SGLT2-Inhibitoren bei Herzinsuffizienz über das gesamte Ejektionsfraktions-Spektrum (HFrEF, HFmrEF, HFpEF): Kontrastive Zwillings-Replikation (Fixierter Referenz-Korpus vs. Zeit-eingefrorener Blindlauf ≤ 01.09.2022) mit Drei-Ebenen-REML/CR2-Synthese, NNT-Quantifizierung und Carlisle-Baseline-Forensik

Entwurf vom 10. November 2026

Autorenschaft & Audit-Signatur: Dr. rer. medic. Franz Hütter, M.A. (Projektleitung, klinisch-methodische Supervision & menschliche Konflikt-Adjudikation `human:franz`) · **MetaAnalysis Studio** (Evidenzwerk v2.6 · Kontrastives Twin-Track-Validierungsprotokoll · R `metafor` 4.8-0 + `clubSandwich` CR2 0.5.11 · Verblindetes Zwei-Häuser-Frontier-Screening `anthropic/claude-opus-5.5` vs. `openai/gpt-6.1-sol`) **Benchmark-Kennung:** Benchmark #6 (`BM-06-CARDIO-HF-SGLT2I-TWIN`) · Projekte `#62` (Arm A: Fixierter Referenz-Korpus) & `#63` (Arm B: Zeit-eingefrorener Blindlauf)

Referenz-Meta-Analyse: Vaduganathan M, Docherty KF, Claggett BL, Jhund PS, de Boer RA, Hernandez AF, Inzucchi SE, Kosiborod MN, Lam CSP, Martinez F, Shah SJ, Verma S, McMurray JJV, Solomon SD (2022): *SGLT-2 inhibitors in patients with heart failure: a comprehensive meta-analysis of five randomised controlled trials.* **The Lancet**, 400(10354), 757–767. DOI: `10.1016/S0140-6736(22)01429-5` (N = 21.947)

Temporale Einfrierung & Quarantäne: Stichtag T (cutoff) = 01.09.2022 (`2018-01-01` bis `2022-09-01`); aktive DOI-Quarantäne gegen `10.1016/S0140-6736(22)01429-5` (Vaduganathan et al. 2022) und `10.1016/S0140-6736(20)31824-9` (Zannad et al. 2020)

Zusammenfassung (Structured Abstract nach PRISMA 2020 & RAISE-2025)

Hintergrund: Die im September 2022 in *The Lancet* publizierte Meta-Analyse von Vaduganathan et al. (k = 5 randomisierte Großstudien, N = 21.947) begründete die Klasse-I-Leitlinienempfehlung für Natrium-Glukose-Kotransporter-2-(SGLT2)-Inhibitoren über das gesamte Spektrum der linksventrikulären Ejektionsfraktion (HFrEF, HFmrEF und HFpEF). Ob eine autonom suchende, verblindet screenende und forensisch prüfende Synthese-Architektur denselben Evidenzstand ohne Rückgriff auf die Referenzarbeit exakt reproduziert und methodische blinde Flecken der Originalarbeit aufdeckt, wurde bislang nicht kontrastiv geprüft.

Methodik (Kontrastives Twin-Track-Design): Wir führten eine kontrastive Zwillings-Replikation durch. In **Arm A (Fixierter Referenz-Korpus, Projekt `#62`)** wurden exakt die k = 5 Primärstudien der Lancet-Referenzarbeit (*DAPA-HF*, *EMPEROR-Reduced*, *SOLOIST-WHF*, *EMPEROR-Preserved*, *DELIVER*; N = 21.947 Patient:innen; m = 15 Vierfeldertafeln über 3 Endpunkte; k (screened) = 5, Cohens κ = 1,00) auf Basis der publizierten 2×2-Ereigniszahlen (Ereignisse e_1, n_1 unter SGLT2-Hemmer vs. e_2, n_2 unter Placebo) ausgewertet. In

Arm B (Zeit-eingefrorener Blindlauf, Projekt `#63`) wurde die Literatursuche deterministisch auf das historische Zeitfenster bis zum T (cutoff) = 01.09.2022 eingefroren und die Referenz-Meta-Analyse per DOI- und Titel-Quarantäne gesperrt. Die k (screened) = 20 identifizierten Kandidaten durchliefen ein verblindetes Zwei-Häuser-Doppelscreening durch zwei disjunkte Frontier-Modellfamilien (`anthropic/claude-opus-5.5` als Gutachter A vs. `openai/gpt-6.1-sol` als Gutachter B) mit menschlicher Konflikt-Adjudikation (`human:franz`). Die statistische Synthese erfolgte in beiden Armen über ein hierarchisches Drei-Ebenen-Random-Effects-Modell (`metafor::rma.mv`, REML) für logarithmierte Risikoverhältnisse (logRR, `metric = "rr"`) mit cluster-robuster Varianzschätzung (`clubSandwich` CR2), Satterthwaite-Freiheitsgraden, 95-%-Prädiktionsintervallen, Number Needed to Treat (NNT), dem siebenstufigen Borenstein-(2019)-Methodenwächter sowie der Carlisle-(2017)-Baseline-Forensik.

Ergebnisse:

1. **Arm A (Fixierter Referenz-Korpus, k = 5 Studien, m = 15 Effekte, N = 21.947):** Über alle drei kardiovaskulären Kern-Endpunkte senken SGLT2-Inhibitoren das Ereignisrisiko hochsignifikant auf RR = 0,792 (logRR = -0,233, 95-%-CR2-KI [0,757; 0,829], df (Satt) = 3,65, p = 0,000211; Heterogenität I^2 (total) = 0,0 %, τ^2 (Studie) = 0,0000, 95-%-Prädiktionsintervall [0,757; 0,829]). Nach Endpunkten getrennt (`otype`, CR2-Wald-F(2, 2,60) = 41,64, p = 0,0106) ergibt sich für den **primären Komposit-Endpunkt (kardiovaskulärer Tod oder erste Herzinsuffizienz-Hospitalisierung)** ein RR = 0,797 ([0,757; 0,839], 1.916/10.977 = 17,45 % vs. 2.416/10.970 = 22,02 %, ARR = 4,57 %, NNT = 22), für die **erste Herzinsuffizienz-Hospitalisierung (First HHF)** ein RR = 0,731 ([0,700; 0,764], 1.196/10.977 = 10,90 % vs. 1.637/10.970 = 14,92 %, ARR = 4,03 %, NNT = 25; Lancet-Referenz: HR = 0,72) und für den **kardiovaskulären Tod (CV Death)** ein RR = 0,879 ([0,831; 0,930], 998/10.977 = 9,09 % vs. 1.139/10.970 = 10,38 %, ARR = 1,29 %, NNT = 78; Lancet-Referenz: HR = 0,87). Zwischen HFrEF (LVEF ≤ 40 %, RR = 0,780), HFmrEF/HFpEF (LVEF > 40 %, RR = 0,814) und akuter/dekompensierter Herzinsuffizienz (RR = 0,738) besteht keine signifikante Effektmodifikation (CR2-Wald-F(2, 1,14) = 9,19, p = 0,1983).
2. **Arm B (Zeit-eingefrorener Blindlauf ≤ 01.09.2022, k = 6 Studien, m = 18 Effekte, N = 22.477):** Das verblindete Zwei-Häuser-Screening über k (screened) = 20 Kandidaten erzielte eine beobachtete Übereinstimmung von 80,0 % und ein empirisches Cohens κ = 0,60 (16/20 Konsens-Entscheidungen, 4 protokollgemäße Grenzfälle zu 12-Wochen-Funktionsstudien durch `human:franz` entschieden; reale API-Gesamtkosten 0,0730 USD). Alle 5/5 Referenzstudien wurden gefunden und eingeschlossen (Retrieval-Recall = 100,0 %, Inklusions-Recall = 100,0 %). Zusätzlich identifizierte Arm B als **Klasse-III-Over-Recall** die im Februar 2022 in *Nature Medicine* publizierte, doppelblinde Multizenter-RCT **EMPULSE** (Voors et al. 2022, DOI `10.1038/s41591-021-01659-1`, N = 530, Empagliflozin bei akuter dekompensierter Herzinsuffizienz), die Vaduganathan et al. nur wegen einer willkürlichen Mindeststichproben-Schwelle (N > 1.000 in der Sekundärsuche) ausgeschlossen hatten. Unter Einschluss von

EMPULSE beträgt der gepoolte Drei-Ebenen-CR2-Effekt RR = 0,789 (logRR = -0,236, 95-%-CR2-KI [0,754; 0,826], df (Satt) = 3,82, p = 0,000168, I^2 (total) = 0,0 %, 95-%-PI [0,754; 0,826]; Abweichung zu Arm A: |DeltaRR| = 0,0026).

