THE LANCET

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Butler J, Shah SJ, Petrie MC, et al. Semaglutide versus placebo in people with obesity-related heart failure with preserved ejection fraction: a pooled analysis of the STEP-HFpEF and STEP-HFpEF DM randomised trials. *Lancet* 2024; published online April 7. https://doi.org/10.1016/S0140-6736(24)00469-0.

Supplementary Appendix

CONTENTS

SUPPLEMENTAL METHODS 2

Description of the hypothetical trial product estimand 2

Additional information on the statistical analyses and imputation methods to account for missing data 2

Full description of supportive secondary and exploratory endpoints 3

TABLES 5

Table S1. Full eligibility criteria 5

Table S2. List of all endpoints including the dual primary, confirmatory secondary, supportive secondary, and exploratory endpoints 9

Table S3. Dual primary and confirmatory secondary efficacy endpoints (trial product estimand) 12

Table S4. Baseline characteristics of the total population for the STEP-HFpEF and STEP-HFpEF DM trial participants 14

Table S5. Treatment effects of semaglutide versus placebo on change in KCCQ-CSS, change in body weight, change in 6-minute walking distance, the hierarchical composite endpoint, and CRP ratio to baseline by trial (STEP-HFpEF and STEP-HFpEF DM) 16

Table S6. Number of Participants With Missing Data at Week 52 for the Dual Primary and Confirmatory Secondary Endpoints 18

FIGURES 19

REFERENCES 23

SUPPLEMENTAL METHODS

Description of the hypothetical trial product estimand

In line with the hypothetical trial product estimand strategy (secondary estimand strategy) as described in the Statistical Analysis Plan Section 1.1.1, the trial product estimand for the primary endpoints was based on all randomized participants (full analysis set) and the "on-treatment without other anti-obesity therapies" observation period (see Statistical Analysis Plan Section 4). In the trial product estimand, observations after a heart failure event were used if collected. The trial product estimand for the primary endpoints addressed the efficacy of semaglutide 2.4 mg and was assessed using a mixed model for repeated measurements (MMRM). Week 52 assessments from retrieved subjects were not used in this analysis. The MMRM used assessments only from participants who were taking the randomized treatment until the end of treatment or until the first discontinuation of randomized treatment. The derived date of the second consecutive missed dose was used as the latest date for using assessments in the MMRM. The assessment closest in time and before the derived date of the second consecutive missed dose was used as last assessment on randomized treatment. For participants who initiated other anti-obesity therapies, as defined in the Statistical Analysis Plan Section 4, before completion or first discontinuation of randomized treatment, the date of starting the other anti-obesity therapies was used as latest date for using assessments in this MMRM. Similarly, the assessment closest in time and before the date of starting other anti-obesity therapies was used as the last assessment on randomized treatment.

Additional information on the statistical analyses and imputation methods to account for missing data

The primary analysis was a multiple imputation similar to the one described by McEvoy et al.[1, 2] Furthermore, a single imputation approach using an unfavorable value was employed

for the Kansas City Cardiomyopathy Questionnaire Clinical Summary Score (KCCQ-CSS) when participants had missing measurements at week 52 due to cardiovascular death, or for those participants with a heart failure event prior to a missing measurement at week 52 (non-retrieved measurements). The unfavorable value was determined using the minimal value observed during the trial. For participants in the semaglutide and placebo groups, missing primary endpoint measurements at week 52 for non-retrieved participants were imputed using assessments from retrieved participants in each treatment group. This was done according to the timing and the actual value of last available observation during the on-treatment period for KCCQ-CSS and body weight. Furthermore, baseline body mass index category, baseline body weight, baseline KCCQ-CSS (not for change in body weight), and sex were used in the imputation model.

The statistical model and imputation approach for confirmatory secondary endpoints of change in C-reactive protein (CRP) (log transformed) and 6-minute walk distance (6MWD) were the same as for the primary endpoints, using the imputation approach for change in body weight and change in KCCQ-CSS, respectively. Similar baseline variables were used as above with the baseline endpoint variable instead of baseline KCCQ-CSS.

Full description of supportive secondary and exploratory endpoints

Supportive secondary endpoints included change in systolic blood pressure, waist circumference, and KCCQ Overall Summary Score (OSS) from baseline (randomization) to 52 weeks; several thresholds of change in body weight and KCCQ-CSS; and thresholds for improvement in KCCQ-CSS and 6-minute walk distance based on the patient global impression of severity (PGI-S) as an anchor. The PGI-S for KCCQ-CSS asked patients to rate their symptoms of heart failure in the last 2 weeks using a 4-point ordinal scale (no symptoms, mild, moderate, severe). The PGI-S for 6-minute walk distance asked subjects to rate any difficulty they were currently experiencing in walking quickly using a 5-point ordinal

scale (not at all difficult, a little difficult, moderately difficult, very difficult, unable to walk quickly). The derivation of the anchor-based threshold for KCCQ-CSS and 6-minute walk distance are given in footnotes to Table 2 in the main manuscript. For the anchor-based threshold for 6-minute walk distance, the patient global impression of change was also used as a sensitivity analysis of anchor-based assessment of thresholds. An estimated 45·3% of participants in the semaglutide group and 32·1% in the placebo group improved at least 23·6 m, based on the anchor using the patient global impression of change from baseline to week 52 in perception of ability to walk quickly (category “moderately better”). The mean change in 6-min walk distance was calculated (using pooled data across treatment groups) in the participants in the category “moderately better” (n=182) to establish the threshold. The odds ratio for the sensitivity analysis was 1·8 (95% CI 1·4–2·3).

