Dapagliflozin and empagliflozin versus placebo for cardiovascular death or worsening heart failure across the ejection-fraction spectrum: a systematic review and meta-analysis of randomized placebo-controlled trials

***Running title**: SGLT2 inhibitors across the heart-failure ejection-fraction spectrum*

Nawaf AM Alharbi[1,*], Abdullah Ahmed Alfllaj[1], Mohamed Ghaith Al Abdin[2], Lana Musaad Alhatlani[3], Ihab Fathi Ibrahim Suliman[4,5]

[1] Department of Clinical Sciences, College of Medicine, Sulaiman AlRajhi University, AlBukayriyah, Qassim, Saudi Arabia

[2] Department of Internal Medicine, Dr. Sulaiman Al-Habib Medical Group, Qassim, Saudi Arabia

[3] College of Medicine, Qassim University, Qassim, Saudi Arabia

[4] Department of Cardiac Sciences, King Abdulaziz Cardiac Center, King Abdulaziz Medical City, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia

[5] College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

***Corresponding author:** Nawaf AM Alharbi, Department of Clinical Sciences, College of Medicine, Sulaiman AlRajhi University, AlBukayriyah, Qassim, Saudi Arabia. Email: nawafga7222@gmail.com

ORCID iDs: Abdullah Ahmed Alfllaj: https://orcid.org/0009-0002-8005-2419; Mohamed Ghaith Al Abdin: https://orcid.org/0009-0002-9461-3555; Lana Musaad Alhatlani: https://orcid.org/0009-0005-7571-5611; Ihab Fathi Ibrahim Suliman: https://orcid.org/0000-0001-8656-7013

Abstract

Background and aims: Dapagliflozin and empagliflozin are established therapy for chronic heart failure across the ejection-fraction spectrum. Their pivotal trials were powered for a composite of cardiovascular death and worsening heart failure, and the pooled composite hazard ratio is often carried into clinical statements as though every component were reduced to the same degree. The size of each component's contribution to the pooled composite, and the certainty attaching to each, have not been reported together in a single outcome-level synthesis of these four trials. We estimated the pooled effect on the trial-defined composite and on each component separately, and rated certainty for each.

Methods: We conducted a systematic review and meta-analysis of randomized, double-blind, placebo-controlled chronic heart-failure outcome trials of dapagliflozin 10 mg or empagliflozin 10 mg. MEDLINE/PubMed was searched with a documented record-level strategy; CENTRAL was queried as a supplementary count-only source. Screening and data extraction were performed independently and in duplicate. Hazard ratios were pooled on the log scale by random-effects inverse-variance methods with the ad hoc Hartung-Knapp adjustment. Fixed-effect, leave-one-out, and endpoint-definition sensitivity analyses were performed. Risk of bias was assessed with RoB 2 and certainty with GRADE.

Results: Four trials enrolling 20,725 participants were included. Dapagliflozin or empagliflozin reduced cardiovascular death or worsening heart failure/hospitalization for heart failure (HR 0.78, 95% CI 0.70–0.86; I^2 = 0.0%; exploratory 95% prediction interval 0.68–0.90). The component signal was strongest for worsening heart failure/hospitalization (HR 0.73, 95% CI 0.64–0.83; high certainty), whereas cardiovascular death (HR 0.88, 95% CI 0.76–1.02) and all-cause death (HR 0.93, 95% CI 0.82–1.05) were directionally favorable but imprecise (moderate certainty, downgraded for imprecision). Leave-one-out estimates ranged from 0.76 to 0.79. Replacing the DAPA-HF primary composite with its prespecified secondary composite of cardiovascular death or hospitalization for heart failure changed the pooled estimate from 0.78 to 0.78 (95% CI 0.70–0.87), indicating that the pooled effect does not depend on the inclusion of urgent heart-failure visits.

Conclusions: Dapagliflozin and empagliflozin reduce the trial-defined composite of cardiovascular death or worsening heart failure across reduced and above-40% ejection-fraction phenotypes, and the pooled estimate is insensitive to the differing composite wording of the contributing trials. The composite benefit is driven principally by fewer worsening heart-failure events. Mortality estimates are favorable but should not be presented as definitive standalone reductions in this aggregate-data synthesis.

Dapagliflozin and empagliflozin versus placebo for cardiovascular death or worsening heart failure across the ejection-fraction spectrum: a systematic review and meta-analysis of randomized placebo-controlled trials

Keywords: dapagliflozin; empagliflozin; SGLT2 inhibitors; heart failure; ejection fraction; composite endpoint; meta-analysis

How to cite this article: Alharbi NAM, Alfllaj AA, Al Abdin MG, Alhatlani LM, Suliman IFI. Dapagliflozin and empagliflozin versus placebo for cardiovascular death or worsening heart failure across the ejection-fraction spectrum: a systematic review and meta-analysis of randomized placebo-controlled trials. Int J Drug Deliv Technol. 2026;16(69s): 1889-1904. DOI: 10.25258/ijddt.16.69s.187

Clinical perspective

What is already known?	Dapagliflozin and empagliflozin are guideline-directed heart-failure therapies, and prior syntheses support reductions in composite heart-failure outcomes across ejection-fraction phenotypes.
What does this study add?	An outcome-level synthesis of the four pivotal chronic heart-failure trials that reports the composite, its components, and the certainty of each side by side, and shows that the pooled estimate does not depend on the differing composite definitions used by the trials.
How may this inform practice?	The strongest evidence supports prevention of worsening heart failure and hospitalization. Mortality estimates are favorable but should not be presented as definitive standalone reductions.

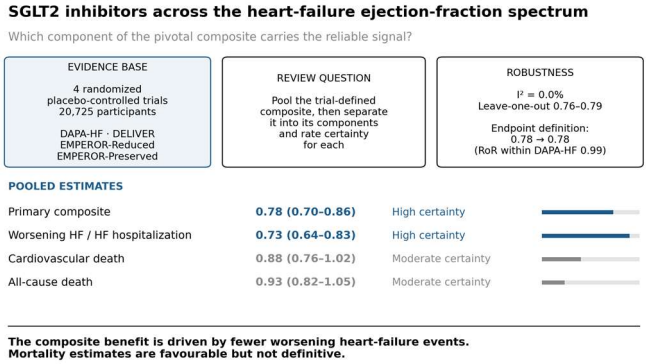


Figure 1. Graphical abstract summarizing the evidence base, review question, pooled estimates, and component-level interpretation.

Introduction

Heart-failure care is organized across a continuum of left ventricular ejection fraction rather than around a reduced-versus-preserved dichotomy, and current guidelines recommend sodium-glucose cotransporter 2 (SGLT2) inhibitors across that continuum [1,2]. This shift has consequences for evidence synthesis. A relative treatment effect observed in heart failure with reduced ejection fraction (HFrEF) does not carry the same absolute implication in heart failure with mildly reduced or preserved ejection fraction (HFmrEF/HFpEF), even when point estimates are directionally similar. Systematic reviews of these therapies should therefore define the target phenotype, the endpoint wording, the effect measure, and the level of inference before interpreting a pooled estimate.

Dapagliflozin and empagliflozin are the two agents tested in large chronic heart-failure placebo-controlled outcome trials spanning reduced and above-40% ejection-fraction populations. DAPA-HF and EMPEROR-Reduced established benefit in HFrEF; DELIVER and EMPEROR-Preserved extended the evidence base to patients with ejection fraction above 40% [3–6]. The four trials share chronic symptomatic heart failure, once-daily 10 mg dosing, placebo control, background standard therapy, and time-to-event composite outcomes anchored in cardiovascular death and worsening heart failure or hospitalization for heart failure.

The interpretive difficulty lies in the composite endpoint. Composite outcomes are efficient and clinically appropriate when their components are patient-important and directionally coherent. They mislead when the pooled effect is described as though each component were reduced to the same degree, and this failure is common in cardiovascular trials, where the more frequent and less severe components typically carry the larger treatment effect [7,8]. The four trials also worded their composites differently: DAPA-HF and DELIVER counted urgent heart-failure visits requiring intravenous therapy within worsening heart failure, whereas EMPEROR-Reduced and EMPEROR-Preserved counted

hospitalization for heart failure. These definitions are close enough to support a trial-defined composite synthesis, but they are not identical, and the difference should not be concealed under a broader label.