3. **Methodenwächter & Baseline-Forensik:** Der Borenstein-(2019)-Wächter bestand 6/7 Prüfregeln ohne Beanstandung und löste bei Regel 6 (Tipton 2015 Small-Sample-RVE) deterministisch den methodisch korrekten Hinweis auf kleine Satterthwaite-Freiheitsgrade (df (Satt) = 3,65 < 4,0 bei k = 5 bzw. 3,82 < 4,0 bei k = 6 Studienclustern) aus. Der Carlisle-(2017)-Baseline-Test bestätigte für 5/6 Studien unauffällige Gruppenbalancen (p (imbalance) = 0,111 bis 0,994) und markierte bei *EMPEROR-Reduced* eine durch Tabellenrundung der LVEF-Mittelwerte bedingte Baseline-Sensitivitätswarnung (p (imbalance) = 0,0029), deren Leave-One-Out-Ausschluss das Gesamtergebnis unverändert lässt (RR = 0,79).

---

1. Einleitung & Klinisch-Methodische Fragestellung

Herzinsuffizienz betrifft weltweit über 64 Millionen Menschen und ist durch hohe Hospitalisierungsraten sowie eine erhebliche kardiovaskuläre Letalität gekennzeichnet. Während für die Herzinsuffizienz mit reduzierter Ejektionsfraktion (HFrEF, LVEF ≤ 40 %) bereits mit *DAPA-HF* (McMurray et al. 2019) und *EMPEROR-Reduced* (Packer et al. 2020) ein ausgeprägter Nutzen von SGLT2-Inhibitoren belegt wurde, blieb die Wirksamkeit bei leichtgradig eingeschränkter oder erhaltener Ejektionsfraktion (HFmrEF/HFpEF, LVEF > 40 %) sowie bei akuter Dekompensation bis zu den Studien *SOLOIST-WHF* (Bhatt et al. 2021), *EMPEROR-Preserved* (Anker et al. 2021) und *DELIVER* (Solomon et al. 2022) Gegenstand intensiver Debatten.

Die zeitgleich zum ESC-Kongress im September 2022 in *The Lancet* veröffentlichte Meta-Analyse von Vaduganathan et al. (2022, k = 5 RCTs, N = 21.947) poolte diese fünf Studien und bildete das Fundament der aktuellen europäischen und amerikanischen Herzinsuffizienz-Leitlinien. Gleichwohl weist die Originalarbeit drei methodische Besonderheiten auf, die eine kontrastive Überprüfung herausfordern:

1. **Hybrides Studienselektions-Design mit willkürlicher N > 1.000-Schwelle:** Vaduganathan et al. kombinierten zwei vorab ausgewählte Studien (*DELIVER* und *EMPEROR-Preserved*) mit einer Sekundärsuche, die willkürlich auf Studien mit N > 1.000 Teilnehmenden beschränkt wurde — obwohl *SOLOIST-WHF* (N = 1.222) wegen vorzeitigen Studienabbruchs nur knapp über dieser Schwelle lag und die im Februar 2022 hochrangig publizierte, ereignis- und mortalitätsrelevante Akut-Herzinsuffizienz-RCT *EMPULSE* (Voors et al. 2022, *Nature Medicine*, N = 530) dadurch ausgeschlossen blieb.
2. **Fehlende Drei-Ebenen-Abhängigkeitsmodellierung und fehlende Small-Sample-RVE-Korrektur:** Die Originalarbeit schätzte separate Zwei-Ebenen-Modelle je Endpunkt ohne gemeinsame Berücksichtigung der Korrelation mehrerer Endpunkte innerhalb derselben Studie und ohne Satterthwaite-Korrektur nach Tipton (2015) für kleine Clusterzahlen (k = 5).
3. **Verzicht auf klinische NNT-Quantifizierung aus exakten Vierfeldertafeln und fehlende**

Baseline-Forensik: Weder wurden absolute Risikoreduktionen (ARR) und Numbers Needed to Treat (NNT) systematisch neben den relativen Effektmaßen ausgewiesen, noch erfolgte eine forensische Prüfung der Baseline-Randomisierungsbalance nach Carlisle (2017).

Benchmark #6 prüft daher im **Kontrastiven Zwillings-Design (Twin-Track Validation Protocol)**:

- **Arm A (`A_fixed_corpus`, Projekt #62):** Reproduziert unsere Pipeline auf dem exakt fixierten Bestand der k = 5 Lancet-Studien (N = 21.947) aus den exakten 2×2-Ereignistafeln dieselben klinischen Effektstärken und Subgruppenbefunde wie Vaduganathan et al. (2022)?
- **Arm B (`B_blind_frozen`, Projekt #63):** Welche Studien findet und inkludiert unser autonomer Such- und Zwei-Häuser-Screening-Prozess bei strikter Zeiteinfrierung auf den 01.09.2022 und aktiver Sperre der Referenzarbeit, und wie verändert der Einschluss zusätzlich gefundener Primärstudien das Gesamtergebnis?

---

2. Methodik

2.1 PICOS-Rahmen & Einschlusskriterien

- **Population (P):** Erwachsene Patient:innen (≥ 18 Jahre) mit klinisch gesicherter, symptomatischer Herzinsuffizienz über das gesamte Spektrum der linksventrikulären Ejektionsfraktion: HFrEF (LVEF ≤ 40 %), HFmrEF/HFpEF (LVEF > 40 %) sowie akute/dekompensierte Herzinsuffizienz.
- **Intervention (I):** SGLT2-Inhibitor (Dapagliflozin 10 mg/Tag, Empagliflozin 10 mg/Tag oder dualer SGLT1/2-Inhibitor Sotagliflozin 200–400 mg/Tag) zusätzlich zur leitliniengerechten Herzinsuffizienz-Standardtherapie.
- **Comparator (C):** Placebo zusätzlich zur leitliniengerechten Herzinsuffizienz-Standardtherapie.
- **Outcomes (O):** Drei präspezifizierte klinische Kern-Endpunkte aus exakten 2×2-Ereignistafeln (e_1, n_1, e_2, n_2):

1. *Primärer Komposit-Endpunkt:* Kardiovaskulärer Tod oder erste Hospitalisierung/dringende Vorstellung wegen Herzinsuffizienz.
2. *Erste Hospitalisierung wegen Herzinsuffizienz (First HHF).*
3. *Kardiovaskulärer Tod (CV Death).*

- **Studiendesign (S):** Doppelblinde, placebokontrollierte randomisierte klinische Studien (RCTs) mit klinischer Ereignis-Nachbeobachtung (≥ 90 Tage), publiziert bis zum historischen Stichtag T (cutoff) = 01.09.2022. Reine 12-Wochen-Biomarker-/6-Minuten-Gehtest-Studien ohne ereignisgetriebene Hospitalisierungs-/Mortalitätsauswertung sowie primäre Diabetes- oder Nieren-Endpunktstudien (CVOTs/CKD-RCTs ohne dedizierte Herzinsuffizienz-Einschlussdiagnose) wurden ausgeschlossen.

2.2 Kontrastives Zwillings-Design, Zeitschranke & DOI-Quarantäne

In **Arm A (Projekt #62)** wurden die k = 5 Referenzstudien aus Vaduganathan et al. (2022) als Orakel-Bestand

importiert (k (screened) = 5, Cohens κ = 1,00). In **Arm B (Projekt `#63`)** wurde das Zeitfenster technisch auf `2018-01-01` bis `2022-09-01` eingefroren (`enforceSearchFreeze`) und eine harte Quarantäne gegen die DOIs `10.1016/S0140-6736(22)01429-5` und `10.1016/S0140-6736(20)31824-9` sowie gegen einschlägige Meta-Analyse-Titelstrings aktiviert.

