Obesity modifies response to IL-23p19 inhibitors in psoriasis: systematic review and meta-analysis

Jacopo Tartaglia, Carlo Garbo', Emanuele Zanardo, Mauro Alaibac, Stefano Piaserico

Reproducibility Materials

1. Search Strategy

1.1 MEDLINE (Ovid)

Search performed from inception to the 1st October 2025.

The following search strategy was used for MEDLINE (Ovid):

1. exp psoriasis/
2. (psorias$ or "plaque psoriasis").ti,ab.
3. 1 or 2
4. exp guselkumab/ or guselkumab.ti,ab. or tremfya.ti,ab.
5. exp risankizumab/ or risankizumab.ti,ab. or skyrizi.ti,ab.
6. exp tildrakizumab/ or tildrakizumab.ti,ab. or ilumetri.ti,ab. or ilumya.ti,ab.
7. exp mirikizumab/ or mirikizumab.ti,ab. or omvoh.ti,ab.
8. 4 or 5 or 6 or 7
9. 3 and 8
10. limit 9 to humans

1.2 Embase (Ovid)

Search performed from inception to the 1st October 2025.

The following search strategy was used for Embase (Ovid):

1. 'psoriasis'/exp
2. (psorias* OR 'plaque psoriasis'):ti,ab
3. 1 OR 2
4. 'guselkumab'/exp OR guselkumab:ti,ab OR tremfya:ti,ab
5. 'risankizumab'/exp OR risankizumab:ti,ab OR skyrizi:ti,ab
6. 'tildrakizumab'/exp OR tildrakizumab:ti,ab OR ilumetri:ti,ab OR ilumya:ti,ab
7. 'mirikizumab'/exp OR mirikizumab:ti,ab OR omvoh:ti,ab
8. 4 OR 5 OR 6 OR 7
9. 3 AND 8
10. limit 9 to human

1.3 Rationale

To maximise sensitivity, the search combined subject headings and free-text terms for psoriasis and IL-23p19 inhibitors (guselkumab, risankizumab, tildrakizumab, mirikizumab). BMI- or obesity-related terms were intentionally omitted to avoid filtering out studies that reported obesity-stratified outcomes only within subgroup analyses. No date restrictions were applied. Duplicates were removed before screening.

2. Study Selection

Studies were eligible for inclusion if they met all of the following criteria:

1. original research published in English;
2. adult patients (≥18 years) with plaque psoriasis;
3. treatment with an approved IL-23p19 inhibitor (guselkumab, risankizumab, mirikizumab or tildrakizumab) at standard dosing;
4. obesity stratification using BMI ≥30 kg/m^2 versus <30 kg/m^2; and
5. extractable PASI90 and/or PASI100 response rates at induction (approximately weeks 12–24) and/or maintenance (approximately weeks 44–60).

We included both randomized controlled trials (RCTs) and real-world evidence (RWE) studies (prospective and retrospective cohorts, and registry-based analyses). Post-hoc analyses of RCTs were eligible only if obesity-stratified outcomes were reported. Studies that grouped overweight and obese patients together without separate reporting for BMI ≥30 kg/m^2, or that used different BMI thresholds (e.g., BMI ≥25 kg/m^2), were excluded. Case reports, case series, reviews, and studies reporting only patient-reported outcomes without PASI were excluded.

For studies in which an eligible outcome was reported at multiple timepoints within the predefined windows, we selected the assessment closest to week 16 for induction and to week 52 for maintenance. When analyses stratified by weight (<90 kg vs ≥90 kg) or by BMI quartiles were presented in addition to BMI categories, only the BMI ≥30 kg/m^2 versus <30 kg/m^2 comparison was used. If subgroup denominators at the follow-up visit were not available, baseline subgroup counts were used as denominators; in such cases, this was documented in the extraction form. A pre-planned sensitivity analysis was performed excluding all studies for which baseline denominators were used.

Two reviewers (JT and CG) independently screened titles and abstracts, followed by full-text evaluation. Disagreements were resolved by consensus. No restrictions were applied on geographic setting, enrollment period, or prior biologic exposure.

3. Data Extraction and Outcomes

Data were extracted using a standardized form collecting study design, biologic agent, sample size, analysis population, follow-up timepoints, and the number of obese (BMI ≥30 kg/m^2) and non-obese patients achieving PASI90 and/or PASI100.

Induction was defined as the first post-baseline efficacy assessment within weeks 12–24; when multiple assessments were available, the timepoint closest to week 16 was used. Maintenance was defined as an assessment within weeks 44–60, with preference for the timepoint closest to week 52. RCT-derived analyses and RWE cohorts were analyzed separately.

Responder counts were taken as reported. When only percentages were available, responder numbers were derived by applying the reported percentage to the stated denominator. If subgroup-specific denominators at follow-up were unavailable, baseline subgroup counts were used. In studies reporting overall attrition without subgroup-specific losses, dropouts were allocated proportionally to the baseline obese/non-obese distribution. All such instances were flagged in the extraction form, and a pre-planned sensitivity analysis was conducted excluding studies that required denominator substitution or derived responder counts. Two reviewers (JT and CG) independently extracted data form the retrieved papers. Disagreements were resolved by consensus.

The primary outcomes were PASI90 and PASI100 at induction and maintenance. PASI90 was defined as a ≥90% reduction from baseline PASI, and PASI100 as complete clearance.

4. Statistical Analysis

All meta-analyses were performed in R (version 4.5.1) using the *meta* package. For each study, risk ratios (RRs) with 95% confidence intervals (CIs) were calculated comparing obese with non-obese patients. Pooled estimates were obtained using Mantel–Haenszel random-effects models. Heterogeneity was quantified using the I^2 statistic and τ^2.

Several pre-planned sensitivity analyses were conducted for the RWE data:

1. exclusion of studies requiring substitution of baseline denominators at follow-up or in which responder counts were derived from percentages or figures rather than directly reported;
2. exclusion of studies with potential overlap in patient populations; and
3. leave-one-out analyses when models were estimable.

For RCTs, additional sensitivity analyses were performed by repeating the RCT meta-analyses after excluding individual trial arms (tildrakizumab 100 mg, tildrakizumab 200 mg, guselkumab 100 mg, or risankizumab 150 mg).

Small-study effects were assessed for outcomes with ≥10 studies using funnel plots and Harbord and Egger regression tests. When asymmetry was detected, trim-and-fill analyses were used to explore the impact of potentially missing studies. No adjustment for multiple comparisons was applied.