Existing syntheses of these agents are either broad class-level analyses or patient-level pooled analyses of a subset of the trials [9–13]. They establish the class effect and are larger than any trial-level synthesis can be. What they do not provide is a single outcome-level account of the four pivotal chronic heart-failure trials in which the composite, each of its components, the certainty of each, and the sensitivity of the pooled estimate to the trials' differing composite wording are reported together. Safety-focused syntheses of these agents have translated adverse-event risks into clinically interpretable benefit-risk summaries [14]; efficacy reporting has not been held to the same standard, and the clinical message is frequently reduced to a single hazard ratio. This is the gap the present review addresses.

We therefore conducted a systematic review and meta-analysis of randomized, double-blind, placebo-controlled chronic heart-failure trials of dapagliflozin or empagliflozin. We aimed to estimate the pooled effect on cardiovascular death or worsening heart failure/hospitalization for heart failure across the ejection-fraction spectrum; to determine whether that effect is driven principally by worsening heart failure or by mortality; to test whether the pooled estimate depends on the differing composite definitions used by the trials; and to rate the certainty of evidence for each outcome using GRADE.

Methods

Protocol and reporting framework

This review was prepared according to PRISMA 2020 for reporting, PRISMA-S for search reporting, PRISMA-P logic for protocol elements, Cochrane methods for intervention reviews, RoB 2 for randomized trials, and GRADE for outcome-level certainty [15–22]. The protocol framework, search details, PRISMA flow accounting, extraction sheet, RoB 2 rationale, GRADE evidence profile, and statistical derivations are provided in the Supplementary Appendix.

The eligibility criteria, outcome hierarchy, endpoint-harmonization rules, and synthesis plan were specified a priori and are archived in the Supplementary Appendix. The review was not prospectively registered. The project began as a focused synthesis of a small, predefined set of pivotal chronic heart-failure outcome trials rather than as an open-ended evidence map, and the absence of prospective registration is stated here and in the Limitations as a reproducibility weakness rather than a design choice. The final search, extraction, statistical analysis, and manuscript package were completed on 25 February 2026. To limit post hoc bias, the final analysis preserves the prespecified PICO, keeps the primary composite and component outcomes separate, reports the record-level search pathway transparently, and confines all conclusions to the locked eligibility and synthesis plan.

Eligibility criteria

Eligible studies were randomized, double-blind, placebo-controlled trials enrolling adults with symptomatic chronic heart failure and comparing dapagliflozin 10 mg once daily or empagliflozin 10 mg once daily with placebo plus usual care. Studies had to report a time-to-event hazard ratio with a 95% confidence interval for cardiovascular death or worsening heart failure/hospitalization for heart failure, or an endpoint judged clinically equivalent for the trial-defined composite.

Trials were excluded from the primary synthesis if they evaluated another SGLT2 inhibitor or a dual SGLT1/2 inhibitor, enrolled acute decompensated heart-failure-only populations, studied post-myocardial-infarction patients without chronic symptomatic heart failure as the target syndrome, reported only mechanistic or biomarker endpoints, or were non-randomized. Sotagliflozin trials, acute heart-failure trials, diabetes cardiovascular-outcome trials, and chronic kidney-disease trials were not pooled because they address related but distinct clinical questions [23–25].

Information sources and search strategy

The record-level search, screening, extraction, and final data analysis were completed and locked on 25 February 2026. MEDLINE/PubMed was used as the exportable bibliographic source, with a primary concept-block search and a broader sensitivity search. The searches combined heart-failure terms, dapagliflozin or empagliflozin terms, and randomized placebo-controlled trial terms. The two PubMed searches produced 833 records before deduplication. After removal of 151 duplicates, 682 unique records were screened. CENTRAL was also queried and returned 545 records; because export and deduplication were unavailable in the documented workflow, CENTRAL is reported as a count-only supplementary source and was not merged into the PRISMA denominator. The reliance on a single exportable database is a search limitation and is stated as such in the Limitations.

The search strategy was intentionally broad enough to capture trial reports, secondary analyses, and adjacent studies, whereas the eligibility criteria were intentionally narrow. Reference lists of the pivotal trials, major SGLT2 inhibitor meta-analyses, and

heart-failure guidelines were checked for potentially eligible chronic heart-failure placebo-controlled outcome trials. Exact strings and the search ledger are provided in the Supplementary Appendix.

Study selection and data collection

Records were screened against the PICO criteria using a prespecified eligibility sheet. Full-text assessment was performed for reports that clearly represented, or could plausibly represent, chronic symptomatic heart-failure randomized placebo-controlled outcome trials of dapagliflozin or empagliflozin. Four reports met all eligibility criteria: DAPA-HF, DELIVER, EMPEROR-Reduced, and EMPEROR-Preserved [3–6]. No full-text report was excluded after full-text assessment in the documented PubMed/MEDLINE pathway.

Extracted fields included trial acronym, registry number, drug and dose, ejection-fraction phenotype, randomized sample size by group, follow-up duration, primary endpoint definition, primary-composite event counts, hazard ratio and 95% confidence interval for the primary composite, and hazard ratios for worsening heart failure/hospitalization, cardiovascular death, and all-cause death. Extraction was cross-checked against the trial reports and tabulated to preserve source traceability.

Records were screened and data were extracted independently and in duplicate by two reviewers (N.A.M.A. and A.A.A.), each working without access to the other's decisions. The two sets of decisions were then compared. The reviewers disagreed on 33 of the 682 screened records (4.8%; simple agreement 95.2%), and every disagreement was adjudicated by the senior author (I.F.I.S.), whose decision was final. All four trials included in the quantitative synthesis were selected by both reviewers independently.

Outcome harmonization

The primary outcome was cardiovascular death or trial-defined worsening heart failure/hospitalization for heart failure. This wording preserves the original trial definitions while maintaining clinical coherence. DAPA-HF and DELIVER included urgent heart-failure visits in the worsening heart-failure definition; EMPEROR-Reduced and EMPEROR-Preserved counted hospitalization for heart failure. The primary meta-analysis therefore uses the trial-defined composite, component analyses are presented separately, and an endpoint-definition sensitivity analysis quantifies the consequence of the difference.

Secondary outcomes were worsening heart failure/hospitalization for heart failure, cardiovascular death, and all-cause death. Event counts are presented descriptively because the prespecified effect measure was the hazard ratio and follow-up differed across trials. Approximate absolute effects per 1000 treated participants were calculated from pooled hazard ratios and pooled placebo risks to support clinical interpretation; these estimates are not substitutes for trial-specific time-to-event analyses.

Risk of bias and certainty assessment

Risk of bias was assessed at outcome level using the five RoB 2 domains [19]. All included trials were large randomized, double-blind, placebo-controlled outcome trials with published protocols and adjudicated clinical endpoints, and the primary composite outcome was judged at low risk of bias in each. Domain-level justifications are provided in the Supplementary Appendix.

Certainty of evidence was rated with GRADE [20–22]. Randomized evidence started at high certainty, and downgrading was considered for risk of bias, inconsistency, indirectness, imprecision, and publication bias. The primary composite and worsening heart-failure outcomes were rated high certainty. Cardiovascular death and all-cause death were downgraded one level for imprecision because their confidence intervals crossed the null and did not exclude a small or absent effect. The publication-bias domain was recorded as not assessed for every outcome: with four trials, asymmetry testing is uninformative, and no outcome was downgraded or reassured on that basis.

Statistical analysis

Hazard ratios were transformed to the natural-log scale and pooled by inverse-variance methods [26,27]. Standard errors were derived from reported 95% confidence intervals as SE = (ln upper − ln lower) / (2 × 1.96). The primary model was random-effects inverse-variance pooling with Paule-Mandel estimation of τ^2 and the Hartung-Knapp adjustment, which was applied in its ad hoc form: where the weighted mean-square residual fell below one, it was truncated at one so that the random-effects interval could not be narrower than the corresponding fixed-effect interval [28]. This variant is reported explicitly because it widens the primary interval relative to the unadjusted Hartung-Knapp calculation and because only four studies contributed to the main analysis. Fixed-effect inverse-variance pooling was performed as a sensitivity analysis.

