this review by a wide margin (3,523 women), found GDM incidence significantly higher among women with untreated periodontitis (11.21%) than without periodontitis (4.79%), with periodontal treatment during gestation associated with a lower incidence (7.32%), and *P. gingivalis* presence in gingival crevicular fluid positively correlated with serum tumour necrosis factor-alpha and interleukin-8.

3.4.3 Mixed mechanistic and interventional evidence

One 2026 study (Gao et al.) sits apart from the other 13 in scope and cannot be reduced to a single design label. Its human observational component longitudinally profiled 534 pregnant women within a broader cohort infrastructure exceeding 2,500 volunteers and reported a progressive shift from a *Streptococcus*-dominated to a *Prevotella*/*Porphyromonas*-enriched oral community associated with GDM. Its mechanistic component, in mouse and cellular models, linked this dysbiotic shift to periodontal inflammation, systemic interleukin-17 and interleukin-1beta responses, and suppression of glucagon-like peptide-1 and insulin.

Its interventional component was a registered, double-blind randomized controlled trial (ChiCTR2400080741) of topical gingival docosahexaenoic acid in 40 pregnant women with GDM, reporting a smaller rise in fasting glucose with active treatment than placebo (median change 0.10 versus 0.27 mmol/L) alongside improved probing depth. Because the human cohort, animal/cell, and trial components require three different risk-of-bias tools and full methodological detail for the observational component was not available in the early-access article page retrieved for this review, this study is presented narratively rather than folded into the risk-of-bias-stratified counts above; it is nonetheless the only included study to test an intervention on the oral microbiome-GDM pathway rather than only observing it.

3.5 Risk of Bias

Table 2 summarises the risk-of-bias tool applied and overall judgement for each study; full domain-level judgements and their evidentiary basis are provided in Supplementary Table S3. Three studies carried a high risk of bias in at least one consequential domain. Crusell (2020), once full text was obtained, was found to have recruited both the GDM and comparator groups only because they already carried a GDM risk factor and had been referred for an oral glucose tolerance test on that basis (a limitation the original authors themselves state explicitly), alongside concurrent (not prediagnostic) sampling; the same full text also showed formal body-mass-index adjustment and an explicit, favourable check that postpartum follow-up attrition was not differential between groups.

Xu (2020) carried a high risk of bias from case-control patient selection and from flow-and-timing concerns under QUADAS-2, since its reported discrimination reflects concurrent separation of already-diagnosed cases rather than prospective prediction. Li (2021) carried a high risk of bias under PROBAST from case-control participant selection, non-blinded predictor timing, and a sample size (n=111) that is small relative to the complexity of the multivariable machine-learning model being fitted, with no confirmed internal or external validation. The remaining ten classifiable studies were rated unclear with some concerns, reflecting abstract-level reporting that did not state confounder adjustment, blinding, or precise diagnostic criteria, rather than confirmed methodological soundness. One study (Gao et al., 2026) remains not yet classifiable pending full-text extraction of its three components.

Table 2. Risk-of-bias tool and overall judgement for each included study.

No.	Study	Tool applied	Overall judgement

1	Dasanayake, 2008	NOS-Cohort	Unclear – some concerns
2	Gogeneni, 2015	NOS-Case-control	Unclear – some concerns
3	Wang, 2018	JBI Cross-sectional*	Unclear – some concerns
4	Cortez, 2019	JBI Cross-sectional	Unclear – some concerns
5	Yao, 2019	NOS-Case-control	Unclear – some concerns
6	Ganiger, 2019	NOS-Case-control	Unclear – some concerns
7	Crusell, 2020	NOS-Cohort	High risk in ≥2 domains (full-text confirmed)
8	Xu, 2020	QUADAS-2	High risk in ≥1 domain
9	Li, 2021	PROBAST	High risk in ≥1 domain
10	Zhang, 2021	JBI Cross-sectional	Unclear – some concerns
11	Cömert, 2026	JBI Cross-sectional*	Unclear – some concerns
12	Bandi, 2026	JBI Cross-sectional	Unclear – some concerns
13	Gao, 2026	Pending (3 components)	Not assessable – structure confirmed, methods pending
14	Liu, 2022	NOS-Cohort	Unclear – some concerns