5. PRISMA flow diagram

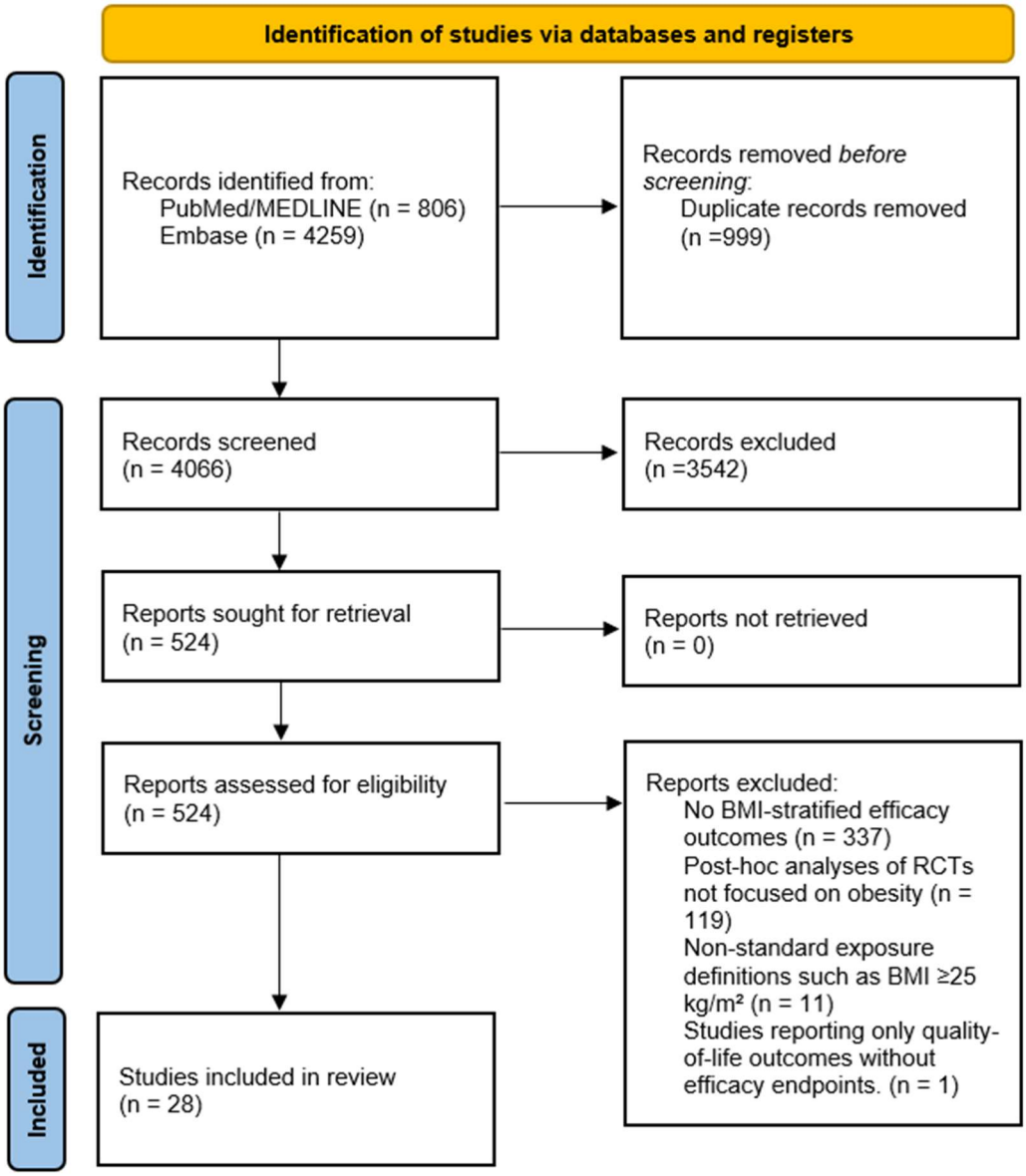

Fig. S0 PRISMA flow diagram of study identification, screening, eligibility assessment, and inclusion for the systematic review and meta-analysis.

6. Risk of bias assesment

Risk of bias was assessed using RoB 2 for randomized trials and ROBINS-I for RWE studies.

For RCTs, RoB 2 domains were: randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. An overall judgement was assigned per trial.

For RWE, ROBINS-I domains were: confounding, selection of participants, classification of interventions, deviations from intended interventions, missing data, measurement of outcomes, and selection of the reported result. An overall judgement was assigned per study. No study was excluded from quantitative synthesis based on risk-of-bias assessment alone.

Across the included RCT datasets, risk of bias was judged as low for domains related to randomization, deviations from intended intervention, missing outcome data, and outcome measurement, while “some concerns” were assigned to selection of the reported result (D5). Accordingly, the overall RoB 2 judgement was “some concerns” for all included trials [Fig. S1].

Across the included RWE cohorts, ROBINS-I ratings were split between moderate and serious overall risk of bias. Studies judged at serious risk were mainly driven by residual confounding (limited adjustment for baseline differences between obese and non-obese patients) and missing outcome data (incomplete follow-up with insufficient information on attrition mechanisms). In contrast, classification of exposure and outcome measurement were generally rated at low risk, while selective reporting was typically rated as moderate [Fig. S2].

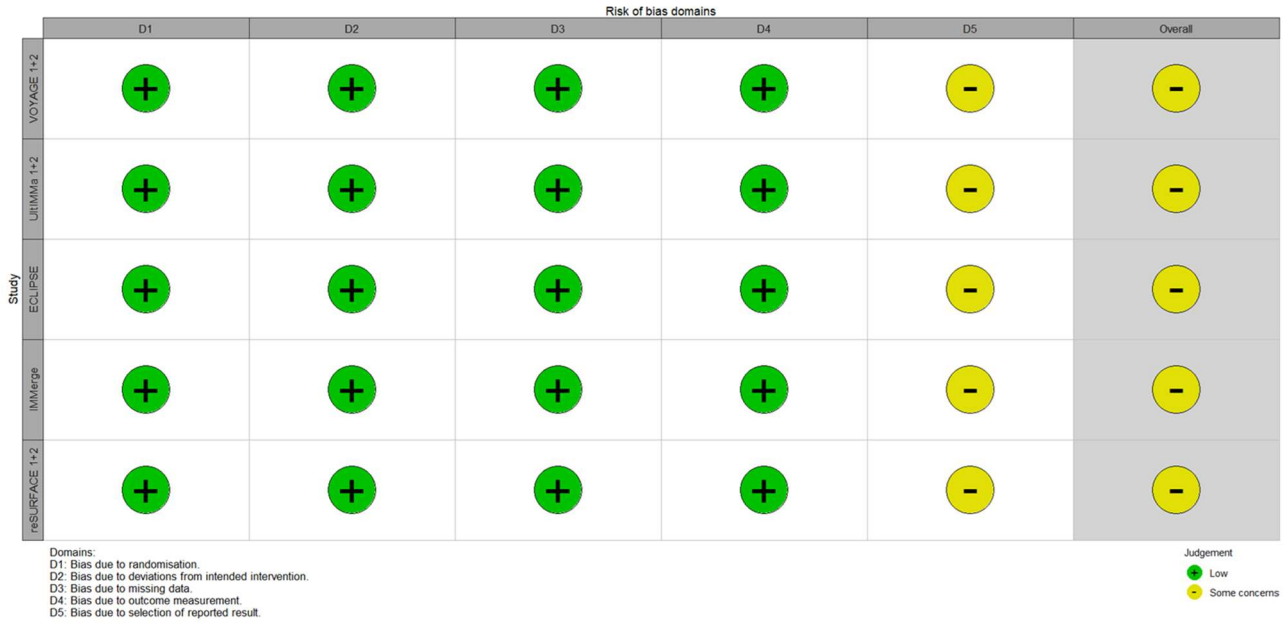

Fig. S1 Risk of bias assessment for the included randomized trials using the Cochrane RoB 2 tool, reported by domain (D1–D5) and overall judgement.

