Open Access Indonesian Journal of Medical Reviews
Volume 6, Issue 5, 2026 | Article ID: 932 | Pages: 378-391 | e-ISSN 2807-6257
OAIJMR

SYSTEMATIC REVIEW AND META-ANALYSIS

Baseline Optical Coherence Tomography Biomarkers Predict Failure to Dry the Macula on Anti-VEGF Therapy in Diabetic Macular Oedema — but Only When Response Is Defined Anatomically: A Systematic Review and Meta-Analysis

Angel Lim[1*], I Gusti Ayu Stiti Sadvika[2]

[1]Medical Doctor, Faculty of Medicine, Universitas Warmadewa, Denpasar, Indonesia
[2]Medical Doctor, Faculty of Medicine, Universitas Udayana, Denpasar, Indonesia

ARTICLE INFO

Article history
Received: July 10, 2026
Revised: August 8, 2026
Accepted: August 29, 2026
Available Online: September 4, 2026
Published: September 6, 2026

Keywords
Anti-VEGF therapy; diabetic macular oedema; optical coherence tomography biomarkers; prognosis; treatment response

***Corresponding author**
Angel Lim
E-mail: ngellim1909@gmail.com

DOI
https://doi.org/10.37275/oaijmr.v6i5.932

ABSTRACT

Background: Baseline optical coherence tomography (OCT) biomarkers are used to counsel patients before anti-vascular endothelial growth factor (anti-VEGF) therapy for diabetic macular oedema, but the two published syntheses covered visual acuity only.

Objective: To estimate the association between adverse baseline OCT biomarkers and poor anatomical response, and to test whether it depends on the response definition.

Methods: Europe PMC and MEDLINE were searched for cohort studies relating a baseline OCT biomarker to an anatomical outcome after anti-VEGF therapy. Odds ratios were pooled by correlated hierarchical effects with robust variance estimation, the unit of analysis being the biomarker-outcome estimate. Endpoints were pre-specified into clinically anchored and percentage-reduction strata; risk of bias used ROBINS-E.

Result: Seventeen studies (1,758 eyes) gave 54 estimates. In the primary stratum (10 studies, 26 estimates, 1,058 eyes) an adverse biomarker carried an odds ratio of 3.75 (95% CI 2.27 to 6.19; p = 0.0004) for failing to dry the macula: against a background rate near 35 in 100, about 67 in 100 (55 to 77). Ellipsoid zone disruption was best supported (3.55, 1.43 to 8.84). The direction reversed against percentage-reduction endpoints (0.84, 0.31 to 2.30; interaction p = 0.006); subretinal fluid from 3.33 to 0.18 (0.06 to 0.53). It held across 20 sensitivity analyses (3.29 to 4.90). Certainty was low.

Conclusion: About two-thirds of eyes with adverse baseline morphology remain wet after loading, against one-third without it — but only where response means the retina became dry. Percentage-reduction endpoints are coupled to baseline thickness and invert it.

1. Introduction

Diabetic macular oedema remains the commonest cause of central vision loss in working-age adults with diabetes, and its burden is growing fastest in the populations least equipped to absorb it. A prospective cohort of Indonesian adults with type 2 diabetes documented substantial incidence and progression of diabetic retinopathy and blindness over follow-up[1], and long-term Asian population data show that vision-threatening retinopathy continues to accumulate even where screening programmes exist[2]. Population studies outside Asia report the same trajectory[3]. Intravitreal anti-vascular endothelial growth factor (anti-VEGF) therapy is the established first-line treatment, and the newest agents and regimens have extended durability without sacrificing efficacy[4-6].

The difficulty is not whether anti-VEGF works on average, but that it does not work equally in every eye. A substantial minority of eyes remain wet after a standard loading course, and these eyes carry the treatment burden, the injections and the residual visual loss. Artificial-intelligence fluid analyses of refractory cohorts confirm that persistent fluid, rather than a lack of anatomical change, characterises these eyes[7], and prediction models built on baseline data have repeatedly identified persistent oedema as a distinct and forecastable phenotype[8]. Identifying such eyes at the first visit would change what clinicians say to patients and when they escalate to a corticosteroid.

Optical coherence tomography (OCT) is the obvious place to look, because the same scan that establishes the diagnosis also displays the structural features that plausibly govern the response. Disorganisation of retinal inner layers, disruption of the ellipsoid zone and external limiting membrane, hyperreflective foci, subretinal fluid, epiretinal membrane and foveal eversion are all reported as prognostic. The biological rationale is reasonably secure: hyperreflective foci are read as aggregates of activated microglia and extravasated lipoprotein, and eyes rich in them behave as though inflammation, rather than VEGF alone, is driving the oedema[9]. Studies of aqueous and vitreous mediators in diabetic macular oedema support that reading, showing that the inflammatory milieu differs systematically between eyes and tracks with imaging phenotype[10-13]. Trial-based analyses of anatomical control across agents have made the same point from the other direction, showing that the anatomical response is not a simple function of drug potency[14].

Two systematic syntheses of baseline OCT biomarkers in diabetic macular oedema were published in the last year. The first pooled 75 biomarkers from 27 studies and found five associated with worse visual acuity at moderate certainty; the second, using vote counting across 96 studies, reached no association above low certainty[15,16]. Both were explicitly restricted to visual acuity. Neither pooled a single anatomical endpoint, and both named non-standardised biomarker definitions and uncontrolled confounding as the reasons certainty stayed low. That restriction matters clinically, because the retreatment decision at the slit lamp is anatomical: an eye that will not dry after loading is switched or escalated largely irrespective of how many letters it has gained. It also matters methodologically, because anatomical endpoints are less vulnerable to the confounding by baseline acuity that forced both reviews to downgrade.

There is a further, less obvious problem that the present review was designed to expose. The primary studies do not agree on what an anatomical response is. Some ask whether the retina became and stayed dry. Others ask whether central thickness fell by a fixed percentage from baseline. These sound interchangeable

and are not, because a thicker and more diseased retina has further to fall and therefore clears a percentage bar more easily while remaining wet. If that coupling is real, the two endpoint families should give opposite answers about the same biomarkers in the same disease, and the published literature should look contradictory for reasons that have nothing to do with biology.

The novelty of this study lies in being the first quantitative synthesis of baseline OCT biomarkers against *anatomical* response to anti-VEGF therapy in diabetic macular oedema, in analysing the biomarker-outcome estimates with correlated hierarchical effects and robust variance estimation rather than treating multiple biomarkers from one cohort as independent, and in formally testing — and demonstrating — that the direction of every biomarker reverses according to how anatomical response was defined. **The aim of this study was to** estimate the pooled odds of poor anatomical response to anti-VEGF therapy in eyes with an adverse baseline OCT biomarker, to rank the individual biomarkers by the strength of their supporting evidence, and to quantify the extent to which the choice of anatomical endpoint definition determines the answer.

2. Methods

2.1. Design and reporting

This systematic review and meta-analysis was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement[17]. The review was not prospectively registered. Because the review analysed only published aggregate data, ethical approval was not required.

2.2. Eligibility criteria

Studies were eligible if they enrolled adults with diabetes mellitus and centre-involving diabetic macular oedema; assessed one or more qualitative structural OCT biomarkers on a pre-treatment scan; treated the eyes with an intravitreal anti-VEGF agent as the index therapy; and reported an anatomical outcome from which an odds ratio for poor response, with a confidence interval or a reconstructable two-by-two table, could be obtained. Prospective and retrospective cohorts and post-hoc analyses of randomised trials were all eligible.

Studies were excluded if the only reported outcome was visual acuity; if the biomarker was measured after treatment had begun; if the index therapy was a corticosteroid alone or the cohort was defined by vitrectomy; if no extractable effect estimate was available; or if the cohort overlapped with a larger included report. Reviews, editorials and protocols were excluded.

Because the analysis rests on judgements about whether a feature is present, each biomarker is defined here as it was applied. *Disorganisation of retinal inner layers* denoted an inability to identify the boundaries between the ganglion cell–inner plexiform layer complex, the inner nuclear layer and the outer plexiform layer within the central subfield. *Ellipsoid zone disruption* denoted any discontinuity of the hyperreflective ellipsoid band beneath the fovea, and *external limiting membrane disruption* the corresponding discontinuity of the thin hyperreflective line internal to it. *Hyperreflective foci* denoted discrete, well-circumscribed intraretinal dots of reflectivity comparable to the retinal pigment epithelium and without a posterior shadow. *Subretinal fluid* denoted an optically empty space between the neurosensory retina and the retinal pigment epithelium, reported by some cohorts as serous or neurosensory retinal detachment. *Epiretinal membrane* denoted a hyperreflective layer on the inner retinal surface. *Foveal eversion* denoted loss of the normal foveal depression with an outward, convex displacement of the Müller cell cone and the surrounding inner retina.

These definitions were not imposed on the primary studies. Each estimate was extracted as the study itself classified the feature, and no attempt was made to re-grade images or to harmonise thresholds. Where a biomarker was reported as a count rather than as present or absent — which applied to hyperreflective foci in every cohort that reported them — the study's own dichotomisation was accepted. The consequences of that decision are examined in the discussion, and they are not trivial.

2.3. Information sources and study selection

Europe PMC and MEDLINE were searched using two structured searches combining terms for diabetic macular oedema, optical coherence tomography, the individual biomarkers, and anti-VEGF agents, supplemented by targeted digital object identifier retrieval of reports cited in the eligible articles. Records were screened on title and abstract and then in full text. Each included report was read in full and re-extracted directly from its published tables; no estimate was carried forward from an abstract alone.

The searches combined five concept blocks: the population (diabetic macular oedema and its variant spellings), the imaging modality (optical coherence tomography and its abbreviations), the biomarkers (each feature listed in section 2.2, with its abbreviations and synonyms), the intervention (anti-VEGF as a class and each agent by name), and the outcome (terms for response, non-response, refractory and persistent disease, and for retinal thickness). Blocks were combined with the Boolean AND and terms within each block with OR. A verbatim listing of the executable strings was not retained in the analysis record; the concept blocks above are reported so that the search can be reconstructed, and the authors will supply the strings on request.

2.4. Outcomes and endpoint strata

The outcome was poor anatomical response to anti-VEGF therapy, expressed as an odds ratio comparing eyes with the adverse biomarker present against eyes with it absent. Because the primary studies defined that outcome in materially different ways, every estimate was assigned before analysis to one of four pre-specified endpoint strata: clinically anchored endpoints, which ask whether the retina became and stayed dry (refractory or persistent oedema, failure to reach a flat retina, or a clinician's decision to switch to a corticosteroid); relative percentage-reduction endpoints, which ask whether central thickness fell by a fixed proportion of its baseline value; absolute-change endpoints, expressed in micrometres; and recurrence after an initial response. The clinically anchored stratum was pre-specified as primary because it corresponds to the decision clinicians actually make.

