INPLASY

Vitamin D Status and Oral and Oropharyngeal Neoplasia: A Systematic Review and Meta-Analysis

INPLASY2026110111
doi: 10.37766/inplasy2026.11.0111
Received: 27 November 2026
Published: 27 November 2026

Corresponding author:
YaoHui Yu

148267943@qq.com

Author Affiliation:
School and Hospital of Stomatology, Wenzhou Medical University.

Yaohui , Y; Yu, Z.

ADMINISTRATIVE INFORMATION

Support - The Zhejiang Provincial Health Industry Science and Technology Plan Project (Grant No. 2026772367).

Review stage at the time of initial registration - Completed but not published.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY2026110111

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 27 November 2026 and was last updated on 27 November 2026.

INTRODUCTION

Review question / Objective To quantify the association between lower vitamin D exposure and oral or oropharyngeal lesions and to test whether the disagreement between studies follows from prespecified lesion strata.

Rationale Oral potentially malignant disorders (OPMDs) precede many oral squamous cell carcinomas and are the stage at which prevention is still feasible. Vitamin D signalling helps maintain epithelial differentiation, but epidemiological findings have not converged; neither of the two earlier reviews pooled estimates.

Condition being studied Oral squamous cell carcinoma (OSCC) arises from the stratified squamous epithelium of the lip, tongue, floor of mouth, buccal mucosa, gingiva, and palate, and is the most common malignant tumour of the oral cavity.2 Early disease is often silent: a sore that does not heal, a red or white patch, a lump, loose teeth, or pain on swallowing may be attributed to trauma, dentures, or infection, so diagnosis is frequently delayed.2 Treatment centres on surgical resection with neck dissection, followed by radiotherapy or chemoradiotherapy according to stage and pathological risk features. Because the tongue, floor of mouth, and mandible govern speech, chewing, and swallowing, both the disease and its treatment can leave lasting problems with eating, speaking, and appearance, and recurrence or a second primary tumour remains common after apparently successful therapy.2

A recognised precursor stage precedes many of these cancers. The WHO Collaborating Centre consensus uses oral potentially malignant disorder (OPMD) for a heterogeneous group of mucosal conditions that carry an increased risk of malignant transformation.6 Oral leukoplakia, a white plaque that cannot be wiped off or explained by another definable condition, is the most common; erythroplakia, oral lichen planus, oral submucous fibrosis, and mixed red-and-white lesions are also

Yaohui et al. INPLASY protocol 2026110111. doi:10.37766/inplasy2026.11.0111 Downloaded from https://inplasy.com/inplasy-2026-11-0111/

included.6 Transformation risk is not uniform. It rises with epithelial dysplasia graded on biopsy, with non-homogeneous or erythroplakic appearances, and with lesion site, yet many lesions remain stable for years and some regress, which makes risk estimation for an individual patient difficult.6 Because OPMDs are visible, accessible to biopsy, and usually managed by surveillance, this stage is where a chemopreventive intervention could be tested before invasive disease develops.5,16

The distinction matters here for a practical reason. The studies included in this review drew case groups from different points along this sequence: some enrolled patients with established oral or oropharyngeal cancer, others recruited patients with oral lichen planus or leukoplakia, and at least one combined invasive and premalignant lesions in a single comparison.9,12-14 An odds ratio from a premalignant case group and one from an invasive cancer case group therefore describe different health conditions at different stages, even when both are reported under the heading of oral neoplasia.6.

METHODS

Search strategy PubMed, Embase, and Web of Science.

Participant or population Three were European, 3 Asian, and 1 South American. The 1,759 case participants and 3,368 controls include laryngeal and pharyngeal cases from the 2 cohort studies, and 3 studies contributed.

Intervention Reviewer screened, extracted, and appraised studies and checked each value against its source in a separate pass; no second independent reviewer was available. Log ORs were pooled under a DerSimonian–Laird random-effects model with prespecified stratification by lesion type, region, design, and deficiency definition. Analyses added later are labelled post hoc exploratory and reported separately from the primary analysis.