Exploratory endpoints included a change in N-terminal pro-brain natriuretic peptide (NTproBNP) from screening (week –2) to week 52, achievement of ≥15-point and ≥20-point improvement in KCCQ-CSS, and time to the first adjudicated event of heart failure hospitalization or an urgent visit requiring intravenous therapy (Table S2).

TABLES

Table S1. Full eligibility criteria

Inclusion Criteria

- Informed consent obtained before any trial-related activities. Trial-related activities are any procedures that are carried out as part of the trial, including activities to determine suitability for the trial
- Male or female, age ≥18 years at the time of signing informed consent
- BMI ≥30.0 kg/m^2
- NYHA class II–IV
- LVEF ≥45% at screening
- No hospitalizations due to HF between screening (visit 1) and randomization (visit 2)
- Able to perform the 6MWT at screening with a minimum distance of 100 meters
- KCCQ Clinical Summary Score <90 at screening
- At least one of the following:
    - Mean PWP ≥15 mmHg or LVEDP ≥15 mmHg documented during catheterization at rest *or* PA diastolic pressure measured by implantable monitor ≥15 mmHg *or* PWP or LVEDP ≥25 mmHg documented during catheterization at exercise
    - If BMI <35.0: NT-proBNP ≥220 pg/mL (for patients with sinus rhythm) or NTproBNP ≥660 pg/mL (for patients with persistent/permanent atrial fibrillation); if BMI ≥35.0: NT-proBNP ≥125 pg/mL (for patients with sinus rhythm) or NTproBNP ≥375 pg/mL (for patients with persistent/permanent atrial fibrillation) at screening (NT-proBNP analysed by the central laboratory) in combination with at least one of the following (documented by echocardiography within 12 months prior to or at screening):
        - (i) septal é <7 cm/sec or lateral é <10 cm/sec or average E/é ≥15;
        - (ii) PA systolic pressure >35 mmHg,

        - (iii) LA enlargement, (width ≥3.8 cm or length ≥5.0 cm or area ≥20.0 cm^2 or volume ≥55 mL or volume index ≥29 mL/m^2)
        - (iv) LV hypertrophy with septal thickness or posterior wall thickness ≥1.2 cm
    - Hospitalization with a primary diagnosis of decompensated HF requiring IV loop diuretic treatment within the previous 12 months combined with ≥2 of the following (documented by echocardiography within 12 months prior to or at screening):
        - (i) septal é <7 cm/sec or lateral é <10 cm/sec or average E/é ≥15;
        - (ii) PA systolic pressure >35 mmHg,
        - (iii) LA enlargement, (width ≥3.8 cm or length ≥5.0 cm or area ≥20.0 cm^2 or volume ≥55 mL or volume index ≥29 mL/m^2)
        - (iv) LV hypertrophy with septal thickness or posterior wall thickness ≥1.2 cm,
        - (v) ongoing use of diuretic therapy for ≥30 days before screening
- *For STEP HFpEF DM trial only*: Diagnosed with T2D ≥90 days prior to the day of screening
- *For STEP HFpEF DM trial only:* Subject treated with diet, exercise, and/or glucose-lowering treatment such as OADs (unchanged drug(s), dose and dosing frequency) or insulins (**unchanged regimen [basal, basal+bolus, premix combination] with stable total daily insulin dose as judged by the investigator**), according to local label in stable dosing for at least 30 days prior to screening
- *For STEP HFpEF DM trial:* HbA_{1c} of ≤10.0% as measured at the screening visit

Exclusion Criteria

- MI, stroke, hospitalization for HF, unstable angina pectoris, or TIA within 30 days prior to the day of screening
- SBP >160 mmHg at screening