Figure 2 presents the domain-level risk-of-bias summary as a traffic-light plot. Across the 14 studies, the exposure-measurement domain was rated low risk in most cases (11 of 14), reflecting the use of objective laboratory-based assays; two studies (Xu, 2020; Li, 2021) were rated unclear in this domain owing to insufficient reporting of assay standardisation, and one (Gao, 2026) was not assessable pending full-text extraction. By contrast, the confounding/comparability and outcome/reference-standard domains were predominantly rated as unclear, confirming that abstract-level reporting did not provide sufficient detail to determine whether key confounders were adjusted for. The three studies rated high overall (Crusell, 2020; Xu, 2020; Li, 2021) each carried high risk in at least one of two critical domains: selection/representativeness or temporality/flow-and-timing, both of which directly affect the interpretability of any reported association or discrimination statistic.

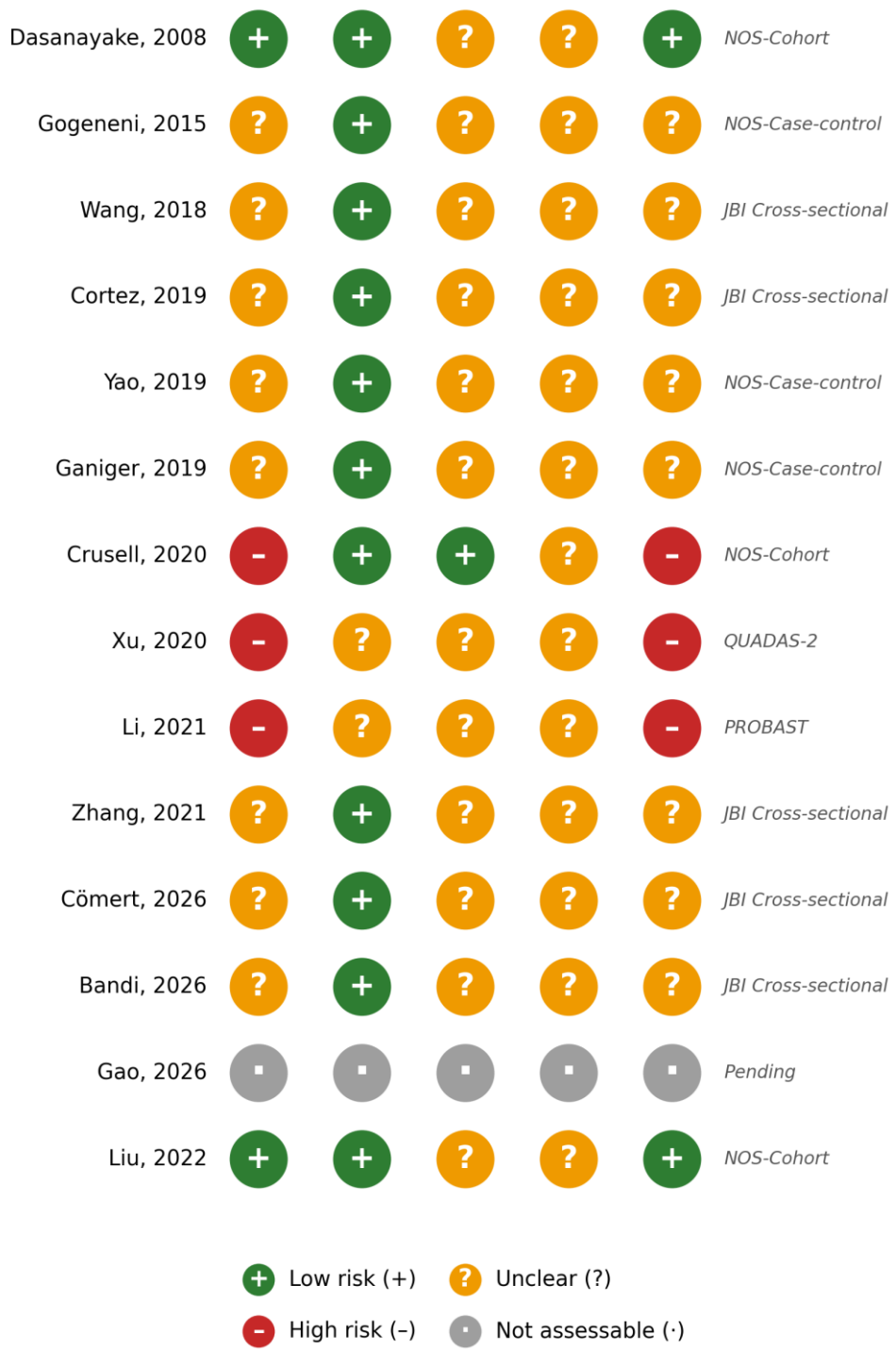

Figure 2. Risk-of-bias summary: domain-level judgements for each included study (n = 14). Green (+) = low risk; yellow (?) = unclear; red (–) = high risk; grey (·) = not assessable. The risk-of-bias tool applied to each study is indicated on the right.

3.6 Evidence Map

Table 3 consolidates the preceding synthesis into a single evidence map, classifying each study by evidence tier (prospective/prediagnostic association, cross-sectional association, cross-sectional discrimination model, or mixed mechanistic/interventional) alongside its key oral signal and risk-of-bias judgement. Of the 14 studies, two provide prospective, prediagnostic association evidence; nine provide cross-sectional or concurrent association evidence; two provide a cross-sectional discrimination model; and one provides mixed mechanistic and interventional evidence not yet fully

classifiable. No study occupies the tier this review's title most directly invokes (a prospective, prediagnostic discrimination or prediction model with reported accuracy), which is itself the clearest single finding of this evidence map.

