NICE Cardiology Guidance Under the Microscope: A Data-Driven Statistical Critique

Zaki Khan
Ahmadiyya Hospital, Kenema, Sierra Leone
Synthēsis · E156 Research Letter · Vol 2 No 7 · 2026 · Published 10 July 2026 · doi: synthesis.v160.r2

Correction and Editorial Note (Version 2 — 10 July 2026)

This corrected version repairs reference and arithmetic errors in the original while confirming its core result. Seven of twelve reference PMIDs were wrong and are corrected; the original E-value (~2.1) was arithmetically incorrect and is corrected to 1.85; and the unverifiable claim that NICE was the “last of six” bodies on a specific named technology appraisal is withdrawn as not independently verifiable. The reproducible core is unchanged and independently regenerated by the companion script: the SGLT2-inhibitor class effect on cardiovascular death or heart-failure hospitalisation is HR 0.79 (I^2=8%, Q=7.61), and a new analysis shows no difference between dedicated heart-failure trials and diabetes cardiovascular-outcome trials (interaction p=0.72).

This corrected and expanded Version 2 supersedes Version 1, which remains available in this article's version history. Every quantitative claim is regenerated by a deterministic companion verification script, and every retained reference PMID was re-verified against PubMed metadata.

Abstract

Background. Sodium–glucose co-transporter-2 (SGLT2) inhibitors are among the most thoroughly evidenced advances in cardiovascular pharmacotherapy of the past decade. We ask two questions: (i) how statistically robust and reversal-resistant is the class effect on the composite of cardiovascular death or heart-failure hospitalisation (CV-death/HHF), and (ii) can guideline-adoption timing across national bodies be quantified from the public record?

Methods. Eight pivotal randomised trials reporting a CV-death/HHF composite were pooled on the log-hazard-ratio scale using REML (primary), Paule–Mandel and DerSimonian–Laird τ^2 estimators, each with a Hartung–Knapp–Sidik–Jonkman (HKSJ) t-interval (q≥1 floor) and, for like-for-like reproduction, a random-effects z-interval. We report Cochran Q, I^2, a 95% prediction interval, leave-one-out influence, a **new** dedicated-HF-trial-versus-diabetes-CVOT subgroup interaction test, a seeded Monte-Carlo reversal probability, and E-values. Guideline dates were taken from the primary guideline publications; unverifiable claims are flagged, not asserted.

Results. Pooled **HR 0.79** (REML; RE z-CI 0.749–0.827; HKSJ 0.739–0.837), **I^2=8.0%**, **Q=7.61** (df=7, p=0.37), $\tau^2 \approx 0$. The 95% prediction interval (0.74–0.84) excludes 1. All eight leave-one-out estimates lie in 0.78–0.80 with CIs excluding 1. The class effect does **not** differ between the four dedicated HF trials (HR 0.781) and the four diabetes CVOTs (HR 0.795): interaction p=0.72. Monte-Carlo P(HR≥1) < 0.001%. E-value 1.85 (point), 1.71 (CI bound).

Conclusion. The SGLT2-inhibitor CV-death/HHF class effect is statistically mature, homogeneous across trial designs, and highly reversal-resistant. The *existence* of measurable inter-body adoption lags is defensible from public dates; the specific claim that NICE was "last of six" is **not** independently verifiable from the substrate here and is flagged accordingly.

1. Introduction

The SGLT2-inhibitor class was developed for type 2 diabetes but, between 2015 and 2022, a succession of large outcome trials demonstrated consistent reductions in heart-failure hospitalisation and cardiovascular death across the ejection-fraction spectrum and in chronic kidney disease. When several guideline bodies

cover the same population and the same evidence base, their adoption timelines become externally comparable, and delays become publicly attributable. This paper does two things a single narrative review cannot: it *reproduces* the class-effect estimate from a transparent script, and it separates what the public record can support from what it cannot.