2.5. Effect measure

Both the predictor and the outcome were binary throughout, so the odds ratio was the appropriate effect measure. A standardised mean difference was not computed, because it is undefined for a dichotomous outcome and would have required discarding the design of every included study. Where a study reported only an adjusted odds ratio with a confidence interval, that estimate was used and flagged; where a two-by-two table was printed, the crude odds ratio was computed from the counts. A Haldane–Anscombe correction of 0.5 was available for zero cells but was not required in any of the 54 estimates.

2.6. Synthesis

Fifty-four estimates were nested within 17 cohorts, and several cohorts contributed five or six biomarkers each. Treating those estimates as independent would have understated every standard error, so the primary model used correlated hierarchical effects with robust variance estimation, applying the CR2 small-sample adjustment and Satterthwaite degrees of freedom, with a working within-study correlation of 0.6. Biomarker-specific subtotals were additionally fitted with a DerSimonian–Laird random-effects model with the Knapp–Hartung adjustment, and were displayed only where at least four estimates were available. Heterogeneity was summarised with the I-squared statistic and the between-study variance τ^2. Subgroup analyses were pre-specified for endpoint stratum,

biomarker, anti-VEGF agent and country, and were tested with a joint Wald test under the same robust variance framework.

The unit of analysis was the biomarker-outcome estimate, not the eye. Each estimate contrasted eyes with a given feature against eyes without it, within its own cohort, and a single cohort could contribute up to six such contrasts from the same set of eyes. The pooled estimate is therefore the average strength of association across biomarker contrasts. It is not, and could not be, the risk carried by an individual eye, because eyes commonly display several of these features at once and an eye-level synthesis would require individual participant data that are not available. This distinction is carried through the wording of the results and is revisited in the limitations.

The I-squared statistic is not uniquely defined under a correlated hierarchical effects model. It is reported here as the proportion of the total variance of the estimates that is attributable to the combined between-study and within-study random components rather than to sampling error, with the two components reported separately where they differ materially. A prediction interval was not computed: with a between-study variance of zero it would collapse toward the confidence interval and would misrepresent the uncertainty that actually exists, which is within cohorts rather than between them.

2.7. Sensitivity analyses and small-study effects

Twenty sensitivity analyses were run: a leave-one-out fit for every study in the primary stratum; restriction to crude estimates; exclusion of cohorts with concomitant corticosteroid exposure; exclusion of clinician-driven switch outcomes; exclusion of studies at high risk of bias; variation of the assumed within-study correlation from 0.2 to 0.8; and a re-coding of one study whose printed count and printed percentage disagreed. Small-study effects were assessed with Egger regression, the Begg rank correlation and trim-and-fill. Both asymmetry tests assume independent observations and were applied to the 26 clustered estimates of the primary stratum; their results are therefore interpreted qualitatively rather than as formal inference.

2.8. Risk of bias and certainty of evidence

Risk of bias was assessed with ROBINS-E across its seven domains[18]. The randomised-trial tool RoB 2.0 was not used and does not apply here, because no study randomised the biomarker; these are prognostic-factor cohort studies even where the underlying dataset came from a randomised trial. Newcastle–Ottawa Scale scores were recorded alongside for comparability with the existing biomarker literature. Certainty of evidence was rated using the GRADE framework for prognostic factors.

3. Results

3.1. Study selection

The searches identified 187 records, of which 13 were duplicates. Of the 174 records screened, 89 were excluded on title and abstract and 85 full texts were retrieved; none was unobtainable. Sixty-eight reports were excluded at full-text assessment: 22 did not address an anti-VEGF treatment-response question, 14 reported a visual-acuity endpoint only, 12 contained no qualitative baseline OCT biomarker, 11 offered no extractable two-by-two table or odds ratio with a confidence interval, six were reviews, editorials or protocols, two were corticosteroid-only or vitrectomy cohorts, and one was a cohort already represented within a larger included report. Seventeen studies contributing 54 biomarker-outcome estimates were included, of which 10 studies and 26 estimates fell in the primary clinically anchored stratum. The selection process is set out in Figure 1.

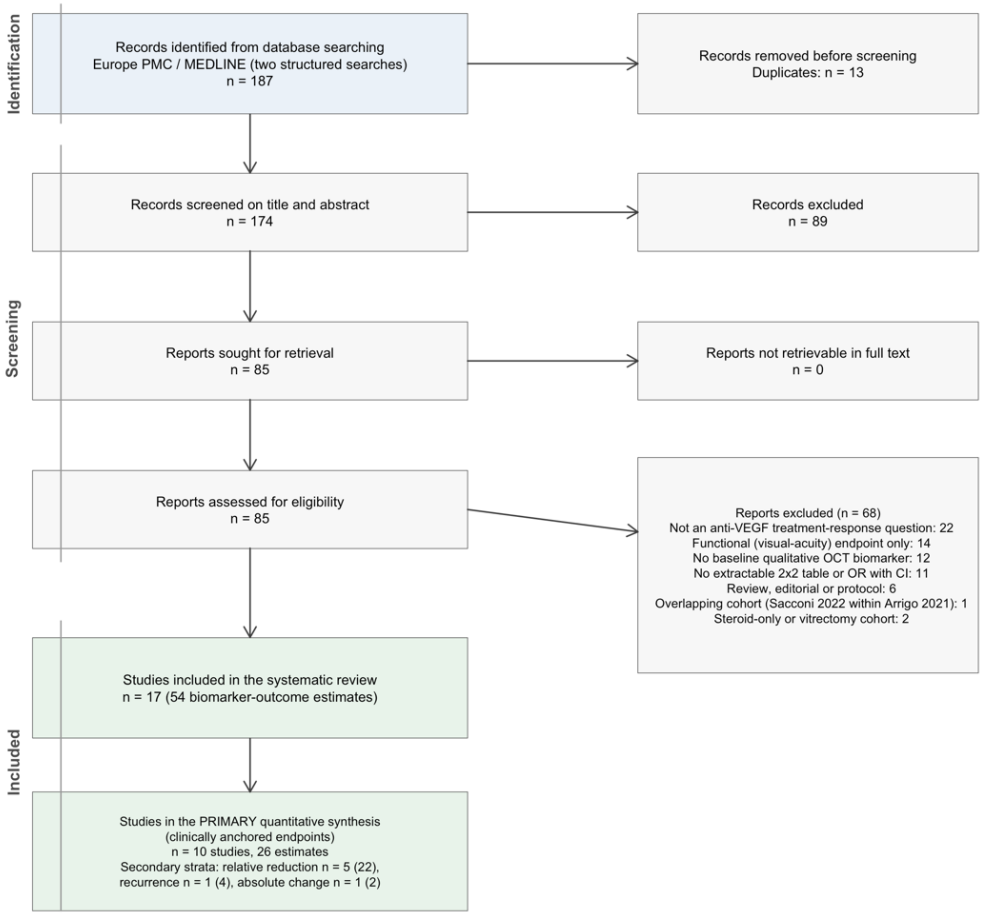

Figure 1. PRISMA 2020 flow diagram for the selection of studies relating baseline OCT biomarkers to anatomical response after anti-VEGF therapy in diabetic macular oedema.

3.2. Characteristics of the included studies

The 17 included cohorts contributed 1,758 eyes and were published between 2019 and 2026, with 14 appearing from 2022 onward[19-35]. Cohort size ranged from 30 eyes to 275. Seven cohorts were Chinese, three Korean, two Taiwanese, and one each from Italy, Romania, Spain, Türkiye and the United States; no Indonesian or wider South-East Asian cohort met the inclusion criteria. Seven studies used a mixed or switched anti-VEGF regimen, four aflibercept, four bevacizumab and two ranibizumab. Forty-three of the 54 estimates were crude odds ratios computed from printed two-by-two tables and 11 were adjusted odds ratios taken from published multivariable models. Table 1 sets out the characteristics of each study.

The timing of the outcome assessment varied and is embedded in each study's response definition. Eight cohorts assessed the outcome at a stated calendar time point, ranging from four weeks to twelve months; four assessed it at the end of a defined loading course; and five assessed it at a point determined by the clinical course — the final recorded visit, or the moment a clinician decided to switch therapy. Whether eyes were treatment-naïve was not stated consistently enough across the reports to be extracted as a variable, and this is noted as a limitation.

Table 1. Characteristics of the 17 included studies. Estimates = number of biomarker-outcome estimates contributed. CMT, central macular thickness; CRT, central retinal thickness; CST, central subfield thickness; DRIL, disorganisation of retinal inner layers; ELM, external limiting membrane; ERM, epiretinal membrane; EZ, ellipsoid zone; HRF, hyperreflective foci; SRF, subretinal fluid.

Study	Country	n	Anti-VEGF agent	Endpoint stratum	Anatomical response definition as reported	OCT biomarkers contributed	Estimates	ROBINS-E
Arrigo 2021	Italy	68	Mixed	Clinically anchored	Persistent oedema at 1 year	Foveal eversion	1	Some concerns
Cakmak Cengiz 2026	Türkiye	48	Bevacizumab	Clinically anchored	Anatomical failure after 3 injections	DRIL, EZ, SRF, ERM	4	Some concerns
Chang 2023	Taiwan	65	Ranibizumab	Absolute change	Final CRT improvement <50 µm	DRIL, SRF	2	Some concerns
Choi 2019	Republic of Korea	119	Bevacizumab	Clinically anchored	Refractory to anti-VEGF and to dexamethasone	HRF, EZ	2	Some concerns
Ergin 2025	Romania	104	Aflibercept	Relative reduction	<10% CMT reduction at 1 month	DRIL, SRF, HRF	3	Some concerns
Gu 2023	China	130	Aflibercept	Clinically anchored	Ineffective at 12 months	DRIL, EZ, ELM	3	Some concerns
Han 2023	Republic of Korea	30	Aflibercept	Clinically anchored	Suboptimal response after loading	ERM	1	High
Hsieh 2023	Taiwan	72	Mixed	Relative reduction	<10% CMT reduction at 4 months	EZ, SRF, HRF, ERM	4	Some concerns
Lee 2023	Republic of Korea	144	Bevacizumab	Clinically anchored	Switched to a dexamethasone implant	SRF	1	High
Li 2025	China	112	Ranibizumab	Clinically anchored	Non-response after 3 injections	DRIL, EZ	2	High
Li 2026	China	81	Mixed	Clinically anchored	Poor-response group	DRIL, EZ, SRF, HRF	4	Some concerns
Lu 2025	China	204	Mixed	Relative reduction	<20% CRT reduction	DRIL, EZ, ELM, SRF	4	Some concerns
Ruiz-Medrano 2025	Spain	275	Mixed	Clinically anchored	Switched to a dexamethasone implant	DRIL, SRF, HRF	3	High
Turski 2022	United States	52	Bevacizumab	Relative reduction	<10% CST reduction at 4–6 weeks	DRIL, EZ, ELM, SRF, HRF, ERM	6	High
Yu 2023	China	81	Mixed	Relative reduction	<10% CST reduction at 6 months	DRIL, EZ, SRF, HRF, ERM	5	Low
Zhou 2025	China	51	Aflibercept	Clinically anchored	Refractory at 6 months	DRIL, EZ, ELM, SRF, foveal eversion	5	High
Zhu 2025	China	122	Mixed	Recurrence	Recurrence within 12 months	EZ, ELM, SRF, HRF	4	Some concerns

Endpoint stratum was assigned before analysis from the response definition printed in each report.