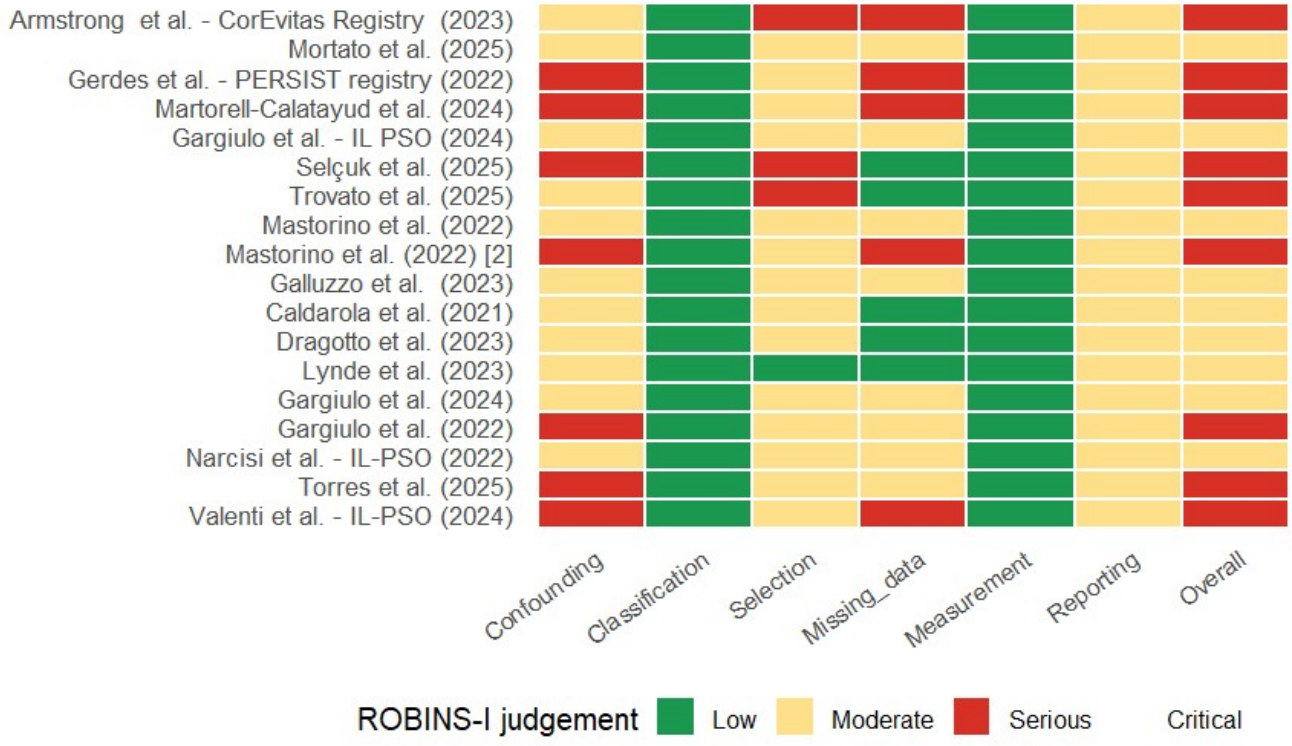

Fig. S2 Domain-level risk of bias for the included real-world cohorts using ROBINS-I, including confounding, exposure classification, participant selection, missing data, outcome measurement, reporting, and overall judgement.

7. Results

Table S1 summarizes PASI90 and PASI100 responder counts by BMI category across registrational RCTs of guselkumab (VOYAGE 1/2, ECLIPSE), risankizumab (UltIMMa 1/2, IMMerge), and tildrakizumab (reSURFACE 1/2). Timepoints include early assessment (induction; 12–24 weeks) and longer follow-up (maintenance; later than week 28). Responder counts are reported separately for PASI90 and PASI100 at each available timepoint.

Table S2 includes maintenance-phase PASI90 and PASI100 responder counts from real-world cohorts treated with guselkumab, risankizumab, or tildrakizumab. In addition to outcome counts by BMI category, the table reports key cohort descriptors (mean BMI, proportion of obese patients, mean age, baseline PASI, and biologic-naïve proportion when stated), providing context for cross-study differences in populations and follow-up.

TABLE S1: Counts of PASI 90 and PASI 100 events and corresponding study populations. Parentheses indicate the total number of obese and non-obese patiens at a certain timepoint.

Trial ID	***Drug***	***Timepoint***	***PASI 90***		***PASI 100***		***Timepoint***	***PASI 90***		***PASI 100***	
			BMI $\geq$ 30	BMI $<$ 30	BMI $\geq$ 30	BMI $<$ 30		BMI $\geq$ 30	BMI $<$ 30	BMI $\geq$ 30	BMI $<$ 30
VOYAGE 1 and 2	Guselkumab 100 mg	24w			151 (333)	279 (492)					
ECLIPSE	Guselkumab 100 mg						48w	184 (223)	266 (310)	117 (223)	194 (310)
UltIMMa 1 and 2	Risankizumab 150 mg	16w	188 (255)	212 (270)	105 (255)	119 (270)	52w	214 (255)	238 (270)	143 (255)	176 (270)
IMMerge	Risankizumab 150 mg						52w	66 (76)	76 (88)		
reSURFACE 1 and 2	Tildrakizumab 100 mg	12w	85 (266)	143 (348)	29 (266)	52 (348)	52w	165 (266)	271 (348)	64 (266)	139 (348)
reSURFACE 1 and 2	Tildrakizumab 200 mg	12w	61 (243)	163 (379)	27 (243)	53 (379)	52w	165 (243)	303 (379)	83 (243)	186 (379)

TABLE S2: Counts of PASI 90 and PASI 100 events during maintenance and characteristics of the corresponding populations in real-life studies. The parentheses indicate the total number of obese and non-obese patients at a given time.