- Planned coronary, carotid, or peripheral artery revascularization
    - Any other condition judged by the investigator to be the primary cause of dyspnea (such as heart failure due to restrictive cardiomyopathy or infiltrative conditions [e.g. amyloidosis], hypertrophic obstructive cardiomyopathy, primary pulmonary arterial hypertension, chronic obstructive pulmonary disease, right heart failure due to pulmonary disease, complex congenital heart disease, anaemia, or more than moderate heart valve disease)
- Bariatric surgery prior to screening or planned within the trial time course
- Self-reported changed in body weight >5 kg (11 Ibs) within 90 days before screening irrespective of medical records
- *For STEP HFpEF trial:* HbA_{1c} ≥6.5% (48 mmol/mol) based on latest available value from medical records, no older than 3 months or if unavailable at local measurement at screening
- *For STEP HFpEF trial:* History of T1D or T2D (history of gestational diabetes is allowed)
- *For STEP HFpEF DM trial:* History of T1D (history of gestational diabetes is allowed)
- Treatment with any GLP-1 RA within 90 days prior to the day of screening
    - Uncontrolled and potentially unstable diabetic retinopathy or maculopathy. Verified by a fundus examination performed within 90 days prior to screening or in the period between screening and randomization. Pharmacological pupil-dilation is a requirement unless using a digital fundus photography camera specified for non-dilated examination
- *For STEP HFpEF DM trial:* Recurrent severe hypoglycemic episodes within the last year as judged by the investigator
- *For STEP HFpEF DM trial:* Treatment with continuous subcutaneous insulin infusion
- Personal or first-degree relative(s) history of MEN2 or MTC

- Acute pancreatitis within the last 180 days prior to screening or history or presence of chronic pancreatitis
- End-stage renal disease or chronic or intermittent hemodialysis or peritoneal dialysis
- Presence or history of malignant neoplasm within 5 years prior to the day of screening. Basal and squamous cell cancer and any carcinoma in situ are allowed
- Known or suspected hypersensitivity to trial product(s) or related products
- Participation in any clinical trial of an approved or non-approved device for the treatment of HF, diabetes, or obesity within 30 days before screening
- Receipt of any investigational medicinal product within 30 days before screening
- Female who is pregnant, is breast-feeding, or intends to become pregnant or is of child-bearing potential and not using a highly effective contraceptive method
- Major surgery scheduled for the duration of the trial, affecting walking ability in the opinion of the investigator
- Any disorder, including severe psychiatric disorder, suicidal behavior within 90 days before screening, and suspected drug abuse, which in the investigator's opinion might jeopardize the subject's safety or compliance with the protocol

The criteria will be assessed at the investigator's discretion unless otherwise stated. Echocardiographic features must be documented within 12 months of screening.

6MWT denotes 6-minute walk test, AF atrial fibrillation, BMI body mass index, GLP-1 RA glucagon-like peptide-1 receptor agonist, HbA_{1c} glycated hemoglobin, HF heart failure, HFpEF heart failure with preserved ejection fraction, IV intravenous, KCCQ Kansas City Cardiomyopathy Questionnaire, LA left atrial, LV left ventricular, LVEDP left ventricular end diastolic pressure, LVEF left ventricular ejection fraction, MEN2 multiple endocrine neoplasia type 2, MI myocardial infarction, MTC medullary thyroid carcinoma, NTproBNP N-terminal pro-brain natriuretic peptide, NYHA New York Heart Association, OAD oral anti-diabetic, PA pulmonary artery, PWP pulmonary wedge pressure, SBP systolic blood pressure, T1D type 1 diabetes, T2D type 2 diabetes, and TIA transient ischemic attack.

Table S2. List of all endpoints including the dual primary, confirmatory secondary, supportive secondary, and exploratory endpoints

Dual Primary Endpoints

- Change in KCCQ Clinical Summary Score from baseline (week 0) to end of treatment (week 52)
- Change in body weight from baseline (week 0) to end of treatment (week 52)

Confirmatory Secondary Endpoints

- Change in 6MWD
- Hierarchical composite (assessed using the win ratio) of:
  - Time to all-cause death
  - Number of HF events requiring hospitalization or urgent HF visit
  - Time to first HF event requiring hospitalization or urgent HF visit
  - Difference at least 15 in KCCQ Clinical Summary Score change from baseline (week 0) to 52 weeks
  - Difference at least 10 in KCCQ Clinical Summary Score change from baseline (week 0) to 52 weeks
  - Difference at least 5 in KCCQ Clinical Summary Score change from baseline (week 0) to 52 weeks
  - Difference at least 30 meters in 6MWD change from baseline (week 0) to 52 weeks
- Change in C-reactive protein from baseline (week -2) to 52 weeks

Supportive Secondary Endpoints

- Subjects achieving ≥10%, ≥15%, or ≥20% weight loss
- Subjects improving by ≥5 or ≥10 points in KCCQ Clinical Summary Score
- Change in KCCQ Overall Summary Score

- Subject achieving threshold for clinically meaningful within-subject change in KCCQ Clinical Summary Score
- Subject achieving threshold for clinically meaningful within-subject change in 6MWD
- Change in SBP
- Change in waist circumference
- *STEP-HFpEF DM only:* Change in HbA_{1c}
- Number of treatment emergent severe or clinically significant hypoglycemia episodes

