ESC HEART FAILURE
ESC Heart Failure 2022; **9**: 885–893
Published online 13 January 2022 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/ehf2.13785

ORIGINAL ARTICLE

Robustness of outcomes in trials evaluating sodium–glucose co-transporter 2 inhibitors for heart failure

Muhammad Shariq Usman[1†], Muhammad Shahzeb Khan[2†], Gregg C. Fonarow[3], Stephen J. Greene[2], Tim Friede[4,5], Muthiah Vaduganathan[6], Gerasimos Filippatos[7], Andrew J. Stewart Coats[8,9], Stefan D. Anker[10] and Javed Butler[1*]

[1]*Department of Medicine, University of Mississippi, Jackson, MS, USA;* [2]*Division of Cardiology, Duke University Medical Center, Durham, NC, USA;* [3]*Division of Cardiology, Ronald Reagan UCLA Medical Center, Los Angeles, CA, USA;* [4]*Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany;* [5]*DZHK (German Centre for Cardiovascular Research), partner site Göttingen, Göttingen, Germany;* [6]*Brigham and Women's Hospital Heart & Vascular Center, Boston, MA, USA;* [7]*National and Kapodistrian University of Athens School of Medicine, Athens University Hospital Attikon, Athens, Greece;* [8]*Department of Cardiology, IRCCS San Raffaele Pisana, Rome, Italy;* [9]*University of Warwick, Coventry, UK; and* [10]*Department of Cardiology (CVK); and Berlin Institute of Health Center for Regenerative Therapies (BCRT); DZHK (German Centre for Cardiovascular Research), partner site Berlin, Charité—Universitätsmedizin Berlin, Berlin, Germany*

Abstract

Aims Recent trials have evaluated sodium–glucose co-transporter 2 inhibitors in patients with heart failure (HF). We sought to assess the robustness of findings from these trials using the fragility index (FI).

Methods and results Fragility index is defined as the minimum number of patients that must be moved from the 'non-event' to the 'event' group to turn a statistically significant result to non-significant. In addition to FI, fragility quotient [(FQ); FI divided by the sample size] was calculated to assess the proportion of events that must be moved to change the significance. For statistically non-significant outcomes, reverse fragility index (RFI) and reverse fragility quotient (RFQ) were calculated. Robustness of findings after pooling data from all three trials was also assessed. A robust reduction in first HF hospitalization or cardiovascular mortality was seen with dapagliflozin (FI = 62 and FQ = 0.013), empagliflozin (FI = 50 and FQ = 0.013), and sotagliflozin (FI = 60 and FQ = 0.049). Dapagliflozin nominally improved all-cause and cardiovascular mortality, with modest FI (*n* = 8 and 5) and FQ (0.002 and 0.001). Empagliflozin and sotagliflozin did not demonstrate statistically significant reductions in all-cause mortality, with modest RFI (empagliflozin: RFI = 26 and RFQ = 0.007; sotagliflozin: RFI = 6 and RFQ = 0.005). A similar trend was seen with cardiovascular mortality (empagliflozin: RFI = 24 and RFQ = 0.006; sotagliflozin: RFI = 7 and RFQ = 0.006). Upon meta-analysis, the result for first HF hospitalization or cardiovascular mortality was robust (FI = 95 and FQ = 0.010). The reductions in all-cause (FI = 12 and FQ = 0.001) and cardiovascular mortality (FI = 9 and FQ = 0.001), while statistically significant, were fragile.

Conclusion Improvement in the composite outcome of first HF hospitalization or cardiovascular death was highly concordant and robust across sodium–glucose co-transporter 2 inhibitor trials. In contrast, secondary endpoints of all-cause and cardiovascular mortality were statistically fragile, underscoring the need to power trials for mortality to fully understand the benefit of therapies on fatal events.

Keywords Fragility index; Robustness; Sodium–glucose co-transporter 2 inhibitors; Cardiac failure

Received: 28 May 2021; Revised: 25 October 2021; Accepted: 13 December 2021
**Correspondence to: Javed Butler, Department of Medicine, University of Mississippi Medical Center, 2500 N State Street, Jackson, MS 39216, USA. Tel: 601 984-5600; Fax: 601 984-5608. Email: jbutler4@umc.edu*
† Muhammad Shariq Usman and Muhammad Shahzeb Khan are co-primary authors.

Introduction

Recent trials show that sodium–glucose co-transporter 2 (SGLT2) inhibitors improve heart failure (HF) outcomes.[1–3] The 'Dapagliflozin and Prevention of Adverse-Outcomes in HF' (DAPA-HF)[1] showed that dapagliflozin reduced the composite endpoint of first HF hospitalization, urgent HF visit, or cardiovascular mortality among patients with HF with reduced ejection fraction (HFrEF). The 'Empagliflozin Outcome Trial in Patients With Chronic HFrEF' (EMPEROR-Reduced) enrolled higher risk HFrEF patients and demonstrated a similar benefit.[2] Recently, the 'Effect of Sotagliflozin on Cardiovascular Events in Patients With Type 2 Diabetes Post Worsening HF' (SOLOIST-WHF) trial evaluated the effects of sotagliflozin in HF patients with diabetes and recent hospitalization for worsening HF[3] and included patients with HFrEF as well as HF with preserved ejection fraction (HFpEF). While this trial was terminated early, SOLOIST-WHF also showed a reduction in first HF hospitalization or cardiovascular mortality.

