ESC European Society of Cardiology

European Journal of Heart Failure (2023)
doi:10.1002/ejhf.2868

RESEARCH ARTICLE

Sex differences in the generalizability of randomized clinical trials in heart failure with reduced ejection fraction

Megan Schroeder[1,2]*, Yvonne Mei Fong Lim[3,4], Gianluigi Savarese[2], Kiliana Suzart-Woischnik[5], Claire Baudier[6], Tomasz Dyszynski[5], Ilonca Vaartjes[3], Marinus J.C. Eijkemans[3], Alicia Uijl[2,7], Ronald Herrera[5], Eleni Vradi[8], Jasper J. Brugts[7], Hans-Peter Brunner-La Rocca[9], Vanessa Blanc-Guillemaud[9], Sandra Waechter[10], Fabrice Couvelard[9], Benoit Tyl[9], Samuel Fatoba[5], Arno W. Hoes[3], Lars H. Lund[2], Christoph Gerlinger[11,12], Folkert W. Asselbergs[7,13], Diederick E. Grobbee[3,14], Maureen Cronin[15], and Stefan Koudstaal[16]

[1]Institute for Medical Information Processing, Biometry, and Epidemiology Pettenkofer School of Public Health, LMU Munich, Munich, Germany; [2]Division of Cardiology, Department of Medicine, Karolinska Insitutet, Stockholm, Sweden; [3]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands; [4]Institute for Clinical Research, National Institutes of Health, Selangor, Malaysia; [5]Medical Affairs & Pharmacovigilance, Berlin, Germany; [6]Department of Cardiology, Erasmus MC University Medical Centre, Rotterdam, The Netherlands; [7]Department of Cardiology, Amsterdam University Medical Centers, Cardiovascular Sciences, University of Amsterdam, Amsterdam, The Netherlands; [8]Biomedical Data Science II, Bayer AG, Berlin, Germany; [9]Institut de Recherches Internationales Servier (I.R.I.S.), Suresnes, France; [10]Vifor Pharma Ltd, Glattbrugg, Switzerland; [11]Statistics and Data Insights, Bayer AG, Berlin, Germany; [12]Gynecology, Obstetrics and Reproductive Medicine, University Medical School of Saarland, Saar, Germany; [13]Health Data Research UK and Institute of Health Informatics, University College London, London, UK; [14]Julius Clinical, Zeist, The Netherlands; [15]Cronin Pharma Consulting, Lucerne, Switzerland; and [16]Department of Cardiology, Green Heart Hospital, Gouda, The Netherlands