Exploratory Endpoints

- Change in antihypertensive medication
- Change in loop diuretic medication
- Change in NTproBNP
- Subject improving by ≥15 points in KCCQ Clinical Summary Score*
- Change in EQ-5D-5L score
- Subject worsening by ≥5, ≥10, or ≥15* points in KCCQ Clinical Summary Score
- Subject improving by ≥5, ≥10, or ≥15* points in KCCQ Overall Summary Score
- Subject worsening by ≥5, ≥10, or ≥15* points in KCCQ Overall Summary Score
- Change in subscales of KCCQ (total symptom score, physical limitations score, social limitations score, and health-related quality of life)
- Subject experiencing improvement in NYHA class
- Subject experiencing deterioration in NYHA class
- Time to first HF event (hospitalization or urgent visit)
- Subject achieving HbA_{1c} below 7.0%
- Subject achieving HbA_{1c} below 6.5%

Echocardiographic sub-study†

- Change in left atrial volume
- Change in left ventricular filling pressure (diastolic function) (E/e´)
- Change in global longitudinal strain

*Identified as an exploratory endpoint only in the Statistical Analysis Plan. †End points for the echocardiographic sub-study are not reported in this manuscript.

6MWD denotes 6-minute walk distance, EQ-5D-5L European Quality of Life 5 Dimensions 5 Level, HbA_{1c} glycated hemoglobin, HF heart failure, HFpEF heart failure with preserved ejection fraction, KCCQ Kansas City Cardiomyopathy Questionnaire, NTproBNP N-terminal pro-brain natriuretic peptide, NYHA New York Heart Association, and SBP systolic blood pressure.

Table S3. Dual primary and confirmatory secondary efficacy endpoints (trial product estimand)

Endpoint	**Semaglutide 2.4 mg** (N=573)	**Placebo** (N=572)	**Estimated difference between semaglutide and placebo (95% CI)**	**Ratio (95% CI)***	**P value**
Dual primary endpoints					
Change in KCCQ-CSS from baseline to week 52, points	17.7	9.0	8.7 (6.6, 10.8)	—	<0·0001
Percent body weight change from baseline to week 52, %	–12.9	–2.8	–10.1 (–10.9, –9.3)	—	<0·0001
Confirmatory secondary endpoints					
Change from baseline to week 52					
6MWD, meters	25.0	5.7	19.3 (12.2, 26.4)	—	<0·0001
Hierarchical composite endpoint†	—	—	—	1.96 (1.68, 2.29)	<0·0001
CRP, ratio (week 52/baseline)	0.52	0.92	—	0.57 (0.51, 0.63)	<0·0001

Data are for the trial product estimand and full analysis set. The trial product estimand assessed treatment effect under the assumption that participants received their assigned treatment for the duration of the trial, without rescue intervention. Treatment effects were estimated using a MMRM with treatment (semaglutide and placebo) and BMI stratum as fixed factors and baseline endpoint value as a covariate nested within trial visits with an unstructured covariance matrix. Missing values were predicted from the MMRM.

Data showing proportions or events are observed data from the on-treatment period, which was the time from randomization to last contact with a trial site, regardless of treatment discontinuation or rescue intervention.

*Odds ratio, unless otherwise stated. For supportive secondary and exploratory endpoints, the widths of confidence intervals have not been adjusted for multiplicity and should not be used to infer treatment effects.

†The hierarchical endpoint is a composite of death, number of heart failure events, time to first heart failure event (from baseline to week 57), a difference of at least 15, 10, and 5 points in KCCQ-CSS from baseline to week 52 and a difference of at least 30 meters in 6MWD from baseline to week 52 using the on-treatment period. Missing value was predicted using the MMRM model from above for KCCQ-CSS and 6MWD. This was assessed using a win ratio approach. All patients randomized to semaglutide 2.4 mg were compared with all patients randomized to placebo within each stratum of BMI (<30 kg/m^2 vs ≥30 kg/m^2).

6MWD, 6-minute walk distance; BMI, body mass index; CI, confidence interval; CRP, C-reactive protein; KCCQ-CSS, Kansas City Cardiomyopathy Questionnaire Clinical Summary Score; MMRM, mixed model for repeated measures.

Table S4. Baseline characteristics of the total population for the STEP-HFpEF and STEP-HFpEF DM trial participants