Author	*Drug*	*Mean BMI*	*Obese population (%)*	*Mean age*	*Baseline PASI*	*Biologic naive (%)*	*Timepoint*	*PASI 90*		*PASI 100*	
								BMI $\geq$ 30	BMI $<$ 30	BMI $\geq$ 30	BMI $<$ 30
Armstrong et al.	Guselkumab	32.6	56.1	50.2	10.5	27.2	52w	46 (101)	38 (77)	33(101)	30 (77)
Caldarola et al.	Tildrakizumab	27.5		53.5	14.2	59.3	28w	14 (21)	24 (38)		
Dragotto et al.	Risankizumab	27.4	19.6	56.0	14.9	18.6	36w	14 (20)	61 (82)	10 (20)	43 (82)
Galluzzo et al.	Guselkumab	28.2	32.8	51.1	16.2	35.2	52w	24 (30)	56 (63)	18 (30)	49 (63)
Gargiulo et al.	Risankizumab	28.8	38.6	51.0	13.5	60.3	52w	20 (32)	41 (51)	16 (32)	31 (51)
Gargiulo et al.	Guselkumab	25.5	21.6	52.4	11.2	51.0	52w	12 (21)	56 (74)	12 (21)	41 (74)
Gargiulo et al.	Risankizumab	28.0	27.1	51.4	15.7	57.4	52w	190 (231)	501 (611)	152 (231)	407 (611)
Gerdes et al.	Guselkumab	29.7		49.7	16.4		28w	44 (99)	103 (167)	25 (99)	52 (167)
Lynde et al.	Guselkumab	29.0	36.7	44.2	14.6	41.3					
Lynde et al.	Risankizumab	28.6	35.0	44.1	15.4	57.1					
Martorell et al.	Risankizumab	30.2	44.5	51.7	11.4	18.1	52w			73 (132)	108 (164)
Mastorino et al.	Risankizumab	27.0	22.0	65.1	12.5	42	40w	4 (8)	18 (22)	1 (8)	15 (22)
Mastorino et al.	Risankizumab	26.8	23.3	52.0	14.8	55.5	52w	18 (26)	78 (91)	17 (26)	71 (91)
Mortato et al.	Guselkumab		24.5	49.8	15.2	63.1	52w	165 (230)	564 (701)	139 (230)	482 (701)
Narcisi et al.	Tildrakizumab	26.2	19.6	48.6	14.4	81.0	52w	15 (19)	66 (77)	14 (19)	50 (77)
Selçuk et al.	Risankizumab		35.9	45.3	17.0	22.0	52w			7 (14)	15 (27)
Selçuk et al.	Guselkumab		32.7	43.0	18.2	51.0	52w			10 (16)	17 (33)
Torres et al.	Risankizumab	27.3	27.7	46.8	15.0	56.0	52w			74 (124)	203 (331)
Trovato et al.	Tildrakizumab	29.0	39.4	53.9	14.8	33.6	36w	37 (54)	64 (83)	24 (54)	51 (83)
Valenti et al.	Guselkumab	26.9	18.3	53.5	9.2	0	52w	14 (25)	80 (112)	7 (25)	73 (112)

Forest plots for the additional RCT outcomes (induction PASI90 and PASI100, and maintenance PASI100) are reported in Figures S3–S4.

At induction, four phase-3 programmes (VOYAGE, UltIMMa, and reSURFACE 1–2 at 100 and 200 mg) contribute data for both PASI90 and PASI100. For PASI90, the pooled RR for obese versus non-obese patients is below 1 and of similar magnitude to the primary maintenance PASI90 analysis, with marked between-trial heterogeneity (I^2=83.9%; τ^2=0.0200). For PASI100, the pooled RR is below 1, and study estimates are tightly grouped with no relevant heterogeneity (I^2=0%; τ^2=0).

For maintenance PASI100, four arms (ECLIPSE, UltIMMa 1–2, and reSURFACE 1–2 at 100 and 200 mg) contributed. The pooled RR remains below 1, with moderate-to-high heterogeneity (I^2=65.1%; τ^2=0.0152) [Fig. S4].

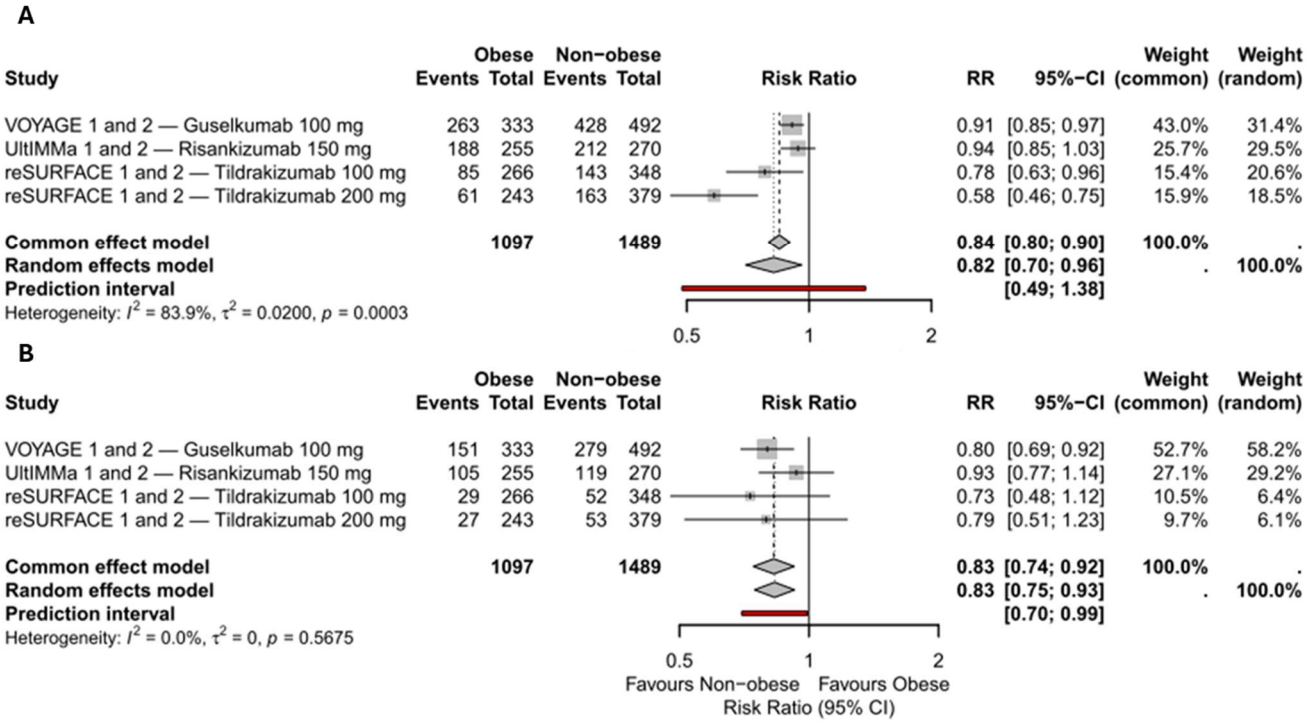

Fig. S3 Induction outcomes in randomized trials. (A) PASI90 response and (B) PASI100 response at induction in phase-3 trials of guselkumab, risankizumab and tildrakizumab, comparing obese with non-obese patients. Random-effects Mantel–Haenszel models: pooled RR 0.82 (95% CI 0.70–0.96; I^2=83.9%) for PASI90 and 0.83 (95% CI 0.75–0.93; I^2=0%) for PASI100.