2. Methods

Outcome definition. The composite of cardiovascular death **or** hospitalisation for heart failure (CV-death/HHF), as reported in each trial's primary publication. For DAPA-HF this is the CV-death-or-HHF composite (HR 0.75), distinguished from the trial's broader primary endpoint that also counted urgent HF visits (HR 0.74); we use the CV-death/HHF form throughout for cross-trial commensurability.

Eligibility. Randomised, placebo-controlled trial of an approved SGLT2 inhibitor reporting a CV-death/HHF composite; eight trials met this (Table 1).

Statistics. Effects pooled on the log-HR scale. Per small-k rules, REML is the primary τ^2 estimator, with Paule–Mandel and DerSimonian–Laird for contrast; DL is shown but not used for inference at $k<10$. Confidence intervals use HKSJ with t_{k-1} and a $q \geq 1$ variance floor (so HKSJ cannot spuriously narrow below the z-interval); the RE z-interval is also reported because it is what v1 quoted. Heterogeneity: Cochran Q, I^2, and the Q p-value. A 95% prediction interval uses t_{k-1}. Influence is assessed by leave-one-out. Reversal probability is the Monte-Carlo P(HR≥1) over 200,000 seeded draws from the fitted random-effects distribution $N(\mu, \tau^2+\mathrm{Var}(\mu))$, cross-checked analytically. E-values follow VanderWeele & Ding (2017). All figures are emitted by the companion script.

3. Results

3.1 Trial-level data (Table 1)

Trial	Agent	Type	HR (95% CI)	PMID
EMPA-REG OUTCOME	Empagliflozin	CVOT	0.66 (0.55–0.79)	26378978
CANVAS	Canagliflozin	CVOT	0.78 (0.67–0.91)	28605608
DECLARE-TIMI 58	Dapagliflozin	CVOT	0.83 (0.73–0.95)	30415602
DAPA-HF	Dapagliflozin	HF	0.75 (0.65–0.85)	31535829
EMPEROR-Reduced	Empagliflozin	HF	0.75 (0.65–0.86)	32865377
VERTIS CV	Ertugliflozin	CVOT	0.88 (0.75–1.03)	32966714
EMPEROR-Preserved	Empagliflozin	HF	0.79 (0.69–0.90)	34449189
DELIVER	Dapagliflozin	HF	0.82 (0.73–0.92)	36027570

"CVOT" = diabetes cardiovascular-outcome trial (CV-death/HHF was a secondary composite); "HF" = dedicated heart-failure trial (CV-death/HHF was the primary or co-primary endpoint). Aggregate HRs are verifiable from each primary publication; patient-level event counts are not independently reconstructable from public aggregate data and are therefore not tabulated as if they were.

3.2 Pooled class effect

Across eight trials the pooled hazard ratio is **HR 0.787** (REML). Heterogeneity is low and non-significant: **I^2 = 8.0%**, **Q = 7.61** (df = 7, **p = 0.37**), with τ^2 estimated at 0 (REML), 0.00056 (PM) and 0.00045 (DL) — i.e. negligible. The REML random-effects z-interval is **0.749–0.827**; the more conservative HKSJ t-interval is **0.739–0.837**. Rounded to two decimals both give the familiar **0.79 (0.74–0.84)**, reproducing the v1 headline exactly and confirming that the choice of τ^2 estimator or interval method does not move the substantive result at this near-zero heterogeneity.

3.3 Prediction interval, influence, and a new subgroup test

The **95% prediction interval** is 0.74–0.84 (REML) and, under the more generous PM τ^2, 0.72–0.86 — in every case excluding 1. A future trial drawn from this distribution would, with high probability, again show benefit.

Leave-one-out (Table 2 in script output) moves the pooled HR only between **0.777 and 0.798**, and every reduced-set CI excludes 1; no single trial — not even the null-trending VERTIS CV (0.88) — drives the result.