Table 3. Evidence map: tier of evidence, key oral signal, and risk-of-bias judgement for each included study.

No.	Study	Evidence tier	Key oral signal	Risk-of-bias judgement
1	Dasanayake, 2008	Prospective, prediagnostic association	No clear pathogen-level difference; other clinical/inflammatory factors predicted GDM	Unclear – some concerns
2	Gogeneni, 2015	Cross-sectional association (post-diagnosis)	Increased key periodontal pathogen infection with GDM + gingivitis	Unclear – some concerns
3	Wang, 2018	Cross-sectional association (post-diagnosis)	*Proteobacteria*/*Firmicutes*/*Neisseria*/*Leptotrichia* shifts	Unclear – some concerns
4	Cortez, 2019	Cross-sectional association (post-diagnosis, null)	No significant oral-specific difference (negative finding)	Unclear – some concerns
5	Yao, 2019	Cross-sectional association (post-diagnosis)	Higher periodontal indices; altered anaerobe detection	Unclear – some concerns
6	Ganiger, 2019	Cross-sectional association (post-diagnosis)	Higher *P. gingivalis* and *P. intermedia* burden	Unclear – some concerns
7	Crusell, 2020	Longitudinal association (concurrent + postpartum)	Minor dysbiosis; BMI abolished most raw associations	High risk in ≥2 domains
8	Xu, 2020	Cross-sectional discrimination model	*Leptotrichia*; AUC ≈ 0.70, not externally validated	High risk in ≥1 domain
9	Li, 2021	Cross-sectional discrimination model	*Lautropia*/*Neisseria*/*Streptococcus*; AUC up to 0.89	High risk in ≥1 domain
10	Zhang, 2021	Cross-sectional association (factorial)	GDM raised diversity; periodontitis affected structure	Unclear – some concerns
11	Cömert, 2026	Cross-sectional association (post-diagnosis)	Mixed/counter-directional pathobiont shift; *T. forsythia* raised GDM risk	Unclear – some concerns
12	Bandi, 2026	Cross-sectional association (periodontitis-controlled)	Increased pathogenic burden independent of periodontitis	Unclear – some concerns
13	Gao, 2026	Mixed: association + mechanistic + RCT	*Streptococcus*→*Prevotella*/*Porphyromonas* shift; DHA trial benefit	Not assessable (pending)
14	Liu, 2022	Prospective, prediagnostic association	*P. gingivalis* and untreated periodontitis raised GDM incidence	Unclear – some concerns