While the magnitude of benefit for the primary endpoint was consistent in all trials, the observed effects on the secondary endpoint of mortality varied. Mortality benefit was nominally significant with dapagliflozin but not with empagliflozin or sotagliflozin, raising questions regarding differences in baseline risk profiles or play of chance. None of these trials however were designed to assess mortality impact by itself. Robustness of statistically significant and non-significant dichotomous outcomes can be evaluated using fragility index (FI) and reverse fragility index (RFI).[4,5] FI can help evaluate trial results in addition to *P*-values and effect size estimates. In this study, we sought to assess the robustness of the results across the SGLT2 inhibitors trials in HF by assessing the FI and RFI for clinical outcomes.

Methods

Study populations, definitions, and outcomes of interest

Publicly available data were utilized, and thus, institutional review board approval was not applicable. All placebo-controlled trials designed to evaluate outcomes in HF patients using SGLT2 inhibitors were included. Three randomized controlled trials met these criteria. DAPA-HF and EMPEROR-Reduced included HFrEF outpatients with or without diabetes.[1,2] EMPEROR-Reduced enrolled a higher risk population with lower ejection fraction (EF) and estimated glomerular filtration rate, and higher natriuretic peptides. SOLOIST-WHF enrolled patients with diabetes hospitalized for worsening HF, regardless of EF.[3] Sotagliflozin, studied in SOLOIST-WHF, differs from other SGLT2 inhibitors as it also has SGLT1-inhibiting activity.[3]

The primary outcome varied in all three trials. In EMPEROR-Reduced, it was a composite of first HF hospitalization or cardiovascular mortality. DAPA-HF had a similar primary composite but included urgent outpatient visits for intravenous HF therapy. The number of urgent visits was few, and excluding them resulted in no meaningful change in the effect size. The primary outcome in SOLOIST-WHF was a composite of total (first and recurrent) HF hospitalizations, urgent HF visits, and cardiovascular mortality. We could not evaluate the FI/RFI for this outcome without patient-level data access, but SOLOIST-WHF also reported the composite of first HF hospitalization or cardiovascular mortality.

Fragility index and RFI were assessed for the (i) composite of first HF hospitalization or cardiovascular mortality, (ii) first HF hospitalization, (iii) cardiovascular mortality, and (iv) all-cause mortality. FI for subgroup of the primary outcome was also assessed, but this was not possible for the SOLOIST-WHF trial because the primary outcome included recurrent events.

Statistical analysis

For significant outcomes, FI was calculated in the manner described by Walsh *et al*.[6] In the treatment arm with a lower event rate, patients were added to the event group while subtracting patients from the non-event group. Fisher's exact test was used to recalculate the two-sided *P*-value, while iteratively adding events until the *P*-value became $\geq$0.05. For non-significant outcomes, RFI was calculated. The total number of events in each group over the entire follow-up was considered. Lower FI/RFI indicates less statistical robustness; however, there is no standardized cut-off defined for acceptable fragility. Loss of follow-up was compared with FI/RFI for each trial as it affects both the number of participants at risk and the number of events. When loss to follow-up exceeds the FI or RFI, results should be cautiously interpreted as events of interest may occur in patients lost to follow-up and factoring these may shift the results.

Fragility quotient (FQ),[7,8] which is the FI divided by the sample size, was also calculated to assess what proportion of patients must change status to change the significance of results. For instance, trial X has an FI of 2 and sample size of 500 while trial Y has an FI of 2 and sample size of 1000. Although both trials have the same FI, FQ can gauge which trial is 'relatively' more fragile. Trial X has an FQ of 0.004, meaning that four events per 1000 patients will be needed to change the results significance; while trial Y has an FQ of 0.002, indicating that the non-significance of trial Y is contingent on ~2 events per 1000 patients, suggesting trial Y as more fragile. For statistically non-significant outcomes, RFQ was calculated by dividing the RFI by the sample size. FI, RFI, FQ, and RFQ were calculated using the R Version 3.51 (R Project for

Statistical Computing) and Excel, Version 14.1.3 (Microsoft 160 Corp).

The FI was also calculated after pooling data from all three trials. The previous meta-analysis utilized HRs for which FI or RFI cannot be calculated.[9] For this study, (logarithm of the) risk ratios (RRs) were pooled from each study, which were calculated from dichotomous endpoints, ignoring the event times. A random-effects model was used for meta-analysis. Weights were assigned using the Mantel–Haenszel method. Fragility of meta-analysis results was assessed using the technique described by Atal *et al.*[10] Review Manager (V.5.3) was used to conduct the meta-analysis, and the calculator available at http://clinicalepidemio.fr/fragility_ma/ was used to calculate FI of meta-analysis.

Results

Patient population

The baseline characteristics of patients are shown in *Table 1*. The three studies included a total of 9696 patients (DAPA-HF, n = 4744; EMPEROR-Reduced, n = 3730; and SOLOIST-WHF, n = 1222). The median follow-up time was 18 months in DAPA-HF, 16 months in EMPEROR-Reduced, and 9 months in SOLOIST-WHF. The number of patient's lost to follow-up in DAPA-HF, EMPEROR-Reduced, and SOLOIST-WHF was 36, 42, and 43, respectively.

Fragility index, reverse fragility index, and fragility quotient

Table 2 summarizes the findings from each trial and meta-analysis and displays the FI/RFI and FQ. *Figures 1* and *2* visually represent the FI/RFI and FQ for each outcome of interest.

