Supplementary methods and sensitivity analyses

Orforglipron and danuglipron systematic review. PROSPERO CRD420261508176. Revised 5 November 2026. Table S1 retains the original search strategies. Dataset S2 contains the traceable inputs and executable analysis.

Methods for reproducibility

Data were organized by molecule, development phase, population, comparator, dose, follow-up and estimand. In accordance with the registered phase-stratified approach, phase 2 evidence was presented separately from phase 3 evidence. The targeted phase 3 meta-analyses selected one 36-mg orforglipron versus placebo contrast per trial, avoiding repeated use of a shared placebo arm. This dose-specific restriction was an analytic choice, not a restriction in the registered eligibility criteria, and does not summarize the entire dose-response relation. The presence of a dulaglutide arm in the phase 2 T2D trial does not invalidate its placebo contrast; the previous explanation to that effect has been removed. Phase 2 efficacy is now reported separately. Across-molecule contrasts, active-comparator trials and weight-maintenance trials were not combined with the placebo-controlled phase 3 treatment-initiation contrasts.

Adjusted treatment differences were pooled using generic inverse-variance REML random-effects models; SE=(upper CI-lower CI)/(2×1.959964) for reported 95% CIs. Randomized N is descriptive and does not enter the weight calculation. Wald 95% CIs are accompanied by DerSimonian-Laird [15] and common-effect estimates, ordinary HKSJ and modified HKSJ analyses using t(k-1) and variance scale max[1,Q_RE/(k-1)], and leave-one-out analyses. We report Q, I^2 and tau^2, recognizing that their estimation is unreliable with few studies. The original two-trial weight subset is retained as a sensitivity analysis; the corrected main analysis includes the previously omitted, extractable ACHIEVE-1 weight contrast. Published 90% CIs for danuglipron are labeled as 90% CIs and are not silently treated as 95% CIs. Safety log-risk ratios and their sampling variances were calculated from treated arm counts; a 0.5 correction was applied to all four cells only when a zero cell occurred. Safety was pooled separately for the three phase 3 T2D placebo comparisons, with other phases, populations, comparators, maintenance/rescue designs and danuglipron titration cohorts tabulated separately. Hepatic events were described by trial and source rather than pooled as a common DILI risk. This restores part of the planned safety synthesis; residual non-pooling is declared and justified by these specific differences. The primary efficacy GRADE assessment [18] concerns the selected placebo contrasts, not equivalence between molecules. Analyses were reproduced in Python with NumPy and SciPy; executable code is provided in Dataset S2. Few-study interpretation follows contemporary Cochrane guidance [33]. Non-pooled findings were organized by molecule, phase, population and outcome, consistent with the SWiM reporting framework [17].

The original trial CIs are rounded, so reconstructed SEs and pooled estimates are approximate at the reported source precision. Adjusted model-based contrasts are not reconstructed from the separate arm means or final observed sample sizes. Selected randomized N is a descriptor of assigned arms, not the number of complete cases at the final visit. Safety event counts represent people with an event, not numbers of episodes. A blank count is unavailable, not zero.

Table S2 Efficacy sensitivity analyses

Outcome	Dataset	Model / CI	k	Estimate (95% CI)	I^2 %	Tau^2
HbA1c	corrected complete set	DL / Wald	3	-1.14 (-1.25 to -1.02)	0.0	0.000
HbA1c	corrected complete set	REML / Wald	3	-1.14 (-1.25 to -1.02)	0.0	0.000
HbA1c	corrected complete set	CE / Wald	3	-1.14 (-1.25 to -1.02)	0.0	0.000
HbA1c	corrected complete set	REML / HKSJ	3	-1.14 (-1.37 to -0.91)	0.0	0.000