3.3. Risk of bias

One study was judged at low risk of bias overall, ten raised some concerns, and six were at high risk[25,27,28,31-34]. Newcastle–Ottawa scores ranged from five to eight with a median of seven. The two domains that performed worst were confounding, where nine studies raised some concerns and three were high risk because baseline retinal thickness — which is both a determinant of the biomarker and a determinant of the outcome — was left unadjusted, and post-exposure interventions, where five studies were high risk because rescue corticosteroid or laser was administered during follow-up. Exposure measurement raised some concerns in 14 studies, reflecting the absence of an agreed grading protocol for qualitative OCT features. Figure 2 presents the study-level judgements.

Those overall figures understate the weakness of the analysis that matters most. Reading the risk-of-bias judgements against the endpoint stratum, **the single study judged at low risk of bias falls in the relative-reduction stratum and therefore contributes nothing to the primary analysis**[33]. The primary stratum consists of five studies raising some concerns and five at high risk, and contains no low-risk study at all. This is stated here rather than left to be reconstructed, because it is the single most important qualification on the primary estimate. Its principal counterweight is that removing the high-risk studies strengthens rather than weakens the pooled result, which is reported in section 3.7.

Figure 2. ROBINS-E risk-of-bias judgements for the 17 included studies across seven domains. Green denotes low risk, amber some concerns and red high risk.

3.4. Primary analysis: clinically anchored endpoints

In the primary stratum of 10 studies contributing 26 estimates from 1,058 eyes, the average association between an adverse baseline OCT biomarker and failure to dry the macula on anti-VEGF therapy was a pooled odds ratio of 3.75 (95% confidence interval 2.27 to 6.19; p = 0.0004), with residual heterogeneity of I^2 = 70.5%. This is an average across biomarker contrasts and not the risk carried by an individual eye; the biomarker-specific estimates below are the ones that bear on a particular scan. Figure 3 shows the study-level estimates and the biomarker subtotals.

The structure of that heterogeneity is more informative than its magnitude. The between-study variance was estimated at zero (τ^2 = 0.00), which means that none of the residual heterogeneity sat between cohorts. All of it sat *within* cohorts — across the different biomarkers measured in the same eyes. Cohorts were therefore behaving consistently as cohorts; the biomarkers were behaving differently from one another. This is precisely the pattern that robust variance estimation exists to handle, and it is the pattern that a conventional random-effects model applied to 54 supposedly independent estimates would have concealed.

Ranked by the strength of their evidence, ellipsoid zone disruption was the only biomarker individually significant on both models, at odds ratio 3.55 (1.43 to 8.84; p = 0.016) under robust variance estimation, and it was also the most internally consistent (I^2 = 39.5%). Disorganisation of retinal inner layers reached significance under the robust-variance model at 3.22 (1.07 to 9.70; p = 0.042) although its random-effects interval crossed unity. Subretinal fluid pointed strongly in the same direction at 3.94 (0.92 to 16.96) but did not reach significance at five estimates. External limiting membrane disruption, epiretinal membrane and foveal eversion all produced large point estimates with intervals so wide as to be uninformative, because only two estimates were available for each. Hyperreflective foci were the outlier, with a pooled estimate near unity and the highest heterogeneity in the dataset (I^2 = 88.0%, τ^2 = 3.664). Table 2 gives the biomarker-specific results under both models.

One estimate deserves to be named rather than left in the tables. The Spanish multicentre cohort[31] supplied the only strongly protective estimate in the entire primary stratum, a hyperreflective foci odds ratio of 0.06 (0.01 to 0.47), which sits in direct opposition to the two other hyperreflective foci estimates of 8.72[22] and 3.71[29]. That single estimate accounts for most of the null pooled result for hyperreflective foci, and for a substantial share of the residual heterogeneity of the whole stratum: omitting this cohort was the single most influential leave-one-out scenario, moving the pooled estimate from 3.75 to 4.35 and reducing I^2 from 70.5% to 42.7%, a larger effect on heterogeneity than any other omission. No other biomarker showed comparable disagreement between cohorts.

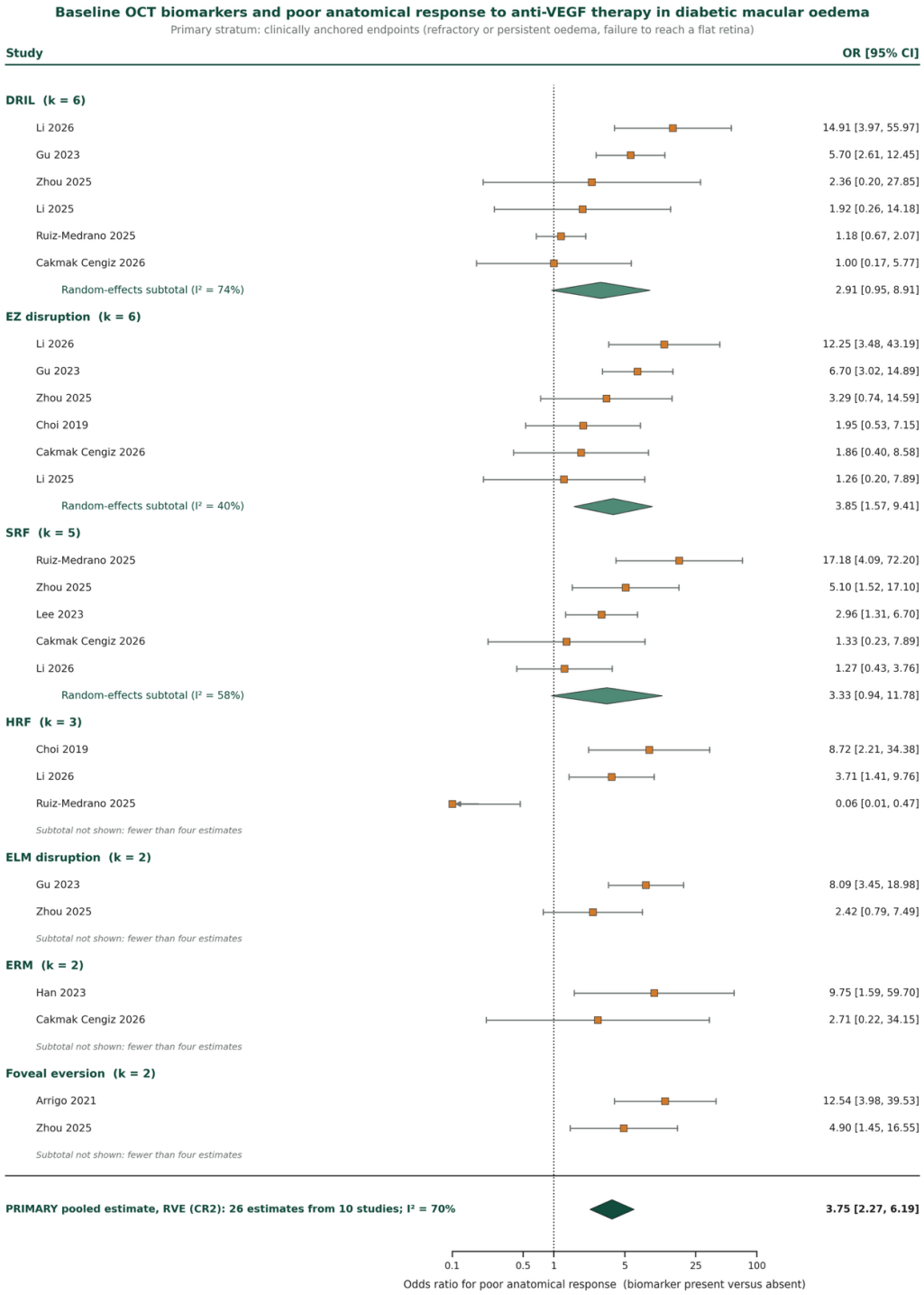

Figure 3. Forest plot of the primary stratum. Individual estimates are shown as squares with 95% confidence intervals; arrows denote intervals extending beyond the plotted range. Green diamonds are random-effects subtotals, displayed only where at least four estimates were available; the bottom diamond is the pooled robust-variance estimate.

Table 2. Biomarker-specific pooled estimates within the primary clinically anchored stratum. Random-effects estimates used DerSimonian–Laird with the Knapp–Hartung adjustment; robust-variance estimates used correlated hierarchical effects with the CR2 adjustment.

Biomarker	Studies	Estimates	Random-effects OR (95% CI)	I^2	τ^2	Robust-variance OR (95% CI)	p
Ellipsoid zone disruption	6	6	3.85 (1.58 to 9.41)	39.5%	0.280	3.55 (1.43 to 8.84)	0.016
Disorganisation of retinal inner layers	6	6	2.91 (0.95 to 8.92)	73.9%	0.932	3.22 (1.07 to 9.70)	0.042
Subretinal fluid	5	5	3.34 (0.94 to 11.78)	58.2%	0.509	3.94 (0.92 to 16.96)	0.060
External limiting membrane disruption	2	2	4.70 (0.00 to 9586.89)	64.2%	0.466	4.40 (0.01 to 3899.21)	0.226
Epiretinal membrane	2	2	6.32 (0.00 to 13815.45)	0%	0.000	7.55 (0.16 to 363.00)	0.097
Foveal eversion	2	2	8.02 (0.02 to 3109.88)	17.4%	0.077	8.02 (0.03 to 1916.53)	0.131
Hyperreflective foci	3	3	1.46 (0.00 to 838.22)	88.0%	3.664	1.56 (0.01 to 516.52)	0.770

Estimates based on two studies are reported for completeness; their intervals are too wide to support any inference and are not interpreted.