Between-study heterogeneity was summarized by Cochran Q, I^2, and τ^2 [29]. An exploratory 95% prediction interval was calculated for the primary composite [30,31]. Because only four trials contributed and τ^2 was estimated as zero, the prediction interval is interpreted as supportive rather than definitive and was not used to support subgroup-

level claims. Funnel-plot asymmetry testing was not performed; such testing is unreliable with fewer than ten studies [32]. Leave-one-out analyses evaluated whether the primary conclusion depended on any single trial.

An endpoint-definition sensitivity analysis addressed the differing composite wording. DAPA-HF reports, as a prespecified secondary outcome, the composite of cardiovascular death or hospitalization for heart failure, which matches the EMPEROR composite definition [3]. The primary model was refitted with that estimate substituted for the DAPA-HF primary composite, holding the other three trials unchanged, and the two pooled estimates were compared. DELIVER does not report an equivalent time-to-first-event composite restricted to hospitalization, so full harmonization across all four trials was not possible; the analysis is therefore a one-trial substitution and is interpreted accordingly.

Drug and ejection-fraction-phenotype subgroup analyses were descriptive rather than inferential, because the review was not powered to compare dapagliflozin with empagliflozin or HFrEF with HFmrEF/HFpEF.

Results

Study selection

The two PubMed/MEDLINE searches identified 833 records before deduplication. After 151 duplicates were removed, 682 unique records were screened; 678 were excluded at title/abstract level because they were not randomized chronic heart-failure placebo-controlled trials of dapagliflozin or empagliflozin, or did not report an eligible time-to-event outcome. Four full-text reports were assessed and all four were included in the quantitative synthesis. CENTRAL produced 545 count-only records and is reported as a search limitation because export and deduplication were unavailable (Figure 2).

Study characteristics

The four eligible trials were DAPA-HF, DELIVER, EMPEROR-Reduced, and EMPEROR-Preserved. They randomized 20,725 participants: 10,364 assigned to dapagliflozin or empagliflozin and 10,361 assigned to placebo. Two trials enrolled HFrEF populations and two enrolled patients with ejection fraction above 40%. Median follow-up ranged from 16 months to 2.3 years. All four used once-daily 10 mg dosing, placebo control, background standard therapy, and time-to-first-event composite endpoints (Table 1).

The primary composite occurred in 1,674 of 10,364 participants assigned to dapagliflozin or empagliflozin and in 2,085 of 10,361 assigned to placebo, based on extracted trial-level event counts. These crude proportions are descriptive; the prespecified pooled effect measure was the hazard ratio. The direction of effect favored active treatment in every trial.

Risk of bias

For the primary composite outcome, all four trials were judged at low risk of bias. Randomization and allocation concealment were adequately described; placebo-controlled blinding minimized deviations from intended interventions and outcome-measurement bias; missing outcome data were low relative to sample size; and the primary endpoint was prespecified in the trial protocols. The RoB 2 traffic-light summary and domain-level rationale are provided in the Supplementary Appendix.

Primary composite outcome

Compared with placebo, dapagliflozin or empagliflozin reduced cardiovascular death or worsening heart failure/hospitalization for heart failure (pooled HR 0.78, 95% CI 0.70–0.86; I^2 = 0.0%; τ^2 = 0; exploratory 95% prediction interval 0.68–0.90; Figure 3). Trial estimates were directionally consistent: DAPA-HF 0.74, DELIVER 0.82, EMPEROR-Reduced 0.75, EMPEROR-Preserved 0.79. The fixed-effect sensitivity estimate was identical in point estimate and narrower in interval (HR 0.78, 95% CI 0.73–0.83).

Leave-one-out estimates were stable. Omitting DAPA-HF yielded 0.79; omitting DELIVER, 0.76; omitting EMPEROR-Reduced, 0.79; omitting EMPEROR-Preserved, 0.77. No single trial changed the direction of the pooled estimate or removed the clinically meaningful reduction.

Endpoint-definition sensitivity

Substituting the DAPA-HF prespecified secondary composite of cardiovascular death or hospitalization for heart failure (HR 0.75, 95% CI 0.65–0.85) for its primary composite left the pooled estimate essentially unchanged (HR 0.78, 95% CI 0.70–0.87; I^2 = 0.0%). Within DAPA-HF itself, the two definitions differed by a ratio of hazard ratios of 0.99. The inclusion of urgent heart-failure visits in the dapagliflozin composites therefore accounts for a negligible part of the pooled effect, and pooling the four trials under a trial-defined composite label does not materially distort the estimate.

Component outcomes

The component synthesis identifies the clinical source of benefit. The pooled effect was strongest and most precise for worsening heart failure/hospitalization for heart failure (HR 0.73, 95% CI 0.64–0.83). The mortality components were directionally favorable but imprecise: cardiovascular death HR 0.88 (95% CI 0.76–1.02) and all-cause death HR 0.93 (95% CI 0.82–1.05) (Figure 4).

The approximate absolute-effect translation matched the component interpretation. Per 1000 treated participants, the pooled estimate corresponded to approximately 45 fewer primary-composite events and 39 fewer worsening heart-failure/hospitalization events. Approximate absolute reductions for cardiovascular death and all-cause death were smaller and their ranges included little or no benefit (Figure 5).

Subgroups and certainty of evidence

Descriptive subgroup analyses did not suggest inconsistency by agent or ejection-fraction phenotype. Fixed-effect estimates were similar for the dapagliflozin trials (HR 0.78, 95% CI 0.72–0.86) and the empagliflozin trials (HR 0.77, 95% CI 0.70–0.85), and directionally consistent in the HFrEF trials (HR 0.74, 95% CI 0.68–0.82) and the above-40% trials (HR 0.81, 95% CI 0.74–0.88). These analyses are not evidence of comparative differences between drugs or phenotypes.

GRADE certainty was high for the primary composite and for worsening heart failure/hospitalization, and moderate for cardiovascular death and all-cause death because of imprecision. No outcome was downgraded for risk of bias, inconsistency, or indirectness within the defined chronic heart-failure question, and the publication-bias domain was not assessed. The certainty pattern is summarized in Figure 6 and Table 2.

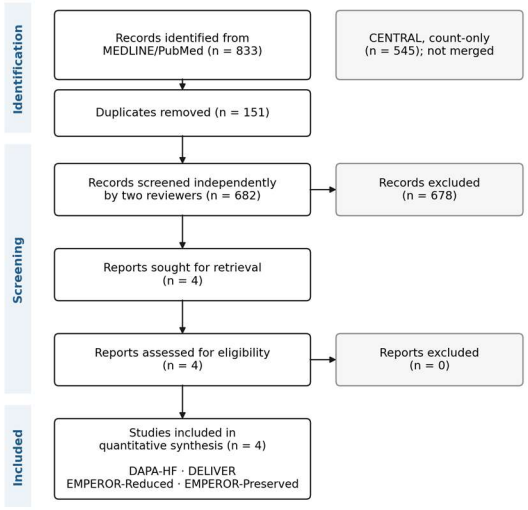


Figure 2. PRISMA 2020 flow diagram for the documented search ledger. CENTRAL is reported as a count-only supplementary source because export and deduplication were unavailable.

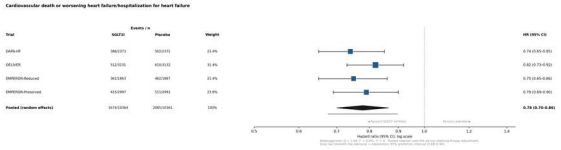


Figure 3. Primary composite forest plot. Hazard ratios are pooled on the log scale; the pooled interval uses the ad hoc Hartung-Knapp adjustment. The exploratory prediction interval is shown because the review includes only four trials.

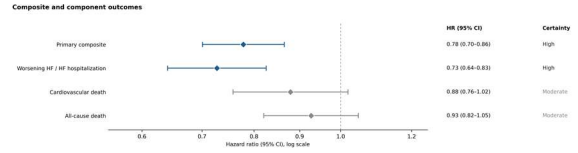


Figure 4. Composite and component outcomes. Worsening heart failure/hospitalization is the strongest component signal; the mortality components are favorable but imprecise.