Received 16 October 2022; revised 6 March 2023; accepted 17 April 2023

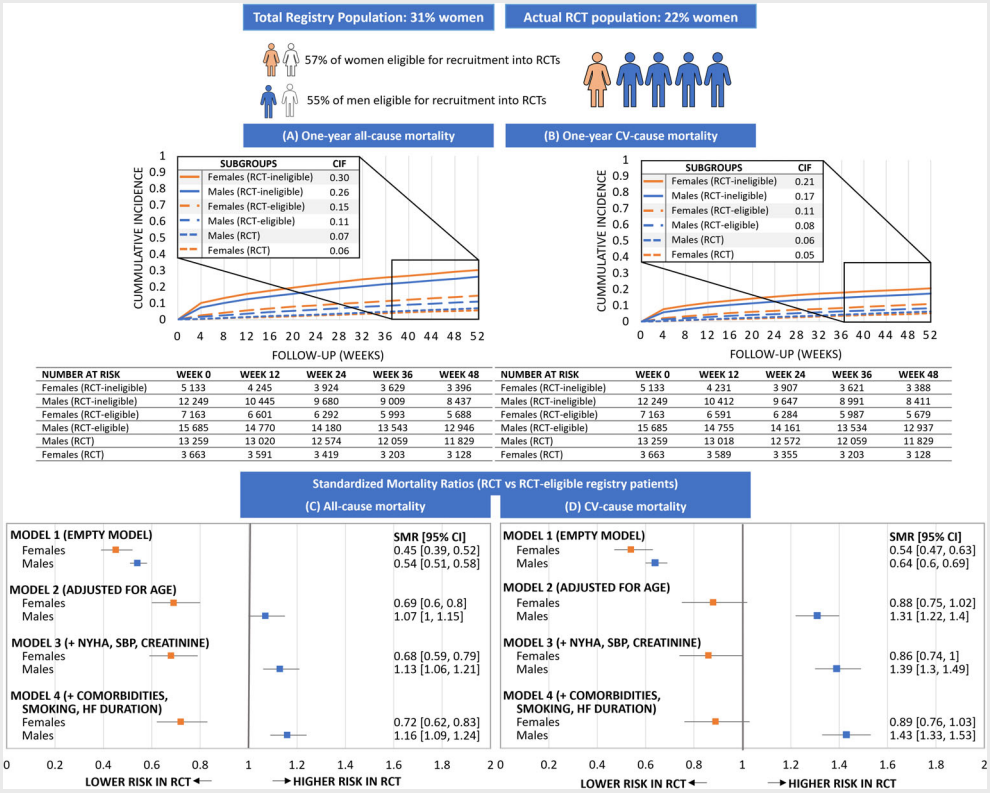
Aims In order to understand how sex differences impact the generalizability of randomized clinical trials (RCTs) in patients with heart failure (HF) and reduced ejection fraction (HFrEF), we sought to compare clinical characteristics and clinical outcomes between RCTs and HF observational registries stratified by sex.

Methods and results Data from two HF registries and five HFrEF RCTs were used to create three subpopulations: one RCT population ($n = 16\,917$; 21.7% females), registry patients eligible for RCT inclusion ($n = 26\,104$; 31.8% females), and registry patients ineligible for RCT inclusion ($n = 20\,810$; 30.2% females). Clinical endpoints included all-cause mortality, cardiovascular mortality, and first HF hospitalization at 1 year. Males and females were equally eligible for trial enrolment (56.9% of females and 55.1% of males in the registries). One-year mortality rates were 5.6%, 14.0%, and 28.6% for females and 6.9%, 10.7%, and 24.6% for males in the RCT, RCT-eligible, and RCT-ineligible groups, respectively. After adjusting for 11 HF prognostic variables, RCT females showed higher survival compared to RCT-eligible females (standardized mortality ratio [SMR] 0.72; 95% confidence interval [CI] 0.62–0.83), while RCT males showed higher adjusted mortality rates compared to RCT-eligible males (SMR 1.16; 95% CI 1.09–1.24). Similar results were also found for cardiovascular mortality (SMR 0.89; 95% CI 0.76–1.03 for females, SMR 1.43; 95% CI 1.33–1.53 for males).

Conclusion Generalizability of HFrEF RCTs differed substantially between the sexes, with females having lower trial participation and female trial participants having lower mortality rates compared to similar females in the registries, while males had higher than expected cardiovascular mortality rates in RCTs compared to similar males in registries.

*Corresponding author. Global Epidemiology, Boehringer Ingelheim, Binger Strasse 173, 55216 Ingelheim, Germany. Tel: +49 175 4026604, Email: megan.schroeder@outlook.de

Graphical Abstract



Sex differences in randomized clinical trial (RCT) recruitment and clinical outcomes at 1 year. Cumulative incidence functions and adjusted standardized mortality ratios (SMRs) for all-cause and cardiovascular mortality at 1 year stratified by sex. (*A*) Cumulative incidence for all-cause mortality between RCT and registry patients. (*B*) Cumulative incidence for cardiovascular mortality between RCT and registry patients. (*C*) SMRs for all-cause mortality. (*D*) SMRs for cardiovascular mortality. Pooled SMRs estimated from five trials with their 95% confidence intervals were reported.