Composite of first heart failure hospitalization or cardiovascular mortality

SGLT-2 inhibitors reduced the risk of the composite endpoint of first HF hospitalization or cardiovascular mortality in DAPA-HF (hazard ratio [HR] 0.74 [0.65–0.85]), EMPEROR-Reduced (HR: 0.75 [0.65–0.86]), and SOLOIST-WHF (HR: 0.71 [0.56–0.89]) trials. The results were robust (dapagliflozin: FI = 62 and FQ = 0.013; empagliflozin: FI = 50 and FQ = 0.013; and sotagliflozin: FI = 60 and FQ = 0.049), and FI was greater than patients lost to follow-up. *Table 3* shows the FI/RFI of the primary outcomes of DAPA-HF and EMPEROR-Reduced stratified according to different subgroups. Meta-analysis demonstrated significant (RR: 0.75 [0.69–0.81]; $P < 0.001$; I^2 = 20%) and robust (FI = 95 and FQ = 0.010) benefit.

First heart failure hospitalization

Both DAPA-HF (HR: 0.70 [0.59–0.83]) and EMPEROR-Reduced (HR: 0.69 [0.59–0.81]) showed a significant reduction in HF hospitalization. The FI was high in DAPA-HF (n = 43; FQ = 0.009) and EMPEROR-Reduced (n = 50; FQ = 0.013), and FI was higher than the number of patients lost to follow-up. This outcome was not reported in SOLOIST-WHF. Meta-analysis demonstrated a significant reduction in HF hospitalization with SGLT2 inhibitors (RR: 0.72 [0.65–0.81]; $P < 0.001$; I^2 = 0%) with an FI of 61, and the FQ was 0.007.

Mortality

In DAPA-HF, nominally significant reduction in all-cause (HR: 0.83 [0.71–0.97]) and cardiovascular mortality (HR: 0.82 [0.69–0.98]) was observed. The FI was 8 for all-cause and 5 for cardiovascular mortality, with FQs of 0.002 and 0.001, respectively. In EMPEROR-Reduced, no statistically significant reduction in all-cause (HR 0.92 [0.77–1.10]) or cardiovascular mortality (HR 0.92 [0.75–1.12]) was seen (RFI and RFQ were 26 and 0.007 for all-cause and 24 and 0.006 for cardiovascular mortality). In SOLOIST-WHF trial, no statistically significant difference in all-cause (HR: 0.82 [0.59–1.14]) and cardiovascular mortality (0.84 [0.58–1.22]) was seen; results were fragile for in both cases (RFI = 6 and RFQ = 0.005 for all-cause mortality; RFI = 7 and RFQ = 0.006 for cardiovascular mortality). Meta-analysis demonstrated a significant, but fragile, reduction in all-cause (RR: 0.88 [0.79–0.98]; P = 0.02; I^2 = 0%; FI = 12 and FQ = 0.001) and cardiovascular mortality (RR: 0.87 [0.78–0.98]; P = 0.02; I^2 = 0%; FI = 9 and FQ = 0.001).

Discussion

The DAPA-HF, EMPEROR-Reduced and SOLOIST-WHF trials all reported highly concordant and statistically robust results for the primary endpoint of time to first HF hospitalization or cardiovascular death.[1–3] The FI for this endpoint was higher than the FI for outcomes in trials of other drugs, for example, therapies referenced in diabetes treatment guidelines (FI = 16) and anti-thrombotic therapy (FI = 5) in venous thromboembolism guidelines.[11,12] The FI and RFIs of outcomes in landmark HF trials are displayed in *Table 4*.

In the SOLOIST-WHF trial, the initial primary endpoint was the composite of first HF hospitalization or cardiovascular mortality; however, this was later changed to a composite of first and recurrent HF hospitalization, urgent HF visits, and cardiovascular mortality. Despite enrolling only 1222 patients, this trial showed a significant and robust (FI = 60) reduction in first HF hospitalization or cardiovascular mortality, reinforcing the benefit of SGLT2 inhibitors in HF. These effects remained consistent across a range of subgroups including patients with reduced and preserved EF, in-hospital vs. post-discharge initiation, and use of sacubitril/valsartan.[3,13]