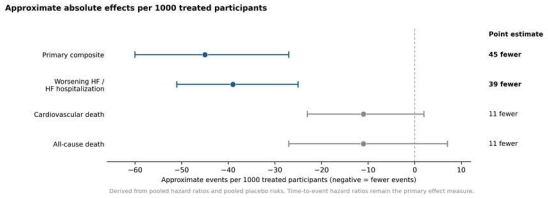


Figure 5. Approximate absolute effects per 1000 treated participants. Estimates are derived from pooled hazard ratios and pooled placebo risks; time-to-event hazard ratios remain the primary effect measure.

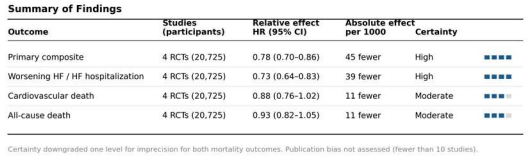


Figure 6. GRADE Summary of Findings figure. Certainty is shown together with relative effects and approximate absolute effects per 1000 treated participants; numerical details are provided in Table 2.

Tables

Table 1. Characteristics of included randomized placebo-controlled trials.

Trial	Drug	Population	Active / placebo	Follow-up	Primary endpoint	HR (95% CI)
DAPA-HF	Dapagliflozin	HFrEF; LVEF ≤40%	2373 / 2371	18.2 mo	CV death or worsening HF (HF hospitalization or	0.74 (0.65–0.85)

Trial	Drug	Population	Active / placebo	Follow-up	Primary endpoint	HR (95 % CI)
					urgent HF visit requiring IV therapy)	
DELIVER	Dapagliflozin	HFmrEF/HFpEF; LVEF >40%	3131 / 3132	2.3 y	CV death or worsening HF (unplanned HF hospitalization or urgent HF visit)	0.82 (0.73 – 0.92)
EMPEROR-Reduced	Empagliflozin	HFrEF; LVEF ≤40%	1863 / 1867	16 mo	CV death or hospitalization for worsening HF	0.75 (0.65 – 0.86)
EMPEROR-Preserved	Empagliflozin	HFmrEF/HFpEF; LVEF >40%	2997 / 2991	26.2 mo	CV death or hospitalization for HF	0.79 (0.69 – 0.90)

Table 2. Summary of Findings using GRADE.

Outcome	No. of studies	Participants	Relative effect HR (95 % CI)	Approximate absolute effect per 1000	Certainty	Interpretation
Primary composite	4 RCTs	20,725	0.78 (0.70–0.86)	45 fewer per 1000 (range 27 to 60 fewer)	High	Consistent and precise reduction; no downgrade.
Worsening HF/HF hospitalization	4 RCTs	20,725	0.73 (0.64–0.83)	39 fewer per 1000 (range 25 to 51 fewer)	High	Consistent and precise reduction; no downgrade.
Cardiovascular death	4 RCTs	20,725	0.88 (0.76–1.02)	11 fewer per 1000 (range 2 more to 23 fewer)	Moderate	Downgraded one level for imprecision; CI crosses the null.
All-cause death	4 RCTs	20,725	0.93 (0.82–1.05)	11 fewer per 1000 (range 7 more to 27 fewer)	Moderate	Downgraded one level for imprecision; CI crosses the null.

Abbreviations: CI, confidence interval; GRADE, Grading of Recommendations Assessment, Development and Evaluation; HR, hazard ratio; RCT, randomized controlled trial. Absolute effects are approximate because follow-up differed across trials and hazard ratios were the primary effect measure.

The publication-bias domain was not assessed for any outcome (fewer than ten studies).

Discussion

Principal findings

Dapagliflozin and empagliflozin reduce the trial-defined composite of cardiovascular death or worsening heart failure/hospitalization for heart failure across chronic heart-failure trials spanning reduced and above-40% ejection-fraction phenotypes. The pooled estimate was consistent across all four pivotal placebo-controlled outcome trials, stable in leave-one-out analyses, and unchanged when the DAPA-HF composite was replaced by its hospitalization-restricted secondary composite. The composite benefit is driven mainly by fewer worsening heart-failure events rather than by a definitive standalone reduction in cardiovascular or all-cause mortality.

The clinical message should therefore distinguish the composite from its components. The pooled composite HR of 0.78 is robust and guideline-concordant. It should not be translated into an unqualified claim that dapagliflozin and empagliflozin reduce cardiovascular mortality across all ejection-fraction phenotypes: cardiovascular death and all-cause death were directionally favorable, but their confidence intervals crossed the null. The evidence supports interpretation of the treatment effect primarily through reductions in worsening heart-failure events and, through them, in the primary composite used by the pivotal trials.

Comparison with prior evidence

The findings agree with broad SGLT2 inhibitor meta-analyses and patient-level pooled analyses [9–13]. Those analyses are larger and answer wider questions, including class-level effects, agent-specific considerations, and effects across diabetes, kidney disease, and different heart-failure phenotypes. The present review asks a narrower question at the outcome level: what each component contributes to the pooled composite in the four pivotal chronic heart-failure trials, how certain each of those estimates is, and whether the pooled estimate survives the trials' differing composite wording.

The component pattern matches how current guidelines discuss these agents. ESC and AHA/ACC/HFSA recommendations emphasize reductions in heart-failure events across phenotypes and interpret mortality effects more cautiously above the HFrEF range [1,2]. Displaying the composite, the worsening heart-failure component, and the mortality components side by side makes that distinction explicit rather than leaving it implicit within a composite endpoint.

The endpoint-definition analysis addresses a recurring objection to pooling these four trials. The concern that dapagliflozin composites are inflated by urgent heart-failure visits is reasonable in principle, since composite endpoints are known to be driven by their more frequent and less severe components [7,8]. In these trials it does not hold: within DAPA-HF the two definitions differ by a ratio of hazard ratios of 0.99, and substituting the hospitalization-restricted composite leaves the pooled estimate at 0.78. Pooling under a trial-defined composite label is defensible here, and we report the quantification rather than the assurance.

Clinical and methodological interpretation

For clinicians, dapagliflozin and empagliflozin should be understood primarily as therapies that reliably prevent worsening heart-failure events across chronic heart-failure phenotypes. Hospitalization and urgent heart-failure visits are patient-important outcomes associated with symptom burden, functional decline, cost, and subsequent risk. The absolute-effect display adds clinical context to the relative estimates while reinforcing that absolute benefit depends on baseline risk and follow-up duration.

For methodologists, the review illustrates that endpoint harmonization is testable rather than merely assertable. A pooled composite is defensible when endpoint wording differs modestly across trials, provided the differences are acknowledged, the components are presented, and the consequence of the difference is quantified where the trials report enough to permit it. Pooling should not erase distinctions between urgent visits, hospitalizations, cardiovascular death, and all-cause death, and certainty must be rated separately for each outcome even when the composite and its components point in the same direction.

Strengths

The review uses a clinically narrow PICO rather than a broad class-level approach that would combine heterogeneous populations and interventions. It retains the hazard ratio as the effect measure, matching the time-to-event design of the trials. It separates the composite from its components and rates certainty for each, which prevents overstatement of mortality benefit. It quantifies, rather than assumes, the consequence of the trials' differing composite definitions. Screening and extraction were independent and duplicated, which limits selection and transcription error. The PRISMA flow uses one internally consistent record-level path with a transparent supplementary source note for CENTRAL. The Supplementary Appendix contains the search ledger, extraction sheet, endpoint-

harmonization rules, RoB 2 rationale, GRADE evidence profile, and statistical provenance.

The visual reporting supports the same structure. The graphical abstract states the evidence base and the main inference; the PRISMA diagram communicates the record flow; the forest plot includes trial context, event counts, and weights; and the absolute-effect figure translates relative effects onto a patient-centered scale.

Limitations

Several limitations require explicit qualification. The synthesis includes only four trials, which suits the narrow question but limits subgroup inference and assessment of small-study effects. The exploratory prediction interval remained below 1.0, but a prediction interval based on four trials with τ^2 estimated as zero is fragile and is offered as supportive rather than definitive.

The record-level screening pathway rested on MEDLINE/PubMed alone, with CENTRAL reported as count-only because export and deduplication were unavailable in the documented workflow. Embase and Scopus were not searched. The review was also not prospectively registered. These two features are reproducibility weaknesses; readers should weigh the synthesis accordingly, and we anticipate addressing both in a registered, multi-database update.