Figure 3 visualises this evidence map as a bubble chart, plotting each study by the timing of oral sampling (horizontal axis) against its evidence tier (vertical axis), with bubble size proportional to sample size and colour indicating risk-of-bias judgement. The most conspicuous feature of this plot is the empty upper-left quadrant: no study occupies the prediction-model tier with prediagnostic timing, which is the evidence class most

directly relevant to this review's title question. The two studies reporting discrimination models (Xu, 2020; Li, 2021) cluster in the post-diagnosis column and are marked as high risk of bias, while the two prediagnostic studies (Dasanayake, 2008; Liu, 2022) remain at the plain-association tier. Liu (2022), as the largest study by a wide margin (n = 3,523), stands out visually, yet it too does not report an AUC-based model. This spatial separation between discrimination strength and temporal suitability is the central gap identified by this evidence map.

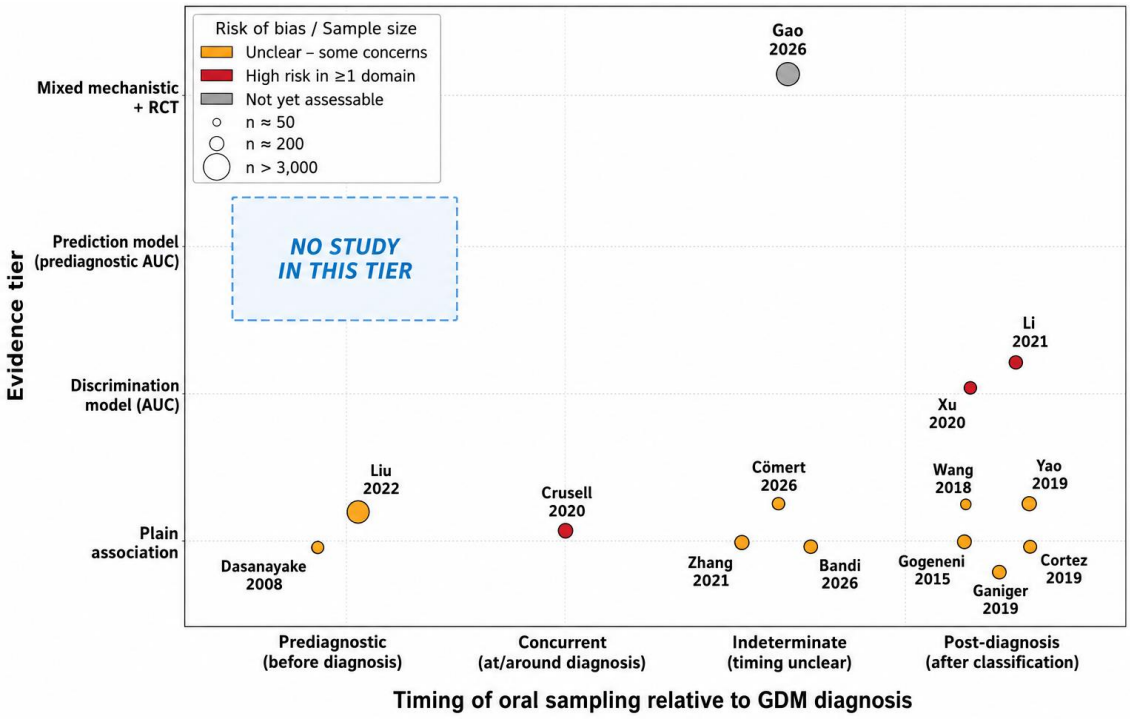

Figure 3. Evidence map: timing of oral sampling versus evidence tier for each included study (n = 14). Bubble size is proportional to total sample size; colour indicates overall risk-of-bias judgement (yellow = unclear, red = high risk, grey = not yet assessable). The dashed box highlights the prediction-model tier with prediagnostic timing, which no study occupies.