Outcome	Dataset	Model / CI	k	Estimate (95% CI)	I² %	Tau²
HbA1c	corrected complete set	REML / mHKSJ	3	-1.14 (-1.40 to -0.88)	0.0	0.000
Weight T2D	corrected complete set	DL / Wald	3	-6.40 (-7.36 to -5.44)	29.3	0.218
Weight T2D	corrected complete set	REML / Wald	3	-6.39 (-7.38 to -5.39)	29.3	0.269
Weight T2D	corrected complete set	CE / Wald	3	-6.50 (-7.28 to -5.72)	29.3	0.000
Weight T2D	corrected complete set	REML / HKSJ	3	-6.39 (-8.43 to -4.34)	29.3	0.269
Weight T2D	corrected complete set	REML / mHKSJ	3	-6.39 (-8.58 to -4.20)	29.3	0.269
Weight T2D	original two-study subset	DL / Wald	2	-6.54 (-7.96 to -5.11)	49.8	0.560
Weight T2D	original two-study subset	REML / Wald	2	-6.54 (-7.96 to -5.11)	49.8	0.560
Weight T2D	original two-study subset	CE / Wald	2	-6.72 (-7.63 to -5.81)	49.8	0.000
Weight T2D	original two-study subset	REML / HKSJ	2	-6.54 (-15.77 to 2.70)	49.8	0.560
Weight T2D	original two-study subset	REML / mHKSJ	2	-6.54 (-15.77 to 2.70)	49.8	0.560

The original two-study weight subset is shown because modified HKSJ intervals include zero. It is not the complete corrected T2D dataset. The complete three-study set includes all extractable treatment-regimen 36-mg placebo weight contrasts in the identified phase 3 initiation trials. This correction was made during revision, not selected on statistical significance.

Table S3 Registration and revision decisions

Registered framework	Revised implementation	Explanation
Phase 2/3 trials eligible; both published and unpublished evidence sought	Retained phase 2 trials; added previously identified ACHIEVE-4 as eligible sponsor-only evidence	Correction of trial disposition, not a new search
Separate pooling by development phase	Phase 2 results reported separately; phase 3 placebo effects pooled	The added dulaglutide arm is not a reason to discard a valid placebo contrast
Any eligible dose/formulation	36-mg phase 3 contrasts selected once per trial	Dose-specific analytic restriction; not a registered eligibility restriction
DL or REML random effects	REML main estimate; DL, common effect, HKSJ and modified HKSJ checks	Additional sensitivity analyses requested during peer review
HbA1c and weight at longest reported follow-up	ACHIEVE-1 weight contrast added at 40 weeks; other trials 40 or 72 weeks	Corrects extractable-outcome omission; original subset retained
Dichotomous safety as risk ratios	Restored selected phase 3 T2D safety pooling; detailed other arms tabulated	Residual deviation: no pooled effects across phases, populations, molecules, titration cohorts or rescue designs

Registered framework	Revised implementation	Explanation
Hepatic outcomes as proportions/descriptive synthesis	Liver abnormalities, adjudicated injury and external sponsor signals separated	No pooled DILI estimate or claim of comparative hepatic safety
GRADE for primary efficacy outcomes	Corrected contrast N and interpretive scope; new safety pools exploratory	No assumption that registration/publication excludes reporting bias
Multiple information sources	Embase search was performed in addition to the sources selected in the PROSPERO form	Additional source, not an omitted registered source

Table S4 Safety sensitivity analyses

Outcome	Model / CI	Risk ratio (95% CI)	I^2 %
AE-related discontinuation	DL / Wald	2.52 (1.66 to 3.81)	0.0
AE-related discontinuation	REML / Wald	2.52 (1.66 to 3.81)	0.0
AE-related discontinuation	CE / Wald	2.52 (1.66 to 3.81)	0.0
AE-related discontinuation	REML / mHKSJ	2.52 (1.02 to 6.25)	0.0
Nausea	DL / Wald	4.69 (3.58 to 6.15)	0.0
Nausea	REML / Wald	4.98 (3.41 to 7.28)	0.0
Nausea	CE / Wald	4.69 (3.58 to 6.15)	0.0
Nausea	REML / mHKSJ	4.98 (2.17 to 11.45)	0.0
Vomiting	DL / Wald	8.02 (3.59 to 17.93)	29.3
Vomiting	REML / Wald	7.21 (3.95 to 13.16)	29.3
Vomiting	CE / Wald	6.61 (4.36 to 10.02)	29.3
Vomiting	REML / mHKSJ	7.21 (1.63 to 31.97)	29.3
Diarrhoea	DL / Wald	2.09 (1.58 to 2.76)	15.4
Diarrhoea	REML / Wald	2.14 (1.55 to 2.97)	15.4
Diarrhoea	CE / Wald	2.01 (1.61 to 2.53)	15.4
Diarrhoea	REML / mHKSJ	2.14 (1.05 to 4.38)	15.4
Constipation	DL / Wald	2.81 (2.08 to 3.80)	0.0
Constipation	REML / Wald	2.81 (2.08 to 3.80)	0.0
Constipation	CE / Wald	2.81 (2.08 to 3.80)	0.0
Constipation	REML / mHKSJ	2.81 (1.45 to 5.46)	0.0