Comparator Primary: the pooled OR for any oral/oropharyngeal lesion. Secondary: stratum-specific estimates, heterogeneity statistics, and Grading of Recommendations Assessment, Development, and Evaluation (GRADE) certainty.

Study designs to be included I^2, τ^2, and Cochran's Q assess heterogeneity; Egger's regression and Begg's rank correlation test small sample effects.

Eligibility criteria Records excluded: Irrelevant topic, non-clinical studies, non-original research, conference abstracts, and mismatched interventions/outcomes; Effect sizes unavailable/uncalculable, Ineligible study type, no control group or mismatched exposure/control.

Information sources Pubmed:130; EMBASE:436; Web of science:60.

Main outcome(s) Seven studies contributed 10 estimates. The pooled OR was 1.26 (95% CI, 0.95–1.68; I^2 = 51.7%; P for heterogeneity = .03). Stratification by lesion type accounted for the heterogeneity (P = .04): premalignant/mixed lesions, 1.51 (95% CI, 1.08–2.09; I^2 = 0%); oral cancer, 1.39 (0.62–3.12); oropharyngeal cancer, 0.90 (0.70–1.16). Region (P = .34), design, and threshold (both P = .71) showed no subgroup differences; the last two were collinear. Funnel-plot tests were descriptive only. GRADE certainty ranged from very low to low; post hoc analyses were excluded from inference.

Quality assessment / Risk of bias analysis Quality assessment was conducted at three separate levels, each with its own unit and instrument, so that no rating is inferred from another. At study level, the six case–control and nested case–control studies were appraised with the Newcastle–Ottawa Scale across selection, comparability, and exposure (maximum 9; ≥7 denotes high quality), and the single cross-sectional study with the 11-item Agency for Healthcare Research and Quality checklist (≥8 of 11 denotes high quality). Scores ranged from 7 to 9 (median 8.5), with 9 for three studies, and the cross-sectional study met 8 of 11 criteria. A high study-level score does not imply low outcome-level risk of bias: four of the seven studies adjusted for no confounder at all, including tobacco, alcohol, and betel quid, and none standardised sunlight exposure or season of blood draw completely. At outcome level, certainty was judged with GRADE across five domains. Because every included design is observational, all outcomes started at low certainty and were downgraded for serious study limitations (four unadjusted studies, blood usually collected after diagnosis), serious inconsistency ($I^2 \geq 50\%$, $\tau^2 = 0.083$, effect directions crossing the null), serious indirectness (dietary intake alongside serum assays, deficiency thresholds from 25 to approximately 87 nmol/L, invasive carcinoma combined with premalignant lesions), and serious imprecision (a confidence interval crossing the null). No outcome met upgrading criteria; final ratings were very low for the overall estimate and for the oral cancer and

Yaohui et al. INPLASY protocol 2026110111. doi:10.37766/inplasy2026.11.0111 Downloaded from https://inplasy.com/inplasy-2026-11-0111/

other/precancerous strata, and low for the oropharyngeal stratum. At review level, AMSTAR-2 was kept distinct from both. Selection, extraction, and appraisal were performed by one reviewer, with a separate pass that verified every value against the source PDF. No second independent assessor was available, so duplicate screening, inter-rater agreement (Cohen κ or PABAK), and a domain-level ROBINS-I V2 assessment cannot be reported; the aggregate publications lacked the individual-level data needed to reconstruct a target-trial contrast and domain confounding judgments. These fields remain blank, the limitation is disclosed in the cover letter, and its consequence is carried into the GRADE study-limitations downgrade instead of being given a numeric score. Funding sources and conflicts of interest of the primary studies were not systematically collected.