Baseline characteristic	STEP HFpEF (N=529)	STEP HFpEF DM (N=616)*
Female, n (%)	297 (56.1)	273 (44.3)
Age, years	69 (62.0–75.0)	69.0 (63.0–75.0)
Race, n (%)†		
White	507 (95.8)	519 (84.3)
Asian	N/A	76 (12.3)
Black/African American	21 (4.0)	18 (2.9)
Other race	1 (0.2)	3 (0.5)
Ethnicity, n (%)†		
Hispanic or Latino	36 (6.8)	76 (12.3)
Not Hispanic or Latino	493 (93.2)	540 (87.7)
Body weight, kg	105.1 (92.4–120.8)	102.7 (90.5–117.7)
Body mass index, kg/m^2‡	37.0 (33.7–41.4)	36.9 (33.5–41.3)
Body mass index stratification, n (%)		
Body mass index 30 to <35 kg/m^2	180 (34.0)	220 (35.7)
Body mass index ≥35 kg/m^2	349 (66.0)	396 (64.3)
Waist circumference, cm	119.4 (110.5–128.0)	120.3 (112.0–130.0)
Systolic blood pressure, mmHg	133 (121.0–144.0)	135.0 (125.0–144.0)
Diastolic blood pressure, mmHg	78.0 (71.0–85.0)	78.0 (70.0–84.0)
NT-proBNP, pg/mL	450.8 (218.2–1015.0)	493.0 (245.7–1018.6)
CRP, mg/L	3.8 (1.9–7.7)	3.3 (1.7–8.4)
LVEF, %	57.0 (50.0–60.0)	56.0 (50.0–60.0)
LVEF stratification, n (%)		
LVEF ≥45 to <50%§	85 (16.1)	106 (17.2)
LVEF ≥50 to <60%	215 (40.6)	259 (42.0)
LVEF ≥60%	229 (43.3)	251 (40.7)
KCCQ-CSS, score	58.9 (41.7–72.9)	59.4 (43.8–72.0)
6MWD, metres	320.0 (240.0–389.0)	280.0 (203.5–350.0)
HF hospitalisation within prior 1 year, n (%)	81 (15.3)	112 (18.2)
Comorbidities at screening, n (%)		
Atrial fibrillation	275 (52.0)	243 (39.4)
Hypertension	433 (81.9)	526 (85.4)
Coronary artery disease	98 (18.5)	148 (24.0)
Obstructive sleep apnoea	66 (12.5)	53 (8.6)
NYHA functional class, n (%)		

Class II	350 (66.2)	435 (70.6)
Class III	178 (33.6)	180 (29.2)
Class IV	1 (0.2)	1 (0.2)
Concomitant medications, n (%)		
s Beta-blockers	418 (79.0)	510 (82.8)
Diuretics	427 (80.7)	498 (80.8)
Loop diuretics	329 (62.2)	373 (60.6)
Mineralocorticoid receptor antagonists	184 (34.8)	200 (32.5)
Thiazides	90 (17.0)	85 (13.8)
ACE inhibitor/ARB	397 (75.0)	502 (81.5)
Angiotensin receptor-neprilysin inhibitors (ARNI)	27 (5.1)	31 (5.0)
Biguanides/metformin	9 (1.7)	443 (71.9)
Sulfonylureas	N/A	108 (9.4)
SGLT2 inhibitors	19 (3.6)	202 (32.8)
DPP-4 inhibitors	N/A	92 (8.0)
Insulins	N/A	128 (11.2)

Percentages may not equal 100% due to rounding. Data are median (Q1–Q3) unless otherwise stated and are from the full analysis set. ACE=angiotensin-converting enzyme. ARB=angiotensin II receptor blocker. ARNI= angiotensin receptor/neprilysin inhibitor. DPP-4=dipeptidyl peptidase 4. HF=heart failure. KCCQ-CSS=Kansas City Cardiomyopathy Questionnaire Clinical Summary Score. LVEF=left ventricular ejection fraction. N/A=not applicable. NT-proBNP=N-terminal pro B-type natriuretic peptide. NYHA=New York Heart Association. Q=quartile. SGLT2=sodium-glucose cotransporter-2.

*A total of 617 participants in STEP-HFpEF DM trial were randomised; however, one participant was randomised in error such that the full analysis set comprises 616 participants.

[†]Race and ethnic group were reported by the investigator.

[‡]Body mass index is the weight (kg) divided by the square of the height (m).

[§]Includes one participant with an LVEF of 33.

Table S5. Treatment effects of semaglutide versus placebo on change in KCCQ-CSS, change in body weight, change in 6-minute walking distance, the hierarchical composite endpoint, and CRP ratio to baseline by trial (STEP-HFpEF and STEP-HFpEF DM)

Subgroup	Semaglutide 2.4 mg		Placebo		ETD [95% CI]	Interaction p-value	I^2	Heterogeneity p-value
	n	Change from baseline to Week 52	n	Change from baseline to Week 52				
Change in KCCQ-CSS (points)								
STEP-HFpEF	243	16.4	237	8.5	7.9 [4.6, 11.1]	0.7911	0.0	0.8081
STEP-HFpEF DM	281	13.8	272	6.6	7.3 [4.2, 10.3]			
Change in body weight (kg)								
STEP-HFpEF	246	–11.9	237	–2.6	–10.9 [–11.9, –9.4]	<0.0001	95.7	<0.0001
STEP-HFpEF DM	286	–9.7	272	–3.4	–6.4 [–7.6, –5.2]			
Change in 6MWD (metres)								
STEP-HFpEF	240	24.1	225	3.7	20.4 [8.8, 32.0]	0.4429	0	0.4518
STEP-HFpEF DM	281	10.5	265	–3.8	14.2 [3.5, 25.0]			
Hierarchical composite endpoint‡								
STEP-HFpEF	—	—	—	—	1.72 [1.37, 2.15]	0.7334	0	0.5867
STEP-HFpEF DM	—	—	—	—	1.61 [1.31, 1.97]			
CRP ratio to baseline								
STEP-HFpEF	241	0.57*	243	0.94	0.61 [0.51, 0.73]†	0.4877	0	0.4896
STEP-HFpEF DM	286	0.58*	277	0.87	0.67 [0.56, 0.80]†			