4. DISCUSSION

4.1 Summary of Findings

This review identified 14 studies directly measuring the maternal oral microbiome in relation to GDM, roughly evenly split between community-level sequencing and targeted pathogen quantification, plus one large mixed observational-mechanistic-interventional study. Across methods, GDM was repeatedly associated with a shift away from health-associated taxa such as *Streptococcus* and *Neisseria* and toward periodontitis-associated taxa and pathobionts, most consistently *Porphyromonas gingivalis* and *Prevotella* species; this direction was not, however, universal; Cömert and colleagues (2026) found two of three assessed pathobionts paradoxically lower in the GDM group, a partial exception this review treats as a genuine signal of biological heterogeneity rather than as noise to be averaged away. This broader pattern is directionally consistent with the two prior narrower reviews in this area (Corrêa et al., 2023; Camoni et al., 2024); the three 2026 studies and one 2022 population-based cohort identified through this review's supplementary search extend rather than overturn that pattern. However, consistency of direction should not be mistaken for consistency of magnitude or for readiness to support detection or prediction claims; distinguishing those is the main purpose of the risk-of-bias stratification presented below.

4.2 Strength of Evidence for Detection and Prediction Claims

Only two of the 14 studies reported a formal discrimination statistic. Xu and colleagues found oral *Leptotrichia* abundance discriminated GDM from control status with an area under the curve of approximately 0.70, and Li and colleagues reported a machine-learning model reaching an area under the curve of up to 0.89 when microbial and clinical features were combined. Both, however, measured the oral microbiome after GDM was already diagnosed and used case-control sampling, which QUADAS-2 and PROBAST both flag as a structural source of inflated accuracy; neither reported external validation. The two studies that did sample before diagnosis (Dasanayake and colleagues in 2008 and Liu and colleagues in 2022) did not report a discrimination statistic at all. Dasanayake's plaque pathogen measures were, if anything, unremarkable predictors relative to other clinical and inflammatory factors in that cohort, a finding that complicates any assumption that the oral microbiome will prove a straightforward early biomarker. Liu's very large cohort showed a clear association between periodontitis, *P. gingivalis*, and later GDM incidence, but as an incidence comparison across periodontitis strata rather than as a combined predictive model with a reportable accuracy statistic.

In summary, the studies with the strongest discrimination statistics are precisely the studies least suited, by timing, to support a prediction claim, while the studies best suited by timing to support prediction do not yet report a discrimination statistic. The distinction between detection and prediction implied by this review's title is therefore accurate only if it is kept visible throughout any application of these findings, rather than treated as interchangeable.

4.3 Confounding by Pre-Pregnancy Body Mass Index

Full-text access was obtained for only one included study during this review (Crusell et al., 2020, open access), and what it showed is instructive beyond that single study. Pre-pregnancy body mass index, once adjusted for, abolished most of the raw salivary microbiome associations with GDM status that the same authors had found before adjustment, leaving BMI – not glycaemic status as such – as the stronger correlate of community composition. Obesity and BMI are themselves associated with both GDM risk and, independently, with oral microbiome composition, so this is not a minor technical caveat. None of the other 13 included studies reported confounder-adjustment status with enough clarity in their abstracts for us to confirm whether BMI, smoking, periodontal treatment history, or oral hygiene practices were accounted for; most risk-of-bias domains concerning confounding were therefore rated unclear rather than low, which should be read as an honest gap in the literature rather than an assumption of good practice.

Two studies did structurally address one specific confounder, periodontitis itself, by design: Zhang and colleagues used a four-group factorial design separating periodontitis from GDM status, and Bandi and colleagues restricted both comparison arms to women with periodontitis. Both approaches are more informative than a simple two-group comparison, and both deserve more imitation in future study designs than they have so far received.