New analysis — trial-type interaction. Because four trials were diabetes CVOTs (where CV-death/HHF was secondary) and four were dedicated HF trials (where it was primary), a legitimate concern is that the pooled effect is an artefact of mixing designs and populations. It is not. The dedicated-HF pool is **HR 0.781** (0.732–0.833) and the diabetes-CVOT pool is **HR 0.795** (0.737–0.858); the interaction (ratio of HRs 0.98) is **not significant (z=−0.35, p=0.72)**. The CV-death/HHF benefit is therefore statistically indistinguishable between the two trial designs — a genuinely reassuring, and previously unreported, homogeneity finding.

3.4 Reversal resistance and E-value

The seeded Monte-Carlo reversal probability **P(HR≥1) is < 0.001%** (0 of 200,000 draws; analytic value ≈10^{-8}). This restates, probabilistically, that the evidence base is statistically mature.

For unmeasured confounding, the **E-value for the point estimate HR 0.79 is 1.85**, and for the null-closest CI bound (0.827) it is **1.71**. (v1 reported "≈2.1"; that value is arithmetically incorrect — 1/0.79 + $\sqrt{[(1/0.79)(1/0.79-1)]}$ = 1.85.) Because the eight studies are randomised, unmeasured confounding is not a plausible threat in the first place; the E-value is reported for completeness and is most relevant to any observational extension of the evidence.

3.5 Absolute benefit (NNT) with an explicit denominator

v1's "NNT ≈ 38/year" carried no stated event rate. Using the DAPA-HF placebo-arm CV-death/HHF rate of **≈21.2 per 100 patient-years** (McMurray 2019) and the pooled HR 0.787, the absolute risk reduction is 0.212×(1−0.787) = **4.5 per 100 patient-years**, i.e. **NNT ≈ 22 patient-years** to prevent one event. The NNT is sensitive to baseline risk: in lower-risk HFpEF/CVOT populations (control rates nearer 5–8%/year) the NNT is proportionally larger (≈60–95). No single NNT should be quoted without its baseline risk.

3.6 Guideline-adoption timing — what is and isn't verifiable

The primary guideline publications are date-verifiable: the **2021 ESC HF guidelines** (McDonagh, *Eur Heart J* 2021; PMID 34447992) and the **2022 AHA/ACC/HFSA HF guideline** (Heidenreich, *J Am Coll Cardiol* 2022; PMID 35379503) both post-date DAPA-HF (Sept 2019) by roughly 23 and 31 months respectively, and both give SGLT2 inhibitors a Class I recommendation for HFrEF. **These dates are VERIFIED.**

The further v1 claims — that NICE issued a specific technology appraisal ("TA599") on a specific date and was "the last of six bodies" — are **NOT independently verifiable from the substrate available here** (primary NICE/CCS/NHFA appraisal documents were not on disk at compute time, and TA numbering is easy to misquote). Rather than assert a possibly-wrong appraisal identifier, these specifics are **flagged UNVERIFIED and withdrawn from the quantitative claims**. What survives is the defensible, weaker statement: *guideline bodies covering the same evidence diverged by roughly a year in formal recommendation timing, and this divergence is measurable from public publication dates.* The precise ranking of any one body requires primary-source extraction of every appraisal/guideline date and a pre-specified adoption metric — a worthwhile but separate study.

4. Certainty (GRADE-style summary)

Claim	Certainty	Basis
Class effect HR ≈ 0.79 on CV-death/HHF	**High**	8 RCTs, I^2=8%, PI excludes 1, LOO-robust, reproduced by script
Effect homogeneous across HF trials vs CVOTs	**Moderate–High**	Prespecified-plausible subgroup, interaction p=0.72
Reversal probability negligible	**High**	τ^2≈0, Monte-Carlo P(HR≥1)<0.001%
Inter-body adoption lags exist (~1 yr scale)	**Moderate**	Two guideline dates verified; others from public record

Claim	Certainty	Basis
NICE was "last of six" on a named TA/date	**Very low / UNVERIFIED**	Primary appraisal documents not verified; withdrawn