Strategy of data synthesis Data synthesis uses the log odds ratio as the effect measure, with seven observational studies contributing ten estimates from 1,759 cases and 3,368 controls. The effect estimate, not the study, is the unit of analysis, because distinct subsites or lesions within one study address distinct outcomes. The most fully adjusted published estimate takes precedence over crude ones. Standard errors were derived from confidence limits as [ln(upper 95% CI) − ln(lower 95% CI)]/(2 × 1.96); category frequencies yielded unconditional odds ratios from 2 × 2 tables, and published prevalence proportions yielded reconstructed odds ratios. Two per-doubling odds ratios from the EPIC study were inverted under log-linearity so that higher odds correspond to lower vitamin D status; the dietary estimate of 0.83 (0.67–1.04) per 1-SD higher intake is retained in the primary analysis as published, and its direction is examined only in post hoc work. Every derivation is logged cell by cell. Log odds ratios were pooled with a DerSimonian–Laird random-effects model, with a fixed-effect model fitted for comparison; the DerSimonian–Laird τ^2 estimator reproduces the original R analysis and the supplied forest plots. Heterogeneity is summarised by I^2 and Cochran Q ($P < .10$), read alongside τ^2 and the spread and direction of individual estimates, and subgroup differences use the Cochran-type Q test on stratum summaries. Because only seven independent studies are available, small-sample guidance led to Hartung–Knapp–Sidik–Jonkman intervals, τ^2 from the Paule–Mandel and restricted maximum likelihood estimators, a Q-profile confidence interval for τ^2, and 95% prediction intervals, all flagged as unstable at this k. Robust variance estimation and three-level models could handle the dependent estimates in principle, but the earlier log values could not be reproduced and were removed instead of published unverified. Three studies each contributed two comparisons against shared controls, so one-estimate-per-study and leave-one-estimate-out analyses address non-independence. Funnel-plot asymmetry was examined with Egger regression and Begg rank correlation, descriptive only at this k; an optimal information size was not estimable and no trial sequential analysis was performed. Two-sided $P < .05$ defines significance. The original R meta (metagen) pooling was reproduced independently in Python 3.13.14 under the same estimator, agreeing to the reported precision for every pooled value and subgroup Q test. With no common baseline risk, Table 4 gives illustrative absolute effects per 1000 at assumed control risks of 10, 50, and 200 per 1000, derived as $p_1 = OR \times p_0/(1 - p_0 + OR \times p_0)$ and not by multiplying risk by the odds ratio. Prespecified analyses and those added after the primary result are reported separately; the latter are labelled post hoc exploratory and support no conclusion.

Subgroup analysis Four stratifications were fixed in the written plan before pooling: lesion type, region, study design, and deficiency definition. Subgroup differences were tested with the Cochran-type Q test on stratum summaries. Lesion type had three strata: oral cancer, oropharyngeal cancer, and other/precancerous (OSCC, OLP, leukoplakia, and mixed leukoplakia–OSCC). It accounted for the inconsistency in the primary result ($\chi^2_2 = 6.25$; P for subgroup = .04): other/precancerous gave 1.51 (95% CI, 1.08–2.09; P = .01; $I^2 = 0\%$; P = .67), oral cancer 1.39 (0.62–3.12; $I^2 = 27.5\%$) from two estimates, and oropharyngeal cancer 0.90 (0.70–1.16; $I^2 = 12.8\%$) from three, with neither of the latter excluding the null. Region was Europe versus Asia, with Saudi Arabia assigned to Asia and the Chilean study left out because it fits neither stratum; nine estimates entered. Because the source forest plot placed the Saudi study in Europe, both groupings are reported. With the corrected grouping, the overall estimate was 1.23 (0.93–1.63; $I^2 = 52.8\%$), Europe 1.12 (0.74–1.70; $I^2 = 66.4\%$), and Asia 1.46 (1.04–2.04; $I^2 = 0\%$), with $\chi^2_1 = 0.92$ and P = .34; the original grouping gave P = .74. Design compared prospective nested case–control with case–control/cross-sectional: four prospective comparisons gave 1.37 (0.97–1.94; $I^2 = 16.7\%$) and six others 1.24 (0.85–1.82; $I^2 = 54.9\%$), with $\chi^2_1 = 0.14$ and P = .71. Deficiency definition compared the strict <25 versus ≥50 nmol/L contrast with every other definition and returned identical numbers. All four strict contrasts came from the