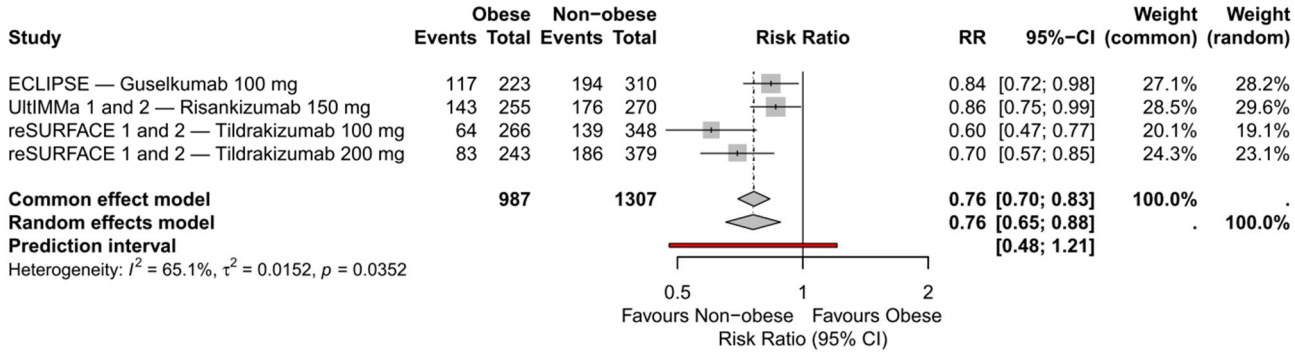

Fig. S4 PASI100 maintenance response in randomized trials. Forest plot of PASI100 at maintenance in phase-3 trials of guselkumab, risankizumab and tildrakizumab, comparing obese with non-obese patients. Random-effects Mantel–Haenszel model: pooled RR 0.76 (95% CI 0.65–0.88; I^2=65.1%).

Supplementary forest plots for PASI100 in the real-world cohorts are shown in Figures S5.

For induction PASI100, 17 cohorts are included. The random-effects pooled RR for obese versus non-obese patients is 0.86 (95% CI 0.79–0.94), with no detectable heterogeneity (I^2=0%; τ^2=0).

For maintenance PASI100, 17 cohorts contribute data. The pooled RR is 0.91 (95% CI 0.85–0.97), with low heterogeneity (I^2=16.6%; τ^2=0.0014).

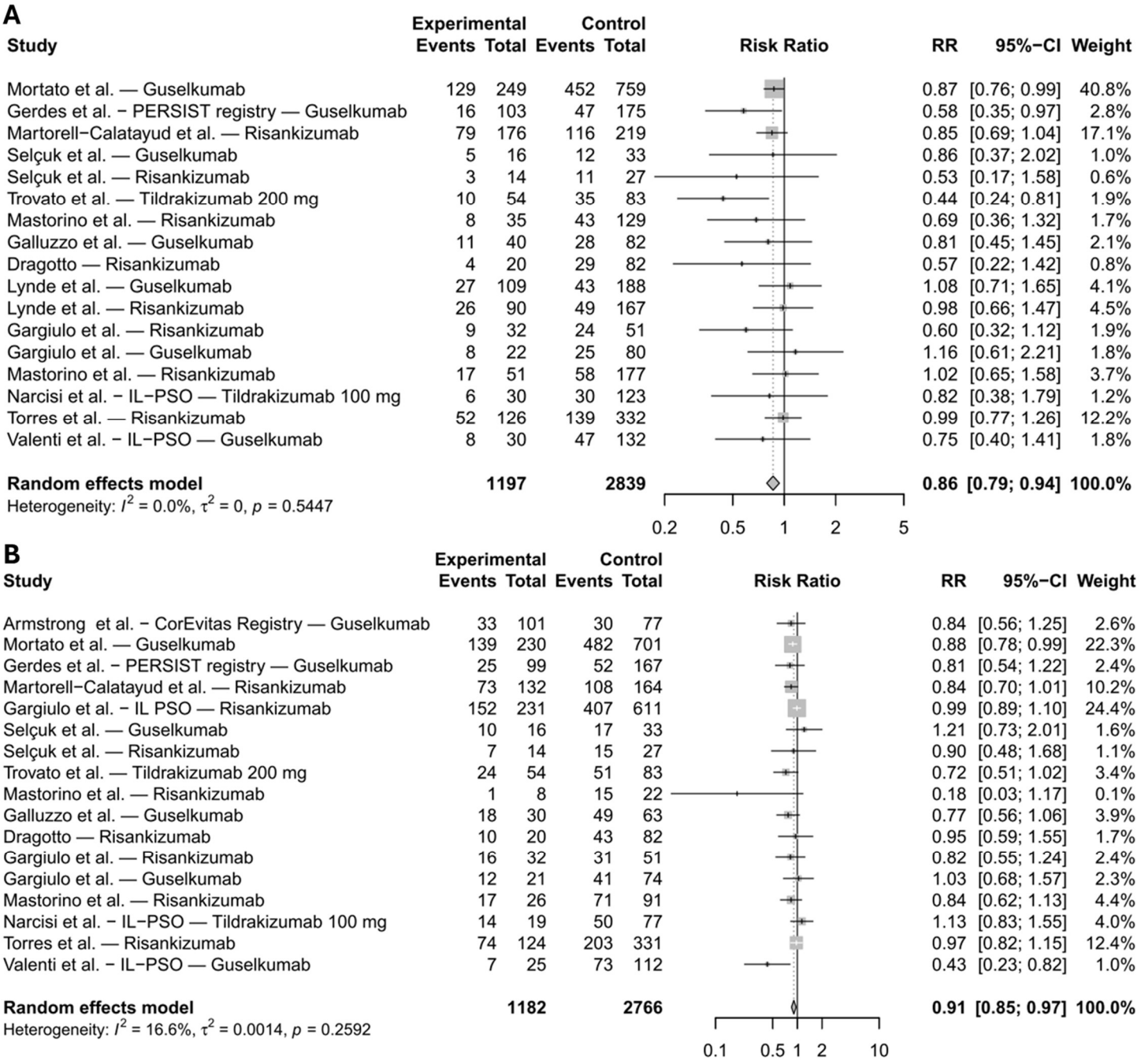

Fig. S5 PASI100 response in real-world cohorts. (A) Induction PASI100 and (B) maintenance PASI100 in real-world studies of guselkumab, risankizumab and tildrakizumab, comparing obese with non-obese patients. Random-effects Mantel–Haenszel models: pooled RR 0.86 (95% CI 0.79–0.94; I^2=0%) at induction and 0.91 (95% CI 0.85–0.97; I^2=16.6%) at maintenance.

8. Sensitivity analyses

8.1 Sensitivity analyses for RCTs

Sensitivity analyses were performed by excluding individual trial arms from the RCT meta-analyses.

For induction PASI90, exclusion of the tildrakizumab 100 mg arm left three trial datasets (k=3) and gave a pooled RR of 0.83 (95% CI 0.69–0.99; I^2=87.4%). Heterogeneity remained high.

For induction PASI100, removal of the tildrakizumab 100 mg arm also left three datasets (k=3), with an RR of 0.84 (95% CI 0.75–0.94) and no detectable between-study heterogeneity (I^2=0%).

For maintenance PASI90, exclusion of the tildrakizumab 100 mg arm (k=4) gave an RR of 0.94 (95% CI 0.88–1.00), with I^2 decreasing from 73.7% in the main analysis to 50.8% (prediction interval 0.79–1.12; Figure S6). Excluding the tildrakizumab 200 mg arm (k=4) produced an RR of 0.93 (95% CI 0.85–1.01; I^2=74.6%). Removal of all guselkumab data (k=4) gave an RR of 0.90 (95% CI 0.81–1.00; I^2=77.8%), and exclusion of all risankizumab arms (k=3) resulted in an RR of 0.87 (95% CI 0.77–0.98; I^2=79.3%).