3.5. The endpoint-definition reversal

The single most consequential result of this review is that the direction of every biomarker depended on how anatomical response had been defined. Against clinically anchored endpoints the pooled odds ratio was 3.75. Against relative percentage-reduction endpoints, in five studies contributing 22 estimates, it was 0.84 (0.31 to 2.30; p = 0.65) — no association at all, and a point estimate on the protective side of unity. The formal interaction test under robust variance estimation gave F(1, 9.2) = 12.3, p = 0.006, and the joint test across all four endpoint strata gave F(3) = 105.8, p = 0.006.

The denominator degrees of freedom of that interaction test are small — 9.2 under the Satterthwaite approximation — and robust variance estimation is known to be anti-conservative in that region even with the CR2 correction. The formal p value should therefore not carry the argument on its own. It does not need to: the reversal held in the same direction for every biomarker for which both endpoint families were represented, which is corroboration that does not depend on the small-sample properties of a single test.

The reversal was not an artefact of which biomarkers happened to be studied in which cohorts, because it held biomarker by biomarker. Subretinal fluid is the clearest illustration: odds ratio 3.33 against a clinically anchored endpoint, and 0.18 (0.06 to 0.53) against a relative-reduction endpoint — a statistically significant effect in the opposite direction. Disorganisation of retinal inner layers moved from 2.91 to 0.42 and ellipsoid zone disruption from 3.85 to 1.47. Figure 4 displays the contrast, and Table 3 summarises all four strata.

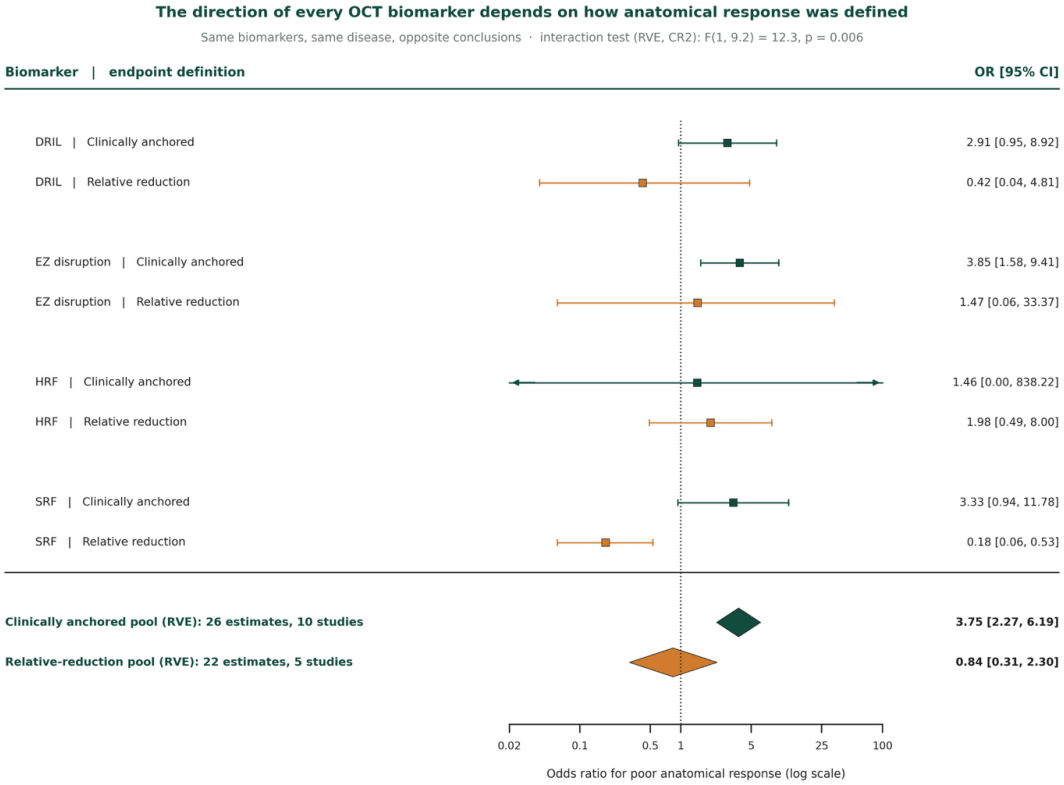

Figure 4. Odds ratios for poor anatomical response by biomarker and endpoint definition. Green denotes clinically anchored endpoints and orange relative percentage-reduction endpoints. Diamonds are the pooled robust-variance estimates for each stratum.

Table 3. Pooled odds ratios by endpoint stratum. The combined estimate across all strata is reported for completeness only and is not interpretable, because it averages effects that point in opposite directions.

Endpoint stratum	Studies	Estimates	What the endpoint asks	Pooled OR (95% CI)	p	I^2
Clinically anchored (primary)	10	26	Did the retina become and stay dry? Refractory or persistent oedema, failure to reach a flat retina, or a switch to a steroid	3.75 (2.27 to 6.19)	0.0004	70.5%
Relative percentage reduction	5	22	Did central thickness fall by a fixed percentage from baseline?	0.84 (0.31 to 2.30)	0.65	89.2%
Recurrence	1	4	Did oedema return after an initial response?	3.03 (3.01 to 3.05)	0.0003	—
Absolute change	1	2	Did central thickness fall by a fixed number of micrometres?	0.09 (0.01 to 0.57)	0.038	—
All strata combined (not interpretable)	17	54	Mixture of the four questions above	1.80 (1.01 to 3.18)	0.045	86.9%

The recurrence and absolute-change strata each rest on a single study and are shown for transparency, not for inference.

The mechanism is arithmetic rather than biological. A percentage-reduction endpoint divides the change in thickness by the baseline thickness, and baseline thickness is itself a consequence of the adverse morphology being tested. An eye with extensive subretinal fluid starts thicker, so a fixed percentage fall is easier to achieve even if a substantial volume of fluid remains and the retina is still, by any clinical standard, wet. The percentage-reduction endpoint is therefore not a neutral measurement of the same construct; it is a different construct that happens to share a name.

3.6. Subgroup analyses

No moderator other than endpoint definition explained any further heterogeneity. Within the primary stratum, the pooled estimate was numerically largest for aflibercept at 5.56 (1.17 to 26.33) and smallest for ranibizumab at 1.53 (1.10 to 2.14), but the

joint test across agents was not significant (p = 0.220) and each agent subgroup rested on one to three cohorts. Country was likewise non-significant (joint p = 0.472), although the two best-represented settings, China and the Republic of Korea, gave concordant estimates of 4.43 (2.08 to 9.46) and 4.37 (1.30 to 14.69). Restricting the analysis to crude odds ratios left the pooled estimate unchanged, so the inclusion of adjusted estimates did not drive the result. Table 4 reports the subgroup analyses.

Table 4. Subgroup analyses within the primary clinically anchored stratum, except for the endpoint-stratum moderator, which by definition spans all strata. Joint p values are Wald tests across all levels of a moderator under robust variance estimation.

Moderator	Level	Studies	Estimates	OR (95% CI)	p within level	Joint p
Anti-VEGF agent	Aflibercept	3	9	5.56 (1.17 to 26.33)	0.043	0.220
	Bevacizumab	3	7	2.58 (0.63 to 10.64)	0.102	
	Mixed or switched	3	8	3.78 (0.28 to 51.35)	0.143	
	Ranibizumab	1	2	1.53 (1.10 to 2.14)	0.039	
Country	China	4	14	4.43 (2.08 to 9.46)	0.010	0.472
	Republic of Korea	3	4	4.37 (1.30 to 14.69)	0.036	
	Spain	1	3	1.56 (0.03 to 72.21)	0.379	
	Türkiye	1	4	1.47 (0.25 to 8.68)	0.223	
	Italy	1	1	12.54 (single estimate)	—	
Endpoint stratum	Clinically anchored	10	26	3.64 (2.17 to 6.10)	0.0005	0.006
	Relative reduction	5	22	0.84 (0.31 to 2.32)	0.650	
	Recurrence	1	4	3.03 (3.01 to 3.05)	0.0003	
	Absolute change	1	2	0.09 (0.01 to 0.57)	0.038	

Subgroups resting on a single cohort are reported for completeness and should not be compared against one another.

3.7. Sensitivity analyses

The primary result withstood every challenge put to it. Across the 10 leave-one-out fits the pooled odds ratio ranged from 3.29 to 4.35 and the p value never exceeded 0.0018. Excluding the six studies at high risk of bias *strengthened* the estimate to 4.90 (2.29 to 10.46) and reduced heterogeneity to 57.2%, which is the opposite of what would be expected if bias were manufacturing the association. Removing clinician-driven switch outcomes, the most subjective endpoints in the set, likewise strengthened the estimate to 4.56 (2.73 to 7.62) and halved heterogeneity to 42.7%. Excluding cohorts with concomitant corticosteroid exposure gave 4.25 (2.34 to 7.74). Varying the assumed within-study correlation from 0.2 to 0.8 moved the pooled estimate by less than 0.14, confirming that the result does not depend on that working assumption. Table 5 reports all 20 scenarios.

One of those scenarios answers a specific structural objection and is worth stating as such. Two of the ten primary-stratum studies defined the outcome as the clinician's decision to switch to a dexamethasone implant[27,31]. That decision is made by a clinician who has seen the OCT scan, which is the exposure, so the outcome is not ascertained independently of the predictor. This is partial circularity rather than a subtle measurement problem. Removing those studies raised the pooled estimate to 4.56 (2.73 to 7.62) and halved residual heterogeneity to 42.7%, so the association is not an artefact of exposure-dependent outcome ascertainment; if anything, that ascertainment was diluting it. The same exclusion also removes the influential hyperreflective foci estimate discussed in section 3.4, and it is plausible that these two observations are the same observation: foci-rich eyes are precisely the eyes a clinician switches early to a steroid, for reasons of mechanism rather than because anti-VEGF has failed to dry them.

Table 5. Sensitivity analyses for the primary clinically anchored stratum.

