SUPPLEMENTARY MATERIAL

Supplemental Methods:

Preregistered Search Query

The following string was used in the PubMed/MEDLINE prespecified search query to identify potential studies to be included in the meta-analysis:

((((((((((canagliflozin[Text Word]) OR (dapagliflozin[Text Word])) OR (empagliflozin[Text Word])) OR (ertugliflozin[Text Word])) OR (sotagliflozin[Text Word])) OR (SGLT-2 inhibitor[Text Word])) OR (sodium glucose co-transporter-2 inhibitor[Text Word])) AND (((((((heart failure with preserved ejection fraction[Text Word]) OR (heart failure with mildly reduced ejection fraction[Text Word])) OR (acute heart failure[Text Word])) OR (worsening heart failure[Text Word])) OR (heart failure hospitalisation[Text Word])) OR (chronic heart failure[Text Word])) OR (left ventricular ejection fraction[Text Word])))) AND ((((randomized clinical trial[Text Word]) OR (placebo[Text Word])) OR (randomized controlled trial[Text Word])) OR (randomized[Text Word])))) AND ("humans"[Filter])) OR ((((Sodium-Glucose Transporter 2 Inhibitors[MeSH Terms])) AND (((heart failure, diastolic[MeSH Terms]) OR (heart failure[MeSH Terms])) OR (heart failure[MeSH Terms])))) AND (Randomized Controlled Trial [Publication Type]))

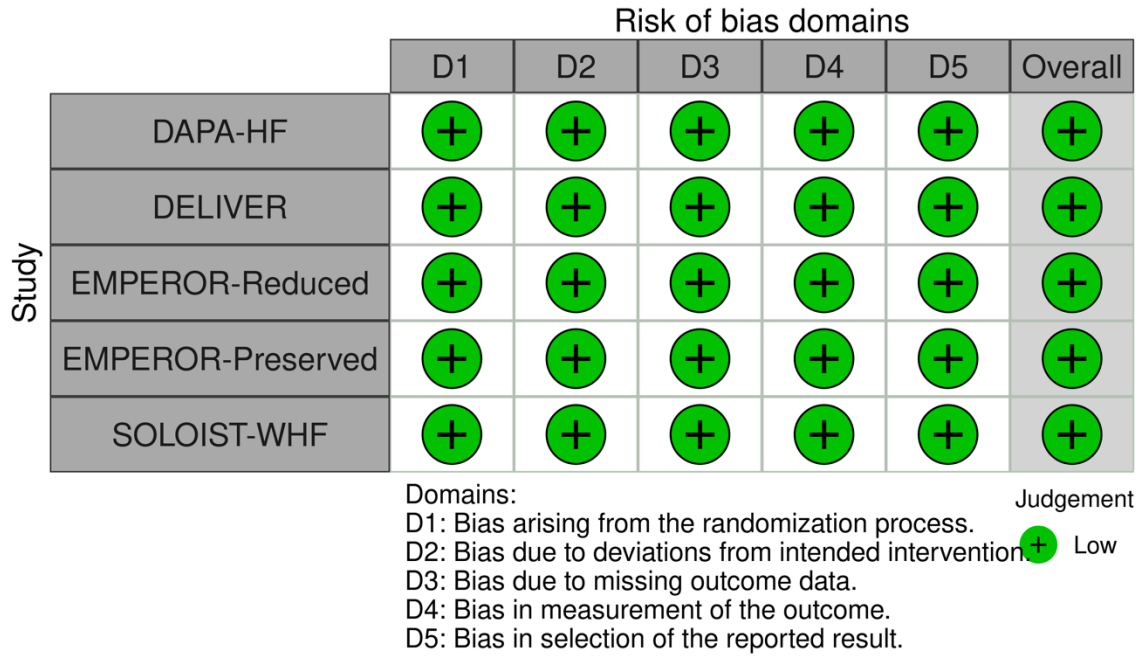
Data Harmonization between DELIVER and EMPEROR-Preserved

The definitions of the endpoints were slightly different and we used participant-level data from DELIVER to match the EMPEROR-Preserved definitions. In addition, to facilitate better alignment of endpoint definitions, we used published estimates from analyses that have applied outcome definitions used in DELIVER to the EMPEROR-Preserved trial.[10] There were 2 key differences in major endpoint definitions and sensitivity analyses reconciled some of these distinctions. First, the primary endpoint for EMPEROR-Preserved (composite of cardiovascular death or first hospitalisation for HF) was different from that in DELIVER (composite of cardiovascular death, first hospitalisation for HF, or urgent HF visit). Since published subgroup data were available for the EMPEROR-Preserved primary endpoint, this was selected as the main endpoint examined in this meta-analysis. Meta-analysis according to the DELIVER primary endpoint definition was additionally reported. Second, cardiovascular death was defined differently in DELIVER and EMPEROR-Preserved trials due to differences in the handling of unknown or undetermined deaths as designated by the clinical endpoints committees. In EMPEROR-Preserved, unknown or undetermined deaths were counted and assumed to be cardiovascular deaths, while in DELIVER, they were not. The harmonized analysis presents cardiovascular death alone[10] (excluding unknown or undetermined deaths) as the primary analysis, while the endpoint of cardiovascular death (including unknown or undetermined deaths) is reported and meta-analyzed separately.

Supplemental Table 1. Definitions of Adverse Events in DELIVER and EMPEROR-Preserved

Adverse Event	DELIVER	EMPEROR-Preserved
Amputation	Any amputation	Investigator defined events and included lower limb amputation only
Diabetic Ketoacidosis	Any definite or probable diabetic ketoacidosis confirmed in blinded adjudication	According to narrow Boehringer Ingelheim customised MedDRA query, and includes events in people without diabetes
Hypoglycemia	Any major hypoglycemic event, which was defined as: AE with the following criteria confirmed by the investigator: i) Symptoms of severe impairment in consciousness or behaviour; ii) need of external assistance; iii) intervention to treat hypoglycemia; iv) prompt recovery of acute symptoms following the intervention.	Events with a plasma glucose value of ≤70 mg/dL or that required assistance
Renal	Based on the narrow Standardised MedDRA query for acute renal failure, with only serious adverse events or adverse events leading to drug discontinuation reported	Based on the narrow Standardised MedDRA query for acute renal failure