For maintenance PASI100, exclusion of the tildrakizumab 100 mg arm left three datasets (k=3), with an RR of 0.81 (95% CI 0.72–0.91) and I^2=36.9%.

Overall, these RCT sensitivity analyses show that the pooled RRs remain close to those of the primary models, with a small disadvantage for obese patients and persistent heterogeneity for PASI90 despite the exclusion of individual arms.

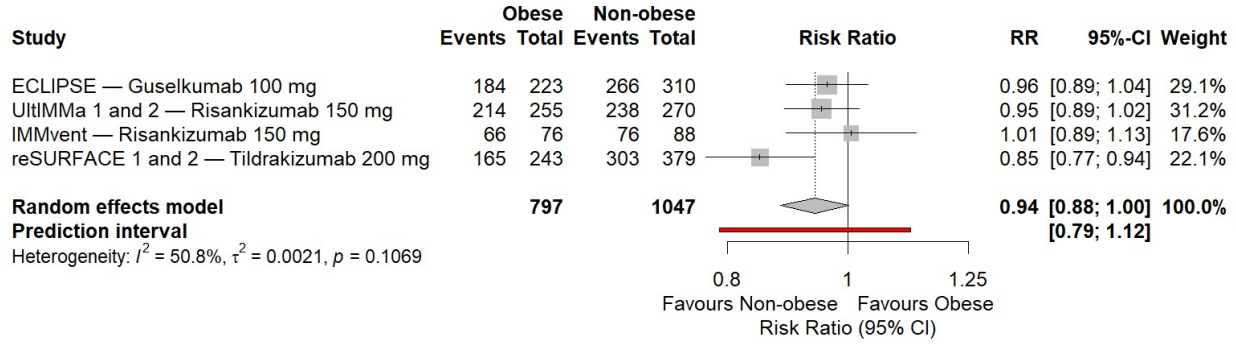

Fig. S6 Sensitivity analysis of PASI90 maintenance response in randomized trials after exclusion of the tildrakizumab 100 mg arm. Random-effects Mantel–Haenszel model; obesity vs non-obesity (RR 0.94; 95% CI 0.88–1.00; I^2=50.8%; prediction interval 0.79–1.12).

Finally, for the primary RCT outcome (maintenance PASI90, k=5), we also applied Hartung–Knapp–Sidik–Jonkman adjustment and inspected a funnel plot (Figure S7). The Hartung–Knapp model gave an RR of 0.91 (95% CI 0.81–1.02; p=0.09). Egger's test did not show evidence of small-study effects (p=0.47).

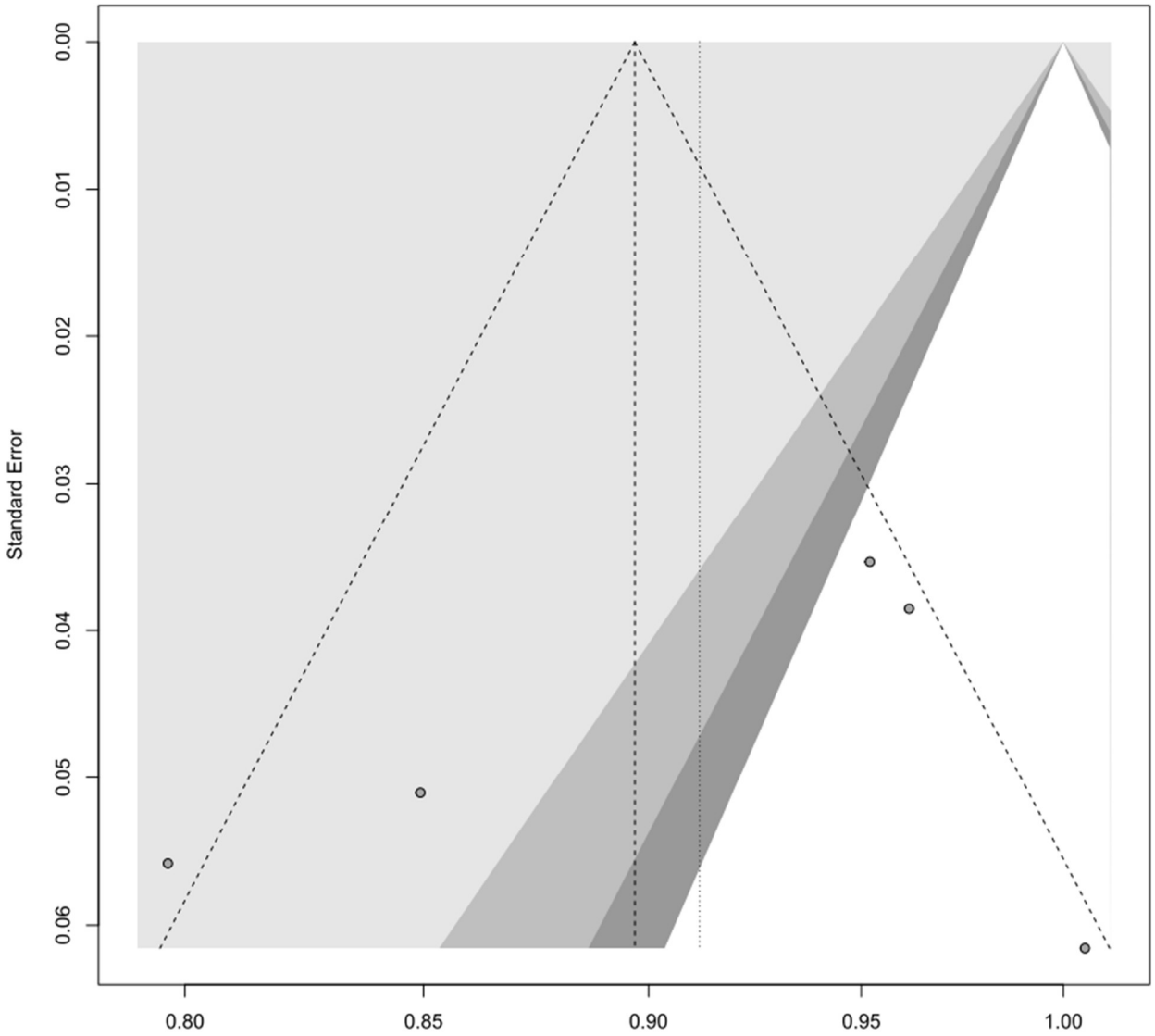

Figure S7. Funnel plot for the primary RCT meta-analysis of PASI90 maintenance response (random-effects Mantel–Haenszel model). Each point represents one trial; the vertical line marks the pooled RR. The plot does not show marked asymmetry.

8.2 Sensitivity analyses for RWE

In RWE analyses, excluding all studies in which responder counts were reconstructed from percentages or baseline denominators did not materially change the results. For induction, the pooled RR for PASI90 was 0.78 (95% CI 0.68–0.91; I^2=28.6%; k=11) and 0.81 (95% CI 0.68–0.96; I^2=2.4%; k=13) for PASI100. For maintenance, the pooled RR for PASI90 was 0.87 (95% CI 0.79–0.95; I^2=37.0%; k=10) and 0.89 (95% CI 0.79–1.00; I^2=35.7%; k=12) for PASI100.