Analyses are based on the treatment policy estimand, reflect the full analysis population, and are from the in-trial period. Responses at week 52 were analysed using ANCOVA and an imputation approach for missing data. *Ratio to baseline at week 52. †Estimated treatment ratio. The ratio to baseline and the corresponding baseline value were log-transformed prior to analysis. The approximate relative changes/differences were derived from estimated ratios by subtracting 1 and multiplying by 100. ‡The hierarchical endpoint is a composite of death from any cause, number of HF events, time to first HF event (from baseline to week 57) using the in-trial period, a difference of at least 15, 10, and 5 points in KCCQ-CSS change from baseline to week 52, and a difference of at least 30 metres in 6MWD change from baseline to week 52 using the in-

trial period. This was assessed using a win-ratio approach. All patients randomised to semaglutide 2.4 mg were compared with all patients randomised to placebo within each stratum of BMI (<35 kg/m2 vs ≥35 kg/m2). Missing data for KCCQ-CSS and 6MWD followed an imputation approach. 6MWD=6-minute walking distance. CI=confidence intervals. CRP=C-reactive protein. KCCQ-CSS= Kansas City Cardiomyopathy Questionnaire Clinical Summary Score.

Note that the treatment effects in this table may differ from the reported individual trial results because they reflect the analyses performed using the pooled dataset from STEP-HFpEF and STEP-HFpEF DM trials, and the models include the treatment by trial interaction terms.

Table S6. Number of Participants With Missing Data at Week 52 for the Dual Primary and Confirmatory Secondary Endpoints

	STEP HFpEF (n=529)		STEP HFpEF DM (n=616)	
	Semaglutide 2.4 mg	**Placebo**	**Semaglutide 2.4 mg**	**Placebo**
Dual Primary end points – change from baseline to week 52 in:				
KCCQ-CSS	27	22	29	34
Body weight	17	24	24	28
Confirmatory secondary end points – change from baseline to week 52 in:				
6MWD	23	41	29	41
CRP	22	23	24	29

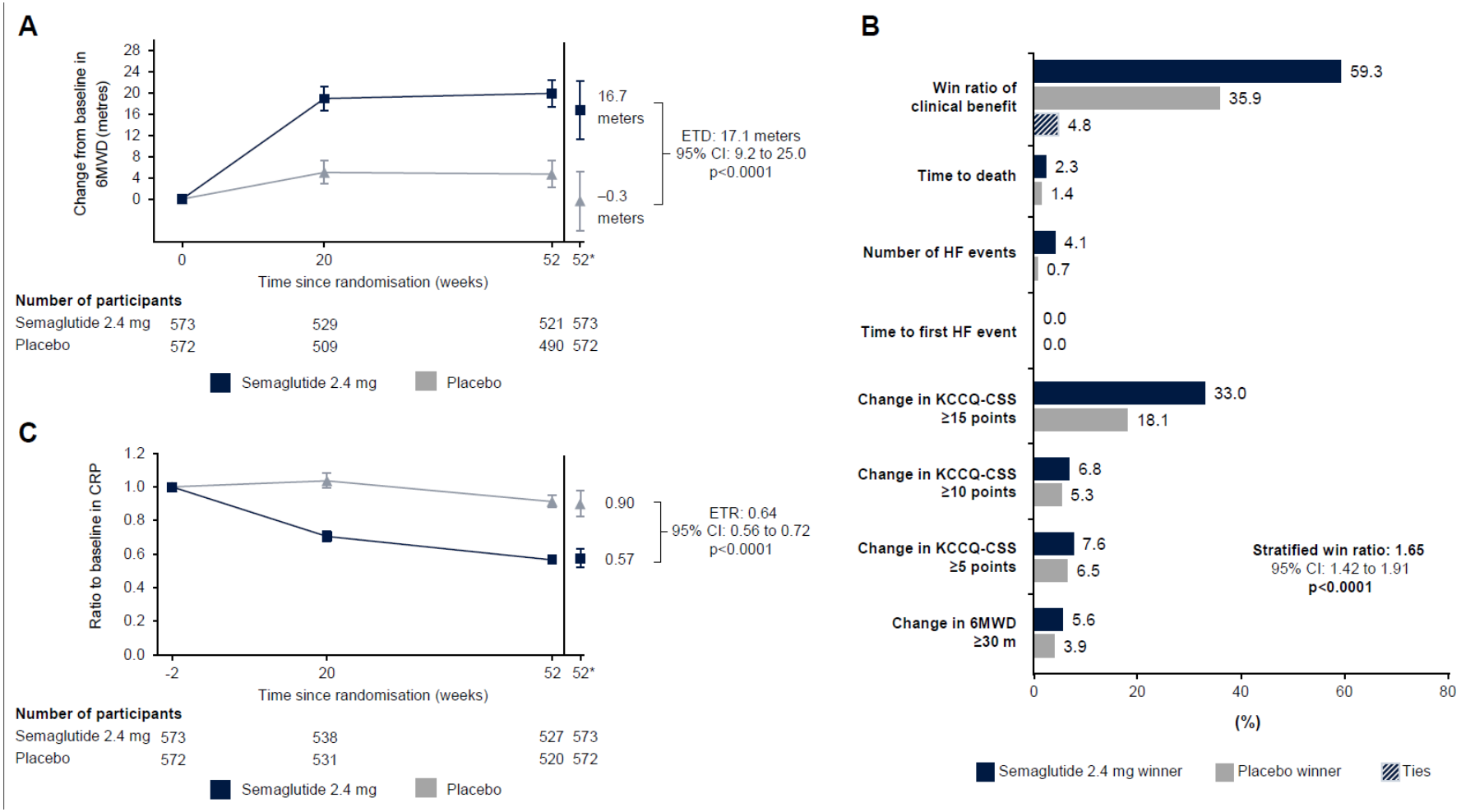
In the STEP HFpEF trial, for KCCQ-CSS, a total of 49 subjects had missing values at week 52, 4 participants had missing values due to a HF event or CV death (3 HF events, 1 CV death) - these were single imputed to a value of 2.083333 (1 participant in semaglutide group; 3 in placebo group), while 45 participants had missing values that were multiple imputed (26 participants in semaglutide group; 19 in placebo group). For body weight, a total of 41 participants had missing values at week 52 and were multiple imputed. For 6MWD, a total of 64 participants had missing values at week 52, 4 participants had missing values due to HF event or CV death (3 HF event, 1 CV death) - these were single imputed to a value of 20 (1 participant in semaglutide group; 3 in placebo group), while 60 subjects were multiple imputed (22 participants in semaglutide group; 38 in placebo group). For CRP, a total of 45 participants had missing values at week 52 and were multiple imputed.