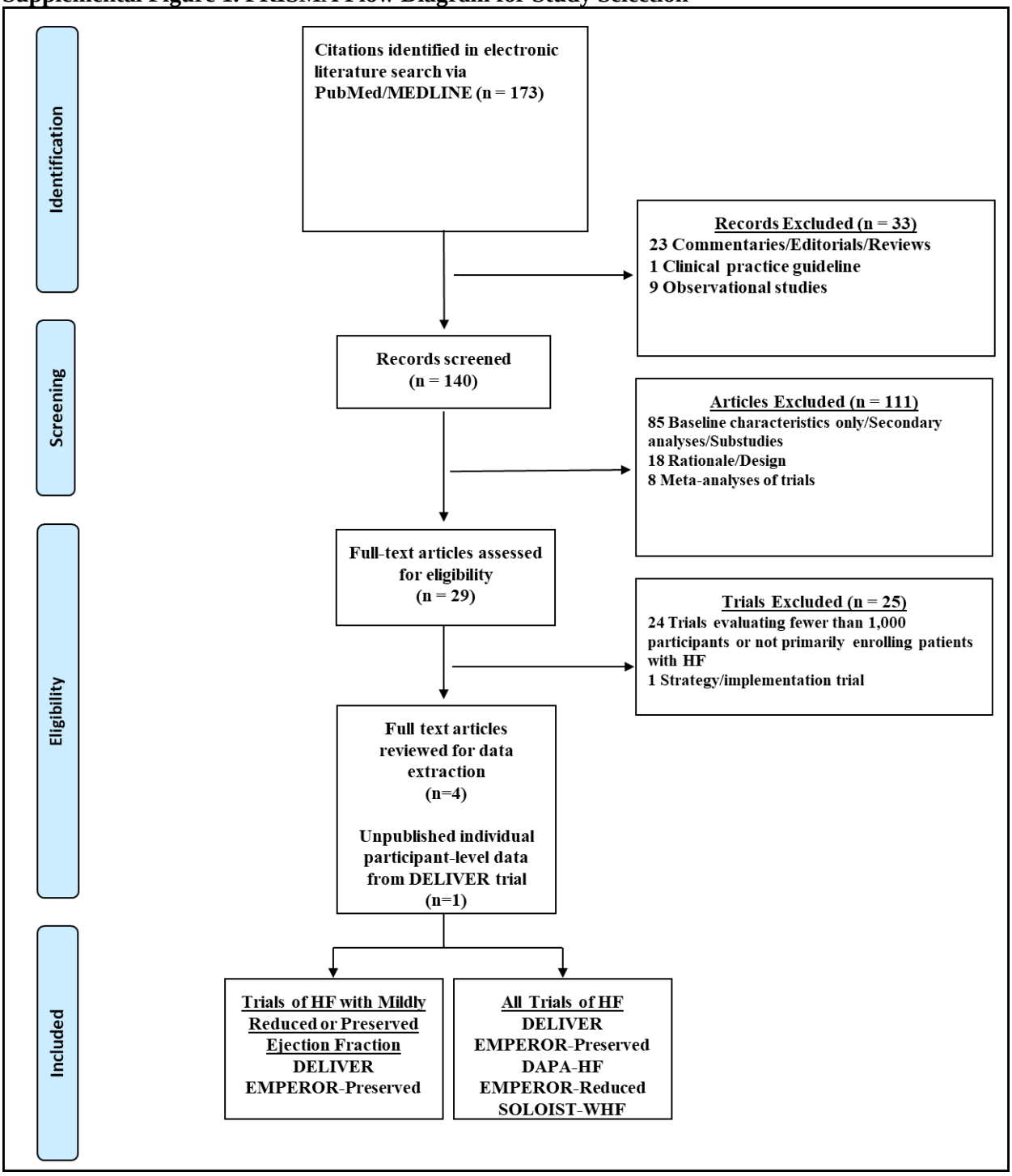
Supplemental Table 2. Risk of Bias Assessment of Selected Trials



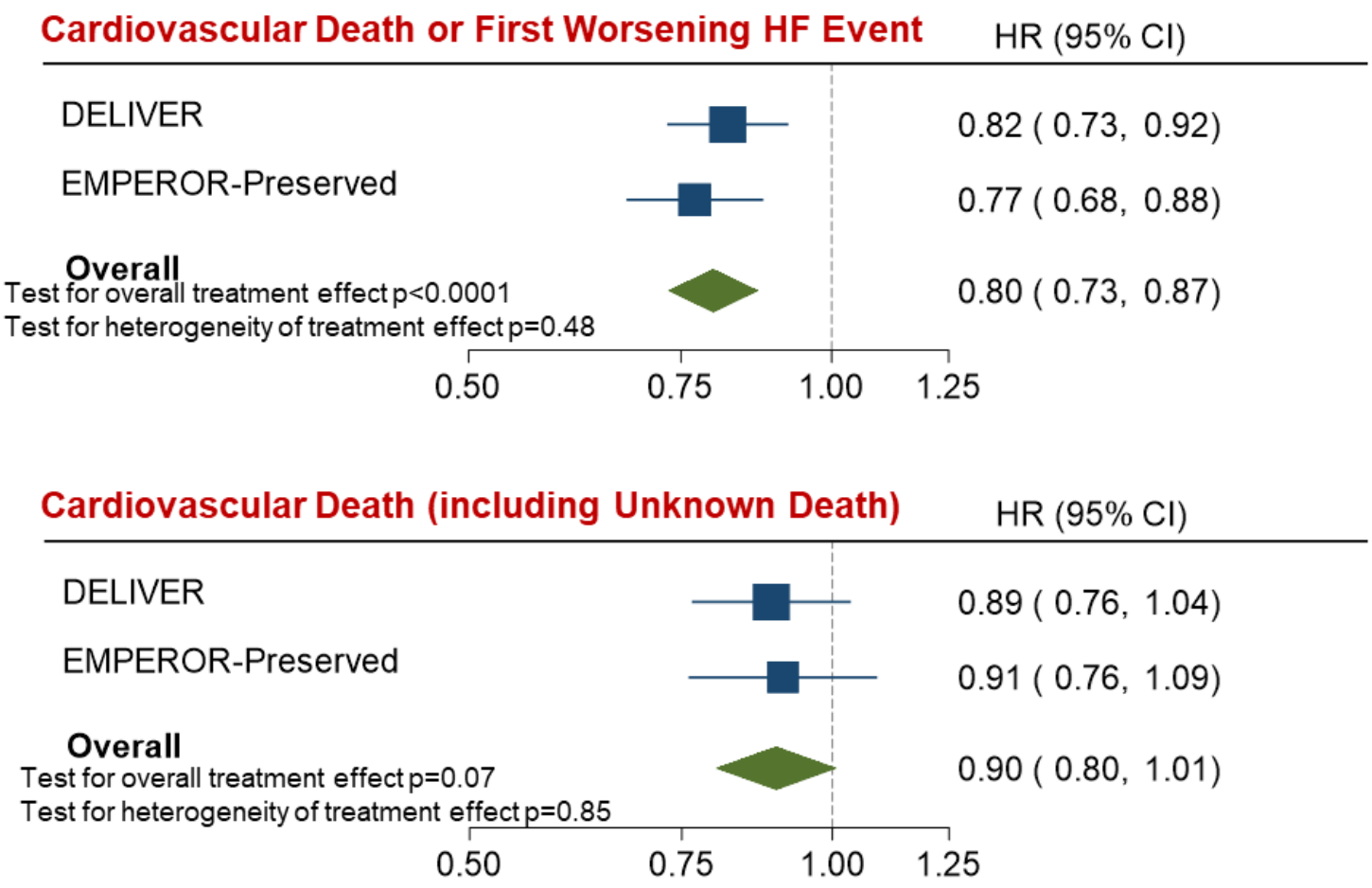
Risk of bias was assessed using the Cochrane risk of bias tool for randomised trials. Version 2.

Figure created using the *robvis* application: McGuinness, LA, Higgins, JPT. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. Res Syn Meth. 2020; 1- 7.