Sensitivity scenario	Studies	Estimates	OR (95% CI)	p	I^2
Primary model	10	26	3.75 (2.27 to 6.19)	0.0004	70.5%
Omit Arrigo 2021	9	25	3.48 (2.06 to 5.86)	0.0009	70.2%
Omit Cakmak Cengiz 2026	9	22	4.08 (2.43 to 6.83)	0.0004	75.2%
Omit Choi 2019	9	24	3.71 (2.09 to 6.58)	0.0012	71.7%
Omit Gu 2023	9	23	3.29 (1.94 to 5.59)	0.0012	70.7%
Omit Han 2023	9	25	3.59 (2.12 to 6.10)	0.0008	71.5%
Omit Lee 2023	9	25	3.80 (2.19 to 6.61)	0.0008	70.9%
Omit Li 2025	9	24	3.95 (2.33 to 6.70)	0.0005	72.7%
Omit Li 2026	9	22	3.58 (1.95 to 6.55)	0.0018	69.7%
Omit Ruiz-Medrano 2025	9	23	4.35 (2.75 to 6.90)	0.0002	42.7%
Omit Zhou 2025	9	21	3.69 (1.98 to 6.87)	0.0018	77.0%
Exclude estimates available only as adjusted odds ratios	10	26	3.75 (2.27 to 6.19)	0.0004	70.5%
Exclude cohorts with concomitant steroid exposure	6	18	4.25 (2.34 to 7.74)	0.0026	41.7%
Exclude clinician-driven switch outcomes	8	22	4.56 (2.73 to 7.62)	0.0005	42.7%
Exclude studies at high risk of bias	5	14	4.90 (2.29 to 10.46)	0.0059	57.2%
Working correlation ρ = 0.2	10	26	3.65 (2.19 to 6.08)	0.0005	64.4%
Working correlation ρ = 0.4	10	26	3.71 (2.23 to 6.15)	0.0004	67.3%
Working correlation ρ = 0.8	10	26	3.78 (2.30 to 6.22)	0.0003	73.5%
Zhou 2025 DRIL coded from the printed count rather than the percentage	10	26	3.75 (2.27 to 6.19)	0.0004	70.6%

ρ is the assumed within-study correlation among estimates contributed by the same cohort; the primary model used ρ = 0.6.

3.8. Small-study effects

No funnel asymmetry was detectable in the primary stratum. Egger regression gave t = 0.016 with p = 0.99, the Begg rank correlation gave τ = −0.16 with p = 0.27, and trim-and-fill imputed no missing studies. This deserves to be stated plainly rather than passed over, because prognostic-biomarker literatures characteristically do show small-study effects. Figure 5 shows the funnel plot.

Two caveats keep this from being stronger evidence than it is. Both tests treat the 26 clustered estimates as independent, which they are not, so the standard errors of the tests themselves are understated. And with ten cohorts the analysis only just meets the conventional minimum for a funnel test, so its power is low. What can reasonably be said is that the Egger intercept was essentially zero rather than merely non-significant, and that trim-and-fill imputed nothing — an absence of evidence of asymmetry, not evidence of its absence.

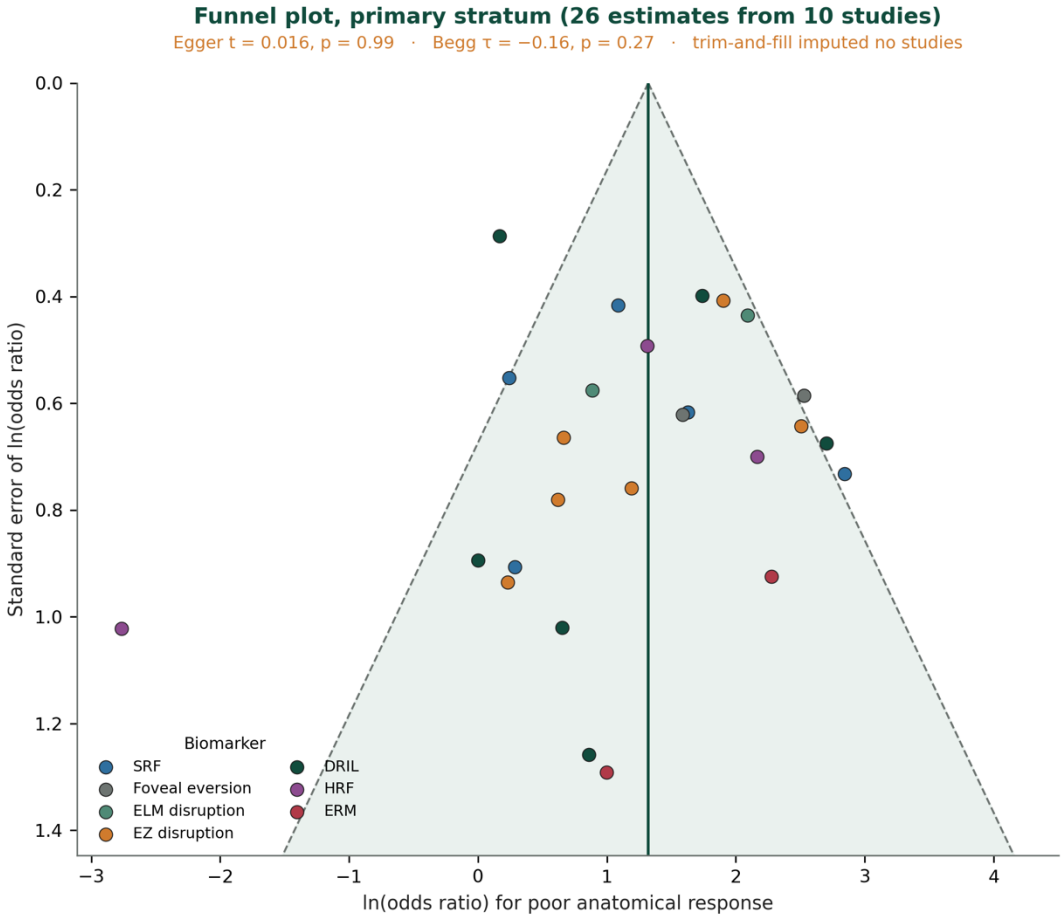

Figure 5. Funnel plot of the 26 estimates in the primary stratum, coloured by biomarker. The vertical line is the pooled robust-variance estimate and the shaded region is the pseudo-95% confidence region.

3.9. Domain-level risk of bias and certainty of evidence

Aggregated across studies, the ROBINS-E judgements concentrated in two domains. Confounding was the weakest: only five studies adequately handled baseline retinal thickness, nine raised some concerns and three were high risk. Post-exposure interventions were the second weakest, with five studies at high risk because rescue therapy was permitted during follow-up. Exposure measurement and selective reporting each raised some concerns in 14 of 17 studies, reflecting the absence of a standardised grading protocol and of published analysis plans. Outcome measurement performed best, at low risk in 10 studies. Figure 6 shows the domain-level distribution.

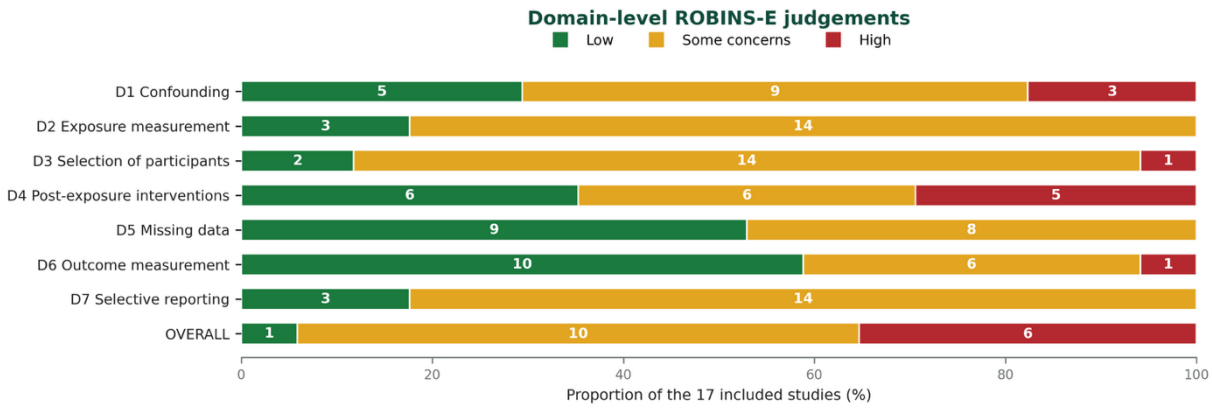

Figure 6. Domain-level distribution of ROBINS-E judgements across the 17 included studies.

Certainty of evidence for the primary comparison was rated LOW. The rating began low because every included study was observational, and was not downgraded further: risk of bias did not explain the association, because excluding high-risk studies strengthened it; inconsistency was high in magnitude but zero between cohorts; indirectness was minimal; the confidence interval excluded unity comfortably; and no small-study effect was detectable. An upgrade for a large effect was considered and not applied. Table 6 sets out the assessment.

Table 6. GRADE assessment of certainty for the association between an adverse baseline OCT biomarker and poor anatomical response in the primary stratum.

Domain	Judgement	Reasoning
Study design	Starts LOW	All 17 studies were observational cohorts of a prognostic factor; no study randomised the biomarker.
Risk of bias	No further downgrade	Six studies were at high risk and ten raised some concerns, but excluding the high-risk studies *strengthened* the pooled estimate to OR 4.90, so bias is unlikely to be generating the association.
Inconsistency	No downgrade	Residual heterogeneity was high (I^2 = 70.5%) but between-study variance was zero (τ^2 = 0.00); the heterogeneity is between biomarkers within cohorts, not between cohorts, and every leave-one-out fit stayed between OR 3.29 and 4.90.
Indirectness	No downgrade	Populations, imaging and treatments matched the review question closely.
Imprecision	No downgrade	The pooled confidence interval excluded 1 comfortably and did not span both appreciable benefit and appreciable harm.
Publication bias	No downgrade	Egger t = 0.016, p = 0.99; Begg τ = −0.16, p = 0.27; trim-and-fill imputed no studies.
Upgrade for large effect	Considered, not applied	The point estimate of 3.75 approaches the conventional threshold, but the observational design and residual confounding by baseline thickness argued against upgrading.
Overall certainty	LOW	Low certainty that the presence of an adverse baseline OCT biomarker increases the odds of failing to dry the macula on anti-VEGF therapy.

4. Discussion

This review provides the first pooled estimate of how baseline OCT morphology relates to whether the macula actually dries on anti-VEGF therapy, and it returns two findings of unequal importance. The first is that an adverse baseline biomarker is associated with nearly four times the odds of anatomical failure, with ellipsoid zone disruption the best-supported individual feature. The second, and the more consequential, is that this association exists only when response is defined by whether the retina became dry, and reverses when response is defined as a percentage fall in thickness.