In the STEP HFpEF DM trial, for KCCQ-CSS, a total of 63 participants had missing values at week 52, 7 participants had missing values due to a HF event or CV death (3 HF events, 4 CV death) - these were single imputed to a value of 0 (3 participants in semaglutide group; 4 in placebo group), while 56 participants had missing values that were multiple imputed (26 participants in semaglutide group; 30 in placebo group). For body weight, 52 participants had missing values at week 52, all missing values at week 52 were multiple imputed (24 participants in semaglutide group; 28 in placebo group). For 6MWD, a total of 70 participants had missing values at week 52, 9 participants had missing values due to HF event or CV death (4 HF events, 5 CV death) - these were single imputed to a value of 40 (3 participants in semaglutide group; 6 in placebo group), while 61 participants had missing values that were multiple imputed (26 participants in semaglutide group; 35 in placebo group). For CRP, a total of 53 subjects had missing values at week 52, all missing values at week 52 and were multiple imputed (24 participants in semaglutide group; 29 in placebo group).

6MWD=6-minute walk distance. CRP=C-reactive protein. CV=cardiovascular. HF=heart failure, KCCQ-CSS=Kansas City Cardiomyopathy Questionnaire Clinical Summary Score.

FIGURES

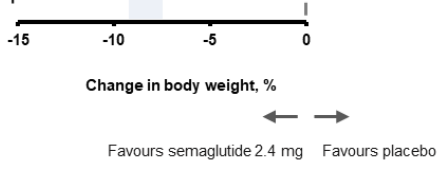
Figure S1. Change in 6MWD; components of the stratified win ratio for the hierarchical composite endpoint; and CRP across the STEP-HFpEF and STEP-HFpEF DM trials, baseline to 52 weeks (semaglutide vs placebo). (A) change in 6MWD; (B) stratified win ratio for the hierarchical composite endpoint (with results for individual components shown); (C) change in CRP.



Analyses are based on the treatment policy estimand, reflect the full analysis population, and are from the in-trial period. Panel A shows the observed (ie, as-measured) mean changes from baseline in the 6MWD (error bars are SE). Panel B shows the stratified win ratio for the composite hierarchical endpoint. Panel C shows the observed mean changes in the CRP (error bars are SE calculated on a logarithmic scale and back transformed to a linear scale). Numbers below the graphs are the number of participants contributing to the mean from the in-trial period. 6MWD=6-minute walk distance. ANCOVA=analysis of covariance. CI=confidence interval. CRP=C-reactive protein. ETD=estimated treatment difference. ETR=estimated treatment ratio. HF=heart failure. KCCQ-CSS=Kansas City Cardiomyopathy Questionnaire Clinical Summary Score. SE=standard error. *Data are estimated mean changes from baseline (from screening at week −2 for CRP) to week 52 for the treatment policy estimand using ANCOVA and an imputation approach for missing data.