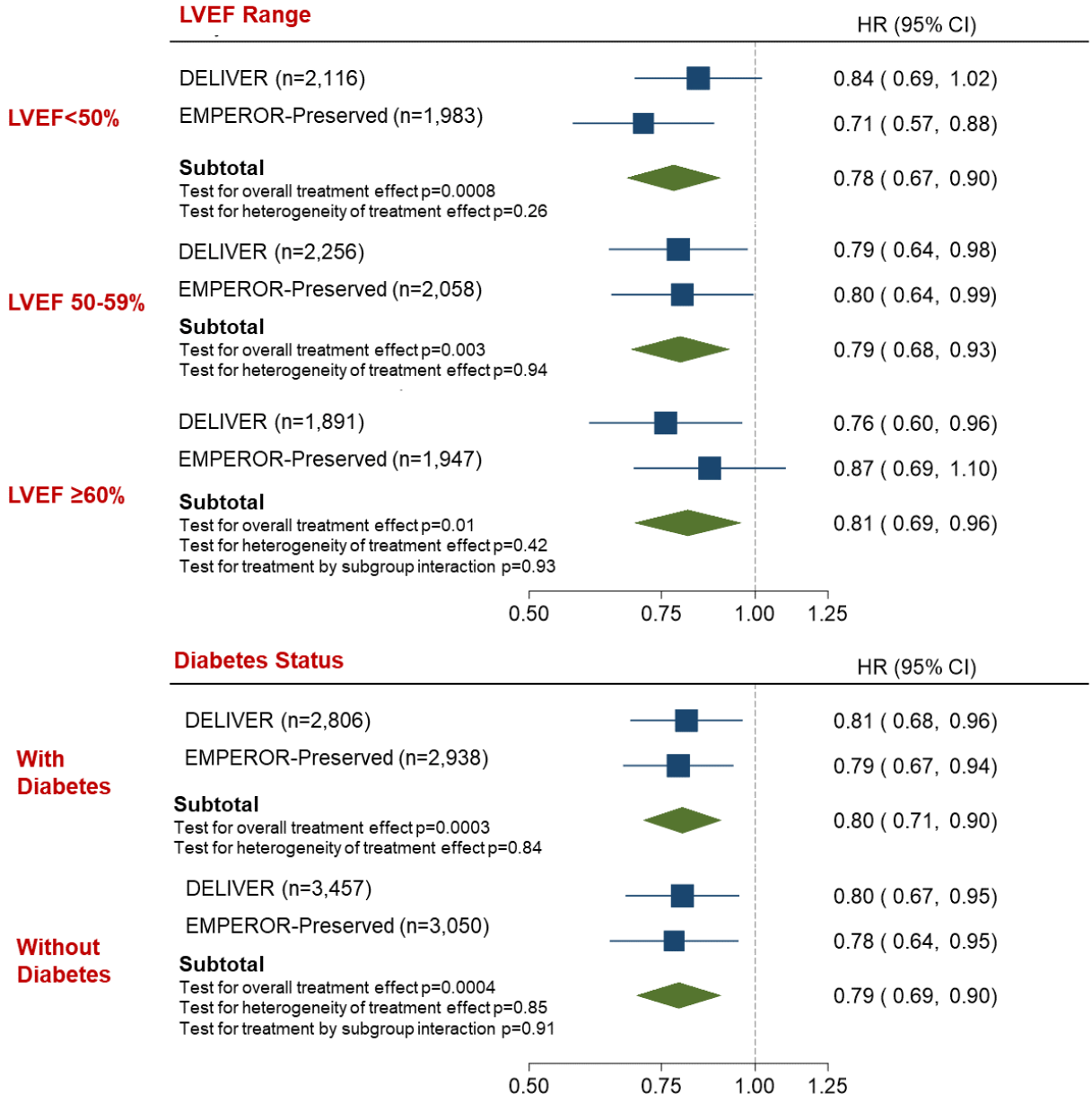
Supplemental Figure 1. PRISMA Flow Diagram for Study Selection