5. Discussion

The SGLT2-inhibitor class delivers a ~21% relative reduction in CV-death/HHF that is strikingly consistent (I^2=8%, $\tau^2 \approx 0$) across agents, ejection-fraction strata, and — as newly shown here — across the dedicated-HF-trial versus diabetes-CVOT divide. In a 100,000-patient HFrEF population with a 10%/year event rate, full uptake of a 21% reduction would avert ≈2,100 events annually; this is a ceiling that realistic uptake scales down proportionally, and it is offered as an order-of-magnitude policy framing, not a claim of realised benefit. Why any individual guideline body lagged is a question public data cannot settle: technology appraisals embed cost-effectiveness modelling and deliberation that are not fully transparent. The honest contribution of this paper is to fix the reproducible science firmly (the class effect and its robustness) while explicitly declining to over-claim the part of the original narrative — a specific "last-of-six" ranking — that the evidence on hand cannot support.

6. Limitations

Aggregate (not individual-participant) data were used; the CV-death/HHF composite is secondary in the four CVOTs, so their contribution is to a secondary rather than primary endpoint. The guideline-timing analysis rests on publication dates, not internal decision records, and the specific NICE appraisal claim is withdrawn as unverified. NNT is baseline-risk dependent and is reported with its assumed control rate. τ^2 is near zero here, so prediction-interval width is dominated by sampling error; with even modest true heterogeneity the PI would widen.

7. Corrected references (PMIDs re-verified against PubMed metadata)

Entries marked **[corrected]** had a wrong PMID in v1; the wrong target is noted.

1. Zinman B, et al. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med.* 2015;373(22):2117-2128. **PMID 26378978.**

2. Neal B, et al. Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med.* 2017;377(7):644-657. **PMID 28605608.**

3. Wiviott SD, et al. Dapagliflozin and cardiovascular outcomes in type 2 diabetes (DECLARE-TIMI 58). *N Engl J Med.* 2019;380(4):347-357. **PMID 30415602. [corrected — v1's 30586774 = a 2018 AHA/ACC guideline]**

4. McMurray JJV, et al. Dapagliflozin in heart failure and reduced ejection fraction (DAPA-HF). *N Engl J Med.* 2019;381(21):1995-2008. **PMID 31535829. [corrected — v1's 31535830 = a kidney ischaemia–reperfusion study]**

5. Packer M, et al. Cardiovascular and renal outcomes with empagliflozin in heart failure (EMPEROR-Reduced). *N Engl J Med.* 2020;383(15):1413-1424. **PMID 32865377.**

6. Cannon CP, et al. Cardiovascular outcomes with ertugliflozin in type 2 diabetes (VERTIS CV). *N Engl J Med.* 2020;383(15):1425-1435. **PMID 32966714. [corrected — v1's 32966728 = a resident work-hours trial]**

7. Anker SD, et al. Empagliflozin in heart failure with a preserved ejection fraction (EMPEROR-Preserved). *N Engl J Med.* 2021;385(16):1451-1461. **PMID 34449189. [corrected — v1's 34780166 = a diesel-exhaust toxicology paper]**

8. Solomon SD, et al. Dapagliflozin in heart failure with mildly reduced or preserved ejection fraction (DELIVER). *N Engl J Med.* 2022;387(12):1089-1098. **PMID 36027570. [corrected — v1's 36093875 = a**

brain-tumour classification review]

9. McDonagh TA, et al. 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* 2021;42(36):3599-3726. **PMID 34447992.**

10. Heidenreich PA, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure. *J Am Coll Cardiol.* 2022;79(17):e263-e421. **PMID 35379503.**

11. Vaduganathan M, et al. Estimating lifetime benefits of comprehensive disease-modifying pharmacological therapies in HFrEF. *Lancet.* 2020;396(10244):121-128. **PMID 32446323. [corrected — v1's 32493620 = an otic-tuberculosis case report; and the journal/pages were wrong]**

12. Wetterslev J, et al. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol.* 2008;61(1):64-75. **PMID 18083463. [corrected — v1's 18078785 = a sports heat-stress commentary]**