The endpoint-definition sensitivity analysis substituted one trial only. DELIVER does not report a time-to-first-event composite restricted to hospitalization, so the pooled estimate could not be recomputed under a fully harmonized definition across all four trials. The finding that the pooled estimate is insensitive to composite wording therefore rests on DAPA-HF and on the identity of the EMPEROR definitions, not on a complete harmonization.

All four pivotal trials were industry sponsored, which is relevant context even though each was double-blind, placebo-controlled, and event-driven. The analysis uses aggregate published estimates rather than individual participant data and cannot test patient-level effect modification by diabetes, kidney function, age, sex, frailty, or continuous ejection fraction; patient-level analysis is stronger for those questions. Approximate absolute effects per 1000 participants are clinical translations of relative time-to-event effects, not reconstructed Kaplan-Meier estimates. Subgroup analyses by drug and phenotype are descriptive only.

The review excludes sotagliflozin, acute heart-failure trials, post-myocardial-infarction trials, and diabetes or kidney outcome trials from the primary synthesis. This limits class-level generalizability and protects the coherence of the chronic symptomatic heart-failure question. The manuscript should be read as a focused synthesis of the pivotal chronic heart-failure trials of dapagliflozin and empagliflozin, not as a definitive meta-analysis of every SGLT-pathway inhibitor in every cardiovascular population.

Implications for future research

Future syntheses could extend this work through individual-participant-data meta-analysis across the four pivotal trials, harmonization of urgent-visit and hospitalization definitions, estimation of absolute risk reductions by baseline-risk strata, and an updated record flow with complete Embase and CENTRAL export. Future trials and secondary analyses should report component outcomes consistently and provide time-to-event estimates for both first and recurrent worsening heart-failure events. Reporting the composite alongside a hospitalization-restricted definition would allow the sensitivity analysis attempted here to be completed for every trial, and we anticipate that such reporting could in time become routine.

Conclusions

Dapagliflozin and empagliflozin reduce trial-defined cardiovascular death or worsening heart failure/hospitalization for heart failure across chronic heart-failure trials enrolling reduced and above-40% ejection-fraction phenotypes, and the pooled estimate does not depend on the differing composite wording of the contributing trials. Component-level synthesis indicates that the treatment effect is driven primarily by fewer worsening heart-failure or hospitalization events. Cardiovascular death and all-cause death are directionally favorable but imprecise as standalone outcomes in this aggregate-data four-trial synthesis. The available evidence supports SGLT2 inhibitor use to prevent worsening heart failure across the ejection-fraction spectrum, without overstating a mortality claim beyond the precision of the component evidence.

Declarations

Ethics approval and consent to participate: Not applicable; this review used published aggregate data.

Consent for publication: Not applicable.

Funding: No specific external funding was received for this work.

Competing interests: The authors declare no competing interests.

Data availability: Extracted aggregate data, the search ledger, risk-of-bias judgments, the GRADE profile, and statistical derivations are provided in the Supplementary Appendix.

Author contributions: Conceptualization: N.A.M.A., M.G.A., and I.F.I.S.; Methodology: N.A.M.A., M.G.A., and I.F.I.S.; Investigation (independent duplicate screening and data

extraction): N.A.M.A. and A.A.A.; Data curation: N.A.M.A., A.A.A., and M.G.A.; Formal analysis: N.A.M.A., A.A.A., and M.G.A.; Visualization: N.A.M.A. and M.G.A.; Writing—original draft: N.A.M.A.; Writing—review and editing: A.A.A., M.G.A., L.M.A., and I.F.I.S.; Supervision and adjudication of screening disagreements: I.F.I.S.; Project administration: N.A.M.A. and M.G.A. All authors have read and approved the final version of the manuscript and agree to be accountable for the integrity of the work.

Abbreviations: CI, confidence interval; CV, cardiovascular; EF, ejection fraction; GRADE, Grading of Recommendations Assessment, Development and Evaluation; HF, heart failure; HFH, hospitalization for heart failure; HFrEF, heart failure with reduced ejection fraction; HFmrEF, heart failure with mildly reduced ejection fraction; HFpEF, heart failure with preserved ejection fraction; HR, hazard ratio; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RCT, randomized controlled trial; SGLT2, sodium-glucose cotransporter 2.

References

1. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2023 Focused Update of the 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2023;44:3627–3639.
2. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure. Circulation. 2022;145:e895–e1032.
3. McMurray JJV, Solomon SD, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, et al. Dapagliflozin in patients with heart failure and reduced ejection fraction. N Engl J Med. 2019;381:1995–2008.
4. Solomon SD, McMurray JJV, Claggett B, de Boer RA, DeMets D, Hernandez AF, et al. Dapagliflozin in heart failure with mildly reduced or preserved ejection fraction. N Engl J Med. 2022;387:1089–1098.
5. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, et al. Cardiovascular and renal outcomes with empagliflozin in heart failure. N Engl J Med. 2020;383:1413–1424.
6. Anker SD, Butler J, Filippatos G, Ferreira JP, Bocchi E, Böhm M, et al. Empagliflozin in heart failure with a preserved ejection fraction. N Engl J Med. 2021;385:1451–1461.
7. Montori VM, Permanyer-Miralda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. BMJ. 2005;330:594–596.
8. Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. BMJ. 2007;334:786.
9. Vaduganathan M, Docherty KF, Claggett BL, Jhund PS, de Boer RA, Hernandez AF, et al. SGLT2 inhibitors in patients with heart failure: a comprehensive meta-analysis of five randomised controlled trials. Lancet. 2022;400:757–767.
10. Jhund PS, Kondo T, Butt JH, Docherty KF, Claggett BL, Desai AS, et al. Dapagliflozin across the range of ejection fraction in patients with heart failure: a patient-level, pooled meta-analysis of DAPA-HF and DELIVER. Nat Med. 2022;28:1956–1964.
11. Zannad F, Ferreira JP, Pocock SJ, Anker SD, Butler J, Filippatos G, et al. SGLT2 inhibitors in patients with heart failure with reduced ejection fraction: a meta-analysis of the EMPEROR-Reduced and DAPA-HF trials. Lancet. 2020;396:819–829.
12. Butler J, Usman MS, Khan MS, Greene SJ, Friede T, Vaduganathan M, et al. Efficacy and safety of SGLT2 inhibitors in heart failure: systematic review and meta-analysis. ESC Heart Fail. 2020;7:3298–3309.
13. Butt JH, Docherty KF, Jhund PS, de Boer RA, McMurray JJV. SGLT2 inhibitors in heart failure across the spectrum of ejection fraction. Eur Heart J. 2022;43:4161–4164.
14. Vukadinović D, Abdin A, Anker SD, Rosano GMC, Mahfoud F, Packer M, et al. Side effects and treatment initiation barriers of sodium-glucose cotransporter 2 inhibitors in heart failure: a systematic review and meta-analysis. Eur J Heart Fail. 2022;24:1625–1632.
15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
16. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA

statement for reporting literature searches in systematic reviews. Syst Rev. 2021;10:39.
17. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;350:g7647.
18. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.5. Cochrane; 2024.
19. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898.
20. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–926.
21. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6: rating the quality of evidence — imprecision. J Clin Epidemiol. 2011;64:1283–1293.
22. Schünemann HJ, Higgins JPT, Vist GE, Glasziou P, Akl EA, Skoetz N, et al. Completing Summary of Findings tables and grading the certainty of the evidence. In: Cochrane Handbook for Systematic Reviews of Interventions. Cochrane; 2024.
23. Bhatt DL, Szarek M, Steg PG, Cannon CP, Leiter LA, McGuire DK, et al. Sotagliflozin in patients with diabetes and recent worsening heart failure. N Engl J Med. 2021;384:117–128.
24. Bhatt DL, Szarek M, Pitt B, Cannon CP, Leiter LA, McGuire DK, et al. Sotagliflozin in patients with diabetes and chronic kidney disease. N Engl J Med. 2021;384:129–139.
25. Voors AA, Angermann CE, Teerlink JR, Collins SP, Kosiborod M, Biegus J, et al. The SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure: the EMPULSE trial. Nat Med. 2022;28:568–574.
26. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. 2nd ed. Wiley; 2021.
27. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36:1–48.
28. IntHout J, Ioannidis JPA, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol. 2014;14:25.
29. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557–560.
30. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ. 2011;342:d549.
31. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open. 2016;6:e010247.
32. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002.