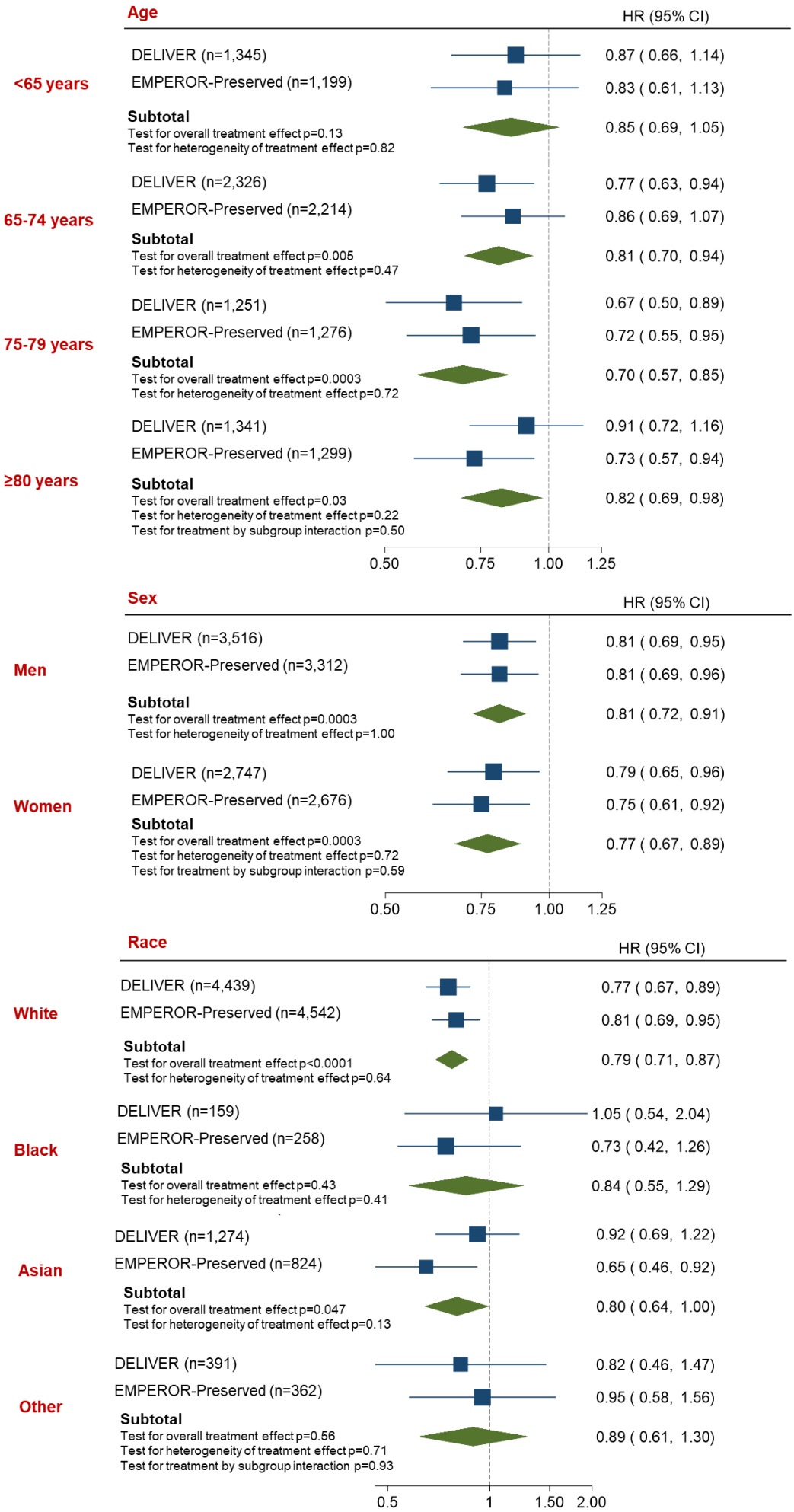


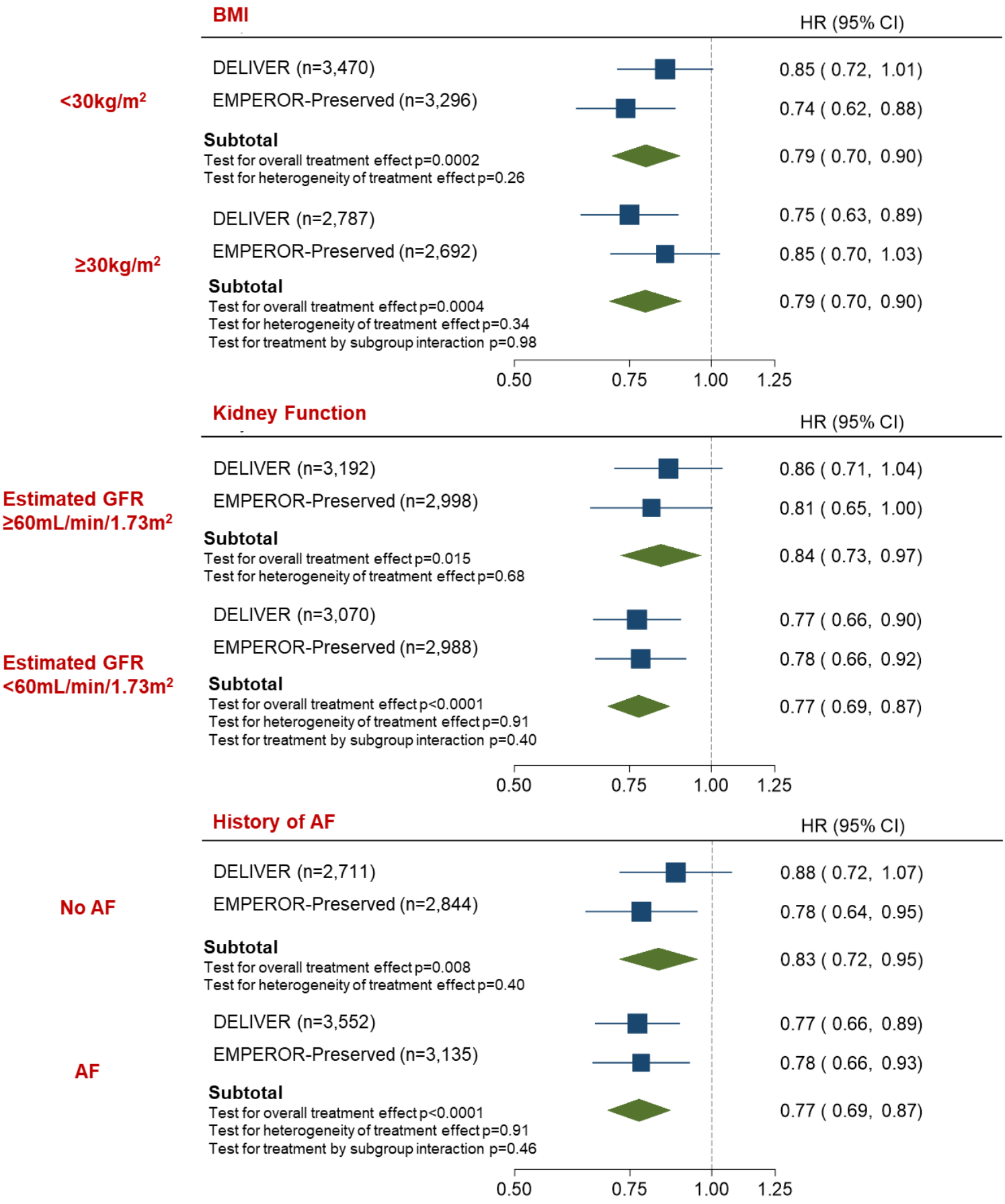
Supplemental Figure 2. Sensitivity Analyses Applying Alternative Definitions in Meta-Analysis

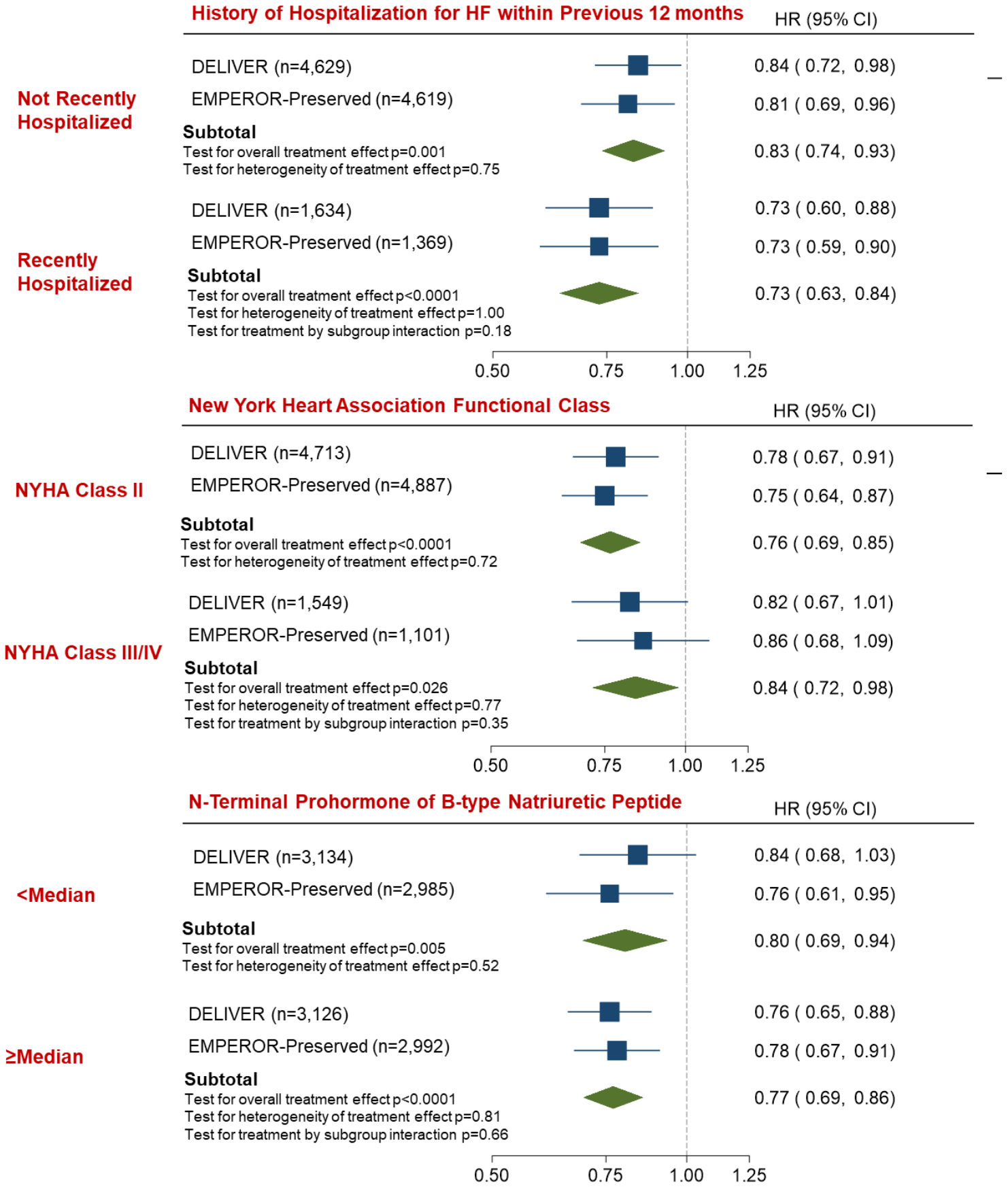


Supplemental Figure 3. Treatment effects of empagliflozin and dapagliflozin on the composite of cardiovascular death or first hospitalization for heart failure across subgroups of interest