Yaohui et al. INPLASY protocol 2026110111. doi:10.37766/inplasy2026.11.0111 Downloaded from https://inplasy.com/inplasy-2026-11-0111/

two prospective cohorts and all six remaining comparisons from case–control or cross-sectional studies, so each off-diagonal cell was empty: design and exposure definition are completely collinear, the two Figure 4 panels are identical, and the variation cannot be credited to either variable alone. Neither region (P = .34), design (P = .71), nor threshold (P = .71) showed a subgroup difference, and tests at seven studies lack power. Restrictions added after the primary pooling, including serum-only, binary-contrast, fully adjusted, malignant-only, NOS = 9, one-estimate-per-study, and direction-harmonised analyses, are not prespecified subgroups; they are reported separately as post hoc exploratory.

Sensitivity analysis Sensitivity work is reported in two groups, and the distinction between them is stated in the review. Small-sample checks were added during revision, registered as a deviation, and leave the primary estimate unchanged. τ^2 was 0.083 (DerSimonian–Laird), 0.038 (Paule–Mandel), and 0.068 (restricted maximum likelihood), with a Q-profile 95% CI of 0 to 0.50; the three pooled odds ratios were 1.26 (0.95–1.68), 1.23 (0.98–1.55), and 1.26 (0.96–1.64), so no estimator changed the inference. The Hartung–Knapp–Sidik–Jonkman interval was 1.26 (0.96–1.67; P = .09), close to the conventional one, while the 95% prediction interval of 0.60 to 2.66 crossed the null and remains unstable at this k. Stratum-level values, including a non-estimable oral cancer prediction interval, are listed without interpretation. All remaining sensitivity analyses were introduced after the primary result showed I^2 = 51.7% and are labelled post hoc exploratory, confined to the lower partition of Table 3, eTable 7, and Figure 5; none supports the Abstract's conclusions or the clinical reading. Excluding the food-frequency questionnaire study left serum and plasma 25(OH)D only and gave 1.44 (1.17–1.79; I^2 = 0%); seven directly comparable binary deficiency contrasts gave 1.48 (1.07–2.04; I^2 = 0%). The five fully adjusted estimates gave 1.12 (0.74–1.70; I^2 = 66.4%), malignant outcomes alone 1.17 (0.84–1.64; I^2 = 60.3%), NOS = 9 studies 1.12 (0.74–1.70), one estimate per study 1.26 (0.87–1.83; I^2 = 64.9%), and a fixed-effect model 1.10 (0.95–1.29). Leave-one-estimate-out pooled odds ratios spanned 1.17 to 1.44, all crossing the null except the analysis omitting the dietary study. Coding every estimate in the same direction by taking the reciprocal of that dietary estimate (1.20; 0.96–1.49) under log-linearity gave 1.32 (1.14–1.54; I^2 = 0%; τ^2 = 0.000; Q = 7.42, P = .59) and the lesion-type difference disappeared (χ^2_2 = 1.49; P = .47); the reciprocal carries 48.4% of the weight in that harmonised model. Two earlier robustness models could not be reproduced and were excluded instead of published unverified. The funnel plot was broadly symmetric, but its Egger and Begg P values were transcribed from the supplied plot instead of reproduced, four discordant sets are documented, and at seven studies with ten dependent estimates these tests are descriptive only and entered no GRADE domain. No inference about publication bias is drawn.

Country(ies) involved China.

Keywords vitamin D; 25-hydroxyvitamin D; oral squamous cell carcinoma; oral potentially malignant disorders; exposure assessment; meta-analysis.

Contributions of each author

Author 1 - YaoHui Yu - Wrote the main manuscript text.
Email: 148267943@qq.com
Author 2 - Yu Zhou.

Yaohui et al. INPLASY protocol 2026110111. doi:10.37766/inplasy2026.11.0111 Downloaded from https://inplasy.com/inplasy-2026-11-0111/