Dapagliflozin and empagliflozin versus placebo for cardiovascular death or worsening heart failure across the ejection-fraction spectrum: a systematic review and meta-analysis of randomized placebo-controlled trials

Supplementary Appendix

Dapagliflozin and empagliflozin versus placebo for cardiovascular death or worsening heart failure across the ejection-fraction spectrum: a systematic review and meta- analysis of randomized placebo-controlled trials

Nawaf AM Alharbi1,*, Abdullah Ahmed Alfllaj 1, Mohamed Ghaith Al Abdin2, Lana Musaad Alhatlani3, Ihab Fathi Ibrahim Suliman4,5

1 Department of Clinical Sciences, College of Medicine, Sulaiman AlRajhi University, AlBukayriyah, Qassim, Saudi Arabia

2 Department of Internal Medicine, Dr. Sulaiman Al-Habib Medical Group, Qassim, Saudi Arabia 3 College of Medicine, Qassim University, Qassim, Saudi Arabia

4 Department of Cardiac Sciences, King Abdulaziz Cardiac Center, King Abdulaziz Medical City, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia

5 College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

ORCID iDs: Abdullah Ahmed Alfllaj: https://orcid.org/0009-0002-8005-2419; Mohamed Ghaith Al Abdin: https://orcid.org/0009-0002-9461-3555; Lana Musaad Alhatlani: https://orcid.org/0009- 0005-7571-5611; Ihab Fathi Ibrahim Suliman: https://orcid.org/0000-0001-8656-7013

This appendix provides supplementary protocol details, search documentation, PRISMA flow, the selection-process record, extraction tables, endpoint-harmonization details, risk-of-bias judgments, certainty assessment, statistical details, and additional figures supporting the main manuscript.

Appendix 1. Protocol framework

Item	Specification
Review type	Systematic review and meta-analysis of randomized placebo-controlled intervention trials.
Population	Adults with symptomatic chronic heart failure across reduced, mildly reduced, and preserved ejection-fraction phenotypes.
Intervention	Dapagliflozin 10 mg once daily or empagliflozin 10 mg once daily.
Comparator	Placebo plus usual background therapy.
Primary outcome	Cardiovascular death or trial-defined worsening heart failure/hospitalization for heart failure.
Secondary outcomes	Worsening heart failure/hospitalization for heart failure; cardiovascular death; all-cause death.
Effect measure	Trial-reported hazard ratio and 95% confidence interval.
Selection process	Independent duplicate screening and extraction by two reviewers (N.A.M.A., A.A.A.); all disagreements adjudicated by the senior author (I.F.I.S.).

Primary model	Random-effects inverse-variance pooling, Paule-Mandel τ^2, ad hoc Hartung-Knapp adjustment.
Prespecified sensitivity analyses	Fixed-effect pooling; leave-one-out; endpoint- definition substitution.
Registration	Not prospectively registered; protocol framework and analysis plan archived in this Supplement before journal submission.

Appendix 2. PICOS eligibility criteria

Domain	Included	Excluded/notes
P	Adults with symptomatic chronic heart failure across the EF spectrum	Acute decompensated HF- only, post-MI without chronic symptomatic HF target syndrome, diabetes/kidney- only populations
I	Dapagliflozin 10 mg or empagliflozin 10 mg once daily	Other SGLT2 inhibitors or SGLT1/2 inhibitors in the primary synthesis
C	Placebo plus usual care	Active comparator or no placebo control
O	CV death or worsening HF/HF hospitalization; components separately	No extractable time-to-event HR for an eligible endpoint
S	Randomized, double-blind, placebo-controlled chronic	Non-randomized, observational, mechanistic-

Domain	Included	Excluded/notes
	HF outcome trials	only, or secondary analysis without a primary trial estimate

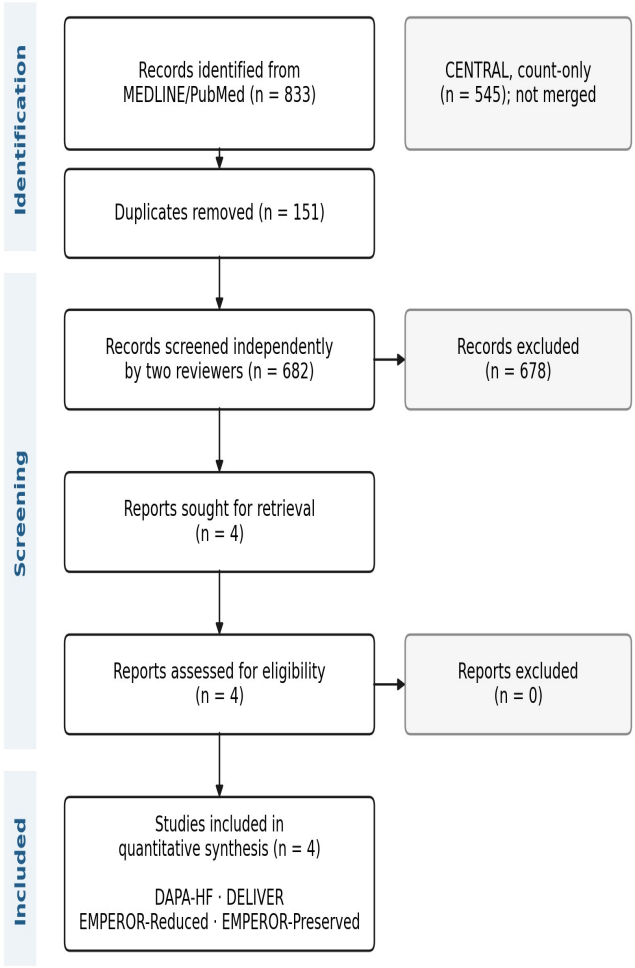
Appendix 3. PRISMA-S search documentation

The record-level flow is anchored to exportable MEDLINE/PubMed results locked on 25 February 2026. CENTRAL was searched as an additional source; because record-level export was unavailable in the documented workflow, CENTRAL is reported as count-only and was not merged into the screened denominator. Embase and Scopus were not searched; this is stated as a limitation in the main manuscript.

Source	Search string/action	Result count

PubMed/MEDLINE main concept-block search (25 February 2026)	(heart failure[Title/Abstract] OR cardiac failure[Title/Abstract]) AND (dapagliflozin[Title/Abstract] OR empagliflozin[Title/Abstract]) AND (randomized[Title/Abstract] OR placebo[Title/Abstract] OR trial[Title/Abstract])	329
PubMed/MEDLINE sensitivity search (25 February 2026)	(dapagliflozin OR empagliflozin) AND (heart failure) AND (placebo OR randomized OR trial)	504
CENTRAL count-only query (25 February 2026)	heart failure AND (dapagliflozin OR empagliflozin) AND placebo	545
Citation chasing	Pivotal trials, prior SGLT2 meta-analyses, ClinicalTrials.gov records, and HF guidelines	Qualitative

Appendix 4. PRISMA 2020 selection flow



Supplementary Figure S1. PRISMA 2020 flow diagram for the documented search ledger.

Stage	**n**
Records identified (two PubMed/MEDLINE searches)	833
Duplicates removed	151
Unique records screened by both reviewers	682
Records excluded at title/abstract	678
Full-text reports assessed	4
Reports excluded after full-text assessment	0

Stage	**n**

Studies included in quantitative synthesis	4
CENTRAL records (count-only, not merged into the denominator)	545

Appendix 5. Selection process and inter-reviewer agreement

Two reviewers (N.A.M.A., A.A.A.) screened all 682 unique records independently, each without access to the other's decisions, and extracted data from the eligible reports independently. The two sets of decisions were then compared.