4.4 Heterogeneity and Rationale for Narrative Synthesis

The included studies used at least four different specimen types (saliva, supragingival plaque, subgingival plaque, gingival crevicular fluid), several different 16S rRNA variable regions and primer sets where sequencing was used, and inconsistent reporting of GDM diagnostic criteria; only two studies (Crusell et al., 2020; Liu et al., 2022) explicitly named their diagnostic standard (IADPSG criteria, International Association of Diabetes and Pregnancy Study Groups Consensus Panel, 2010, and oral glucose tolerance test respectively) in the material available to us. Sample sizes ranged across two orders of magnitude, from 52 to 3,523 participants. Pooling effect estimates across such heterogeneous exposure definitions, outcome ascertainment methods, and population risk profiles would have produced a number without a clear referent, which is why this review presents a stratified narrative synthesis and evidence map instead.

4.5 Strengths and Limitations of This Review

This review's search extended beyond a single database replication, incorporated backward and forward citation chasing, and specifically targeted the 2022-2026 period once an initial four-search protocol was judged, on reflection, to have under-sampled recent Chinese-language-adjacent and population-based literature; this targeted extension is what surfaced the largest and most prediagnostically timed study in the review (Liu et al., 2022). Risk of bias was assessed using design-appropriate tools rather than a single instrument applied uniformly, an approach we adopted here to match the heterogeneous study designs in this evidence base, and every domain-level judgement is traceable to a specific piece of extracted text or an explicit statement that no such text was available.

Equally, several limitations should temper how this review is read, and one is consequential enough to flag as a priority. First, no protocol was registered in advance (e.g., PROSPERO). Second, while the initial ten search iterations did not query Scopus, Web of Science, or Embase through institutional subscription exports, two additional database-equivalent iterations were conducted on 30 June 2026: PubMed/MEDLINE was searched with three independent query strategies (109 unique records after deduplication) and Scite-indexed literature was searched with oral microbiome and gestational diabetes terms (30 records with full metadata screening). Neither iteration identified any additional primary study beyond the 14 already included. This convergence across manually searched, free-database, and PubMed/Scite-indexed sources is supportive of search saturation for this specific, narrow topic, although direct Scopus and Embase institutional exports would further strengthen reproducibility.

Beyond these two priority items, full text could be obtained for only one of the 14 included studies during the conduct of this review, with a second study's structure (but not its full methods) clarified from an early-access journal page and a third study's recruitment setting cross-checked against an independent secondary review; the remaining domain-level risk-of-bias judgements rest on abstract-level reporting and are explicitly marked unclear rather than assumed favourable.

One included study (Gao et al., 2026) combines an observational cohort, mechanistic experiments, and a registered trial that could not be fully disentangled without its complete methods section, and is discussed narratively rather than risk-of-bias-rated for that reason. The ratings and synthesis presented here should be treated as a rigorous first pass whose risk-of-bias ratings merit confirmation against full text and whose search coverage would benefit from direct institutional Scopus and Embase exports in any future update. This manuscript flags these outstanding steps explicitly rather than obscuring them.

4.6 Implications for Research and Practice

For clinical practice, no oral microbial signature in this literature is ready to inform GDM screening or risk stratification: the two studies with the strongest discrimination statistics have the weakest temporal design for a prediction claim, and the two studies with genuinely prediagnostic timing do not yet report a discrimination statistic. For research, three priorities follow directly from the gaps mapped here. First, future studies should sample the oral cavity in the first or early second trimester, before GDM is diagnosed, and should report a discrimination or prediction statistic against that same prediagnostic exposure, combining what Dasanayake-and Liu-style timing did right with what Xu- and Li-style analysis did right. Second, pre-pregnancy BMI and periodontal status should be measured and adjusted for, given the confounding demonstrated in the one study where this was possible to check in full text. Third, any prediction model, once developed, needs external validation in an independent cohort before an area-under-the-curve figure is presented as evidence of clinical utility rather than of internal model fit.