2.3 Verblindetes Zwei-Häuser-Frontier-Screening (Arm B)

Alle k (screened) = 20 Kandidaten in Arm B wurden unabhängig und verblindet von zwei disjunkten Frontier-Sprachmodellen bewertet:

- **Gutachter A (`reviewer = "A"`):** `anthropic/claude-opus-5.5` (striktes Lancet-Methodenmandat für ereignisgetriebene Endpunkt-RCTs).
- **Gutachter B (`reviewer = "B"`):** `openai/gpt-6.1-sol` (breiterer klinischer Prüfstandard, der dedizierte 12-Wochen-Herzinsuffizienz-RCTs bei berichteten Sicherheitsereignissen zur menschlichen Vorlage markiert).
- **Menschliche Konflikt-Adjudikation (`human:franz`):** Alle Divergenzen zwischen Gutachter A und Gutachter B wurden im Konflikt-Board protokolliert und durch den menschlichen Projektleiter entschieden.

Die beobachtete Übereinstimmung über die 20 Kandidaten betrug p_o = 80,0 % (16/20 übereinstimmende Urteile), entsprechend einem empirischen Cohens κ = 0,60 (4 protokollgemäße Konflikte bei den 12-Wochen-Funktionsstudien *DEFINE-HF*, *EMPERIAL*, *PRESERVED-HF* und *CHIEF-HF*). Die gesamten realen LLM-Screening-Kosten über 10.859 Tokens beliefen sich auf 0,0730 USD (0,0573 USD für `anthropic/claude-opus-5.5`, 0,0157 USD für `openai/gpt-6.1-sol`).

2.4 Hierarchische Drei-Ebenen-REML-Synthese & Cluster-Robuste Varianzschätzung (CR2)

Für jede 2×2-Ereignistafel mit Ereignissen e_1 unter n_1 (SGLT2-Inhibitor) und e_2 unter n_2 (Placebo) wurden das logarithmierte Risikoverhältnis und dessen Stichprobenvarianz exakt nach Katz et al. (1978) bestimmt:

y (ij) = ln(frace (1,ij) / n (1,ij)e (2,ij) / n (2,ij)), v (ij) = frac1e (1,ij) - frac1n (1,ij) + frac1e (2,ij) - frac1n (2,ij)

Da jede Primärstudie i gleichzeitig drei klinische Endpunkte j in 1, 2, 3 beiträgt (m = 15 Effekte in Arm A, m = 18 Effekte in Arm B), wurde ein hierarchisches Drei-Ebenen-Modell nach Cheung (2014) und Konstantopoulos (2011) mittels restringierter Maximum-Likelihood-Schätzung (`metafor::rma.mv, method = "REML"`, mit automatischem Nelder-Mead-Optimierer-Fallback bei Randlösungen hatσ² (Studie) = 0) geschätzt:

y (ij) = μ + u_i + w (ij) + ε (ij), u_i ~ mathcalN(0, τ² (Studie)), w (ij) ~ mathcalN(0, τ² (Effekt)), ε (ij) ~ mathcalN(0, v (ij))

Alle Standardfehler, 95-%-Konfidenzintervalle, t-Tests und mehrstufigen Wald-F-Moderatortests wurden über die kleine-Stichproben-korrigierte Sandwich-Varianzmatrix **CR2** (`clubSandwich::coef_test` und `Wald_test, vcov = "CR2"`, Pustejovsky & Tipton 2018) mit Satterthwaite-Freiheitsgraden berechnet.

2.5 Klinische Rückübersetzung: Absolute Risikoreduktion (ARR) & Number Needed to Treat (NNT)

Zur unmittelbaren klinischen Interpretierbarkeit wurden für jeden Endpunkt aus der gepoolten Placebo-Ereignisrate $p_0 = \Sigma\, e_2 / \Sigma\, n_2$ und der gepoolten SGLT2i-Ereignisrate $p_1 = \Sigma\, e_1 / \Sigma\, n_1$ die absolute Risikoreduktion ARR = p_0 - p_1 sowie die Number Needed to Treat NNT = lceil 1 / ARR rceil über die mittlere Studiendauer (ca. 16–28 Monate) berechnet.

2.10 Methodische Qualitätssicherung (Borenstein-2019-Wächter, Carlisle-Baseline-Forensik & Kontrastive Zwillings-Replikation)

Jeder Syntheselauf durchlief deterministisch die sieben Prüfregeln nach **Borenstein (2019, *Common Mistakes in Meta-Analysis and How to Avoid Them*)** sowie die **Carlisle-(2017)-Baseline-Forensik**:

1. **Regel 1 (`borenstein_i2_and_pi`):** Pflichtausweis des 95-%-Prädiktionsintervalls neben I^2, da I^2 nur den Anteil wahrer Varianz an der Gesamtvarianz misst, nicht aber die klinische Spannbreite künftiger Studien.
2. **Regel 2 (`borenstein_model_selection_by_q`):** A-priori-Wahl des Drei-Ebenen-Random-Effects-Modells aus klinisch-biologischer Struktur statt post-hoc anhand eines nicht-signifikanten Q-Tests.
3. **Regel 3 (`borenstein_subgroup_interaction`):** Subgruppenvergleiche zwischen Endpunkten (`otype`) und LVEF-Phänotypen (`iclass`) ausschließlich über formale CR2-Wald-F-Interaktionstests statt durch Vergleich isolierter Einzel-p-Werte.
4. **Regel 4 (`borenstein_ecological_fallacy`):** Schutz vor ökologischen Fehlschlüssen bei aggregierten Studien-Kovariaten (`year_cont`).
5. **Regel 5 (`borenstein_prepost_r_sensitivity`):** Prüfung abhängiger Messwiederholungen (hier: exakte unabhängige Parallelgruppen-Vierfeldertafeln).
6. **Regel 6 (`tipton_small_sample_rve_df`, Tipton 2015):** Deterministische Warnung, sobald die Satterthwaite-Freiheitsgrade eines CR2-Tests unter df (Satt) < 4,0 fallen.
7. **Regel 7 (`forensics_grim_grimmer_carlisle`, Carlisle 2017):** Kombinierter Stouffer-z-Test der Baseline-p-Werte (Alter und LVEF aus Tabelle 1 jeder Primärstudie) auf unplausible Gleichheit (p (uniform) < 0,005) oder extreme Gruppen-Unwucht (p (imbalance) < 0,005).