Keywords Heart failure • Real-world evidence • Females • Enrichment strategies • Standardized mortality ratios • Randomized clinical trial

Introduction

There are sex and gender differences across multiple diseases and clinical syndromes. Some of the most profound differences can be seen in heart failure (HF).[1,2] Females and males differ in HF aetiology, age, risk factors, biomarkers, pathophysiology, comorbidities, and clinical presentation.[2–7] There is increasing awareness on sex differences in HF, however there are still large gaps in knowledge of sex-specific mechanisms, optimal treatment, and prognosis of HF.[2]

One driving factor of the knowledge gap in sex differences is the widespread underrepresentation of females recruited to HF clinical trials. From observational HF registries, the percentage of females with HF and reduced ejection fraction (HFrEF) in the population is around 30–50%,[8,9] whereas the percentage of enrolled females in HFrEF trials is on average 24%.[10] As a consequence, contemporary treatment guidelines are still predominantly based on male-derived data.[10–13] Post-hoc analyses stratified for sex from trials, and observational data, currently suggest that females may need lower dosages.[14,15] However, selection bias amongst other forms of unmeasured confounders warrant caution when critically appraising these results.

Currently, several uncertainties remain to be elucidated, for example (i) are the treatment dosages used in randomized clinical trials (RCTs) the same for females and males? (ii) how does this selective enrolment of females influence the clinical characteristics and prognosis of female trial participants compared to eligible females with HFrEF seen in the broader population?; or (iii) how does selection of females in trials impact the RCT itself with regard to successfully implementing enrichment strategies aimed to enrol patients at higher risk of cardiovascular outcomes of interest

and lower risk for (non-cardiovascular) competing outcomes? Currently, there are few data that shed light on how selective enrolment of females in HFrEF trials impact generalizability to daily clinical practice as well as the success of enrichment strategies in RCTs.

To address these uncertainties, in the present study, we sought to assess differences in clinical characteristics, medication dose and use, and explored unadjusted and case-mix adjusted mortality rates stratified for each of the sexes using individual patient data from five RCTs and two large HF registries.

Methods

Data sources

Detailed information on the methods including data sources, endpoint definitions and collection codes can be found in a previous study.[16] Briefly, five HFrEF RCTs and two HF registries were included in this study. BEAUTIFUL and SHIFT were phase III ivabradine trials ($n = 15\,732$),[17,18] FAIR-HF and CONFIRM were phase III and phase IV studies on intravenous iron supplementation ($n = 763$)[19,20] and PANTHEON was a phase II trial for neladenosone bialanate ($n = 427$).[21] For final analysis, aggregated data from both treatment and placebo arms of each RCT were pooled to represent one RCT population ($n = 16\,917$).

The Dutch CHECK-HF and SwedeHF registries enrol patients with clinician-judged HF and detailed information on the methods can be found elsewhere.[22,23] For the current analysis, only HFrEF patients, defined as those enrolled with left ventricular ejection fraction (LVEF) <40%, were considered. To ensure consistency with CHECK-HF, only outpatients registered between 2000 to 2016 in SwedeHF ($n = 40\,230$) were included. Contrary to Dutch CHECK-HF, SwedeHF contains follow-up data, therefore any analysis of clinical outcomes was restricted to patients from SwedeHF. Ethics approvals were obtained by the original study investigators for the RCTs. CHECK-HF received approval for anonymized analysis of routine clinical data. In SwedeHF, patient consent to enrolment in the registry allows analysis of individual patient data.

Eligibility criteria, study population, and outcomes

The inclusion and exclusion criteria listed in the study protocols of the five RCTs were tabulated to identify common eligibility and ineligibility criteria (*Figure 1*, online supplementary *Table S1*).[16] These common criteria were applied to the SwedeHF and CHECK-HF dataset to identify subgroups of patients who would have been eligible for trial participation or not. Data were then presented by the following groups and additionally stratified by sex: RCT, RCT-eligible, and RCT-ineligible (*Figure 1*). The following clinical outcomes at 1 year were assessed: all-cause mortality, cardiovascular mortality, and first HF hospitalization.