4.1. Relationship to the existing syntheses

The two recent syntheses of OCT biomarkers in diabetic macular oedema restricted themselves to visual acuity and reached, at best, moderate certainty for five biomarkers predicting *worse* acuity[15,16]. The present results are complementary rather than contradictory. Ellipsoid zone disruption, disorganisation of retinal inner layers and hyperreflective foci appear in both the acuity-based and the anatomy-based analyses, which is what would be expected if these features mark a retina that is structurally damaged and functionally impaired as well as pharmacologically less tractable. The value added here is that the anatomical endpoint speaks directly to the retreatment decision, and that its analysis is far less exposed to confounding by baseline acuity, which was the principal reason both earlier reviews downgraded their certainty.

Both earlier reviews also named non-standardised biomarker definitions as a limitation. The present analysis suggests the problem is deeper than definitional drift in the *predictor*. The more damaging inconsistency lies in the *outcome*, and it is not merely noise: it systematically inverts the sign of the association.

4.2. Why the endpoint definition inverts the result

The coupling between baseline thickness and a percentage-reduction endpoint is deterministic rather than probabilistic. Adverse morphology such as extensive subretinal fluid or large intraretinal cysts produces a thicker retina at baseline; the percentage-reduction endpoint then rewards precisely those eyes, because they have more thickness available to lose. An eye can fall by 20% of a 600-micrometre baseline and remain at 480 micrometres — unequivocally wet — while an eye at 300 micrometres that dries to a normal 250 micrometres fails the same bar. The observed inversion for subretinal fluid, from an odds ratio of 3.33 to 0.18, is the quantitative expression of that arithmetic.

This has a practical consequence for how prognostic OCT research should be designed. Percentage-reduction endpoints should be abandoned in prognostic studies of baseline morphology, or, where they must be retained for comparability, they should be adjusted for baseline thickness and reported alongside an absolute anatomical criterion. The same caution applies to reading the existing literature: apparently contradictory reports about whether subretinal fluid is a good or a bad prognostic sign may be reconciled simply by checking which endpoint each study used. Studies that stratified their cohorts by morphological phenotype have reached similarly divergent conclusions depending on the outcome chosen[36,37].

4.3. The individual biomarkers

Ellipsoid zone disruption emerged as the most dependable single predictor, and this is biologically coherent. The ellipsoid zone is a direct proxy for photoreceptor integrity, and its disruption marks a retina in which outer-retinal damage has already occurred. Such an eye has less to gain from suppressing vascular permeability, because the pathology is no longer purely exudative. Region-specific analyses of retinal and choroidal biomarkers support the same reading[38], as do studies relating outer-retinal integrity to long-term restorative capacity after early treatment[39].

Disorganisation of retinal inner layers behaved similarly but with wider intervals, consistent with its greater inter-grader variability. A large multicentre real-world cohort found that baseline disorganisation correlated negatively with both functional and anatomical improvement and positively with systemic lipid parameters[40], which raises the possibility that part of its prognostic signal is metabolic rather than purely structural.

Hyperreflective foci were the conspicuous outlier, with a pooled estimate near unity and by far the highest heterogeneity. Two explanations are plausible and not mutually exclusive. The first is measurement: hyperreflective foci are counted rather than graded present or absent, the counting thresholds differ between studies, and dichotomising a count at different cut-offs will generate incoherent odds ratios. The second is that foci-rich eyes are genuinely a different pharmacological phenotype rather than simply a worse one. Trial data show that foci respond differently to agents with distinct mechanisms[9], and cohorts in which foci-rich eyes failed anti-VEGF but then responded well to a dexamethasone implant point the same way[28]. If hyperreflective foci mark inflammation-driven oedema, then their correct clinical use is not as a marker of general prognosis but as an indication for early corticosteroid escalation — a hypothesis this review can motivate but not test.

4.4. What this means at the slit lamp

An odds ratio on its own cannot be used in a consulting room. Converting it requires a background rate, and this review did not estimate one: because a single cohort contributes up to six biomarker contrasts from the same eyes, an eye-level event rate cannot be derived from the estimate-level dataset without counting the same eyes repeatedly. The conversion below therefore uses an external anchor — the reported range of 30 to 40 per cent of eyes responding poorly after standard loading[8] — and is **illustrative rather than a pooled estimate of absolute risk**. Table 7 sets it out.

Table 7. Illustrative conversion of the pooled odds ratios into absolute risks, using an external background rate of anatomical failure of about 35 in 100 eyes.

Scenario	Odds ratio (95% CI)	Illustrative absolute risk of anatomical failure	Difference from the background rate
Background rate in eyes starting anti-VEGF therapy	—	About 35 in 100 (reported range 30 to 40)	—
Any adverse baseline biomarker present (pooled estimate)	3.75 (2.27 to 6.19)	About 67 in 100 (55 to 77)	About 32 more per 100
Ellipsoid zone disruption present	3.55 (1.43 to 8.84)	About 66 in 100	About 31 more per 100
Disorganisation of retinal inner layers present	3.22 (1.07 to 9.70)	About 63 in 100	About 28 more per 100

Risks are rounded to the nearest whole number per 100 eyes. The interval shown for the pooled estimate is the conversion of the confidence interval of the odds ratio, and carries none of the uncertainty attaching to the background rate itself.

Read that way, the finding becomes usable. Against a background in which roughly one eye in three fails to dry after loading, an eye showing adverse baseline morphology would be expected to fail about two times in three. That is a difference of around thirty eyes per hundred, which is large enough to change how a first visit is conducted without being large enough to change whether treatment is offered.

Which features should change what a clinician says

The biomarkers are not interchangeable, and the review supports a ranking rather than a single statement.

- **Strong enough to influence counselling now: ellipsoid zone disruption.** It was the only feature significant under both models, it was the most internally consistent of the set (I^2 = 39.5%), and it has a coherent explanation — an eye whose photoreceptor layer is already interrupted has less to gain from suppressing vascular permeability. An interrupted ellipsoid band on the baseline scan is a reasonable trigger for closer early review.
- **Suggestive, and worth noting but not yet acting on: disorganisation of retinal inner layers, and subretinal fluid.** Both pointed clearly in the same direction and disorganisation reached significance under the robust-variance model, but each rested on five or six estimates with wide intervals, and disorganisation is among the least reproducible features to grade.
- **Should not currently change practice: hyperreflective foci, external limiting membrane disruption, epiretinal membrane and foveal eversion.** The first was null with extreme heterogeneity; the other three rested on two estimates each and produced intervals spanning several orders of magnitude. Large point estimates from two estimates are not evidence of a large effect.

For the clinician, the finding supports a modest and defensible statement at the first visit: an eye showing ellipsoid zone disruption, inner-layer disorganisation or substantial subretinal fluid on the baseline scan is roughly three to four times more likely to remain wet after a standard loading course than an eye without those features. That is enough to justify closer early review and an earlier discussion about escalation; it is not enough to justify withholding anti-VEGF therapy, and nothing in these data suggests that it should be. Eyes with adverse morphology still improved — they simply improved less reliably.

That last point bears repeating on its own, because it is the sentence most likely to be misquoted. **Eyes with adverse baseline morphology still improved on anti-VEGF therapy. They improved less reliably.** Nothing in this review supports withholding or delaying treatment on the basis of a baseline scan.

The finding is also a caution about how anatomical targets are set. Treat-to-target strategies and automated fluid-monitoring systems increasingly define success in anatomical terms, and the choice between an absolute and a proportional criterion is not a neutral one[7,41]. A service that audits its outcomes against a percentage-reduction target will systematically flatter its results in the very eyes that are doing worst.

4.5. Why a wet macula matters even when vision is preserved

A reader arriving from the two acuity-based syntheses may reasonably ask whether an eye that remains wet but sees well has failed at all. Three arguments say that it has. The first is burden: an eye that does not dry keeps receiving injections, and treatment burden is itself a driver of patient-initiated discontinuation and of loss to follow-up. The second is instability: fluid that never fully resolves recurs, and macular thickness fluctuation over time has been shown to carry independent prognostic weight[35,41]. The third is structural attrition: persistent oedema is accompanied by measurable thinning of the ganglion cell complex and by choroidal change, so an eye that stays wet is not in a steady state even when its acuity is[42,43]. Anatomical failure is therefore not a surrogate that happens to be convenient; it is the process that generates the functional loss the earlier reviews measured.

4.6. Whether these judgements can be made reliably outside a study

Every result in this review depends on someone deciding whether a feature is present. In the included cohorts those judgements were generally made by trained graders, sometimes two and sometimes masked. In a clinic they are made quickly, by clinicians of varying experience, on scans of varying quality. The risk-of-bias assessment recorded concerns about exposure measurement in 14 of 17 studies, which reflects the absence of an agreed grading protocol rather than carelessness by the investigators.

The practical inference is uncomfortable and should be stated: a prognostic marker with poor inter-grader agreement will not perform in practice as it does in a study, however large its pooled odds ratio. Hyperreflective foci illustrate the problem most sharply, because they are counted rather than simply identified, and different counting thresholds applied to the same scans will produce different dichotomies and therefore different odds ratios. No included study reported inter-grader reliability in a form that could be extracted, so this review cannot quantify the problem — only name it. Automated and quantitative approaches to the same features are the obvious answer and are already being developed[44,45].

4.7. Strengths and limitations

The principal strength of this review is that its analytical model matches its data structure. Fifty-four estimates drawn from 17 cohorts are not 54 independent observations, and the correlated-hierarchical-effects model with robust variance estimation prevents the spurious precision that would otherwise follow. A second strength is that every included estimate was taken from a published table in a full text rather than from an

abstract. A third is the breadth of the sensitivity programme: 20 scenarios, none of which overturned the primary result, and two of which — excluding high-risk studies and excluding subjective switch outcomes — strengthened it.

The limitations are real and several. **First**, the certainty of evidence is low, and it is low principally because every included study was observational and because baseline retinal thickness was inadequately controlled in most of them. Baseline thickness is a confounder in the strict sense: it is caused by the adverse morphology and it influences the outcome, so a crude odds ratio necessarily overstates the independent contribution of the biomarker. The direction of that bias is known but its magnitude is not, because too few studies reported thickness-adjusted estimates for an adjusted-only synthesis to be feasible.

Second, a limitation of interpretation rather than of data, and it constrains how the headline number may be used. The unit of analysis was the biomarker-outcome estimate. The pooled odds ratio of 3.75 is the average strength of association across biomarker contrasts, not the risk carried by an eye with adverse morphology, and the two are not the same quantity because eyes commonly display several of these features simultaneously and contribute to several contrasts at once. The illustrative absolute risks in Table 7 inherit that limitation and are offered as a scale, not as a prediction for an individual patient. The biomarker-specific estimates are the ones that bear on a particular scan.