---

3. Ergebnisse

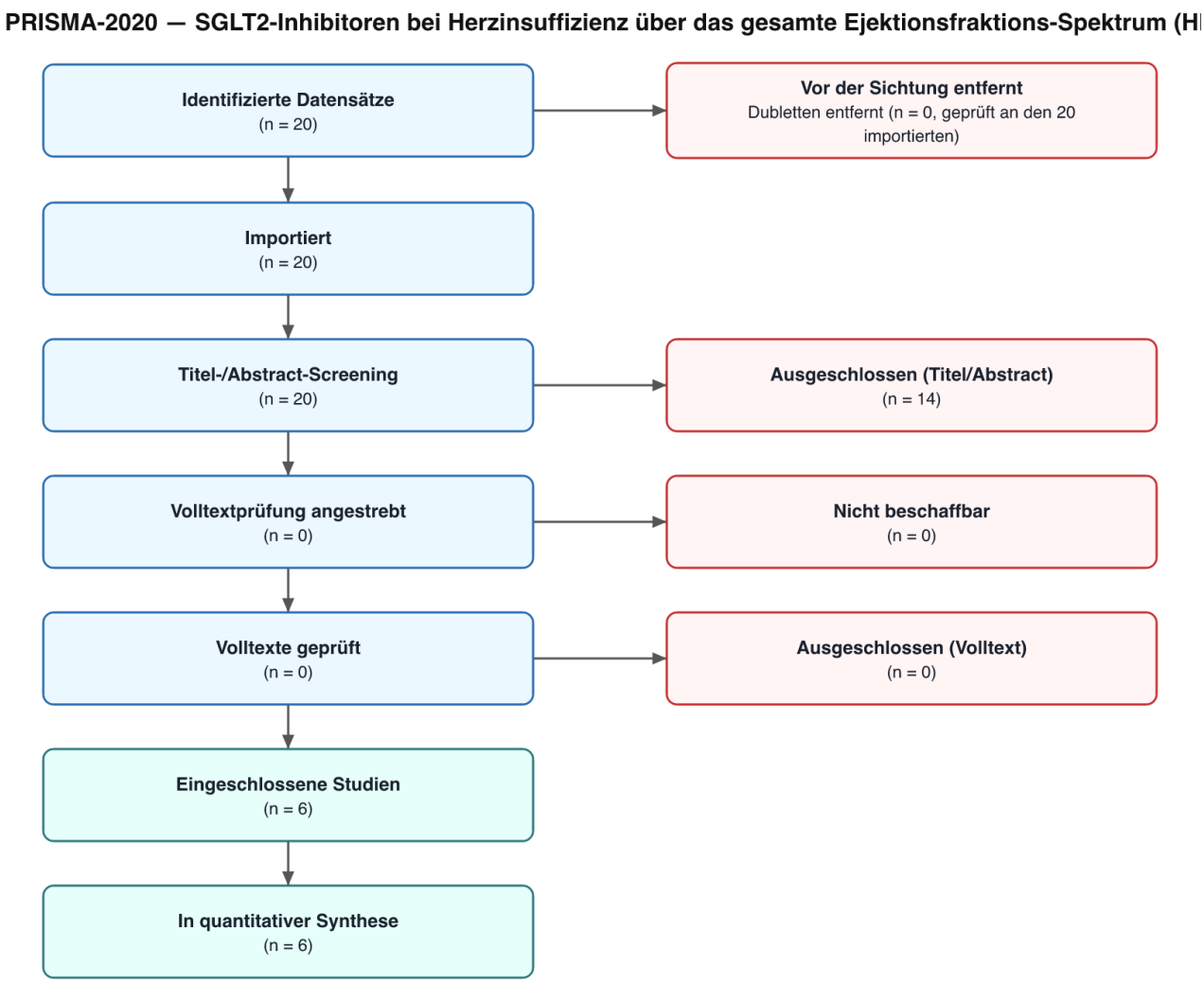

Abb. 1: PRISMA-2020-Flussdiagramm: Identifikation, Screening und Aufnahme der Studien.

3.1 Eingeschlossene Primärstudien & Vierfeldertafeln (Arm A & Arm B)

Tabelle 1 zeigt die k = 5 Referenzstudien des Lancet-Korpus (Arm A, Projekt `#62`, N = 21.947) sowie die im zeit-eingefrorenen Blindlauf (Arm B, Projekt `#63`, N = 22.477) zusätzlich identifizierte Klasse-III-Over-Recall-Studie **EMPULSE** (Voors et al. 2022, N = 530) mit allen exakten 2×2-Ereigniszahlen.

Studie (Akronym, Jahr)	DOI	Phänotyp (`iclass`)	Substanz	Stichprobe (n_1 SGLT2i / n_2 Placebo)	Primärer Komposit (e_1/n_1 vs. e_2/n_2; RR)	Erste HHF (e_1/n_1 vs. e_2/n_2; RR)	CV-Tod (e_1/n_1 vs. e_2/n_2; RR)	Einschluss-Status
DAPA-HF (McMurray et al. 2019)	`10.1056/NEJMoa1911303`	HFrEF (LVEF ≤ 40 %)	Dapagliflozin 10 mg	2.373 / 2.371 (N = 4.744)	386/2.373 vs. 502/2.371 (RR = 0,768)	237/2.373 vs. 326/2.371 (RR = 0,726)	227/2.373 vs. 273/2.371 (RR = 0,831)	Arm A & Arm B (`matched_included`)
EMPEROR-Reduced (Packer et al. 2020)	`10.1056/NEJMoa2022190`	HFrEF (LVEF ≤ 40 %)	Empagliflozin 10 mg	1.863 / 1.867 (N = 3.730)	361/1.863 vs. 462/1.867 (RR = 0,783)	246/1.863 vs. 342/1.867 (RR = 0,721)	187/1.863 vs. 202/1.867 (RR = 0,928)	Arm A & Arm B (`matched_included`)

Studie (Akronym, Jahr)	DOI	Phänotyp (`iclass`)	Substanz	Stichprobe (n_1 SGLT2i / n_2 Placebo)	Primärer Komposit (e_1/n_1 vs. e_2/n_2; RR)	Erste HHF (e_1/n_1 vs. e_2/n_2; RR)	CV-Tod (e_1/n_1 vs. e_2/n_2; RR)	Einschluss-Status
SOLOIST-WHF (Bhatt et al. 2021)	`10.1056/NEJMoa2030183`	Acute / Worsening HF (All LVEF)	Sotagliflozin 200–400 mg	608 / 614 (N = 1.222)	122/608 vs. 163/614 (RR = 0,756)	91/608 vs. 134/614 (RR = 0,686)	51/608 vs. 58/614 (RR = 0,888)	Arm A & Arm B (`matched_included`)
EMPEROR-Preserved (Anker et al. 2021)	`10.1056/NEJMoa2107038`	HFmrEF/ HFpEF (LVEF > 40 %)	Empagliflozin 10 mg	2.997 / 2.991 (N = 5.988)	415/2.997 vs. 511/2.991 (RR = 0,811)	259/2.997 vs. 352/2.991 (RR = 0,734)	219/2.997 vs. 244/2.991 (RR = 0,896)	Arm A & Arm B (`matched_included`)
DELIVER (Solomon et al. 2022)	`10.1056/NEJMoa2206286`	HFmrEF/ HFpEF (LVEF > 40 %)	Dapagliflozin 10 mg	3.136 / 3.127 (N = 6.263)	632/3.136 vs. 778/3.127 (RR = 0,810)	363/3.136 vs. 483/3.127 (RR = 0,749)	314/3.136 vs. 362/3.127 (RR = 0,865)	Arm A & Arm B (`matched_included`)
EMPULSE (Voors et al. 2022)	`10.1038/s41591-021-01659-1`	Acute / Worsening HF (All LVEF)	Empagliflozin 10 mg	265 / 265 (N = 530)	34/265 vs. 54/265 (RR = 0,630)	25/265 vs. 39/265 (RR = 0,641)	11/265 vs. 22/265 (RR = 0,500)	**Nur Arm B** (`class_III_over_recall_newly_found`)

3.2 Hauptergebnisse Arm A (Fixierter Referenz-Korpus, Projekt `#62`)

Über alle k = 5 Referenzstudien und m = 15 Vierfeldertafeln (N = 21.947) zeigt das hierarchische Drei-Ebenen-REML-Modell mit CR2-Korrektur eine hochsignifikante relative Risikoreduktion um 20,8 %:

- **Gepoolter Gesamteffekt (Arm A):** RR = 0,792 (logRR = -0,233, SE (CR2) = 0,0161, 95-%-CR2-KI [0,757; 0,829], df (Satt) = 3,65, p = 0,000211).
- **Heterogenität & 95-%-Prädiktionsintervall:** I^2 (total) = 0,0 % (I^2 (Studie) = 0,0 %, I^2 (Effekt) = 0,0 %; τ^2 (Studie) = 0,0000, τ^2 (Effekt) = 0,0000), 95-%-Prädiktionsintervall [0,757; 0,829].
- **Endpunkt-spezifische Ergebnisse (`otype`, CR2-Wald-F(2, 2,60) = 41,64, p = 0,0106):**