Statistical analysis

Continuous data are presented as mean with standard deviation, while categorical variables are reported in absolute and relative frequencies. Mean and proportion differences between each group were calculated and reported as significant based on their corresponding 99% confidence intervals (CI). Unadjusted outcomes were calculated with cumulative incidence curves for each of the six subgroups outline above. The competing event for cardiovascular mortality was death from other causes whereas for first HF hospitalization, it was all-cause deaths. To test whether the RCT group was more, less, or equally likely to die than the RCT-eligible group, standardized mortality ratios (SMRs) were calculated and stratified by sex. SMRs were calculated by dividing the observed mortality count in the RCT group by expected mortality count in the RCT group. The observed mortality counts were the actual deaths recorded in the RCTs at 1 year. In standard SMR analysis, expected counts are the number of deaths that would be predicted if the study population (RCT group) were to have the same age and/or sex-specific rates as the standard population (RCT-eligible group).[24] However, one limitation in SMR analysis is the inability to account for case-mix between populations.[25] To calculate more precise expected mortality counts in the RCTs, we used a validated prognostic model to apply characteristics of the RCT-eligible group to the RCT group. We first fitted a Poisson model with 11 prognostic indicators from a validated MAGGIC HF risk score (age, sex, LVEF, New York Heart Association [NYHA] class, serum creatinine, chronic obstructive pulmonary disease, diabetes, systolic blood pressure, body mass index, HF duration, smoking status) in a stepwise manner to the RCT-eligible SwedeHF group. Model 1 was the empty model, model 2 included age and sex, model 3 additionally included NYHA class, systolic blood pressure, and creatinine, and model 4 was fully adjusted with all 11 prognostic variables. Each model with derived coefficients from the RCT-eligible population was then applied to each RCT to derive expected counts. If these prognostic factors and their associated risks were similar between the RCT and RCT-eligible group, then the expected deaths in the RCTs would be equal to the observed deaths leading to an SMR value of 1. Therefore, SMRs above 1 indicates that there are more observed deaths in the RCT population than would be expected based on characteristics derived from the RCT-eligible population, and vice versa for SMR ratios below 1.0. The SMRs for all trials were pooled using fixed effect meta-analysis and the corresponding 95% CI was determined using methods described by Breslow and Day.[26] For first HF hospitalization, we did not estimate SMRs because existing prediction models for hospitalization are largely influenced by admission policies within individual health settings and hence have insufficient discriminative model performance.[27]

The largest RCTs (BEAUTIFUL and SHIFT) in this analysis only included patients who were in sinus rhythm and the BEAUTIFUL study included a population who had coronary artery disease; therefore, sensitivity analyses were conducted in subsets of registry patients who were (i) in sinus rhythm or (ii) diagnosed with coronary artery disease. Missing data were multiply imputed by chained equations using the mice package in R. The number of imputations was set at 20.[28] Statistical significance was set at level 0.05. Statistical analysis was performed using the R statistical software version 3.6.1 (R Core Team, 2019) and Stata SE version 15 (StataCorp LP, College Station, TX, USA).[29,30]

Results

Eligibility for potential trial enrolment

Out of 46 914 patients from the registries, 14 584 were females (31.1%). After applying the harmonized set of eligibility and ineligibility criteria, 8294 out of 14 584 (56.9%) females and 17 818 out of 32 330 (55.1%) males in the registries were considered eligible for RCT recruitment for a final RCT-eligible group of 26 104 (31.8% females). Cancer was the most restricting criterion with 27.4% of females and 28.1% of males excluded. The exclusion criteria