5. CONCLUSION

Maternal oral dysbiosis, particularly enrichment of periodontitis-associated taxa such as Porphyromonas gingivalis and Prevotella species alongside depletion of Streptococcus and Neisseria, was reported across several of the 14 studies reviewed here, though not universally (Cömert and colleagues found a partly counter-directional signal), making it a recurring but heterogeneous finding across studies of gestational diabetes published between 2008 and

2026. This directional pattern, however, coexists with considerable heterogeneity in method, timing, and confounder handling, and only two studies offer a formal discrimination statistic, neither of them prediagnostic and neither externally validated. The oral microbiome remains a biologically plausible and non-invasively sampled candidate for future GDM detection or prediction research, but the current evidence base does not yet support that use case in practice. Prospective, early-pregnancy, confounder-adjusted studies with externally validated prediction models are the clearest path from the associational evidence synthesised here to any genuine clinical application.

6. DECLARATIONS

Ethics approval and consent to participate

Not applicable. This is a systematic review of previously published, de-identified data.

Conflict of interest

The authors declare no conflict of interest.

Funding

No external funding was received for this review.

Author contributions

F.A.P. conceived and designed the review, developed the search strategy, conducted the literature searches across all iterations, performed title/abstract and full-text screening, extracted data, carried out risk-of-bias assessment, constructed the evidence map and figures, and drafted and revised the manuscript. M.O.S.K.S. independently screened titles/abstracts and full texts at both stages, verified data extraction and risk-of-bias judgements, contributed to the interpretation of findings, and critically reviewed and approved the final manuscript.

REFERENCES

Abariga, S. A., & Whitcomb, B. W. (2016). Periodontitis and gestational diabetes mellitus: A systematic review and meta-analysis of observational studies. BMC Pregnancy and Childbirth, 16, 344. https://doi.org/10.1186/s12884-016-1145-z

Bandi, D. P., Krishnaswamy, B., & Victor, D. J. (2026). Differential prevalence of periodontal pathogens in pregnant women with gestational diabetes mellitus and periodontitis: A cross-sectional microbiological study. Journal of Oral Biology and Craniofacial Research, 16(4), 101479. https://doi.org/10.1016/j.jobcr.2026.101479

Bokoliya, S., McClellan, S., Zhou, Y., & Fan, N. (2024). Exploring the influence of microbiota on gestational diabetes and its potential as a biomarker. Frontiers in Bacteriology, 3, 1352227. https://doi.org/10.3389/fbrio.2024.1352227

Camoni, N., Conti, G., Majorana, A., Bardellini, E., Salerno, C., Wolf, T. G., Campus, G., & Cagetti, M. G. (2024). Oral microbiota of infants in maternal gestational diabetes: A systematic review. Children, 11(4), 421. https://doi.org/10.3390/children11040421

Cömert, F., Yalçın, F., & Topcuoğlu, N. (2026). Gestational diabetes alters subgingival pathobiont composition. Acta Odontologica Scandinavica, 85, 179–189. https://doi.org/10.2340/aos.v85.45789

Corrêa, J. D., Faria, G. A., & Fernandes, L. L. (2023). The oral microbiota and gestational diabetes mellitus. Frontiers in Clinical Diabetes and Healthcare, 4, 1120920. https://doi.org/10.3389/fcdhc.2023.1120920

Cortez, R. V., Taddei, C. R., Sparvoli, L. G., Ângelo, A. G. S., Padilha, M., Mattar, R., & Daher, S. (2019). Microbiome and its relation to gestational diabetes. Endocrine, 64(2), 254–264. https://doi.org/10.1007/s12020-018-1813-z

Crusell, M. K. W., Brink, L. R., Nielsen, T., Allin, K. H., Hansen, T., Damm, P., Lauenborg, J., Hansen, T. H., & Pedersen, O. (2020). Gestational diabetes and the human salivary microbiota: A longitudinal study during pregnancy and postpartum. BMC Pregnancy and

Childbirth, 20, 69. https://doi.org/10.1186/s12884-020-2764-y

Dasanayake, A. P., Chhun, N., Tanner, A. C. R., Craig, R. G., Lee, M. J., Moore, A. F., & Norman, R. G. (2008). Periodontal pathogens and gestational diabetes mellitus. Journal of Dental Research, 87(4), 328–333. https://doi.org/10.1177/154405910808700421