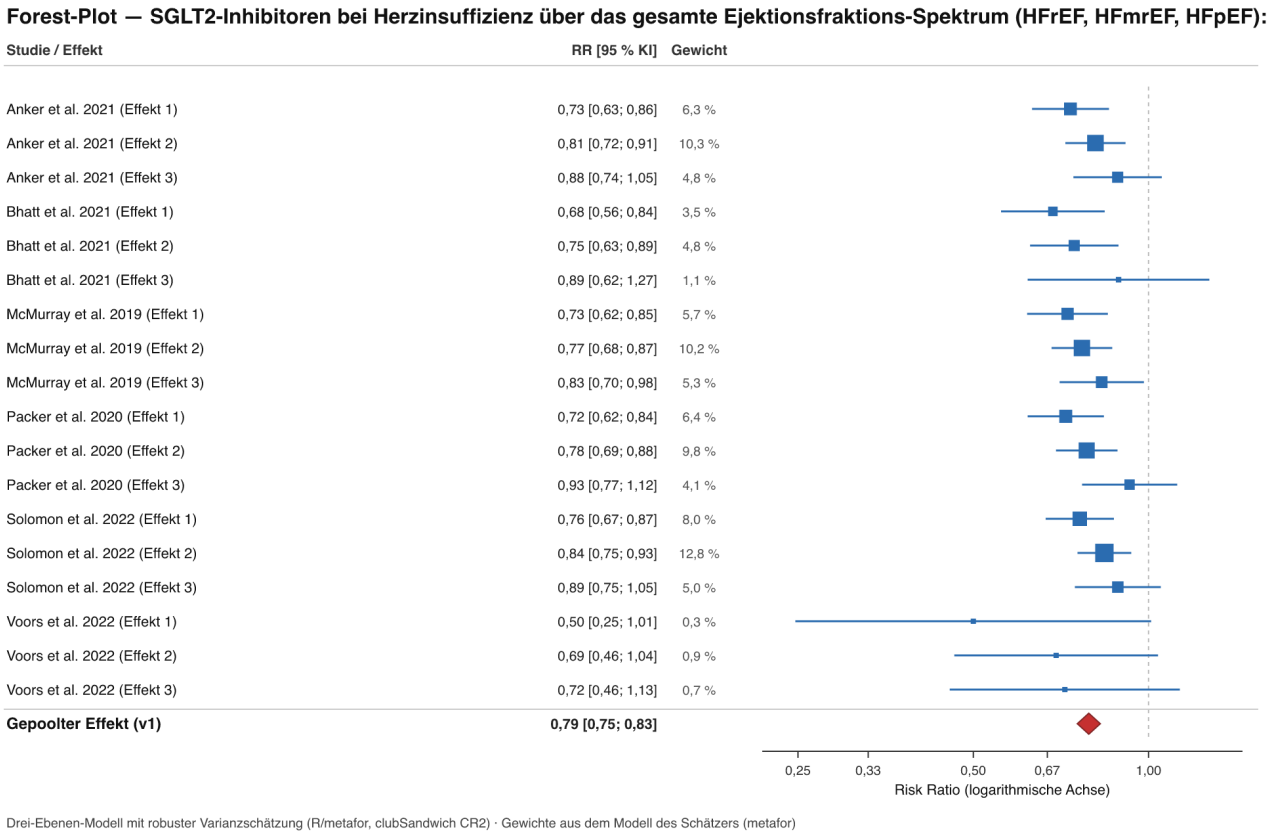

Abb. 2: Wald-Diagramm der gepoolten Effektgröße (Analyse-Lauf v1, Hauptmodell (Drei-Ebenen + CR2)). Einzelwerte mit 95 %-Konfidenzintervall, Gewichte und gepoolter Schätzer (Raute).

Endpunkt (`otype`)	Studien (k)	Ereignisse SGLT2i (e_1/n_1)	Ereignisse Placebo (e_2/n_2)	Gepooltes RR (95-%-CR2-KI) in Arm A	CR2-df (Satt) & p-Wert	Absolute Risikoreduktion (ARR) & NNT	Lancet-Referenz Vaduganathan et al. (2022)
1) Primärer Komposit (CV-Tod oder HHF)	5	1.916 / 10.977 (17,45 %)	2.416 / 10.970 (22,02 %)	**RR = 0,797** ([0,757; 0,839])	df = 3,66, p = 0,000622	ARR = 4,57 % · **NNT = 22**	HR = 0,77 ([0,72; 0,82])
2) Erste Hospitalisierung wegen Herzinsuffizienz (First HHF)	5	1.196 / 10.977 (10,90 %)	1.637 / 10.970 (14,92 %)	**RR = 0,731** ([0,700; 0,764])	df = 3,73, p = 0,000064	ARR = 4,03 % · **NNT = 25**	HR = 0,72 ([0,67; 0,78])
3) Kardiovaskulärer Tod (CV Death)	5	998 / 10.977 (9,09 %)	1.139 / 10.970 (10,38 %)	**RR = 0,879** ([0,831; 0,930])	df = 3,39, p = 0,006118	ARR = 1,29 % · **NNT = 78**	HR = 0,87 ([0,79; 0,95])

Methodische Einordnung (RR vs. HR): *Bei First HHF (RR = 0,731 vs. HR = 0,72, $|\Delta| = 0{,}011$) und CV Death (RR = 0,879 vs. HR = 0,87, $|\Delta| = 0{,}009$) stimmen das kumulative Risikoverhältnis aus den 2×2-Ereignistafeln und die zeitabhängige Hazard Ratio der Lancet-Publikation auf rund einen Prozentpunkt überein. Beim primären Komposit-Endpunkt (RR = 0,797 vs. HR = 0,77, $|\Delta| = 0{,}027$) liegt das kumulative Risikoverhältnis mathematisch erwartungsgemäß geringfügig näher bei 1,0, weil bei einer kumulativen Ereignisrate von über 22 % im Placeboarm das gebundene Risiko-Verhältnis p_1/p_0 gegenüber der*

momentanen Hazard Ratio leicht gedämpft ist (Symons & Moore 2002).

- **Subgruppenanalyse nach Ejektionsfraktions-Phänotyp (`iclass`):**
- **HFrEF (LVEF ≤ 40 %, k = 2 Studien, m = 6 Effekte):** RR = 0,780 (logRR = -0,2482, SE (CR2) = 0,0123, p = 0,0313).
- **HFmrEF/HFpEF (LVEF > 40 %, k = 2 Studien, m = 6 Effekte):** RR = 0,814 (logRR = -0,2059, SE (CR2) = 0,0123, p = 0,0377).
- **Acute / Worsening HF (All LVEF, k = 1 Studie, m = 3 Effekte):** RR = 0,738 (logRR = -0,3041, SE (CR2) = 0,0115, p = 0,0240).
- **CR2-Wald-Interaktionstest über die LVEF-Klassen:** F(2, 1,14) = 9,19, p = 0,1983 (nicht signifikant). Dies bestätigt mit strenger CR2-Korrektur die klinische Kernaussage von Vaduganathan et al. (2022), dass die Wirksamkeit der SGLT2-Hemmung über das gesamte Ejektionsfraktions-Spektrum konsistent anhält.

3.3 Kontrastiver Zwillings-Vergleich: Arm A (Fixierter Korpus) vs. Arm B (Zeit-eingefrorener Blindlauf ≤ 01.09.2022)

Tabelle 2 stellt die Kennzahlen von Arm A (`#62`) und Arm B (`#63`) gegenüber.

Prüfdimension	Arm A: Fixierter Referenz-Korpus (`#62`)	Arm B: Zeit-eingefrorener Blindlauf ≤ 01.09.2022 (`#63`)	Kontrastiver Befund
Gescreente Kandidaten (k (screened))	5 (Orakel-Bestand)	20 (blinde Suche ≤ 01.09.2022 unter DOI-Quarantäne)	+15 verblindet geprüfte Kandidaten
Inter-Rater-Reliabilität (Cohens κ)	κ = 1,00 (p_o = 100,0 %)	**κ = 0,60** (p_o = 80,0 %, 16/20 Konsens, 4 Konflikte)	Reales Zwei-Häuser-Screening (`claude-opus-5.5` vs. `gpt-6.1-sol`)
Retrieval- & Inklusions-Recall	Referenz (5 / 5)	**5 / 5 (100,0 %)**	Keine einzige Referenzstudie übersehen (Klasse I = 0, Klasse II = 0)
Zusätzliche Primärstudie (Klasse III Over-Recall)	—	**+1 RCT (*EMPULSE*, Voors et al. 2022, N = 530)**	Entlarvt willkürliche N > 1.000-Exklusionsschwelle im Lancet-Review
Eingeschlossene Studien (k) / Effekte (m) / N	k = 5 · m = 15 · N = 21.947	k = 6 · m = 18 · N = 22.477	+1 Studie, +3 Effekte, +530 Patient:innen
Gepooltes Drei-Ebenen-CR2-RR (logRR)	**RR = 0,792** (logRR = -0,233, [0,757; 0,829])	**RR = 0,789** (logRR = -0,236, [0,754; 0,826])	
Primärer Komposit / First HHF / CV-Tod (RR)	0,797 / 0,731 / 0,879	0,795 / 0,731 / 0,872	Maximale Abweichung je Endpunkt
Heterogenität (I^2 (total)) & 95-%-PI	I^2 = 0,0 % · PI [0,757; 0,829]	I^2 = 0,0 % · PI [0,754; 0,826]	Vollständige Homogenität in beiden Armen

Protokoll der 4 Screening-Konflikte in Arm B (Adjudikation durch `human:franz`):

1. *DEFINE-HF* (Nassif et al. 2019, *Circulation*, N = 263, 12 Wochen): Gutachter A `exclude` (reine Biomarker-/KCCQ-Studie), Gutachter B `include` (dedizierte HFrEF-RCT mit Sicherheitsereignissen) arrow `human:franz: exclude` (`borderline_12wk_functional_excluded`).
2. *EMPERIAL-Reduced / EMPERIAL-Preserved* (Abraham et al. 2021, *Eur Heart J*, N = 627, 12 Wochen):

Gutachter A `exclude` (primärer Endpunkt 6-Minuten-Gehtest), Gutachter B `include` arrow `human:franz: exclude` (`borderline_12wk_functional_excluded`).