Table 1 Baseline characteristics

	EMPEROR-Reduced		DAPA-HF		SOLOIST-WHF	
	Empagliflozin	Placebo	Dapagliflozin	Placebo	Sotagliflozin	Placebo
Number of participants	1863	1867	2373	2371	608	614
Age, years (SD)	67.2 (10.8)	66.5 (11.2)	66.2 (11.0)	66.5 (10.8)	69 (63–76)	70 (64–76)
Sex, *n* (%)						
Male	1426 (76.5)	1411 (75.6)	1809 (76.2)	1826 (77.0)	410 (67.4)	400 (65.1)
Female	437 (23.5)	456 (24.4)	564 (23.8)	545 (23.0)	198 (32.6)	214 (34.9)
NYHA functional classification (%)						
II	75.1	75.0	67.7	67.4		
III	24.4	24.4	31.5	31.7		
IV	0.5	0.6	0.8	1.0		
Mean LVEF (%)	27.7 (6.0)	27.2 (6.1)	31.2 (6.7)	30.9 (6.9)	35 (28–47)	35 (28–45)
HFpEF (%)					127 (20.9)	129 (21.0)
HFrEF (%)					481 (79.1)	485 (79.0)
NT-proBNP (pg/mL)	1887 (1077–3429)	1926 (1153–3525)	1428 (857–2655)	1446 (857–2641)	1817 (845–3659)	1741 (843–3582)
Hospitalization for heart failure	577 (31.0)	574 (30.7)	1124 (47.4)	1127 (47.5)		
Diabetes	927 (49.8)	929 (49.8)	1075 (45.3)	1064 (44.9)	25 (4.1)	20 (3.3)
Duration of diabetes (years)						
eGFR (mL/min/1.73 m^2)	61.8 (21.7)	62.2 (21.5)	66.0 (19.6)	65.5 (19.3)	49.2 (39.5–61.2)	50.5 (40.5–64.6)
Heart failure medications, *n* (%)						
Renin–angiotensin inhibitor	1644 (88.8)	1673 (90.6)	>90%	>90%	553 (91.0)	563 (91.7)
ACE inhibitor	867 (46.5)	836 (44.8)	1332 (56.1)	1329 (56.1)	254 (41.8)	241 (39.3)
ARB	451 (24.2)	457 (24.5)	675 (28.4)	632 (26.7)	245 (40.3)	270 (44.0)
ARNI	340 (18.3)	387 (20.7)	250 (10.5)	258 (10.9)	93 (15.3)	112 (18.2)
Mineralocorticoid receptor antagonist	1306 (70.1)	1355 (72.6)	1696 (71.5)	1674 (70.6)	403 (66.3)	385 (62.7)
Beta-blocker	1765 (94.7)	1768 (94.7)	2278 (96.0)	2280 (96.2)	564 (92.8)	561 (91.4)
ICD or CRT-D	578 (31.0%)	593 (31.8%)	622 (26.2%)	620 (26.1%)		
CRT-D or CRT-P	220 (11.8%)	222 (11.9%)	190 (8.0%)	164 (6.9%)		

ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; ARNI, angiotensin receptor/neprilysin inhibitor; CRT-D, cardiac resynchronization therapy defibrillator; CRT-P, cardiac resynchronization therapy pacemaker; eGFR, estimated glomerular filtration rate; HF, heart failure; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; ICD, implantable cardiac defibrillator; IQR, inter-quartile range; LVEF, left ventricular ejection fraction; NT-proBNP, N-terminal pro-B-type natriuretic peptide; NYHA, New York Heart Association; SD, standard deviation.
Data are reported as *n* (%), mean (SD), or median (IQR).

Table 2 Fragility of findings from SGLT2i trials and meta-analysis of these trials

	DAPA-HF					EMPEROR-Reduced				
	Events/total SGLT2i	Events/total placebo	HR [CI]	FI/RFI	FQ	Events/total SGLT2i	Events/total placebo	HR [CI]	FI/RFI	FQ
HFH or CVM	386/2373	502/2371	0.74 [0.65–0.85]	62	0.0130	361/1863	462/1867	0.75 [0.65–0.86]	50	0.0130
First HFH	231/2373	318/2371	0.70 [0.59–0.83]	43	0.0090	246/1863	342/1867	0.69 [0.59–0.81]	50	0.0130
CVM	227/2373	273/2371	0.82 [0.69–0.98]	5	0.0010	187/1863	202/1867	0.92 [0.75–1.12]	24[a]	0.0060
ACM	276/2373	329/2371	0.83 [0.71–0.97]	8	0.0020	249/1863	266/1867	0.92 [0.77–1.10]	26[a]	0.0070

ACM, all-cause mortality; CI, confidence interval; CVM, cardiovascular mortality; FI, fragility index; FQ, fragility quotient; HF, heart failure; HFH, heart failure hospitalization; HR, hazard ratio; N/A, not available; RFI, reverse fragility index; SGLT2i, sodium–glucose co-transporter 2 inhibitor.
[a]Reverse fragility index/reverse fragility quotient.

Table 2 (continued)

	SOLOIST-WHF					Meta-analysis				
	Events/total SGLT2i	Events/total placebo	HR [CI]	FI/RFI	FQ	Events/total SGLT2i	Events/total placebo	RR [CI]	FI/RFI	FQ
HFH or CVM	201/608	298/614	0.71 [0.56–0.89]	60	0.0490	948/4844	1262/4852	0.75 [0.69–0.81]	95	0.0098
First HFH	N/A	N/A	N/A	N/A	N/A	477/4236	660/4238	0.72 [0.65–0.81]	61	0.0072
CVM	51/608	58/614	0.84 [0.58–1.22]	7[a]	0.0060	465/4844	533/4852	0.87 [0.78–0.98]	9	0.0009
ACM	65/608	76/614	0.82 [0.59–1.14]	6[a]	0.0050	590/4844	671/4852	0.88 [0.79–0.98]	12	0.0012

ACM, all-cause mortality; CI, confidence interval; CVM, cardiovascular mortality; FI, fragility index; FQ, fragility quotient; HF, heart failure; HFH, heart failure hospitalization; HR, hazard ratio; N/A, not available; RFI, reverse fragility index; SGLT2i, sodium–glucose co-transporter 2 inhibitor.
[a]Reverse fragility index/reverse fragility quotient.

Figure 1 Fragility index and reverse fragility index of outcomes in heart failure (HF)-specific sodium–glucose co-transporter 2 inhibitor trials. CV, cardiovascular. *Reverse fragility index.