Third, two variables that a clinician would want were not extractable. Whether eyes were treatment-naïve was not reported consistently enough across the 17 reports to be coded, so the pooled estimate mixes previously untreated eyes with eyes already partly treated. And the covariate sets of the 11 adjusted models were not uniformly reported and were not extracted, so although restricting the analysis to crude estimates left the pooled result numerically unchanged, it cannot be claimed that the adjusted and crude estimates were estimating the same quantity.

Fourth, several biomarkers rest on very few estimates. External limiting membrane disruption, epiretinal membrane and foveal eversion each contributed only two, and their confidence intervals are correspondingly uninformative. They are reported for completeness and should not be quoted as evidence of a large effect merely because their point estimates are large.

Fifth, the endpoint heterogeneity that this review exposes also constrains it. Six distinct anatomical definitions were in use across the pool, and although the pre-specified stratification handles this transparently, the primary stratum still contains a mixture of refractory-disease definitions, failure-to-flatten definitions and clinician switch decisions. The last of these is the most subjective, and its removal strengthened the estimate rather than weakening it.

Sixth, no cohort from Indonesia or the wider South-East Asian region met the inclusion criteria, despite the regional burden of diabetic retinopathy[1,2]. Whether these odds ratios transport to populations with different diabetes duration, glycaemic control and access to treatment is unknown.

That absence is worth stating as a finding rather than only as a limitation. A literature of 17 cohorts spanning eight countries contains none from a region carrying a large and rising share of the global diabetic retinopathy burden. The gap is also unusually cheap to close. A single tertiary centre could contribute meaningfully with a prospective series of roughly one hundred consecutive treatment-naïve eyes with centre-involving oedema, graded against the definitions in section 2.2 by two independent readers with a reported agreement statistic, treated with a documented loading course, and assessed against an absolute anatomical endpoint at a fixed time point. Nothing in that design requires equipment or funding beyond what a routine retina service already holds, and it would answer the transportability question that this review cannot.

Seventh, no study of faricimab or brolucizumab met the inclusion criteria. This is not because such studies do not exist. Post-hoc analyses of the newer agents have examined OCT biomarkers, but as treatment-effect modifiers rather than as baseline predictors of anatomical failure[6,9,14], so these results describe the anti-VEGF era that preceded them.

4.8. Future research

Three things would move this field forward quickly and none of them is expensive. The first is agreement on a single anatomical outcome definition, anchored on an absolute thickness threshold or on the presence or absence of fluid rather than on a percentage change. The second is routine adjustment for baseline central thickness in every prognostic OCT analysis, which would convert a large volume of existing crude estimates into interpretable ones. The third is a standardised grading protocol for hyperreflective foci, with a published counting method and a pre-specified cut-off; the heterogeneity observed here suggests that most of the disagreement about foci is measurement rather than biology. Beyond that, prospective cohorts pairing baseline imaging with aqueous mediator profiling[12,13] and with automated quantitative reflectivity metrics[44-46] offer the most direct route to separating eyes that need more anti-VEGF from eyes that need something else. Emerging work on choroidal remodelling[43], on ganglion-cell and choroidal thickness profiles in persistent oedema[42], on hard exudate dynamics[47] and on combination approaches in persistent disease[48] indicates how broad that phenotyping effort has already become. Prognostic modelling of the kind now appearing in this literature[8,49-54] will only be as good as the outcome definitions it is trained against, which returns the field to the first recommendation.

5. Conclusion

Averaged across biomarker contrasts, adverse baseline OCT morphology was associated with an odds ratio of 3.75 (95% confidence interval 2.27 to 6.19) for failure of anti-VEGF therapy to dry the macula in diabetic macular oedema. Set against a background failure rate of roughly one eye in three, that scale corresponds to roughly two eyes in three. The estimate is an average across contrasts rather than a risk attaching to an individual eye, and the biomarker-specific results are the ones that bear on a particular scan: ellipsoid zone disruption is the only feature supported well enough to influence counselling today, disorganisation of retinal inner layers and subretinal fluid point the same way without yet being actionable, and the remaining features should not change practice.

The association held across every one of 20 sensitivity analyses and strengthened when studies at high risk of bias were removed. It was, however, entirely conditional on how anatomical response had been defined: against relative percentage-reduction endpoints the same biomarkers in the same disease showed no association and, for subretinal fluid, a significant association in the opposite direction. Percentage-reduction endpoints are coupled to baseline thickness and should be abandoned in prognostic OCT research or, at minimum, adjusted for it. Certainty of evidence is low, and the clinical use of these findings is to identify eyes that warrant closer early review and earlier consideration of escalation — never to withhold treatment from them.

Declarations

Ethical clearance

Ethical approval was not required. This study analysed aggregate data from previously published reports and involved no human participants, no identifiable individual data and no new intervention.

Acknowledgement

The authors thank the investigators of the included cohorts, whose published tables of counts made a quantitative synthesis possible.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of interest

The authors declare that they have no competing interests.

Author contribution

Both authors contributed to the conception and design of the review, to study selection and data extraction, to the interpretation of the analysis, and to drafting and critically revising the manuscript. Both authors have read and approved the final version and agree to be accountable for all aspects of the work.

Data availability

All data analysed in this review were extracted from the published articles cited in the reference list. The extracted dataset, the risk-of-bias assessments and the analysis outputs are available from the corresponding author on reasonable request.

Generative artificial intelligence use

The authors did not use generative artificial intelligence to generate the scientific content, the data extraction or the statistical analysis reported in this manuscript.

6. References

1. Sasongko MB, Febryanto GA, Haryanto S, et al. Incidence and progression of diabetic retinopathy and blindness in Indonesian adults with type 2 diabetes. PLoS One. 2025; 20(8): e0322093. doi: 10.1371/journal.pone.0322093.
2. Koh BMQR, Chong C, Tan G, et al. Long-term incidence and progression of vision-threatening diabetic retinopathy in Asian populations. J Glob Health. 2026; 16: 04174. doi: 10.7189/jogh.16.04174.
3. Keskini C, Coleman AL, Harris A, et al. Prevalence of and risk factors for diabetic retinopathy: The Thessaloniki Eye Study. Eye (Lond). 2026; [Epub ahead of print]. doi: 10.1038/s41433-026-04698-5.
4. Do DV, Wykoff CC, Sivaprasad S, et al. Intravitreal Aflibercept 8 mg for Diabetic Macular Edema: Ninety-Six-Week Results from the Randomized Phase 2/3 PHOTON Trial. Ophthalmology. 2026; 133(5): 577-588. doi: 10.1016/j.ophtha.2025.10.028.
5. Sheth VS, Schlottmann P, Lai TYY, et al. Four-Year Outcomes of Faricimab in Diabetic Macular Edema: Results from the RHONE-X Extension Trial. Ophthalmology. 2026; 133(5): 599. doi: 10.1016/j.ophtha.2026.01.001.
6. Singh RP, Barakat MR, Ip MS, et al. Efficacy and Safety of Brolucizumab for Diabetic Macular Edema: The KINGFISHER Randomized Clinical Trial. JAMA Ophthalmol. 2023; 141(12): 1152-1160. doi: 10.1001/jamaophthalmol.2023.5248.
7. Owlya N, Barbosa M, Bartolomeo N, et al. Intravitreal faricimab in patients with refractory diabetic macular edema: 6-month fluid analysis using artificial intelligence. BMC Ophthalmol. 2026; 26(1): 119. doi: 10.1186/s12886-026-04651-w.
8. Gao G, Chen Y, Zhang X, et al. Machine learning-based prediction of persistent diabetic macular edema using OCT-derived structural biomarkers and clinical features. Eur J Med Res. 2026; 31(1): 458. doi: 10.1186/s40001-026-04064-x.
9. Chakravarthy U, Chaudhary V, Sadda SR, et al. Effect of Faricimab versus Aflibercept on Hyperreflective Foci in Patients with Diabetic Macular Edema from the YOSEMITE/RHINE Trials. Ophthalmol Sci. 2025; 5(5): 100798. doi: 10.1016/j.xops.2025.100798.
10. Kim SJ, Gangaputra S, Choi L, et al. Inflammatory mediators in diabetic macular edema: A Prospective, Controlled Study. Retina. 2026; 46(9): 1513-1519. doi: 10.1097/IAE.0000000000004893.
11. Demirel S, Yanık Ö, Batıoğlu F, et al. Inflammatory and imaging biomarkers in treatment-naïve diabetic macular edema: Comparison with Irvine-Gass and Cataract Controls. Retina. 2026; 46(9): 1532-1539. doi: 10.1097/IAE.0000000000004845.
12. Yang F, Gong Y, Pan X, et al. Development of a predictive model based on aqueous cytokines data for response to anti-VEGF therapy in diabetic macular edema. Eye Vis (Lond). 2026; 13(1): 12. doi: 10.1186/s40662-026-00479-z.
13. Liu J, Zeng J, Ren P, et al. Elevated carbonic anhydrase-1 in the aqueous humor in diabetic macular edema: associations between inflammatory cytokines and retinal vascular dysfunction. Front Med (Lausanne). 2026; 13: 1764980. doi: 10.3389/fmed.2026.1764980.
14. Lim JI, Amador MJ, Dhoot DS, et al. Anatomic Control with Faricimab versus Aflibercept in the YOSEMITE/RHINE Trials in Diabetic Macular Edema. Ophthalmol Retina. 2025; 9(7): 655-666. doi: 10.1016/j.oret.2025.01.017.
15. Nanji K, Grad J, Hatamnejad A, et al. Baseline OCT Biomarkers Associated with Visual Acuity in Diabetic Macular Edema: A Systematic Review and Meta-analysis. Ophthalmology. 2026; 133(1): 75-90. doi: 10.1016/j.ophtha.2025.07.038.
16. Nanji K, Hatamnejad A, Grad J, et al. Visual outcomes associated with optical coherence tomography biomarkers in diabetic macular edema: A systematic review. Surv Ophthalmol. 2026; 71(2): 289-308. doi: 10.1016/j.survophthal.2025.09.009.
17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021; 372: n71. doi: 10.1136/bmj.n71.
18. Higgins JPT, Morgan RL, Rooney AA, et al. A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E). Environ Int. 2024; 186: 108602. doi: 10.1016/j.envint.2024.108602.
19. Arrigo A, Aragona E, Capone L, et al. Foveal Eversion: A Possible Biomarker of Persistent Diabetic Macular Edema. Ophthalmol Ther. 2021; 10(1): 115-126. doi: 10.1007/s40123-020-00324-z.
20. Cakmak Cengiz E, Yucel OE. Systemic and Ocular Predictive Factors in the Treatment of Diabetic Macular Edema with Bevacizumab. Medicina (Kaunas). 2026; 62(2): 283. doi: 10.3390/medicina62020283.
21. Chang YC, Huang YT, Hsu AY, et al. Optical Coherence Tomography Biomarkers in Predicting Treatment Outcomes of Diabetic Macular Edema after Ranibizumab Injections. Medicina (Kaunas). 2023; 59(3): 629. doi: 10.3390/medicina59030629.
22. Choi MY, Jee D, Kwon JW. Characteristics of diabetic macular edema patients refractory to anti-VEGF treatments and a dexamethasone implant. PLoS One. 2019; 14(9): e0222364. doi: 10.1371/journal.pone.0222364.
23. Ergin E, Dascalu AM, Stana D, et al. Predictive Role of Complete Blood Count-Derived Inflammation Indices and Optical Coherence Tomography Biomarkers for Early Response to Intravitreal Anti-VEGF in Diabetic Macular Edema. Biomedicines. 2025; 13(6): 1308. doi: 10.3390/biomedicines13061308.
24. Gu Z, Xi T, Zhang C, et al. Optical coherence tomography classifications of diabetic macular edema and response to aflibercept: One-year follow-up outcomes in a Chinese population. Medicine (Baltimore). 2023; 102(4): e32815. doi: 10.1097/MD.0000000000032815.
25. Han YE, Jo J, Kim YJ, et al. Factors Affecting Intensive Aflibercept Treatment Response in Diabetic Macular Edema: A Real-World Study. J Diabetes Res. 2023; 2023: 1-11. doi: 10.1155/2023/1485059.
26. Hsieh TC, Deng GH, Chang YC, et al. A real-world study for timely assessing the diabetic macular edema refractory to intravitreal anti-VEGF treatment. Front Endocrinol (Lausanne). 2023; 14: 1108097. doi: 10.3389/fendo.2023.1108097.
27. Lee JH, Shin JY, Ahn J. Comparison of Optical Coherence Tomography Biomarkers between Bevacizumab Good Responders and Nonresponders Who were Switched to Dexamethasone Implant in Diabetic Macular Edema. Korean J Ophthalmol. 2023; 37(2): 137-146. doi: 10.3341/kjo.2022.0109.
28. Li S, Li M, Sun A, et al. Hyperreflective Dots on SD-OCT: Implications for Predicting Treatment Outcomes in Diabetic Macular Edema. Diagnostics (Basel). 2025; 15(19): 2539. doi: 10.3390/diagnostics15192539.
29. Li X, Yang L, Long H, et al. Comparative value of retinal structural features versus systemic inflammatory and renal indicators in predicting short-term anti-VEGF response in diabetic macular edema: focus on OCT biomarkers. Front Med (Lausanne). 2026; 13: 1761324. doi: 10.3389/fmed.2026.1761324.
30. Lu W, Xiao K, Zhang X, et al. A machine learning model for predicting anatomical response to Anti-VEGF therapy in diabetic macular edema. Front Cell Dev Biol. 2025; 13: 1603958. doi: 10.3389/fcell.2025.1603958.
31. Ruiz-Medrano J, Udaondo Mirete P, Fernández-Jiménez M, et al. Biomarkers of risk of switching to dexamethasone implant for the treatment of diabetic macular oedema in real clinical practice: a multicentric study. Br J Ophthalmol. 2025; 109(10): 1155-1160. doi: 10.1136/bjo-2024-325665.
32. Turski CA, Jacobs MA, Abou-Jaoude MM, et al. Short-term outcomes in patients with center-involving diabetic macular edema after a single dose of intravitreal bevacizumab. Int J Retina Vitreous. 2022; 8(1): 81. doi: 10.1186/s40942-022-00430-z.
33. Yu L, Hao X, Cheng J, et al. Predictive effect of TCED-HFV grading and imaging biomarkers on anti-VEGF therapy in diabetic macular edema. BMC Ophthalmol. 2023; 23(1): 232. doi: 10.1186/s12886-023-02973-7.
34. Zhou ZH, Zhao L, Wang YL, et al. Predictive impact of serous retinal detachment in refractory diabetic macular edema. BMC Ophthalmol. 2025; 25(1): 177. doi: 10.1186/s12886-025-03993-1.
35. Zhu T, Wang Y, Hua Y, et al. Analysis of OCT-Based Biomarkers and Recurrence in Eyes with Diabetic Macular Edema Following Anti-VEGF Therapy. Int J Gen Med. 2025; 18: 6941-6948. doi: 10.2147/IJGM.S556385.
36. Cicinelli MV, Leonardo B, Maiucci G, et al. What Lies beneath Diabetic Macular Edema: Latent Phenotypic Clustering and Differential