Gao, S., Yin, N., Wei, R., Li, X., Cheng, Q., Zhula, A., Zhou, W., Zhang, Y., Li, S., Zhou, W., Wang, X., Zhang, R., Wang, Q., Fan, H., Peng, S., Zhang, H., Li, K., Hu, Y., Gao, Y., Shi, W., … Wang, J. (2026). Oral microbiome modulation mitigates hyperglycemia exacerbation in gestational diabetes mellitus. Nature Communications. https://doi.org/10.1038/s41467-026-74917-w

Ganiger, K., Sridharan, S., Rahul, A., & Satyanarayana, A. (2019). Quantitative analysis of key periodontopathic bacteria in gestational diabetic and non-diabetic women. Journal of Diabetes & Metabolic Disorders, 18(2), 363–369. https://doi.org/10.1007/s40200-019-00420-3

Gogeneni, H., Buduneli, N., Ceyhan-Öztürk, B., Gümüş, P., Akcali, A., Zeller, I., Renaud, D. E., Scott, D. A., & Özçaka, Ö. (2015). Increased infection with key periodontal pathogens during gestational diabetes mellitus. Journal of Clinical Periodontology, 42(6), 506–512. https://doi.org/10.1111/jcpe.12418

International Association of Diabetes and Pregnancy Study Groups Consensus Panel. (2010). International Association of Diabetes and Pregnancy Study Groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care, 33(3), 676–682. https://doi.org/10.2337/dc09-1848

Li, X., Zheng, J., Ma, X., Zhang, B., Zhang, J., Wang, W., et al. (2021). The oral microbiome of pregnant women facilitates gestational diabetes discrimination. Journal of Genetics and Genomics, 48(1), 32–39. https://doi.org/10.1016/j.jgg.2020.11.006

Liu, F., Sui, W., Zhou, Z. F., Mi, Y., He, T. Q., Li, Z. B., et al. (2022). Development of gestational diabetes mellitus in women with periodontitis in early pregnancy: A population-based clinical study. Journal of Clinical Periodontology, 49(2), 164–176. https://doi.org/10.1111/jcpe.13578

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. Nature Biotechnology, 35(9), 833–844. https://doi.org/10.1038/nbt.3935

Wang, J., Zheng, J., Shi, W., Du, N., Xu, X., Zhang, Y., et al. (2018). Dysbiosis of maternal and neonatal microbiota associated with gestational diabetes mellitus. Gut, 67(9), 1614–1625. https://doi.org/10.1136/gutjnl-2018-315988

Wells, G. A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (n.d.). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa Hospital Research Institute. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., & Bossuyt, P. M. M. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine, 155(8), 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Annals of Internal Medicine, 170(1), 51–58. https://doi.org/10.7326/M18-1376

Xu, Y., Zhang, M., Zhang, J., Sun, Z., Ran, L., Ban, Y., Wang, B., Hou, X., Zhai, S., Ren, L., Wang, M., & Hu, J. (2020). Differential intestinal and oral microbiota features associated with gestational diabetes and maternal inflammation. American Journal of Physiology-Endocrinology and Metabolism, 319(2), E247–E253. https://doi.org/10.1152/ajpendo.00266.2019

Yao, H., Xu, D., Zhu, Z., & Wang, G. (2019). Gestational diabetes mellitus increases the detection rate and the number of oral bacteria in pregnant women. Medicine, 98(11), e14903. https://doi.org/10.1097/MD.0000000000014903

Zhang, X., Wang, P., Ma, L., Guo, R., Zhang, Y., Wang, P., Zhao, J., & Liu, J. (2021). Differences in the oral and intestinal microbiotas in pregnant women varying in periodontitis and gestational diabetes mellitus conditions. Journal of Oral Microbiology, 13(1), 1883382. https://doi.org/10.1080/20002297.2021.1883382