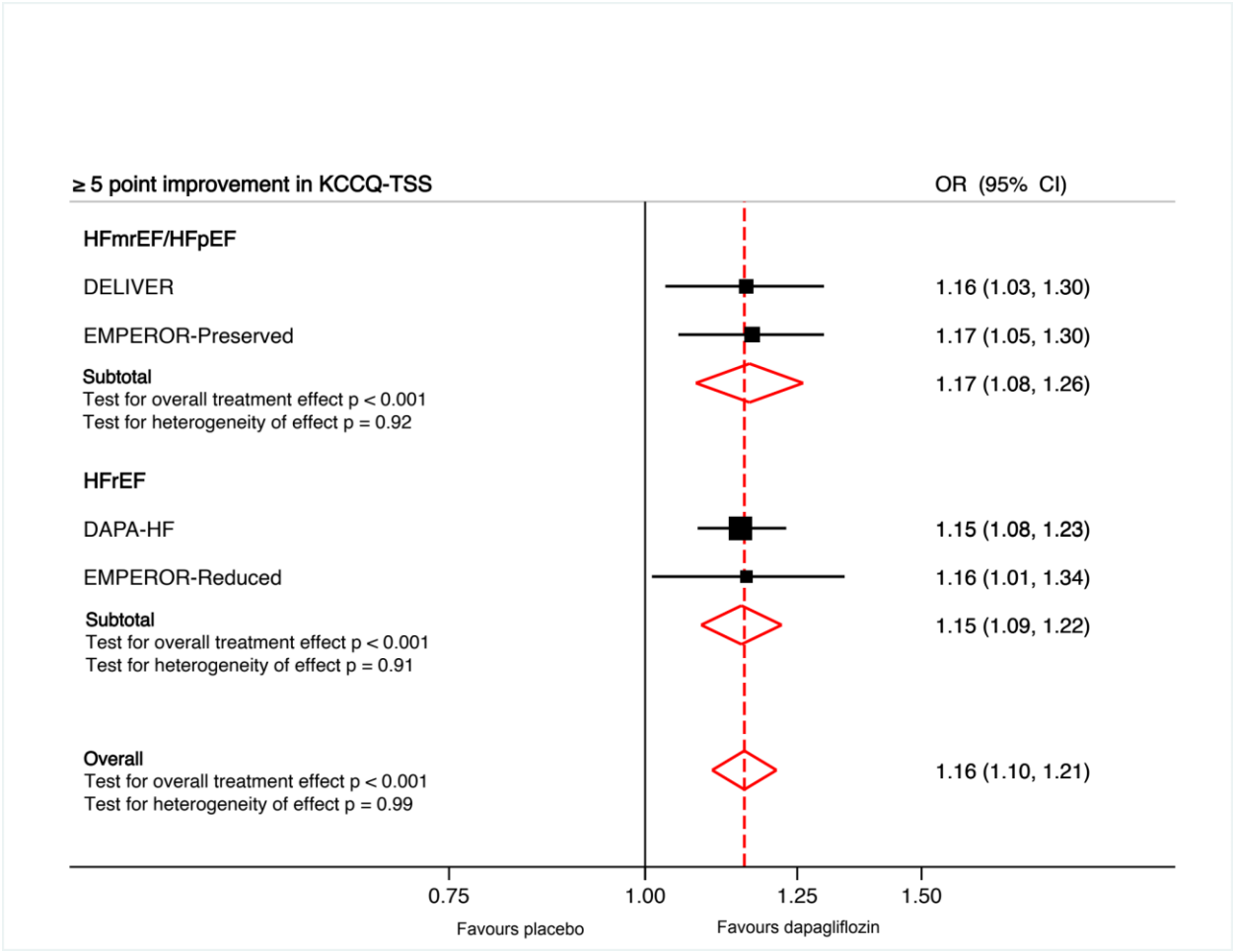

Age
HR (95% CI)
<65 years
DELIVER (n=1,345)
0.87 (0.66, 1.14)
EMPEROR-Preserved (n=1,199)
0.83 (0.61, 1.13)
Subtotal
Test for overall treatment effect p=0.13
Test for heterogeneity of treatment effect p=0.82
0.85 (0.69, 1.05)
65-74 years
DELIVER (n=2,326)
0.77 (0.63, 0.94)
EMPEROR-Preserved (n=2,214)
0.86 (0.69, 1.07)
Subtotal
Test for overall treatment effect p=0.005
Test for heterogeneity of treatment effect p=0.47
0.81 (0.70, 0.94)
75-79 years
DELIVER (n=1,251)
0.67 (0.50, 0.89)
EMPEROR-Preserved (n=1,276)
0.72 (0.55, 0.95)
Subtotal
Test for overall treatment effect p=0.0003
Test for heterogeneity of treatment effect p=0.72
0.70 (0.57, 0.85)
≥80 years
DELIVER (n=1,341)
0.91 (0.72, 1.16)
EMPEROR-Preserved (n=1,299)
0.73 (0.57, 0.94)
Subtotal
Test for overall treatment effect p=0.03
Test for heterogeneity of treatment effect p=0.22
Test for treatment by subgroup interaction p=0.50
0.82 (0.69, 0.98)
0.50
0.75
1.00
1.25
Sex
HR (95% CI)
Men
DELIVER (n=3,516)
0.81 (0.69, 0.95)
EMPEROR-Preserved (n=3,312)
0.81 (0.69, 0.96)
Subtotal
Test for overall treatment effect p=0.0003
Test for heterogeneity of treatment effect p=1.00
0.81 (0.72, 0.91)
Women
DELIVER (n=2,747)
0.79 (0.65, 0.96)
EMPEROR-Preserved (n=2,676)
0.75 (0.61, 0.92)
Subtotal
Test for overall treatment effect p=0.0003
Test for heterogeneity of treatment effect p=0.72
Test for treatment by subgroup interaction p=0.59
0.77 (0.67, 0.89)
0.50
0.75
1.00
1.25
Race
HR (95% CI)
White
DELIVER (n=4,439)
0.77 (0.67, 0.89)
EMPEROR-Preserved (n=4,542)
0.81 (0.69, 0.95)
Subtotal
Test for overall treatment effect p<0.0001
Test for heterogeneity of treatment effect p=0.64
0.79 (0.71, 0.87)
Black
DELIVER (n=159)
1.05 (0.54, 2.04)
EMPEROR-Preserved (n=258)
0.73 (0.42, 1.26)
Subtotal
Test for overall treatment effect p=0.43
Test for heterogeneity of treatment effect p=0.41
0.84 (0.55, 1.29)
Asian
DELIVER (n=1,274)
0.92 (0.69, 1.22)
EMPEROR-Preserved (n=824)
0.65 (0.46, 0.92)
Subtotal
Test for overall treatment effect p=0.047
Test for heterogeneity of treatment effect p=0.13
0.80 (0.64, 1.00)
Other
DELIVER (n=391)
0.82 (0.46, 1.47)
EMPEROR-Preserved (n=362)
0.95 (0.58, 1.56)
Subtotal
Test for overall treatment effect p=0.56
Test for heterogeneity of treatment effect p=0.71
Test for treatment by subgroup interaction p=0.93
0.89 (0.61, 1.30)
0.5
1
1.50
2.00