Item	Value
Records screened independently by each reviewer	682
Records with concordant decisions	649
Records with discordant decisions	33
Simple agreement	95.2%
Disagreements adjudicated by senior author (I.F.I.S.)	33 (100%)
Trials selected independently by both reviewers	4 of 4

All 33 disagreements arose at title/abstract level among records ultimately excluded, and none affected the composition of the final included set. Simple agreement is reported rather than a chance-corrected coefficient because the marginal distribution was extreme (678 exclusions versus 4 inclusions), a configuration in which chance-corrected coefficients are unstable and difficult to interpret.

Appendix 6. Full-text inclusion and adjacent-study exclusion log

Study/report	Decision	Reason
DAPA-HF	Included	Dapagliflozin, chronic HFrEF, placebo-controlled, eligible composite HR
DELIVER	Included	Dapagliflozin, chronic HF with EF >40%, placebo- controlled, eligible composite

Study/report	Decision	Reason
		HR
EMPEROR-Reduced	Included	Empagliflozin, chronic HFrEF, placebo-controlled, eligible composite HR

EMPEROR-Preserved	Included	Empagliflozin, chronic HF with EF >40%, placebo- controlled, eligible composite HR
SOLOIST-WHF	Excluded from primary synthesis	Sotagliflozin is dual SGLT1/2 inhibition and the trial included recent worsening HF; adjacent but outside the dapagliflozin/empagliflozin chronic HF PICO
SCORED	Excluded from primary synthesis	Sotagliflozin CKD/diabetes population; not the chronic HF PICO
EMPULSE	Excluded from primary synthesis	Hospitalized acute HF stabilization trial; not a chronic stable outcome-trial PICO
DAPA-MI / EMPACT-MI	Excluded from primary synthesis	Post-MI populations; not the chronic symptomatic HF target syndrome

Appendix 7. Data extraction sheet

Trial	**Drug**	**Populati on**	**Active n**	**Placebo n**	**Active events**	**Placebo events**	**Follow- up**
DAPA-HF	Dapagliflo zin	HFrEF; LVEF ≤40%	2373	2371	386	502	18.2 mo
DELIVER	Dapagliflo zin	HFmrEF/ HFpEF; LVEF >40%	3131	3132	512	610	2.3 y
EMPERO R-	Empaglifl ozin	HFrEF; LVEF	1863	1867	361	462	16 mo

Trial	**Drug**	**Populati on**	**Active n**	**Placebo n**	**Active events**	**Placebo events**	**Follow- up**
Reduced		≤40%					
EMPERO R-Preserved	Empaglifl ozin	HFmrEF/ HFpEF; LVEF >40%	2997	2991	415	511	26.2 mo

Trial	Primary endpoint	HR	Lower CI	Upper CI	Registry
DAPA-HF	CV death or worsening HF (HF hospitalizatio n or urgent HF visit requiring IV therapy)	0.74	0.65	0.85	NCT03036124
DELIVER	CV death or worsening HF (unplanned HF hospitalizatio n or urgent HF visit)	0.82	0.73	0.92	NCT03619213
EMPEROR-Reduced	CV death or hospitalizatio n for worsening HF	0.75	0.65	0.86	NCT03057977
EMPEROR-Preserved	CV death or hospitalizatio n for HF	0.79	0.69	0.90	NCT03057951

Source of support: DAPA-HF and DELIVER were funded by AstraZeneca; EMPEROR-Reduced and EMPEROR-Preserved were funded by Boehringer Ingelheim and Eli Lilly. Sponsorship was considered as contextual information; risk-of-bias judgments were based on trial design,
conduct, outcome measurement, and reporting.

Appendix 8. Component outcome extraction

Primary composite

Trial	HR	Lower CI	Upper CI	Active events / active n	Placebo events / placebo n
DAPA-HF	0.74	0.65	0.85	386 / 2373	502 / 2371
DELIVER	0.82	0.73	0.92	512 / 3131	610 / 3132
EMPEROR-Reduced	0.75	0.65	0.86	361 / 1863	462 / 1867
EMPEROR-Preserved	0.79	0.69	0.90	415 / 2997	511 / 2991

Worsening HF/HF hospitalization

Trial	HR	Lower CI	Upper CI	Active events / active n	Placebo events / placebo n
DAPA-HF	0.70	0.59	0.83	237 / 2373	326 / 2371
DELIVER	0.79	0.69	0.91	368 / 3131	455 / 3132
EMPEROR-Reduced	0.69	0.59	0.81	246 / 1863	342 / 1867
EMPEROR-Preserved	0.71	0.60	0.83	259 / 2997	352 / 2991

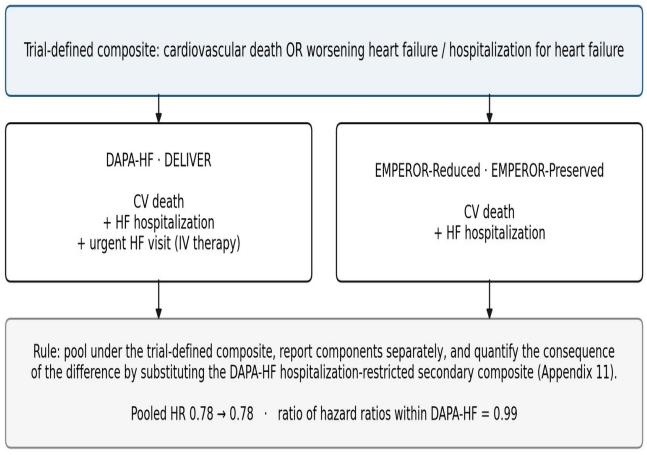
Cardiovascular death

Trial	HR	Lower CI	Upper CI	Active events / active n	Placebo events / placebo n
DAPA-HF	0.82	0.69	0.98	227 / 2373	273 / 2371
DELIVER	0.88	0.74	1.05	231 / 3131	261 / 3132
EMPEROR-Reduced	0.92	0.75	1.12	187 / 1863	202 / 1867
EMPEROR-Preserved	0.91	0.76	1.09	219 / 2997	244 / 2991

All-cause death

Trial	HR	Lower CI	Upper CI	Active events / active n	Placebo events / placebo n
DAPA-HF	0.83	0.71	0.97	276 / 2373	329 / 2371
DELIVER	0.94	0.83	1.07	497 / 3131	526 / 3132
EMPEROR-Reduced	0.92	0.77	1.10	249 / 1863	266 / 1867
EMPEROR-Preserved	1.00	0.87	1.15	422 / 2997	427 / 2991

Appendix 9. Endpoint harmonization



Supplementary Figure S2. Endpoint harmonization rule for the trial-defined composite.
The main manuscript uses the phrase "trial-defined worsening heart failure/hospitalization for heart failure" to preserve the original trial wording. Component outcomes are reported separately to avoid attributing the composite effect directly to mortality components.

Trial	Composite as defined by the trial	Urgent HF visits counted?
DAPA-HF	CV death or worsening HF (HF hospitalization or urgent HF visit requiring IV therapy)	Yes
DELIVER	CV death or worsening HF (unplanned HF hospitalization or urgent HF	Yes

Trial	Composite as defined by the trial	Urgent HF visits counted?
	visit)	

EMPEROR-Reduced	CV death or hospitalization for worsening HF	No
EMPEROR-Preserved	CV death or hospitalization for HF	No

The consequence of this difference is quantified in Appendix 11 rather than assumed to be negligible.

Appendix 10. Statistical provenance

Outcome	**HR (95% CI)**	**Q**	**I^2**	**τ^2**	**p value**
Primary composite	0.778 (0.701–0.865)	1.643	0.0%	0.00000	0.0048
Worsening HF/HF hospitalizatio n	0.727 (0.641–0.825)	2.074	0.0%	0.00000	0.0041
Cardiovascula r death	0.879 (0.758–1.019)	0.943	0.0%	0.00000	0.0689
All-cause death	0.927 (0.821–1.046)	3.115	3.7%	0.00022	0.1401

Standard errors were derived as SE = (ln upper CI − ln lower CI) / (2 × 1.96). Hazard ratios were pooled on the log scale and back-transformed.

The main model used random-effects inverse-variance pooling with Paule-Mandel τ^2 estimation and the Hartung-Knapp adjustment. The adjustment was applied in its ad hoc form: the weighted mean-square residual q was truncated at one where it fell below one, so that the random-effects confidence interval could not be narrower than the corresponding fixed-effect interval. This variant is documented explicitly because it is consequential here. For the primary composite, q = 0.548 before truncation; the unadjusted Hartung-Knapp interval would have been 0.720–0.842, whereas the ad hoc interval reported throughout is 0.701–0.865. The ad hoc form is the conservative choice and was prespecified because only four studies contribute to the main analysis.