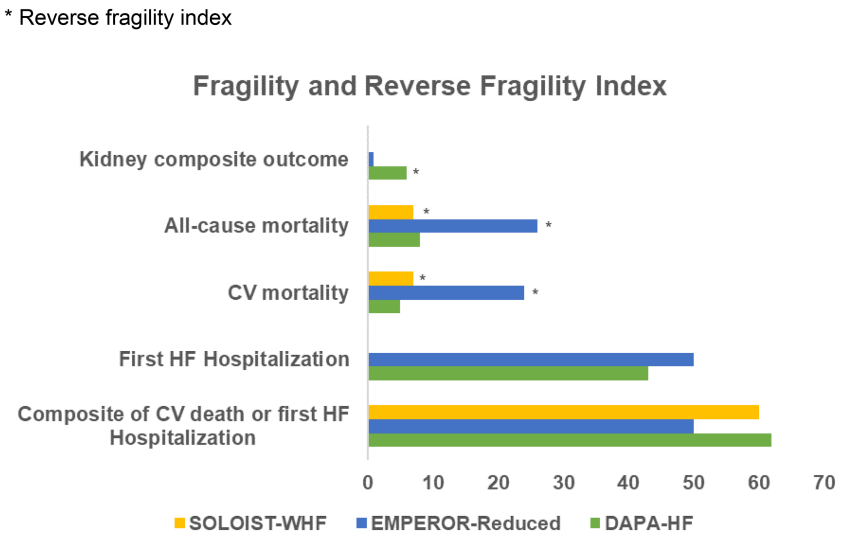
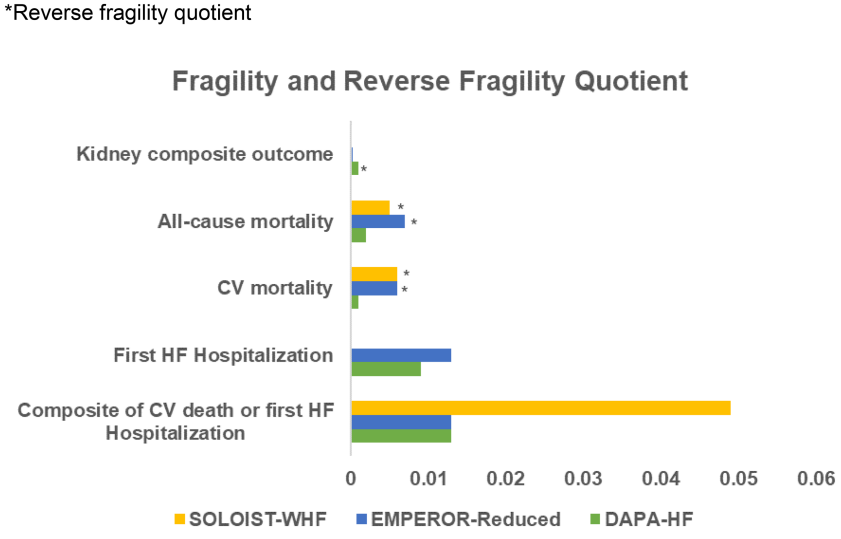


Figure 2 Fragility and reverse fragility quotients of outcomes in heart failure (HF)-specific sodium–glucose co-transporter 2 inhibitor trials. CV, cardiovascular. *Reverse fragility quotient.



The risk of all-cause and cardiovascular mortality was nominally reduced in DAPA-HF but not in EMPEROR-Reduced or SOLOIST-WHF. Mortality outcomes were secondary endpoints and statistically fragile, with significance dependent on less than 10 events per 1000 patients. Meta-analysis showed a robust FI for the composite endpoint of first HF hospitalization or cardiovascular mortality was 95 (FQ = 0.010) but did not result in higher FI for cardiovascular (FI = 9) or all-cause mortality (FI = 12). These findings highlight that although combining studies via random-effects meta-analysis can help detect a statistically significant treatment effect by increasing power, this does not necessarily result in an increased FI and that the FI and RFI are not strictly linked to statistical significance, confidence intervals, and power. While the aforementioned are related concepts, FI offers additional value in interpretation of clinical trials and meta-analyses.

Table 3 Fragility index and reverse fragility index for subgroup analyses of the primary endpoint in DAPA-HF and EMPEROR-Reduced