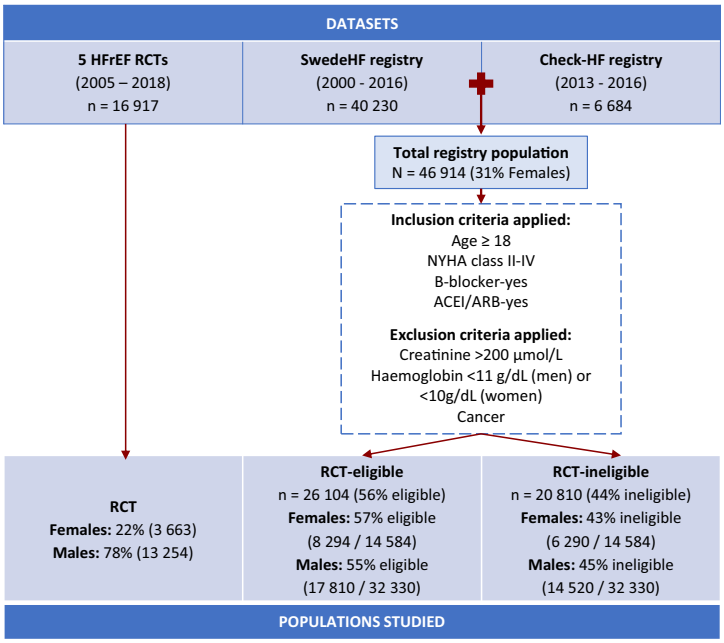


Figure 1 Flowchart of the selection of studied populations from available datasets and the respective proportion of males and females. Common inclusion and exclusion criteria found in the heart failure trials were summarized and applied to the registry populations to determine which registry patients would have been eligible for recruitment into the heart failure randomized clinical trials (RCTs). Corresponding sample size stratified by sex is presented for each subgroup. ACEI, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; HFrEF, heart failure with reduced ejection fraction; NYHA, New York Heart Association.

of angiotensin-converting enzyme inhibitor (ACEI) and angiotensin receptor blocker (ARB) use at baseline differed the most between sexes with 15.7% of females excluded for not taking ACEI or ARBs and 11.6% of males excluded (online supplementary *Table S2*). In the RCT population, the observed number of females was significantly lower with 3663 out of 16 917 (21.7%) patients (*Figure 1*).

Baseline characteristics

Baseline characteristics for the RCT population, RCT-eligible and RCT-ineligible patients stratified by sex are shown in *Table 1*. Overall, patients in the RCTs were younger compared to RCT-eligible and RCT-ineligible patients, with similar directions for both females (66.3 vs. 73.9 vs. 76.7 years) and males (62.8 vs. 69.8 vs. 73.8 years) in the three groups, respectively. Compared to males, females were significantly older in all three groups (*Table 1*). Similarly in females and males, a minority of the RCT population had a LVEF <30% (28.7% and 32.6%) as opposed to both the registry populations of RCT-eligible (47.0% and 56.1%) and RCT-ineligible patients (43.0% and 48.6%) for females and males, respectively. Although females in all three groups had a higher LVEF compared to males, the proportion of patients in NYHA functional class III/IV was also highest in females compared to males in all groups (46.5% vs. 36.4%; 48.1% vs. 42.9%; 46.6% vs. 42.6%; in the RCT, RCT-eligible and RCT-ineligible groups, respectively).

With regard to medical management of HF, the uptake of mineralocorticoid receptor antagonists (MRA) was low for both sexes in all three groups (47.1% and 41.9% in the RCTs, 38.4% and 39.8% in the RCT-eligible, and 33.9% and 32.7% in RCT-ineligible, percentages for females and males, respectively). Overall, loop diuretics were prescribed more often in every female population compared to males, with highest difference in the RCT populations (78.9% in females vs. 69.6% in males). Target dosing did not meaningful differ between the sexes in any medication except for ACEI and ARB where less females received ≥50%–≥100% of target dose for ACEI and ARB compared to males in the RCT-eligible (65.4% vs. 71.6%) and RCT-ineligible groups (36.9% vs. 46.4%), however not in the RCT group (54.7% vs. 56.8%) (online supplementary *Table S3*).