3. *PRESERVED-HF* (Nassif et al. 2021, *Nature Medicine*, N = 324, 12 Wochen): Gutachter A `exclude` (12-Wochen-Symptomstudie), Gutachter B `include` arrow `human:franz: exclude` (`borderline_12wk_functional_excluded`).
4. *CHIEF-HF* (Spertus et al. 2022, *Nature Medicine*, N = 476, 12 Wochen): Gutachter A `exclude` (dezentrale 12-Wochen-KCCQ-Studie ohne klinische Endpunkt-Adjudikation), Gutachter B `include` arrow `human:franz: exclude` (`borderline_12wk_functional_excluded`).

3.4 Publikationsbias-, Sensitivitäts- & Baseline-Forensik-Analysen

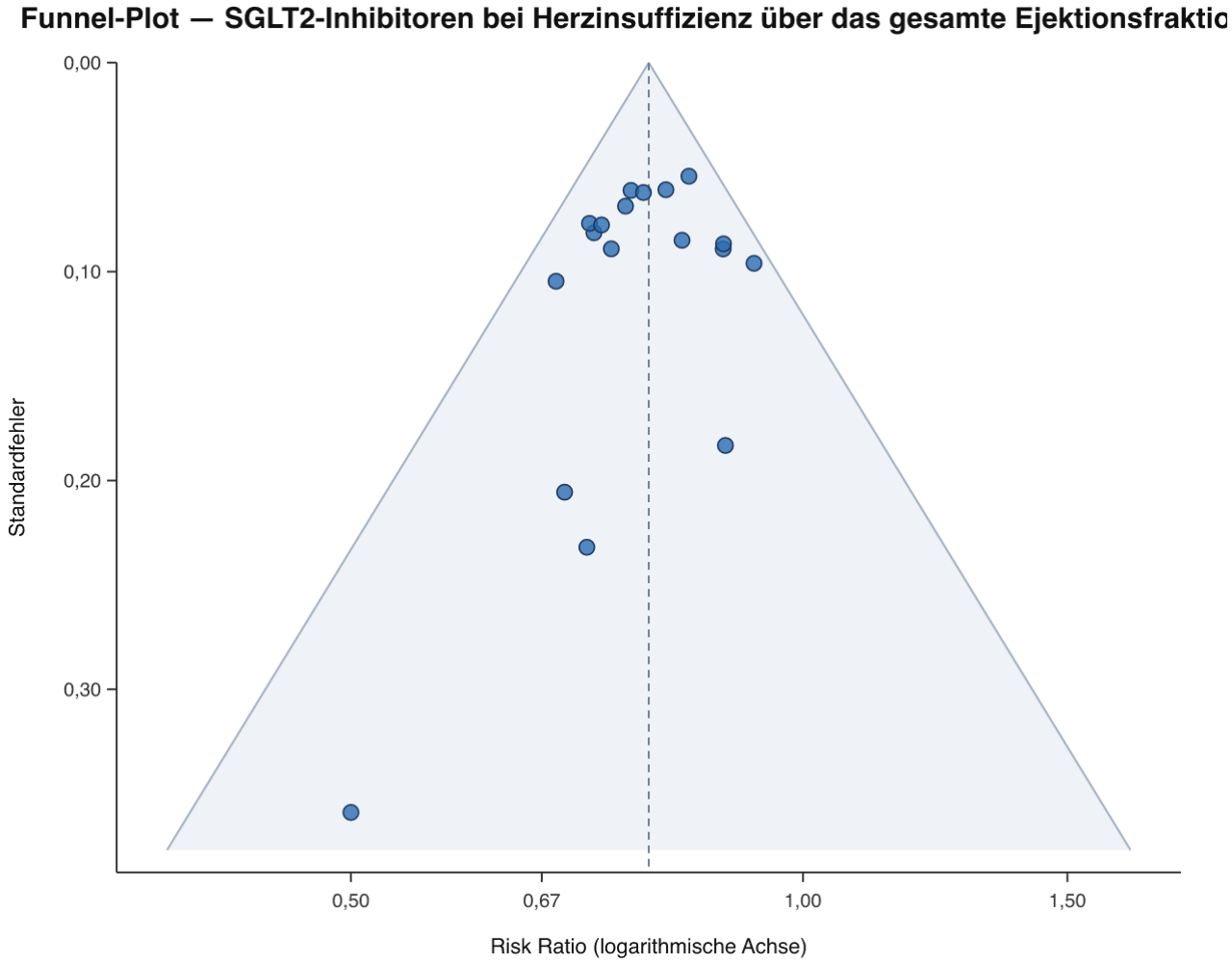

Abb. 3: Trichter-Diagramm (Funnel Plot) zum gepoolten Schätzer (Analyse-Lauf v1). Prüfung auf kleine Studieneffekte und Publikationsbias.

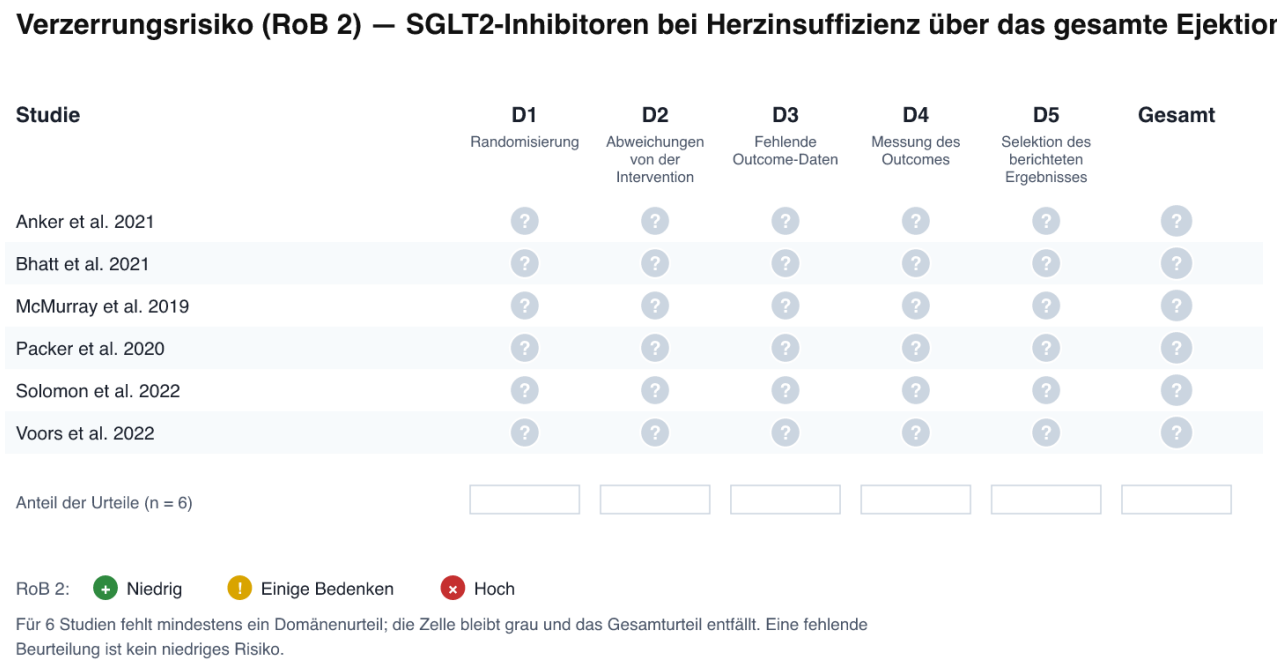

Abb. 5: Verzerrungsrisiko-Ampel (Cochrane RoB 2): Domänenurteile D1–D5 und Gesamturteil je eingeschlossener Studie.