Safety pooling uses 597 treated participants assigned to orforglipron 36 mg and 906 assigned to placebo. Follow-up is 40, 72 and 40 weeks, respectively. For ATTAIN-2, the all-causality safety-table discontinuation counts (34/321 and 29/628) were used, not the separate treatment-period flowchart counts. No continuity correction was needed except for the zero placebo vomiting count in ACHIEVE-5. Sensitivity estimates are exploratory, without adjustment for multiple safety endpoints.

Phase 2 and non-pooled trial evidence

The phase 2 orforglipron T2D trial [31] included a valid placebo contrast despite its additional dulaglutide arm. At 26 weeks, the combined 36-mg cohorts had an HbA1c difference of -1.60 percentage points versus placebo (95% CI -1.96 to -1.25) and a weight difference of -7.4 percentage points (-9.4 to -5.4). These on-treatment phase 2 estimates were reported separately from treatment-regimen phase 3 estimates. The phase 2 obesity trial [30] enrolled 272 participants and reported weight reductions of 9.4% to 14.7% across doses versus 2.3% with placebo at 36 weeks; titration subcohorts were retained in the arm-level safety data. ATTAIN-MAINTAIN [7] concerned continued maintenance after injectable therapy and allowed rescue treatment; its initial clinical state and estimand differ from initiation trials even though additional absolute-weight outcomes were also reported. ACHIEVE-J [8] randomized 401 participants among 3-, 12- and 36-mg doses without a placebo or active-molecule arm; AE-related discontinuations were 7/132, 11/135 and 19/134. ACHIEVE-4 [32] is retained as eligible sponsor-reported evidence. The sponsor-reported efficacy-estimand differences at 52 weeks were -0.66 percentage points for HbA1c (95% CI -0.74 to -0.58) and -10.42 percentage points for weight (-10.92 to -9.93) versus insulin glargine. Its MACE-4 hazard ratio was 0.84 (95% CI 0.59 to 1.20), meeting non-inferiority criteria (prespecified upper bound <1.8), not cardiovascular superiority. No pooled numerical estimate spanning these distinct designs was constructed.

Dose-escalation subcohorts in phase 2 are retained as separate safety rows. The same placebo group is never reused as independent information in a pooled analysis. DAN-01 cohort 3 is compared only with its concurrent placebo in the reported individual efficacy contrast; earlier cohorts had different titration and duration. A point estimate in kilograms is not pooled with a percentage-weight difference. Reported danuglipron 90% intervals remain explicitly labeled.

Sources and availability

Principal efficacy and safety inputs were checked against the supplied trial PDFs. Additional original phase 2 orforglipron reports, ATTAIN-MAINTAIN and the ATTAIN-1 supplementary appendix were retrieved. ATTAIN-1 Table S7 (appendix p37) supplies individual gastrointestinal participant counts for all four arms. ACHIEVE-J descriptive safety counts were completed from ClinicalTrials.gov NCT06010004 results first posted 30 June 2026, before the review cutoff, and accessed 5 November 2026. The registry safety window is baseline to week 54. Its non-serious preferred-term counts for nausea, vomiting, diarrhea and constipation are retained explicitly; there are no serious entries for these same four terms. Hepatic laboratory and hepatobiliary preferred terms are supplied separately, without summing potentially overlapping participants or interpreting them as adjudicated DILI. The full ACHIEVE-J journal article remains unavailable; this does not prevent extraction of these directly reported registry results. These data do not alter the principal three-trial efficacy or safety pools. Dataset S2 includes the public registry JSON and source-labeled arm counts.