A second sensitivity analysis excluding only estimates derived from figures for maintenance PASI90 yielded an RR of 0.90 (95% CI 0.83–0.97; I^2=24.6%; k=12), close to the primary model. These findings indicate that the observed obesity effect in RWE is not driven by back-calculated data.

When potentially overlapping cohorts were excluded, estimates were similar to the primary models. For induction, the pooled RR for PASI90 was 0.76 (95% CI 0.61–0.95; I^2=46.1%; k=6) and 0.86 (95% CI 0.75–0.98; I^2=15.8%; k=10) for PASI100. For maintenance, the pooled RR for PASI90 was 0.90 (95% CI 0.80–1.01; I^2=40.2%; k=6) and 0.92 (95% CI 0.85–1.01; I^2=33.1%; k=10) for PASI100.

A leave-one-out analysis was feasible only for the RWE induction PASI100 model, which was the RWE outcome with the largest number of contributing cohorts and without relevant sparse-data issues. When each cohort was removed in turn, the pooled RR changed only minimally around the primary estimate and the direction of the association did not change [Fig. S8]. For the other RWE outcomes, the number of studies and events was insufficient to support a leave-one-out meta-analysis.

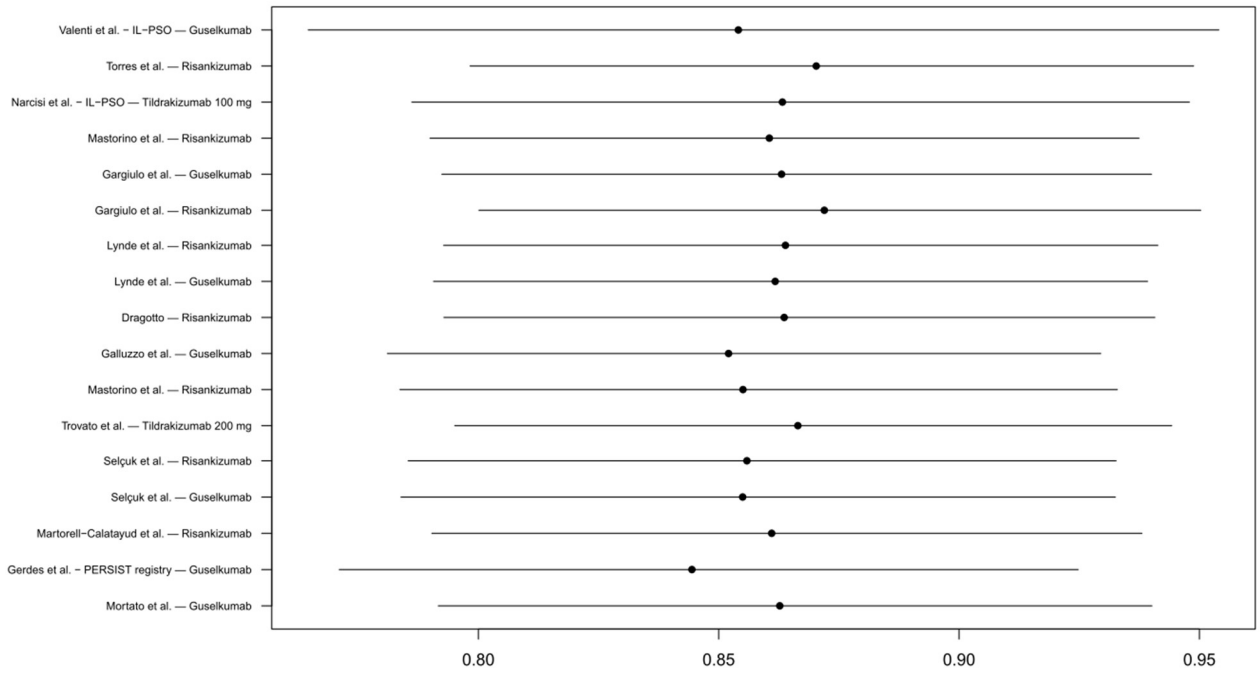

Figure S8. Leave-one-out analysis for real-world induction PASI100 response. Each point represents the pooled RR obtained after excluding one cohort from the meta-analysis. The estimates remain clustered around the primary RR, indicating that no single study drives the overall association between obesity and treatment response.

Funnel plots were examined for all four RWE outcomes (induction and maintenance PASI90 and PASI100). For induction PASI90 and PASI100, the points were approximately symmetrically distributed around the pooled RR. A similar pattern was seen for maintenance PASI100. These plots do not suggest major small-study effects for these outcomes.

In contrast, the funnel plot for the primary RWE outcome (maintenance PASI90) showed asymmetry (Harbord p=0.026; Egger p=0.014). [Fig. S9]. This pattern is compatible with a small-study effect.

To explore this further, a trim-and-fill procedure was applied to the maintenance PASI90 model [Fig. S10]. Additional studies were imputed on the under-represented side of the funnel, and the adjusted pooled RR moved closer to 1 compared with the primary estimate (RR 0.94; 95% CI 0.88–1.00; p=0.067). The direction of the association remained unchanged, but the magnitude of the effect was attenuated.