Figure S2. Treatment effects of semaglutide versus placebo on change in body weight across prespecified subgroups

Subgroup	Semaglutide 2.4 mg n	Semaglutide 2.4 mg Change from baseline to Week 52	Placebo n	Placebo Change from baseline to Week 52	ETD [95% CI]	Interaction *P*-value
All patients	532	–11.4	520	–3.0	–8.4 [–9.2, –7.5]	<0.0001
Sex						
Female	257	–12.6	268	–3.0	–9.6 [–10.9, –8.4]	0.0060
Male	275	–10.2	252	–3.0	–7.2 [–8.4, –5.9]	
BMI group						
<35.0 kg/m²	178	–9.6	187	–2.2	–7.3 [–8.8, –5.8]	0.22
≥35.0 to <40 kg/m²	173	–11.5	179	–2.9	–8.7 [–10.1, –7.2]	
≥40.0 kg/m²	181	–13.1	154	–4.0	–9.1 [–10.6, –7.6]	
Age group						
<65 years	167	–10.3	162	–2.5	–7.7 [–9.2, –6.2]	
≥65 to <75 years	240	–12.1	218	–2.9	–9.2 [–10.5, –7.9]	0.24
≥75 years	125	–11.5	140	–3.8	–7.7 [–9.5, –6.0]	
Race						
White	472	–11.6	473	–2.9	–8.7 [–9.6, –7.8]	0.037
Non-white	60	–9.7	47	–4.0	–5.7 [–8.4, –3.0]	
Region						
North America	86	–11.7	85	–2.1	–9.5 [–11.6, –7.5]	
Europe	373	–11.2	336	–3.3	–7.9 [–8.9, –6.8]	0.28
Other	73	–11.8	99	–2.7	–9.1 [–11.3, –7.0]	
LVEF group						
≥45 to <50%	87	–10.3	92	–3.5	–6.7 [–8.8, –4.7]	
≥50 to <60%	215	–11.6	218	–3.2	–8.4 [–9.8, –7.1]	0.21
≥60%	230	–11.6	210	–2.6	–9.0 [–10.3, –7.6]	
NT-proBNP group						
<Pooled median (475.3 pg/mL)	281	–11.6	247	–2.5	–9.1 [–10.3, –7.9]	0.088
≥Pooled median (475.3 pg/mL)	248	–11.1	271	–3.5	–7.6 [–8.9, –6.4]	
CRP group						
<2 mg/L	155	–11.2	152	–3.0	–8.2 [–9.8, –6.6]	0.81
≥2 mg/L	375	–11.5	366	–3.0	–8.4 [–9.5, –7.4]	
Systolic blood pressure group						
<135 mmHg	283	–11.2	261	–2.8	–8.5 [–9.7, –7.3]	0.83
≥135 mmHg	248	–11.6	259	–3.3	–8.3 [–9.5, –7.0]	
NYHA class						
Class II	379	–11.3	347	–2.9	–8.4 [–9.4, –7.4]	0.95
Class III or IV	153	–11.6	173	–3.2	–8.3 [–9.9, –6.8]	
AFib at baseline						
Yes	232	–11.5	251	–3.0	–8.4 [–9.7, –7.2]	0.86
No	300	–11.3	269	–3.0	–8.3 [–9.5, –7.1]	
Loop diuretics at baseline						
Yes	313	–11.2	328	–3.0	–8.2 [–9.3, –7.1]	0.70
No	219	–11.6	192	–3.0	–8.6 [–9.9, –7.2]	
Heart rate						
<Pooled median (70.0 BPM)	251	–11.6	248	–2.5	–9.1 [–10.3, –7.8]	0.13
≥Pooled median (70.0 BPM)	280	–11.2	272	–3.5	–7.7 [–8.9, –6.5]	
Coronary artery disease						
Yes	121	–11.1	104	–3.3	–7.8 [–9.7, –5.9]	0.51
No	411	–11.5	416	–2.9	–8.5 [–9.5, –7.5]	
RAAS inhibitor use						
Yes	412	–11.0	415	–3.2	–7.8 [–8.7, –6.8]	0.011
No	120	–12.9	105	–2.4	–10.5 [–12.3, –8.6]	



Analyses are based on the treatment policy estimand, reflect the full analysis population, and are from the in-trial period. Responses at week 52 were analysed using ANCOVA and an imputation approach for missing data. AFib=atrial fibrillation. BMI=body mass index. CI=confidence intervals. CRP=C-reactive protein. ETD=estimated treatment difference. LVEF=left ventricular ejection fraction. NT-proBNP=N-terminal pro B-type natriuretic peptide. NYHA=New York Heart Association.

REFERENCES

1 McEvoy BW. Missing data in clinical trials for weight management. *J Biopharm Stat*. 2016;**26**(1):30-6.

2 Wang Y, Tu W, Kim Y, Sinks S, He J, Cambon A, et al. Statistical methods for handling missing data to align with treatment policy strategy. *Pharm Stat*. 2023;**22**(4):650-70.