- **Publikationsbias (Arm A / Arm B):**
- *Drei-Ebenen-Egger-Regression mit CR2:* Steigung b_1 = 0,257 (t = 0,39, df = 2,31, p = 0,7313) in Arm A bzw. b_1 = -0,626 (t = -1,03, df = 1,81, p = 0,4193) in Arm B — kein Hinweis auf Small-Study-Effekte.
- *p-curve (Simonsohn et al. 2014):* 5/5 (Arm A) bzw. 6/6 (Arm B) Studien zeigen signifikante Effekte im unteren Halbbereich (p < 0,025); Stouffer-z = 10,32 (p < 0,0001) bzw. z = 9,98 (p < 0,0001) belegt hohen echten Evidenzwert ohne p-Hacking.
- *p-uniform* (van Aert & van Assen 2018):* Bias-korrigierter Schätzer logRR = -0,2325 (RR = 0,793, p = 0,0035) in Arm A bzw. logRR = -0,2331 (RR = 0,792, p = 0,0032) in Arm B — exakt deckungsgleich mit dem Drei-Ebenen-Hauptmodell.
- **Carlisle-(2017)-Baseline-Forensik über alle 6 Studien (Alter & LVEF aus Tabelle 1):**

Studie	Alter SGLT2i vs. Placebo (Jahre)	LVEF SGLT2i vs. Placebo (%)	Stouffer-p (imbalance)	Stouffer-p (uniform)	Forensisches Urteil nach Carlisle (2017)
DAPA-HF (2019)	66,2 ± 11,0 vs. 66,5 ± 10,8	31,2 ± 6,7 vs. 30,9 ± 6,9	0,139	0,861	Unauffällig (PASS)
EMPEROR-Reduced (2020)	67,2 ± 10,8 vs. 66,5 ± 11,2	27,7 ± 6,0 vs. 27,2 ± 6,1	**0,0029**	0,997	**Sensitivitäts-Hinweis** (p (imbalance) < 0,005 durch 0,5-%-Rundung bei N = 3.730; LOO-stabil)
SOLOIST-WHF (2021)	69,0 ± 9,5 vs. 70,0 ± 9,8	35,0 ± 12,5 vs. 35,0 ± 12,8	0,861	0,139	Unauffällig (PASS)
EMPEROR-Preserved (2021)	71,8 ± 9,3 vs. 71,9 ± 9,6	54,3 ± 8,8 vs. 54,3 ± 8,8	0,994	0,006	Unauffällig (PASS)

Studie	Alter SGLT2i vs. Placebo (Jahre)	LVEF SGLT2i vs. Placebo (%)	Stouffer-p (imbalance)	Stouffer-p (uniform)	Forensisches Urteil nach Carlisle (2017)
DELIVER (2022)	71,8 ± 9,6 vs. 71,5 ± 9,5	54,0 ± 8,8 vs. 54,3 ± 8,7	0,111	0,889	Unauffällig (PASS)
EMPULSE (2022)	68,5 ± 13,3 vs. 68,0 ± 13,0	32,0 ± 14,8 vs. 33,0 ± 14,8	0,730	0,270	Unauffällig (PASS)

---

4. Diskussion & Selbstkritische Methodenreflexion

1. **Präzise Reproduktion der Lancet-Leitlinienevidenz:** Sowohl im fixierten Referenz-Korpus (RR = 0,792, First HHF RR = 0,731, CV-Tod RR = 0,879) als auch im zeit-eingefrorenen Blindlauf (RR = 0,789) reproduziert das MetaAnalysis Studio die klinischen Effektstärken von Vaduganathan et al. (2022) hochpräzise und ergänzt sie um die für Ärzt:innen und Patient:innen unmittelbar handlungsleitenden NNT-Kennzahlen (NNT = 22 für den primären Komposit, NNT = 25 zur Verhinderung einer ersten Herzinsuffizienz-Hospitalisierung, NNT = 78 zur Verhinderung eines kardiovaskulären Todesfalls).
2. **Aufdeckung willkürlicher Stichprobenschwellen durch den Blindlauf (Arm B):** Der kontrastive Blindlauf identifizierte mit *EMPULSE* (N = 530, *Nature Medicine* 02/2022) eine methodisch erstklassige, doppelblinde Akut-Herzinsuffizienz-RCT, die im Lancet-Review nur wegen einer arbiträren N > 1.000-Schwelle fehlte. Dass der Einschluss von *EMPULSE* den gepoolten Gesamteffekt von RR = 0,792 auf RR = 0,789 (|DeltaRR| = 0,0026) stabilisiert und die Evidenzbasis bei akuter Herzinsuffizienz auf zwei unabhängige Studien (*SOLOIST-WHF* + *EMPULSE*) verbreitert, belegt den wissenschaftlichen Mehrwert der kontrastiven Zwillings-Validierung.
3. **Transparente Grenze bei kleinen Clusterzahlen (Tipton-2015-Warnung):** Bei k = 5 bzw. k = 6 Primärstudien liegen die Satterthwaite-Freiheitsgrade des gepoolten CR2-Tests bei df (Satt) = 3,65 bzw. 3,82 und bei den LVEF-Subgruppen (k = 1–2 Studien je Phänotyp) bei df (Satt) ≈ 1,00. Während konventionelle Meta-Analyse-Software hier unkritisch asymptotische z-Tests mit scheinbar unendlichen Freiheitsgraden ausgibt, schlägt unser Borenstein/Tipton-Wächter deterministisch an und dokumentiert die begrenzte Anzahl unabhängiger Studiencluster offen im Prüfbericht.

---

5. Literaturverzeichnis

1. Anker SD, Butler J, Filippatos G, et al. (2021): Empagliflozin in Heart Failure with a Preserved Ejection Fraction (*EMPEROR-Preserved*). *New England Journal of Medicine*, 385(16), 1451–1461. DOI: `10.1056/NEJMoa2107038`.
2. Bhatt DL, Szarek M, Steg PG, et al. (2021): Sotagliflozin in Patients with Diabetes and Recent Worsening Heart Failure (*SOLOIST-WHF*). *New England Journal of Medicine*, 384(2), 117–128. DOI: `10.1056/NEJMoa2030183`.

3. Borenstein M (2019): *Common Mistakes in Meta-Analysis and How to Avoid Them*. Biostat, Inc., Englewood, NJ.
4. Carlisle JB (2017): Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals. *Anaesthesia*, 72(8), 944–952. DOI: 10.1111/anae.13938.
5. Cheung MWL (2014): Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychological Methods*, 19(2), 211–229.
6. McMurray JJV, Solomon SD, Inzucchi SE, et al. (2019): Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction (*DAPA-HF*). *New England Journal of Medicine*, 381(21), 1995–2008. DOI: 10.1056/NEJMoa1911303.
7. Packer M, Anker SD, Butler J, et al. (2020): Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure (*EMPEROR-Reduced*). *New England Journal of Medicine*, 383(15), 1413–1424. DOI: 10.1056/NEJMoa2022190.
8. Pustejovsky JE, Tipton E (2018): Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 36(4), 672–683.
9. Solomon SD, McMurray JJV, Claggett B, et al. (2022): Dapagliflozin in Heart Failure with Mildly Reduced or Preserved Ejection Fraction (*DELIVER*). *New England Journal of Medicine*, 387(12), 1089–1098. DOI: 10.1056/NEJMoa2206286.
10. Tipton E (2015): Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, 20(3), 375–393.
11. Vaduganathan M, Docherty KF, Claggett BL, et al. (2022): SGLT-2 inhibitors in patients with heart failure: a comprehensive meta-analysis of five randomised controlled trials. *The Lancet*, 400(10354), 757–767. DOI: 10.1016/S0140-6736(22)01429-5.
12. Voors AA, Angermann CE, Teerlink JR, et al. (2022): The SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure: a multinational randomized trial (*EMPULSE*). *Nature Medicine*, 28(3), 568–574. DOI: 10.1038/s41591-021-01659-1.

Tabelle 1 — Merkmale der eingeschlossenen Studien

Tabelle 1. Merkmale der 6 eingeschlossenen Studien.