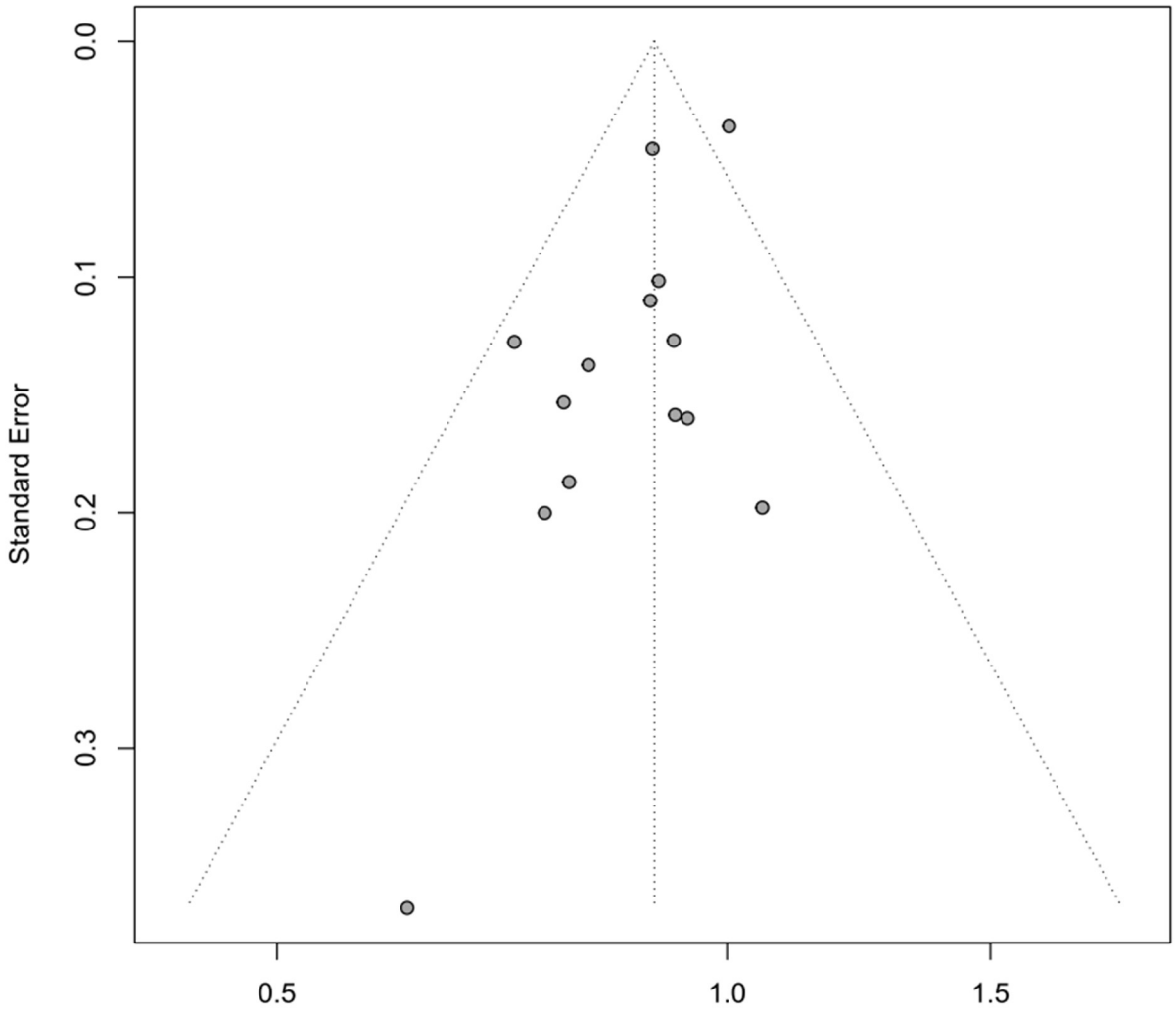

Fig. S9 Funnel plot for the real-world meta-analysis of PASI90 maintenance response (random-effects Mantel–Haenszel model). Each dot represents one cohort; the vertical line marks the pooled risk ratio.

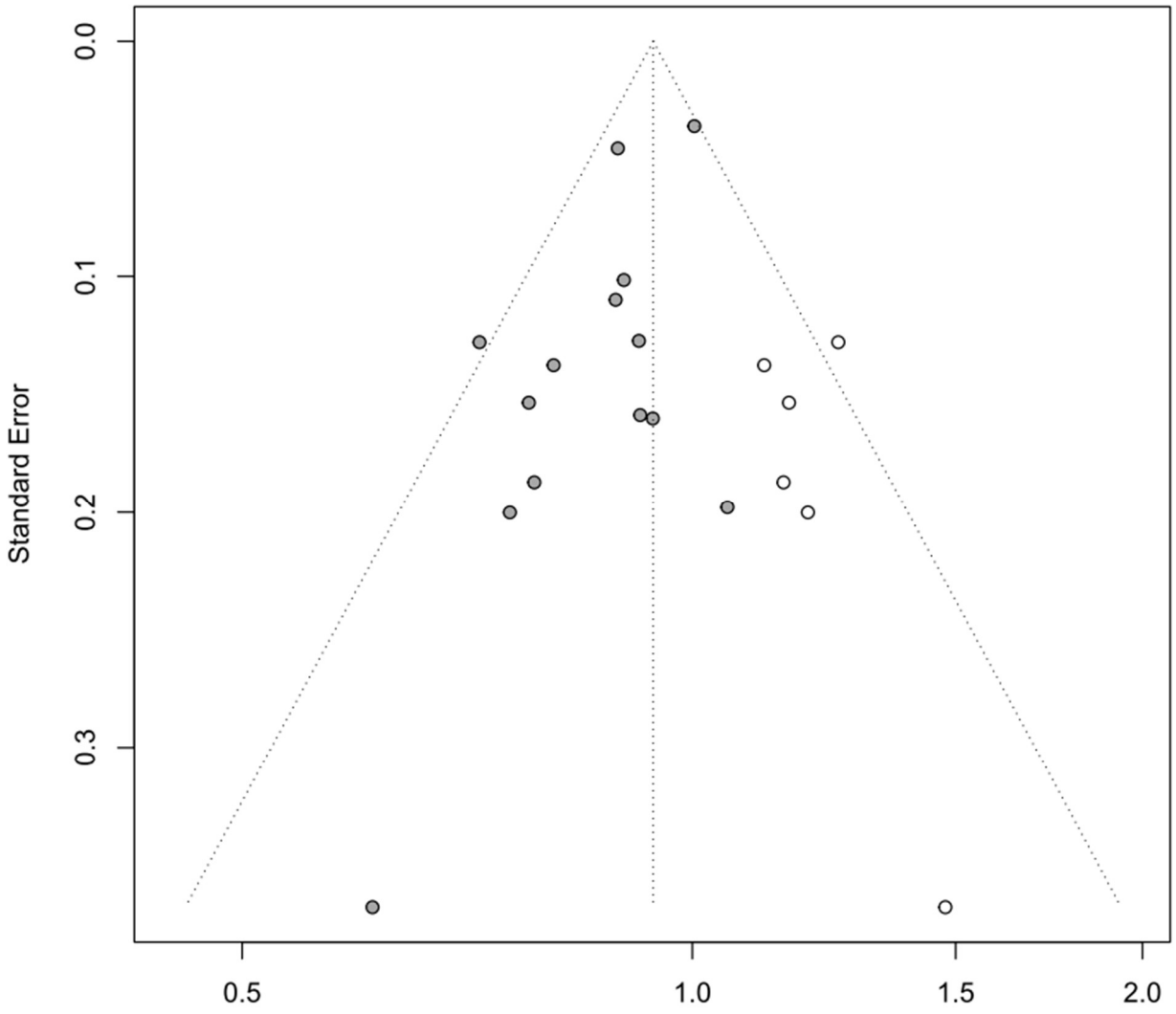

Fig. S10 Trim-and-fill funnel plot for the real-world meta-analysis of PASI90 maintenance response. Solid dots are observed cohorts and open dots are studies imputed by the trim-and-fill procedure; the vertical line marks the pooled risk ratio after trim-and-fill.

9. Certainty of evidence

Evidence	Studies (k)	Participants (Obese / Non-obese)	PASI90 responders (Obese / Non-obese)	RR (95% CI)	Absolute effect (per 1000; obese vs non-obese)	Certainty
RCTs	5 trials	1063 / 1395	794 / 1154	0.91 (0.84–0.99)	753 (695–819) vs 827	Moderate
RWE	14 cohorts	917 / 2249	617 / 1750	0.89 (0.84–0.96)	693 (654–747) vs 778	Low

Table S3. GRADE Summary of Findings for PASI90 response at maintenance. Responder counts are shown as n/N within each stratum. Absolute effects are expressed as expected responders per 1000 obese patients, using the observed risk in non-obese patients as the baseline and applying the pooled RR.