	DAPA-HF					EMPEROR-Reduced				
Subgroup	Events/total in SGLT2 inhibitor group	Events/total in placebo group	HR [95% CI]	FI/RFI	FQ/RFQ	Events/total in SGLT2 inhibitor group	Events/total in placebo group	HR [95% CI]	FI/RFI	FQ/RFQ
Diabetes	215/1075	271/1064	0.75 [0.63–0.90]	20	0.009	200/927	265/929	0.72 [0.60–0.87]	27	0.015
No diabetes	171/1298	231/1307	0.73 [0.60–0.88]	22	0.008	161/936	197/938	0.78 [0.64–0.97]	2	0.001
Men	307/1809	406/1826	0.73 [0.63–0.85]	47	0.013	294/1426	353/1411	0.80 [0.68–0.93]	18	0.006
Women	79/564	96/545	0.79 [0.59–1.06]	3[a]	0.003[a]	67/437	109/456	0.59 [0.44–0.80]	14	0.016
Receiving ARNI	41/250	56/258	0.75 [0.50–1.13]	3[a]	0.006[a]	51/430	93/387	0.64 [0.45–0.89]	11	0.013
Not receiving ARNI	345/2123	446/2113	0.74 [0.65–0.86]	52	0.012	310/1523	369/1480	0.77 [0.66–0.90]	24	0.008
Age <65 years	162/1032	196/998	0.78 [0.63–0.96]	7	0.003	128/675	193/740	0.71 [0.57–0.89]	18	0.013
Age >65 years	224/1341	306/1373	0.72 [0.60–0.85]	34	0.013	233/1188	269/1127	0.78 [0.66–0.93]	10	0.004
History of HF hospitalization	195/1124	279/1127	0.67 [0.56–0.80]	44	0.020	153/577	177/574	0.79 [0.64–0.99]	1	0.001
No history of HF hospitalization	191/1249	223/1244	0.84 [0.69–1.01]	3[a]	0.001[a]	208/1286	285/1293	0.71 [0.60–0.85]	35	0.014
eGFR < 60	191/962	254/964	0.72 [0.59–0.86]	26	0.013	202/893	237/906	0.83 [0.69–1.00]	1[a]	0.001[a]
eGFR > 60	195/1410	248/1406	0.76 [0.63–0.92]	15	0.005	159/969	224/960	0.67 [0.55–0.83]	31	0.016

ARNI, angiotensin receptor/neprilysin inhibitor; CI, confidence interval; eGFR, estimated glomerular filtration rate; FI, fragility index; FQ, fragility quotient; HF, heart failure; HR, hazard ratio; RFI, reverse fragility index; SGLT2, sodium–glucose co-transporter 2.
[a]Reverse fragility index/reverse fragility quotient.

The fact that a few events could change the statistical significance of the all-cause and cardiovascular mortality underscores the importance of powering trials for mortality outcomes. When the primary outcome is a composite including non-fatal events, neither the total number of fatal events nor the duration of trial follow-up supports deriving definitive conclusions regarding mortality. While it is not possible to power trials for all secondary endpoints, considering the risk for mortality in HF, strong consideration should be given to designing trials for confirming mortality results independently by either larger sample size or longer follow-up or both.

Background therapy can potentially influence the FI and RFI of clinical trials. Similar to most contemporary HF trials, patients in EMPEROR-Reduced, DAPA-HF, and SOLOIST-WHF were well treated with guideline-directed medical therapy at baseline (*Table 1*). In all three trials, over 90% of the participants were using beta-blockers and renin–angiotensin–aldosterone inhibitors, and over two-thirds were using mineralocorticoid receptor antagonists. There was some variation in the proportion of patients using a neprilysin inhibitor in addition to a renin–angiotensin–aldosterone inhibitor, with the highest rate of use in EMPEROR-Reduced (20%), followed by SOLOIST-WHF (17%) and then DAPA-HF (11%). Overall, the background therapies across trials were similar and unlikely to influence the FI or RFI.

There are several limitations of this study. FI does not account for the difference in time to event and can give fragile results when the number of events in each group is the same but have a difference in the timing of these events. However, studies have shown no difference when FI was applied to time-to-event vs. frequency data.[4–6] Because trials are powered to detect the effect on primary outcome, interpretability of FI for secondary outcomes and subgroup analyses is limited. The use of Fisher's exact test in calculation of FI and RFI may be limited as DAPA-HF and EMPEROR-Reduced trials analysed data in models with covariates and time-to-event techniques in which the original data if analysed with Fisher's exact test may not yield the same *P*-value as the published trial. While FI may perpetuate the dichotomous *P*-value-oriented data interpretation, it provides a more circumspect view of assessing results than based solely on *P*. Bayesian approaches may provide an alternate option, but the majority of trials currently are based on frequentist approaches.

In conclusion, findings for the composite endpoint of first HF hospitalization or cardiovascular death were highly consistent and robust across trials with dapagliflozin, empagliflozin, and sotagliflozin. In contrast, findings for the all-cause and cardiovascular mortality were overall significant but fragile when meta-analytically assessed, underscoring the need to design trials with adequate power and follow-up to definitely assess the impact of novel interventions on mortality.