Treatment Responses to Intravitreal Therapies. Ophthalmol Sci. 2026; 6(1): 100975. doi: 10.1016/j.xops.2025.100975.
37. Wang Y, Li F, Hao M, et al. Predictive value of hyperreflective foci for anti-VEGF therapeutic outcomes in different subtypes of diabetic macular edema: a retrospective analysis. Front Endocrinol (Lausanne). 2025; 16: 1648828. doi: 10.3389/fendo.2025.1648828.
38. Qu Y, Wan H, Zhou H, et al. Region-Specific Retinal and Choroidal Biomarkers Associated with Visual Outcomes After Anti-VEGF Treatment in Center-Involved Diabetic Macular Edema. Clin Ophthalmol. 2026; 20: 1-16. doi: 10.2147/OPTH.S566570.
39. Okudan S, Acar Duyan S, Erdem A, et al. Optical Coherence Tomography Biomarkers Predict the Long-Term Restorative Effect of Early Anti-VEGF Treatment on Diabetic Macular Edema. Life (Basel). 2025; 15(2): 269. doi: 10.3390/life15020269.
40. Karataş ME, Çakır A, Arıcı M, et al. Evaluation of anti-VEGF response in diabetic macular edema patients with baseline DRIL: a multicenter real-world study (BOSPHORUS-DME Study Group Report No.16). Graefes Arch Clin Exp Ophthalmol. 2026; [Epub ahead of print]. doi: 10.1007/s00417-026-07308-y.
41. Alam T, Das N, Maatouk C, et al. Evaluating the long-term predictive value of macular thickness fluctuations on diabetic macular oedema response to anti-VEGF treatment. Eye (Lond). 2025; 39(16): 2933-2939. doi: 10.1038/s41433-025-03968-y.
42. Condelipes A, Correia D, Fernandes I, et al. Thickness profile of the ganglion cell complex and choroid in patients with persistent diabetic macular edema. Comput Biol Med. 2025; 198(Pt A): 111192. doi: 10.1016/j.compbiomed.2025.111192.
43. Han YE, Choi L, Jung D, et al. Choroidal Vascular Remodeling and Retinal Response After Aflibercept Loading in Diabetic Macular Edema. Ophthalmol Sci. 2026; 6(2): 101025. doi: 10.1016/j.xops.2025.101025.
44. Zhou G, Lin X, Yan X, et al. Quantitative relative reflectivity features based on OCT for predicting the efficacy of anti-VEGF therapy in diabetic macular edema at 3 months. Photodiagnosis Photodyn Ther. 2026; 59: 105445. doi: 10.1016/j.pdpdt.2026.105445.
45. Viggiano P, Carra A, Bruno G, et al. Microaneurysm Reflectivity as a Prognostic Biomarker for Intravitreal Treatment Response in Diabetic Retinopathy. Am J Ophthalmol. 2026; 285: 39-48. doi: 10.1016/j.ajo.2026.01.027.
46. Tekkaya NE, Oruc A, Oruc KY, et al. Association of OCT Angiography and Multifocal Electroretinography Findings with Early Anatomical Response to Intravitreal Bevacizumab in Diabetic Macular Oedema. Photodiagnosis Photodyn Ther. 2026; 105611. doi: 10.1016/j.pdpdt.2026.105611.
47. Pak K, Yoon C, Sadda SR. Hard exudates in diabetic macular edema after intravitreal anti-VEGF therapy: a post-hoc analysis of the DRCR protocol T trial. Graefes Arch Clin Exp Ophthalmol. 2026; 264(2): 427-436. doi: 10.1007/s00417-025-07015-0.
48. Fasolino G, Lazaar M, Della Rocca DG, et al. Predictive Value of Optical Coherence Tomography Biomarkers in Patients with Persistent Diabetic Macular Edema Undergoing Cataract Surgery Combined with a Dexamethasone Intravitreal Implant. Bioengineering (Basel). 2025; 12(5): 556. doi: 10.3390/bioengineering12050556.
49. Mihalache A, Pandya BU, Jhaveri A, et al. Optical Coherence Tomography Biomarkers as Potential Predictors for Visual Function and Response to Intravitreal Therapy in Diabetic Macular Edema. J Vitreoretin Dis. 2026; 24741264261457974. doi: 10.1177/24741264261457974.
50. Kaplan OC, Yazar Z, Onen M, et al. Effect of optical coherence tomography biomarkers on treatment outcomes in patients with diabetic macular edema. Int Ophthalmol. 2026; 46(1): 146. doi: 10.1007/s10792-026-04029-3.
51. El Achour B, Ben Abderrazek A, Zahaf A, et al. Predictive factors for non-response to intravitreal bevacizumab in diabetic macular edema. J Fr Ophtalmol. 2026; 49(4): 104832. doi: 10.1016/j.jfo.2026.104832.
52. Jennings E, Chen D, Geevarghese A, et al. Hispanic Ethnicity and Optical Coherence Tomography-based Biomarkers as Predictive Factors of Diabetic Macular Edema Refractory to Bevacizumab. Ophthalmic Surg Lasers Imaging Retina. 2025; 56(6): 346-353. doi: 10.3928/23258160-20250127-02.
53. Chatziralli I, Dimitriou E, Lambadiari V, et al. The Impact of Laboratory Findings and Optical Coherence Tomography Biomarkers on Response to Intravitreal Anti-VEGF Treatment in Patients with Diabetic Macular Edema. Semin Ophthalmol. 2022; 37(5): 668-675. doi: 10.1080/08820538.2022.2069470.
54. Szeto SK, Hui VWK, Tang FY, et al. OCT-based biomarkers for predicting treatment response in eyes with centre-involved diabetic macular oedema treated with anti-VEGF injections: a real-life retina clinic-based study. Br J Ophthalmol. 2023; 107(4): 525-533. doi: 10.1136/bjophthalmol-2021-319587.