Fixed-effect inverse-variance pooling was used as a sensitivity analysis. For the primary composite the fixed-effect estimate was HR 0.778 (95% CI 0.730–0.831).

The exploratory 95% prediction interval for the primary composite was calculated as the pooled log hazard ratio ± $t(k-2) \times \sqrt{(\tau^2 + SE^2)}$, giving 0.68–0.90. With four trials and τ^2 estimated as zero, this interval is fragile and is reported as supportive only.

Appendix 11. Endpoint-definition sensitivity analysis

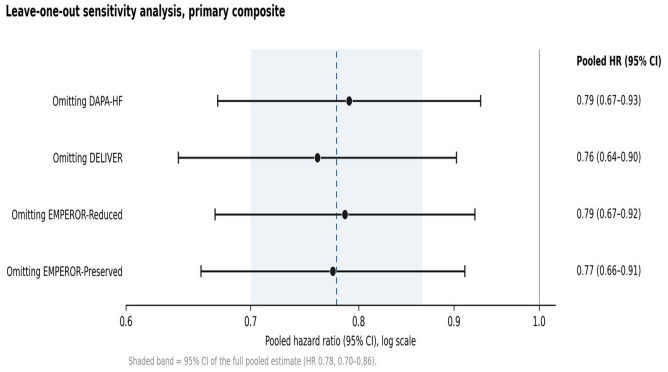
DAPA-HF reports, as a prespecified secondary outcome, the composite of cardiovascular death or hospitalization for heart failure (HR 0.75, 95% CI 0.65–0.85), which matches the composite definition used by both EMPEROR trials. The primary model was refitted with that estimate substituted for the DAPA-HF primary composite, holding the other three trials unchanged.

Model	**DAPA-HF input**	**Pooled HR (95% CI)**	**Q**	**I^2**
As reported (trial-defined composite)	0.74 (0.65–0.85)	0.778 (0.701– 0.865)	1.643	0.0%
Endpoint-definition sensitivity	0.75 (0.65–0.85)	0.781 (0.703– 0.868)	1.382	0.0%

Within DAPA-HF, the ratio of hazard ratios between the two definitions is 0.74 / 0.75 = 0.99. The inclusion of urgent heart-failure visits therefore accounts for a negligible part of the pooled effect.

DELIVER does not report a time-to-first-event composite restricted to hospitalization for heart failure; its prespecified secondary outcomes were the total number of worsening heart-failure events and cardiovascular deaths, the KCCQ total symptom score, cardiovascular death, and death from any cause. Full harmonization across all four trials was therefore not possible, and this analysis is a one-trial substitution. The conclusion that the pooled estimate is insensitive to composite wording rests on DAPA-HF and on the identity of the two EMPEROR definitions.

Appendix 12. Leave-one-out sensitivity analysis



Supplementary Figure S3. Leave-one-out sensitivity analysis for the primary composite.

al omitted	**Pooled HR**
PA-HF	0.79
LIVER	0.76
PEROR-Reduced	0.79
PEROR-Preserved	0.77

All leave-one-out estimates remain directionally consistent with the pooled estimate (HR 0.78, 95% CI 0.70–0.86).

Appendix 13. RoB 2 domain-level assessment

Trial	**D1 randomization**	**D2 deviations**	**D3 missing data**	**D4 measurement**	**D5 reporting**	**Overall**	**Rationale**
DAPA-HF	Low	Low	Low	Low	Low	Low	Randomized, double-blind, placebo-controlled, adjudicated endpoint, prespecified analysis.

DELIVER	Low	Low	Low	Low	Low	Low	Randomized, double-blind, placebo-controlled, adjudicated endpoint, prespecified analysis.

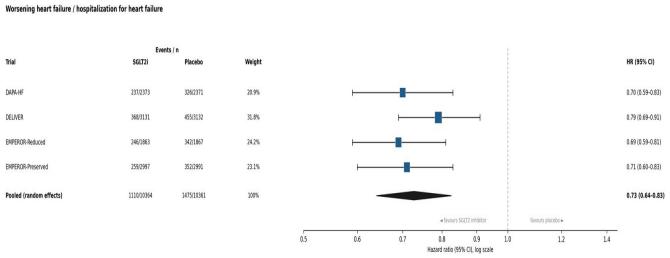
Trial	**D1 randomization**	**D2 deviations**	**D3 missing data**	**D4 measurement**	**D5 reporting**	**Overall**	**Rationale**
EMPEROR-Reduced	Low	Low	Low	Low	Low	Low	Randomized, double-blind, placebo-controlled, adjudicated endpoint, prespecified analysis.
EMPEROR-Preserved	Low	Low	Low	Low	Low	Low	Randomized, double-blind, placebo-controlled, adjudicated endpoint, prespecified analysis.

Appendix 14. GRADE evidence profile

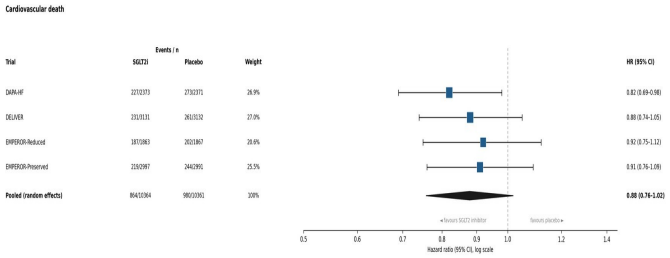
Outcome	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Certainty	Reason for rating
Primary composite	Low	Not serious	Not serious	Not serious	Not assessed	High	Consistent, precise reduction; no downgrade.
Worsening HF/HF hospitalization	Low	Not serious	Not serious	Not serious	Not assessed	High	Consistent, precise reduction; no downgrade.
Cardiovascular death	Low	Not serious	Not serious	Serious	Not assessed	Moderate	Downgraded one level for imprecision; CI crosses the null.
All-cause death	Low	Not serious	Not serious	Serious	Not assessed	Moderate	Downgraded one level for imprecision; CI crosses the null.

The publication-bias domain is recorded as not assessed for every outcome. With four contributing studies, funnel-plot asymmetry testing is uninformative, and no outcome was either downgraded or reassured on that basis.

Appendix 15a. Supplementary forest plot: worsening HF/HF hospitalization

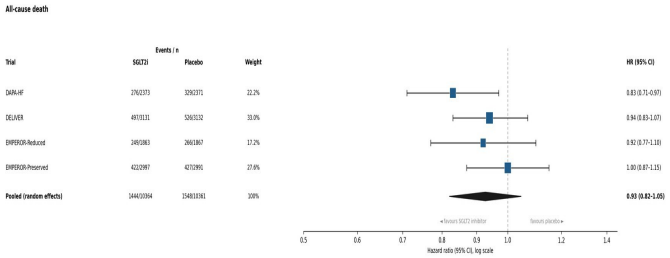


*Supplementary Figure S4. Forest plot for worsening HF/HF hospitalization.***Appendix 15b. Supplementary forest plot: cardiovascular death**



Supplementary Figure S5. Forest plot for cardiovascular death.

Appendix 15c. Supplementary forest plot: all-cause death



Supplementary Figure S6. Forest plot for all-cause death.

Appendix 16. PRISMA checklist location map

PRISMA item area	**Where addressed**
Title	Title page
Structured abstract	Abstract
Rationale/objectives	Introduction
Eligibility criteria	Methods and Appendix 2
Information sources/search	Methods and Appendix 3
Selection process	Methods and Appendices 4, 5
Data items/extraction	Methods and Appendices 7, 8
Risk of bias	Methods and Appendix 13
Synthesis methods	Methods and Appendix 10
Results of individual studies	Table 1 and Appendices 7, 8
Synthesis results	Results, Figures 3–5

PRISMA item area	**Where addressed**
Sensitivity analyses	Results and Appendices 11, 12
Certainty assessment	Table 2 and Appendix 14
Limitations	Discussion
Registration/protocol	Methods and Appendix 1
Funding/conflicts	Declarations