Table 4 Fragility index of outcomes in key heart failure medication trials

Trial	Sample size	Primary endpoint	Primary endpoint			HF hospitalization/CV death			All-cause mortality			CV mortality			HF hospitalization		
			P-value	FI/RFI	FQ/RFQ	*P*-value	FI/RFI	RFQ/RFQ	*P*-value	FI/RFI	RFQ/RFQ	*P*-value	FI/RFI	RFQ/RFQ	*P*-value	FI/RFI	RFQ/RFQ
SGLT2 inhibitors																	
DAPA-HF	4744	HF hospitalization, urgent care visits due to HF or CV death	Sig	62	0.0131	Sig	62	0.0130	Sig	8	0.0020	Sig	5	0.0010	Sig	43	0.0090
EMPEROR-Reduced	3730	HF hospitalization/CV death	Sig	50	0.0134	Sig	50	0.0134	NS	26	0.0070	NS	24	0.0060	Sig	50	0.0130
SOLOIST-WHF	1222	Total (first and recurrent) HF hospitalization/CV death	Sig	NA	NA	Sig	60	0.0490	NS	6	0.0050	NS	7	0.0060	NA	NA	NA
ACE-I																	
CONSENSUS	253	All-cause mortality at 6 months	Sig	7	0.0277	NA	NA	NA	Sig	3	0.0120	Sig	4	0.0158	NA	NA	NA
SOLVD-Treatment	2569	All-cause mortality	Sig	10	0.0039	NA	NA	NA	Sig	10	0.0040	Sig	15	0.0058	Sig	91	0.0354
ARBs																	
Val-HeFT	5010	All-cause mortality/HF hospitalization/resuscitated cardiac arrest/ administration of i.v. inotropic or vasodilator drugs for 4 or more hours	Sig	17	0.0034	NA	NA	NA	NS	63	0.0130	NS	58	0.0116	Sig	59	0.0118
CHARM-Alternative	2028	HF hospitalization/CV death	Sig	29	0.0143	Sig	29	0.0143	NS	10	0.0049	NS	6	0.0030	Sig	40	0.0197
CHARM-Added	2548	HF hospitalization/CV death	Sig	8	0.0031	Sig	8	0.0031	NS	10	0.0039	Sig	3	0.0012	Sig	5	0.0020
Beta-blockers																	
CIBIS II	2647	All-cause mortality	Sig	37	0.0140	NA	NA	NA	Sig	37	0.0140	Sig	11	0.0042	Sig	37	0.0140
MERIT-HF	3991	All-cause mortality	Sig	34	0.0085	NA	NA	NA	Sig	34	0.0085	Sig	38	0.0095	Sig	50	0.0125
COPERNICUS	2289	All-cause mortality	Sig	30	0.0131	NA	NA	NA	Sig	30	0.0131	Sig	22	0.0096	Sig	37	0.0162
SENIORS	2128	All-cause mortality/CV hospital admission	Sig	2	0.0009	NA	NA	NA	NS	11	0.0052	NS	7	0.0033	NS	31	0.0146
MRA																	
RALES	1663	All-cause mortality	Sig	54	0.0325	NA	NA	NA	Sig	54	0.0325	Sig	46	0.0277	Sig	41	0.0247
EMPHASIS-HF	2737	CV death/HF hospitalization	Sig	61	0.0223	Sig	61	0.0223	Sig	5	0.0018	Sig	3	0.0011	Sig	49	0.0179
H-ISDN																	
A-HeFT	1050	Composite score	Sig	NA	NA	NA	NA	NA	Sig	3	0.0029	Sig	2	0.0019	Sig	15	0.0143
Digoxin																	
DIG	6800	All-cause mortality	NS	61	0.0090	NA	NA	NA	NS	61	0.008971	NS	87	0.0128	Sig	191	0.0281
Ivabradine																	
SHIFT	6505	HF hospitalization/CV death	Sig	67	0.0103	Sig	67	0.0103	NS	13	0.0019	NS	18	0.0028	Sig	91	0.0140
ARNI																	
PARADIGM-HF	8399	HF hospitalization/CV death	Sig	118	0.0140	Sig	118	0.0140	Sig	49	0.0058	Sig	66	0.0079	Sig	54	0.0064
Vericiguat																	
VICTORIA	5050	HF hospitalization/CV death	Sig	8	0.0016	Sig	8	0.0016	NS	34	0.0067	NS	25	0.0050	NS	1	0.0002
Omecamtiv mecarbil																	
GALACTIC-HF	8442	HF hospitalization, urgent care visits due to HF or CV death	Sig	1	0.0001	NS	33	0.0040	NS	78	0.0092	NS	79	0.0094	NS	40	0.0047

ACE-I, angiotensin-converting enzyme inhibitor; ARBs, angiotensin receptor blockers; ARNI, angiotensin receptor/neprilysin inhibitor; CV, cardiovascular; FI, fragility index; FQ, fragility quotient; HF, heart failure; H-ISDN, hydralazine–isosorbide dinitrate; MRA, mineralocorticoid antagonist; RFI, reverse fragility index; RFQ, reverse fragility quotient; SGLT2, sodium–glucose co-transporter 2.

Conflict of interest

G.C.F. reports consulting for Abbott, Amgen, AstraZeneca, Bayer, CHF Solutions, Janssen, Medtronic, Merck, and Novartis. S.J.G. was supported by the Heart Failure Society of America/Emergency Medicine Foundation Acute Heart Failure Young Investigator Award funded by Novartis; has received research support from the American Heart Association, AstraZeneca, Amgen, Bristol Myers Squibb, Merck, and Novartis; has served on advisory boards for Amgen and Cytokinetics; and serves as a consultant for Amgen and Merck. M. V. is supported by the KL2/Catalyst Medical Research Investigator Training Award from Harvard Catalyst (NIH/NCATS Award UL 1TR002541); serves on advisory boards for Amgen, AstraZeneca, Baxter Healthcare, Bayer AG, Boehringer Ingelheim, Cytokinetics, and Relypsa; and participates on clinical endpoint committees for studies sponsored by Galmed, Novartis, and the NIH. T.F. reports personal fees from Novartis, Bayer, Janssen, SGS, Roche, Boehringer Ingelheim, Daiichi Sankyo, Galapagos, Penumbra, Parexel, Vifor, Biosense Webster, CSL Behring, Fresenius Kabi, Coherex Medical, and LivaNova. G.F. participated in committees of trials and registries sponsored by BI, Bayer, Medtronic, Servier, Novartis, and Vifor. A.J.S.C. reports honoraria and/or lecture fees from AstraZeneca, Bayer, Menarini, Novartis, Nutricia, Servier, Vifor, Actimed, Cardiac Dimensions, CVRx, Enopace, Faraday, Gore, Impulse Dynamics, Respicardia, Stealth Peptides, V-Wave, Corvia, Arena, and ESN Cleer. S.D.A. has received research support from Vifor International and Abbott Vascular, and fees for consultancy and/or speaking from AstraZeneca, Bayer, Boehringer Ingelheim, Respicardia, Impulse Dynamics, Janssen, Novartis, Servier, and Vifor International. J.B. is a consultant to Abbott, Amgen, Applied Therapeutics, Array, AstraZeneca, Bayer, Boehringer Ingelheim, CVRx, Eli Lilly, G3 Pharmaceutical, Impulse Dynamics, Innolife, Janssen, Luitpold, Medtronic, Merck, Novartis, Novo Nordisk, Relypsa, Sequana, Stealth Peptides, and Vifor. All other authors report no disclosures.