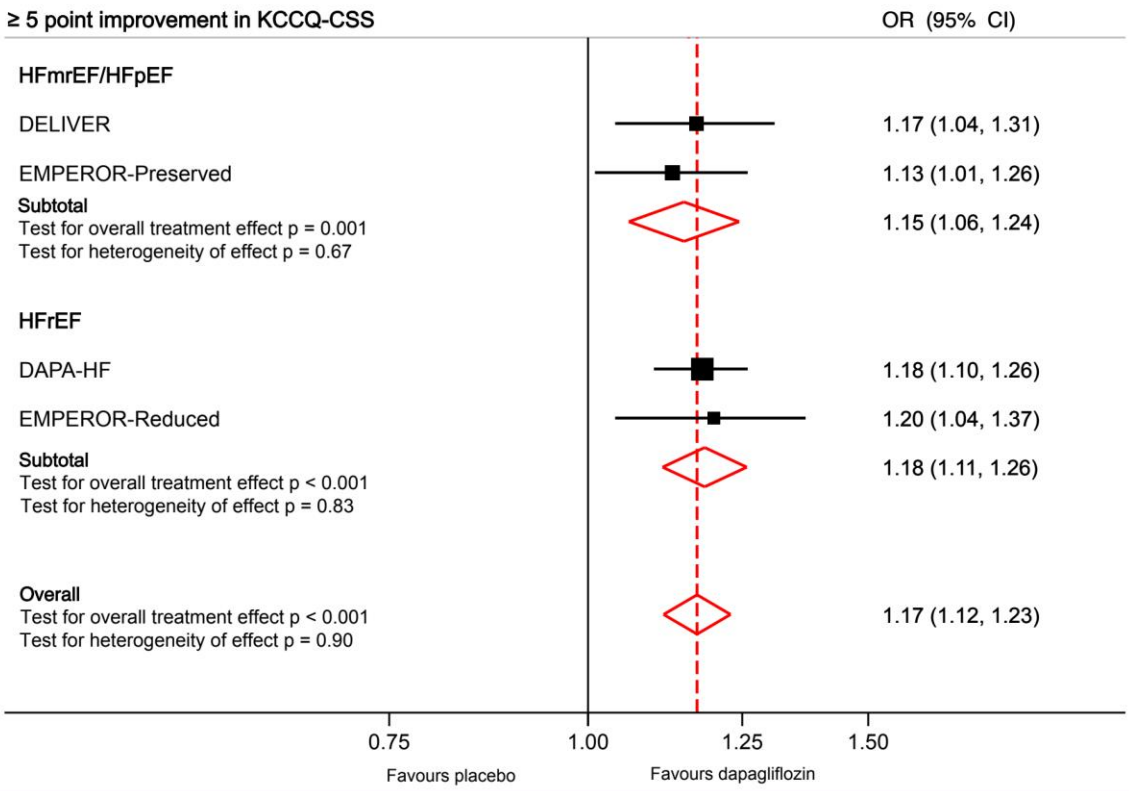

BMI
HR (95% CI)
<30kg/m2
DELIVER (n=3,470)
0.85 (0.72, 1.01)
EMPEROR-Preserved (n=3,296)
0.74 (0.62, 0.88)
Subtotal
Test for overall treatment effect p=0.0002
Test for heterogeneity of treatment effect p=0.26
0.79 (0.70, 0.90)
≥30kg/m2
DELIVER (n=2,787)
0.75 (0.63, 0.89)
EMPEROR-Preserved (n=2,692)
0.85 (0.70, 1.03)
Subtotal
Test for overall treatment effect p=0.0004
Test for heterogeneity of treatment effect p=0.34
Test for treatment by subgroup interaction p=0.98
0.79 (0.70, 0.90)
0.50
0.75
1.00
1.25
Kidney Function
HR (95% CI)
Estimated GFR ≥60mL/min/1.73m2
DELIVER (n=3,192)
0.86 (0.71, 1.04)
EMPEROR-Preserved (n=2,998)
0.81 (0.65, 1.00)
Subtotal
Test for overall treatment effect p=0.015
Test for heterogeneity of treatment effect p=0.68
0.84 (0.73, 0.97)
Estimated GFR <60mL/min/1.73m2
DELIVER (n=3,070)
0.77 (0.66, 0.90)
EMPEROR-Preserved (n=2,988)
0.78 (0.66, 0.92)
Subtotal
Test for overall treatment effect p<0.0001
Test for heterogeneity of treatment effect p=0.91
Test for treatment by subgroup interaction p=0.40
0.77 (0.69, 0.87)
0.50
0.75
1.00
1.25
History of AF
HR (95% CI)
No AF
DELIVER (n=2,711)
0.88 (0.72, 1.07)
EMPEROR-Preserved (n=2,844)
0.78 (0.64, 0.95)
Subtotal
Test for overall treatment effect p=0.008
Test for heterogeneity of treatment effect p=0.40
0.83 (0.72, 0.95)
AF
DELIVER (n=3,552)
0.77 (0.66, 0.89)
EMPEROR-Preserved (n=3,135)
0.78 (0.66, 0.93)
Subtotal
Test for overall treatment effect p<0.0001
Test for heterogeneity of treatment effect p=0.91
Test for treatment by subgroup interaction p=0.46
0.77 (0.69, 0.87)
0.50
0.75
1.00
1.25