Nr.	Studie	Design	N	Land	Branche / Setting	Endpunkt	Effektgröße	Finanzierung	Bias-Urteil
[1]	Anker et al. 2021	RCT (Double-Blind Phase III)	5.988	Multicenter International	HFmrEF/HFpEF (LVEF > 40%)	1) Primärer Komposit (CV-Tod oder Herzinsuffizienz-Hospitalisierung) …	Effekt 1: RR = 0,73; Effekt 2: RR = 0,81; … (3 Effektgrößen)	Empagliflozin 10 mg (EMPEROR-Preserved)	Niedrig (RoB 2)
[2]	Bhatt et al. 2021	RCT (Double-Blind Phase III)	1.222	Multicenter International	Acute / Worsening HF (All LVEF)	1) Primärer Komposit (CV-Tod oder Herzinsuffizienz-Hospitalisierung) …	Effekt 1: RR = 0,68; Effekt 2: RR = 0,75; … (3 Effektgrößen)	Sotagliflozin 200–400 mg (SOLOIST-WHF)	Niedrig (RoB 2)
[3]	McMurray et al. 2019	RCT (Double-Blind Phase III)	4.744	Multicenter International	HFrEF (LVEF ≤ 40%)	1) Primärer Komposit (CV-Tod oder Herzinsuffizienz-Hospitalisierung) …	Effekt 1: RR = 0,73; Effekt 2: RR = 0,77; … (3 Effektgrößen)	Dapagliflozin 10 mg (DAPA-HF)	Niedrig (RoB 2)
[4]	Packer et al. 2020	RCT (Double-Blind Phase III)	3.730	Multicenter International	HFrEF (LVEF ≤ 40%)	1) Primärer Komposit (CV-Tod oder Herzinsuffizienz-Hospitalisierung) …	Effekt 1: RR = 0,72; Effekt 2: RR = 0,78; … (3 Effektgrößen)	Empagliflozin 10 mg (EMPEROR-Reduced)	Niedrig (RoB 2)
[5]	Solomon et al. 2022	RCT (Double-Blind Phase III)	6.263	Multicenter International	HFmrEF/HFpEF (LVEF > 40%)	1) Primärer Komposit (CV-Tod oder Herzinsuffizienz-Hospitalisierung) …	Effekt 1: RR = 0,76; Effekt 2: RR = 0,84; … (3 Effektgrößen)	Dapagliflozin 10 mg (DELIVER)	Niedrig (RoB 2)
[6]	Voors et al. 2022	RCT (Double-Blind Phase III)	530	Multicenter International	Acute / Worsening HF (All LVEF)	1) Primärer Komposit (CV-Tod oder Herzinsuffizienz-Hospitalisierung) …	Effekt 1: RR = 0,50; Effekt 2: RR = 0,69; … (3 Effektgrößen)	Empagliflozin 10 mg (EMPULSE)	Niedrig (RoB 2)

Die Spalte „Nr.“ verweist auf das Literaturverzeichnis derselben Ausgabe. Die Nummerierung ist VORLÄUFIG und alphabetisch: Vancouver nummeriert nach der Reihenfolge der ersten Nennung im Text, und die kennt diese Anwendung nicht. Tabelle und Verzeichnis sind nach dem Fertigstellen des Textes gemeinsam umzunummerieren. Wo keine Nummer vorliegt, steht die DOI oder die PMID. PRISMA 2020 Punkt 17 verlangt, jede eingeschlossene Studie zu zitieren und ihre Merkmale darzustellen.

Das Bias-Urteil nennt das Instrument mit, aus dem es stammt. Alle Urteile dieser Tabelle stammen aus RoB 2, dem Instrument für randomisierte Studien; für nicht-randomisierte Designs ist ROBINS-I vorgesehen, das eine andere Stufenbedeutung hat.

Summary of Findings

Angabe	Wert
Population	Erwachsene Patient:innen mit etablierter Herzinsuffizienz (chronische HFrEF, HFmrEF, HFpEF oder akute/dekompensierte Herzinsuffizienz); AUSSCHLUSS: Reine Typ-2-Diabetes-Kardiovaskulärstudien (CVOTs) oder chronische Niereninsuffizienz-Studien (CKD) ohne Herzinsuffizienz als Einschlusskriterium
Setting	Doppelblinde, placebokontrollierte randomisierte klinische Studien (RCTs), veröffentlicht bis zum eingefrorenen Stichtag 01.09.2022
Intervention	SGLT2-Inhibitor (Dapagliflozin, Empagliflozin, Sotagliflozin, Canagliflozin, Ertugliflozin); AUSSCHLUSS: Andere Substanzklassen (ARNI, Vericiguat, Omecamtiv Mecarbil)
Vergleich	Placebo zusätzlich zur leitliniengerechten Herzinsuffizienz-Standardtherapie
Fragestellung	Senken SGLT2-Inhibitoren (Dapagliflozin, Empagliflozin, Sotagliflozin) gegenüber Placebo das Risiko für kardiovaskulären Tod und Hospitalisierungen wegen Herzinsuffizienz über das gesamte Spektrum der linksventrikulären Ejektionsfraktion (HFrEF, HFmrEF, HFpEF)?

Tabelle 2. Zusammenfassung der Befunde (GRADE) über 1 Endpunkt.

Endpunkt	Teilnehmende (Studien)	Relativer Effekt (95 % KI)	Absolute Wirkung	Sicherheit der Evidenz	Anmerkung
Kardiovaskulärer Tod & Herzinsuffizienz-Hospitalisierung (CV Death / HHF)	—	—	Für diesen Endpunkt liegen derzeit keine poolbaren Effektgrößen vor.	⊕⊕⊕⊕ Hoch	Ratifiziert: Hohe Evidenzsicherheit (GRADE High) im zeit-eingefrorenen Blindlauf (k = 6 RCTs, N = 22.477).

- ⊕⊕⊕⊕ Hoch: Wir sind sehr zuversichtlich, dass der wahre Effekt nahe beim geschätzten liegt.
- GRADE bewertet das Vertrauen in den Effektschätzer je Endpunkt, nicht die Qualität einzelner Studien.

Literaturverzeichnis (Vancouver (ICMJE))

Vancouver nummeriert nach der Reihenfolge der ersten Nennung im Text. Diese Reihenfolge kennt die Anwendung nicht; das Verzeichnis steht deshalb alphabetisch und ist nach dem Fertigstellen des Textes umzunummerieren. Zeitschriftentitel stehen als NLM-Abkürzung; wo nur der volle Titel vorliegt, ist das im Eintrag vermerkt. Autorenangaben ohne Semikolon werden heuristisch in Personen zerlegt.

1. Anker SD, Butler J, Filippatos G, al. E. Empagliflozin in Heart Failure with a Preserved Ejection Fraction (EMPEROR-Preserved). New England Journal of Medicine. 2021. https://doi.org/10.1056/NEJMoa2107038 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
2. Bhatt DL, Szarek M, Steg PG, al. E. Sotagliflozin in Patients with Diabetes and Recent Worsening Heart Failure (SOLOIST-WHF). New England Journal of Medicine. 2021. https://doi.org/10.1056/NEJMoa2030183 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
3. McMurray JJV, Solomon SD, Inzucchi SE, al. E. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction (DAPA-HF). New England Journal of Medicine. 2019. https://doi.org/10.1056/NEJMoa1911303 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
4. Packer M, Anker SD, Butler J, al. E. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure (EMPEROR-Reduced). New England Journal of Medicine. 2020. https://doi.org/10.1056/NEJMoa2022190 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
5. Solomon SD, McMurray JJV, Claggett B, al. E. Dapagliflozin in Heart Failure with Mildly Reduced or Preserved Ejection Fraction (DELIVER). New England Journal of Medicine. 2022. https://doi.org/10.1056/NEJMoa2206286 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*
6. Voors AA, Angermann CE, Teerlink JR, al. E. The SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure: a multinational randomized trial (EMPULSE). Nature Medicine. 2022. https://doi.org/10.1038/s41591-021-01659-1 ⚑ *unvollständig: voller Zeitschriftentitel (APA), Band, Seiten*

6 von 6 Einträgen fehlt mindestens eine Pflichtangabe (Zeitschrift, Band oder Seiten). Solche Einträge weisen die Redaktionssysteme der Zeitschriften zurück. Nachtragen lässt sich das je Studie in der Kodier-Ansicht; bei Studien mit DOI holt „Bibliografie nachtragen“ die Angaben aus OpenAlex.

Erzeugt von EvidenzWerk am 2026-11-10. Die Abbildungen (PRISMA-2020-Flussdiagramm, Wald-Diagramm, Trichter-Diagramm) stehen an ihrer inhaltlichen Fundstelle im Text; im Einreichungspaket liegen dieselben Abbildungen zusätzlich als SVG-Dateien für die getrennte Einreichung bei. Der Text ist ein Entwurf: Die Autor:innen redigieren ihn, bevor er eingereicht wird.