Funding

None.

References

1. McMurray JJV, Solomon SD, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, Ponikowski P, Sabatine MS, Anand IS, Bělohlávek J, Böhm M, Chiang CE, Chopra VK, de Boer RA, Desai AS, Diez M, Drozdz J, Dukát A, Ge J, Howlett JG, Katova T, Kitakaze M, Ljungman CEA, Merkely B, Nicolau JC, O'Meara E, Petrie MC, Vinh PN, Schou M, Tereshchenko S, Verma S, Held C, DeMets DL, Docherty KF, Jhund PS, Bengtsson O, Sjöstrand M, Langkilde AM. Dapagliflozin in patients with heart failure and reduced ejection fraction. *N Engl J Med* 2019; **381**: 1995–2008.
2. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, Januzzi J, Verma S, Tsutsui H, Brueckmann M, Jamal W, Kimura K, Schnee J, Zeller C, Cotton D, Bocchi E, Böhm M, Choi DJ, Chopra V, Chuquiure E, Giannetti N, Janssens S, Zhang J, Gonzalez Juanatey JR, Kaul S, Brunner-La Rocca HP, Merkely B, Nicholls SJ, Perrone S, Pina I, Ponikowski P, Sattar N, Senni M, Seronde MF, Spinar J, Squire I, Taddei S, Wanner C, Zannad F. Cardiovascular and renal outcomes with empagliflozin in heart failure. *N Engl J Med* 2020; **383**: 1413–1424.
3. Bhatt DL, Szarek M, Steg PG, Cannon CP, Leiter LA, McGuire DK, Lewis JB, Riddle MC, Voors AA, Metra M, Lund LH, Komajda M, Testani JM, Wilcox CS, Ponikowski P, Lopes RD, Verma S, Lapuerta P, Pitt B. Sotagliflozin in patients with diabetes and recent worsening heart failure. *N Engl J Med* 2020; **384**: 117–128.
4. Khan MS, Ochani RK, Shaikh A, Usman MS, Yamani N, Khan SU, Murad MH, Mandrola J, Doukky R, Krasuski RA. Fragility index in cardiovascular randomized controlled trials. *Circ Cardiovasc Qual Outcomes* 2019; **12**: e005755.
5. Khan MS, Fonarow GC, Friede T, Lateef N, Khan SU, Anker SD, Harrell FE Jr, Butler J. Application of the reverse fragility index to statistically nonsignificant randomized clinical trial results. *JAMA Netw Open* 2020; **3**: e2012469.
6. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, Molnar AO, Dattani ND, Burke A, Guyatt G, Thabane L, Walter SD, Pogue J, Devereaux PJ. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. *J Clin Epidemiol* 2014; **67**: 622–628.
7. Wayant C, Meyer C, Gupton R, Som M, Baker D, Vassar M. The fragility index in a cohort of HIV/AIDS randomized controlled trials. *J Gen Intern Med* 2019; **34**: 1236–1243.
8. Bowers A, Meyer C, Tritz D, Cook C, Fuller K, Smith C, Diener B, Vassar M. Assessing quality of randomized trials supporting guidelines for laparoscopic and endoscopic surgery. *J Surg Res* 2018; **224**: 233–239.
9. Butler J, Usman MS, Khan MS, Greene SJ, Friede T, Vaduganathan M, Filippatos G, Coats AJ, Anker SD. Efficacy and safety of SGLT2 inhibitors in heart failure: systematic review and meta-analysis. *ESC Heart Fail* 2020; **7**: 3298–3309.
10. Atal I, Porcher R, Boutron I, Ravaud P. The statistical significance of meta-analyses is frequently fragile: definition of a fragility index for meta-analyses. *J Clin Epidemiol* 2019; **111**: 32–40.
11. Chase Kruse B, Matt Vassar B. Unbreakable? An analysis of the fragility of randomized trials that support diabetes treatment guidelines. *Diabetes Res Clin Pract* 2017; **134**: 91–105.
12. Edwards E, Wayant C, Besas J, Chronister J, Vassar M. How fragile are clinical trial outcomes that support the CHEST clinical practice guidelines for VTE? *Chest* 2018; **154**: 512–520.
13. Zannad F, Ferreira JP, Pocock SJ, Anker SD, Butler J, Filippatos G, Brueckmann M, Ofstad AP, Pfarr E, Jamal W, Packer M. 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.