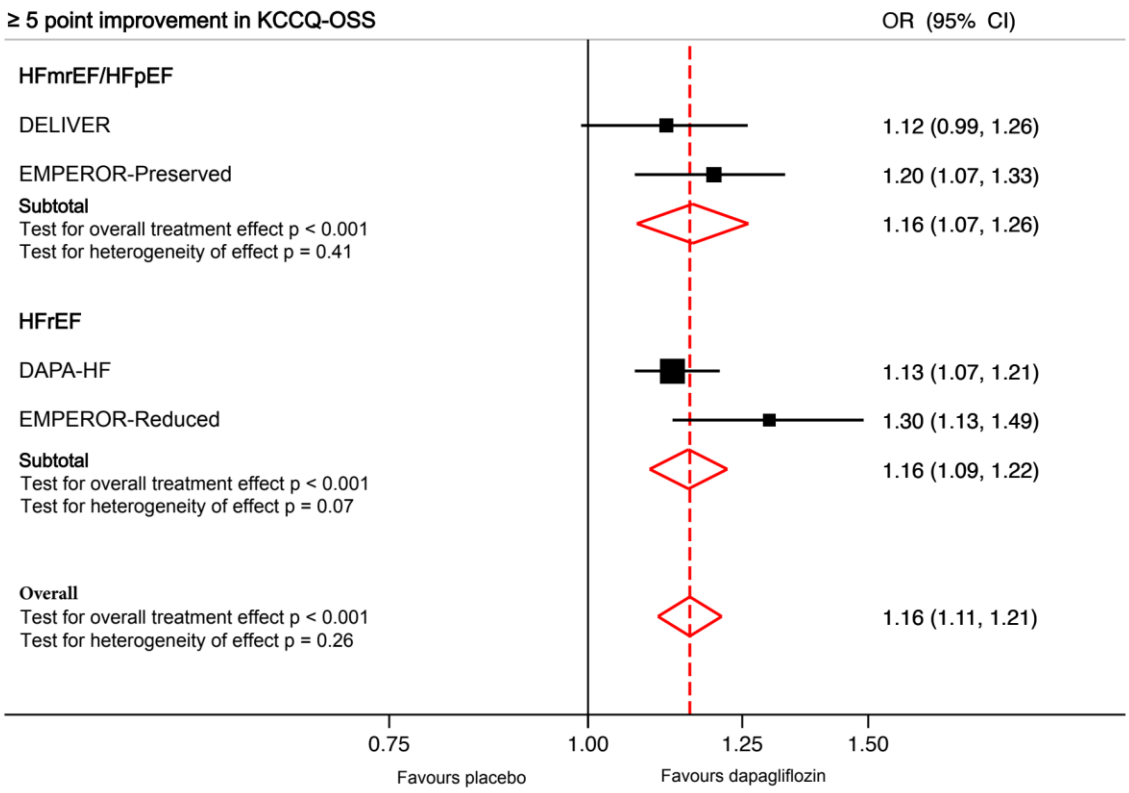

History of Hospitalization for HF within Previous 12 months
HR (95% CI)
Not Recently Hospitalized
DELIVER (n=4,629)
0.84 (0.72, 0.98)
EMPEROR-Preserved (n=4,619)
0.81 (0.69, 0.96)
Subtotal
Test for overall treatment effect p=0.001
Test for heterogeneity of treatment effect p=0.75
0.83 (0.74, 0.93)
Recently Hospitalized
DELIVER (n=1,634)
0.73 (0.60, 0.88)
EMPEROR-Preserved (n=1,369)
0.73 (0.59, 0.90)
Subtotal
Test for overall treatment effect p<0.0001
Test for heterogeneity of treatment effect p=1.00
Test for treatment by subgroup interaction p=0.18
0.73 (0.63, 0.84)
0.50
0.75
1.00
1.25
New York Heart Association Functional Class
HR (95% CI)
NYHA Class II
DELIVER (n=4,713)
0.78 (0.67, 0.91)
EMPEROR-Preserved (n=4,887)
0.75 (0.64, 0.87)
Subtotal
Test for overall treatment effect p<0.0001
Test for heterogeneity of treatment effect p=0.72
0.76 (0.69, 0.85)
NYHA Class III/IV
DELIVER (n=1,549)
0.82 (0.67, 1.01)
EMPEROR-Preserved (n=1,101)
0.86 (0.68, 1.09)
Subtotal
Test for overall treatment effect p=0.026
Test for heterogeneity of treatment effect p=0.77
Test for treatment by subgroup interaction p=0.35
0.84 (0.72, 0.98)
0.50
0.75
1.00
1.25
N-Terminal Prohormone of B-type Natriuretic Peptide
HR (95% CI)
<Median
DELIVER (n=3,134)
0.84 (0.68, 1.03)
EMPEROR-Preserved (n=2,985)
0.76 (0.61, 0.95)
Subtotal
Test for overall treatment effect p=0.005
Test for heterogeneity of treatment effect p=0.52
0.80 (0.69, 0.94)
≥Median
DELIVER (n=3,126)
0.76 (0.65, 0.88)
EMPEROR-Preserved (n=2,992)
0.78 (0.67, 0.91)
Subtotal
Test for overall treatment effect p<0.0001
Test for heterogeneity of treatment effect p=0.81
Test for treatment by subgroup interaction p=0.66
0.77 (0.69, 0.86)
0.50
0.75
1.00
1.25

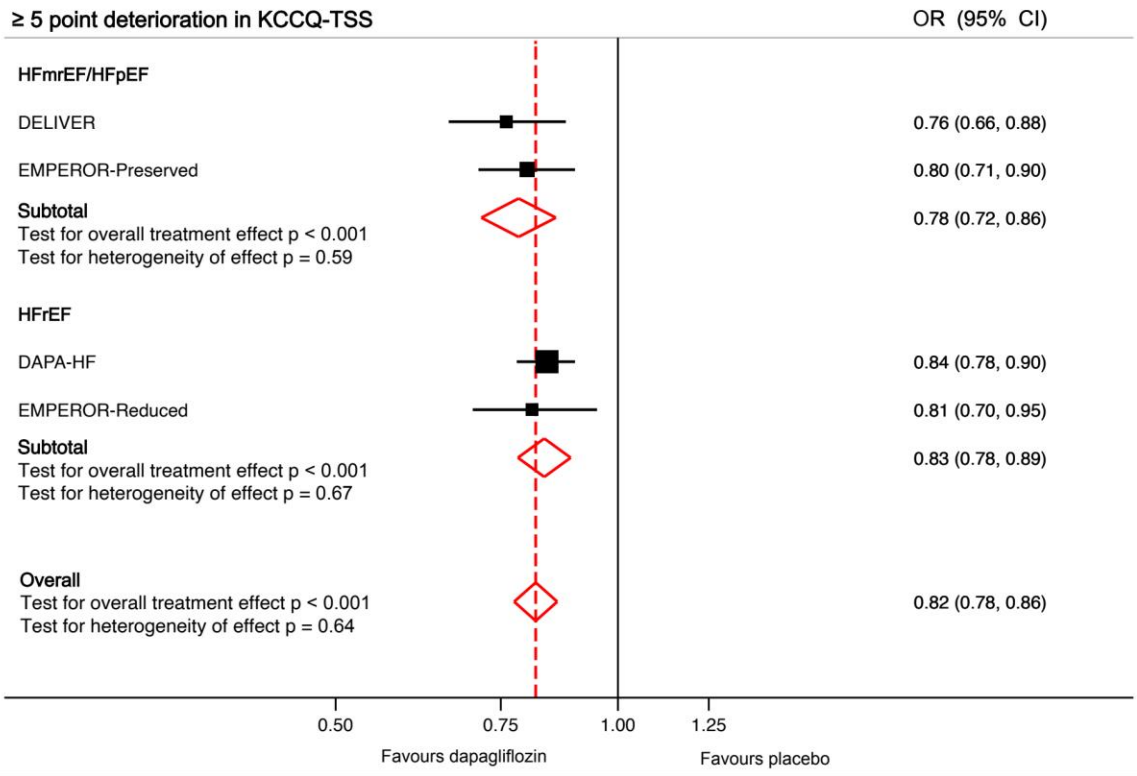
Supplemental Figure 4. Responder analysis for clinically important improvement or deterioration in health status based on the Kansas City Cardiomyopathy Questionnaire (KCCQ)

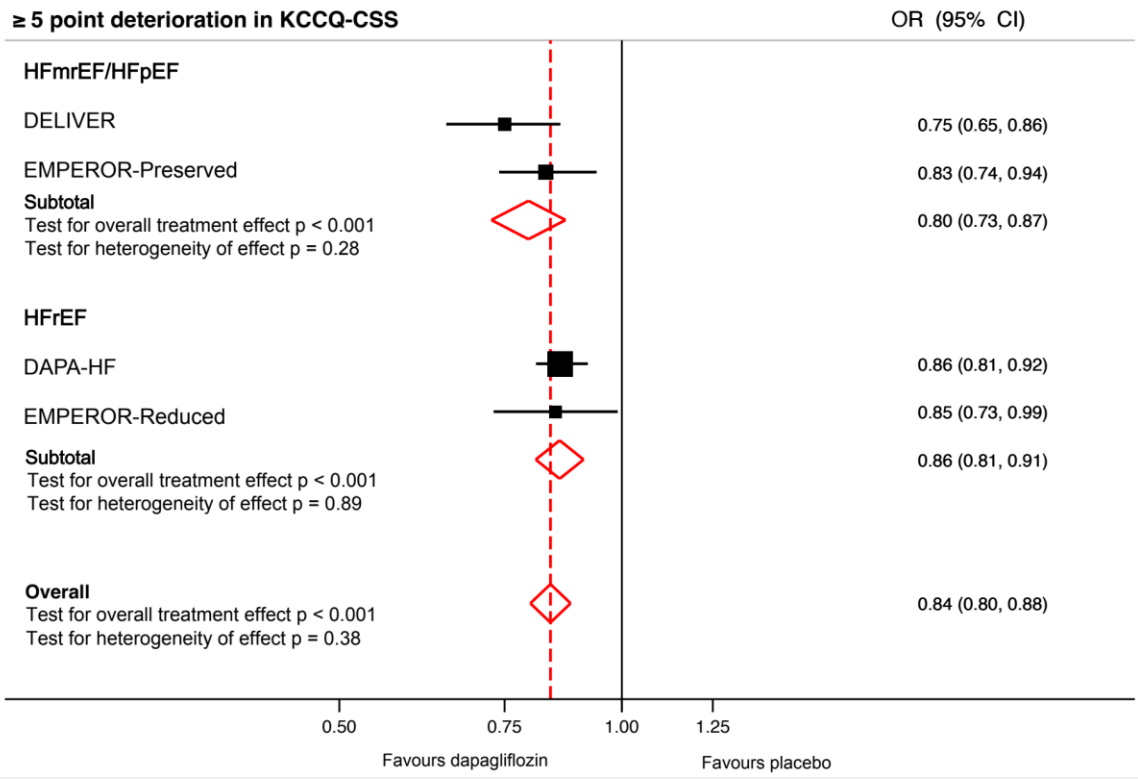
CI = confidence interval; CSS = clinical summary score; OR = odds ratio; OSS = overall summary score; TSS = total symptom score

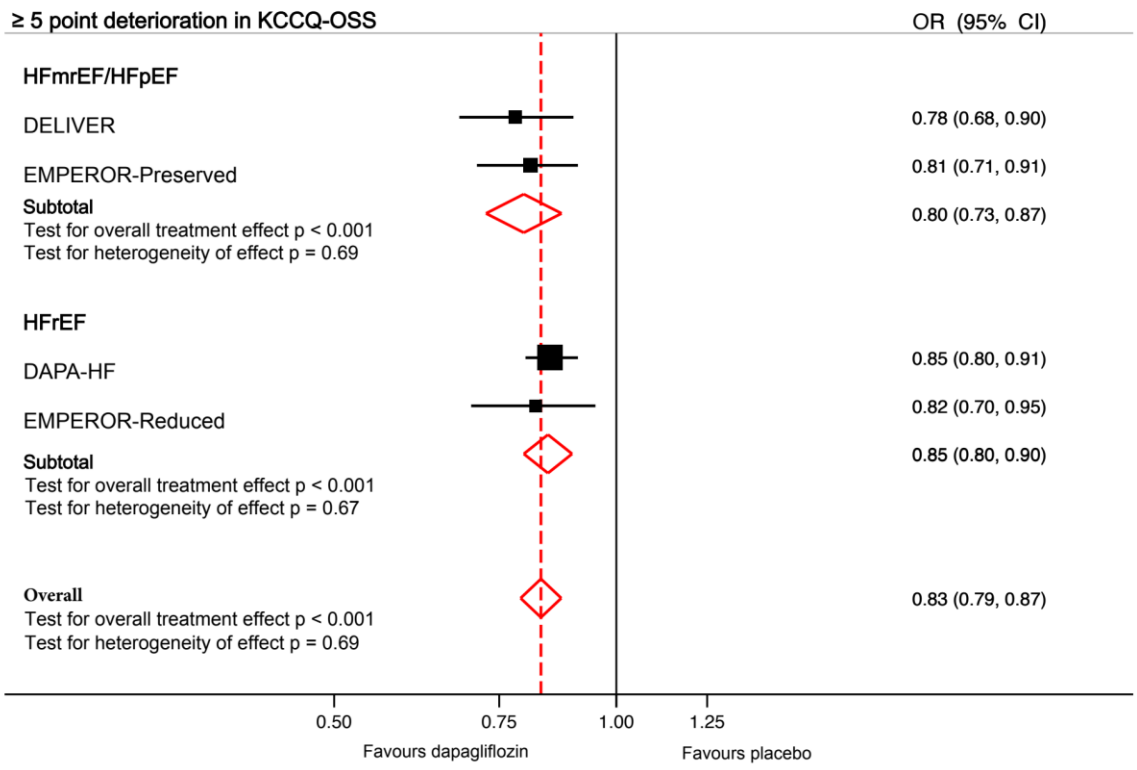



≥ 5 point improvement in KCCQ-CSS
OR (95% CI)
HFmrEF/HFpEF
DELIVER
1.17 (1.04, 1.31)
EMPEROR-Preserved
1.13 (1.01, 1.26)
Subtotal
Test for overall treatment effect p = 0.001
Test for heterogeneity of effect p = 0.67
1.15 (1.06, 1.24)
HFrEF
DAPA-HF
1.18 (1.10, 1.26)
EMPEROR-Reduced
1.20 (1.04, 1.37)
Subtotal
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.83
1.18 (1.11, 1.26)
Overall
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.90
1.17 (1.12, 1.23)
0.75
1.00
1.25
1.50
Favours placebo
Favours dapagliflozin


≥ 5 point improvement in KCCQ-OSS
OR (95% CI)
HFmrEF/HFpEF
DELIVER
1.12 (0.99, 1.26)
EMPEROR-Preserved
1.20 (1.07, 1.33)
Subtotal
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.41
1.16 (1.07, 1.26)
HFrEF
DAPA-HF
1.13 (1.07, 1.21)
EMPEROR-Reduced
1.30 (1.13, 1.49)
Subtotal
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.07
1.16 (1.09, 1.22)
Overall
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.26
1.16 (1.11, 1.21)
0.75
1.00
1.25
1.50
Favours placebo
Favours dapagliflozin


≥ 5 point deterioration in KCCQ-TSS
OR (95% CI)
HFmrEF/HFpEF
DELIVER
0.76 (0.66, 0.88)
EMPEROR-Preserved
0.80 (0.71, 0.90)
Subtotal
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.59
0.78 (0.72, 0.86)
HFrEF
DAPA-HF
0.84 (0.78, 0.90)
EMPEROR-Reduced
0.81 (0.70, 0.95)
Subtotal
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.67
0.83 (0.78, 0.89)
Overall
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.64
0.82 (0.78, 0.86)
0.50
0.75
1.00
1.25
Favours dapagliflozin
Favours placebo


≥ 5 point deterioration in KCCQ-CSS
OR (95% CI)
HFmrEF/HFpEF
DELIVER
0.75 (0.65, 0.86)
EMPEROR-Preserved
0.83 (0.74, 0.94)
Subtotal
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.28
0.80 (0.73, 0.87)
HFrEF
DAPA-HF
0.86 (0.81, 0.92)
EMPEROR-Reduced
0.85 (0.73, 0.99)
Subtotal
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.89
0.86 (0.81, 0.91)
Overall
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.38
0.84 (0.80, 0.88)
0.50
0.75
1.00
1.25
Favours dapagliflozin
Favours placebo


≥ 5 point deterioration in KCCQ-OSS
OR (95% CI)
HFmrEF/HFpEF
DELIVER
0.78 (0.68, 0.90)
EMPEROR-Preserved
0.81 (0.71, 0.91)
Subtotal
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.69
0.80 (0.73, 0.87)
HFrEF
DAPA-HF
0.85 (0.80, 0.91)
EMPEROR-Reduced
0.82 (0.70, 0.95)
Subtotal
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.67
0.85 (0.80, 0.90)
Overall
Test for overall treatment effect p < 0.001
Test for heterogeneity of effect p = 0.69
0.83 (0.79, 0.87)
0.50
0.75
1.00
1.25
Favours dapagliflozin
Favours placebo