Unadjusted clinical outcomes

Cumulative incidence curves for unadjusted cumulative incidence rates for all-cause and cardiovascular mortality, and HF hospitalization rates are shown in *Figure 2* and unadjusted rates are summarized in *Table 2*. Females showed a lower unadjusted 1-year mortality rate in the RCT population compared to males (5.6% vs. 6.9%, $p < 0.01$), whereas females had higher unadjusted 1-year mortality rates compared to males in both the RCT-eligible (14.0% vs. 10.7%, $p < 0.0001$) and RCT-ineligible groups (28.6% vs. 24.6%, $p < 0.0001$). Similar trends were also observed for cardiovascular mortality (*Table 2*). Rate of first HF hospitalization was lowest in the RCTs for both females and males (8.4% and 7.8%, $p > 0.05$), and highest in the registry groups (RCT-eligible: 23.2% and 24.8%,

Table 1 Baseline characteristics compared between sex and stratified by randomized clinical trial (RCT) population, RCT-eligible registry population, and RCT-ineligible registry population

	RCT			RCT-eligible registry			RCT-ineligible registry		
	Females	Males	*p*-value[a]	Females	Males	*p*-value[a]	Females	Males	*p*-value[a]
n	3663	13 254		8294	17 810		6290	14 520	
Demographics and lifestyle, mean (SD) or %									
Age (years)	66.3 (9.9)	62.8 (9.9)	***	73.9 (9.1)	69.8 (12.2)	***	76.7 (9.7)	73.8 (9.3)	***
Smoking history			***			***			***
Never	72.4%	29.0%		53.0%[b]	35.50%[b]		57.0%	37.80%	
Previous/current	27.6%	71.0%		47.0%[b]	64.50%[b]		43.0%	62.20%	
Clinical parameters, mean (SD) or %									
HF duration (months)	42.4 (58.4)	41.6 (58.7)		21.0 (58.6)	30.8 (58.8)	***	21.1 (32.2)	24.4 (31.2)	***
SBP (mmHg)	126.5 (14.9)	124.9 (14.9)	***	126.4 (16.0)	123.7 (20.5)	***	125.7 (17.1)	124.0 (15.2)	***
BMI (kg/m^2)	28.6 (5.5)	28.2 (4.6)	***	26.6 (6.5)	27.2 (6.0)	***	25.6 (4.7)	26.0 (4.3)	***
Creatinine (μmol/L)	90.3 (33.9)	101.6 (41.3)	***	90.6 (37.6)	103.4 (38.2)	***	108.9 (0.3)	129.7 (60.6)	***
LVEF (%)			***			***			***
0–29	28.7%	32.6%		47.0%	56.1%		43.0%	48.6%	
30–39	67.4%	66.6%		53.0%	43.9%		57.0%	51.4%	
≥40	3.9%	0.8%		0.0%	0.0%		0.0%	0.0%	
NYHA functional class			***			***			***
I	0.0%	0.1%		0.0%	0.0%		17.9%	22.9%	
II	53.5%	63.5%		51.9%	57.1%		35.4%	34.5%	
III	45.8%	35.7%		43.5%	39.4%		38.5%	36.2%	
IV	0.7%	0.7%		4.6%	3.5%		8.2%	6.4%	
III/IV	46.5%	36.4%	***	48.1%	42.9%	***	46.6%	42.6%	***
Comorbidities %									
Hypertension	72.0%	67.0%	***	59.1%	54.7%	***	56.6%	54.7%	*
Diabetes mellitus	34.7%	33.5%		25.8%	27.8%	***	25.7%	27.8%	**
CAD	79.3%	87.8%	***	49.0% [a]	48.6%		48.1%[a]	57.4%	***
Valvular heart disease	14.9%	11.0%	***	24.2%	20.2%	***	26.5%	25.0%	*
Stroke or TIA	9.1%	9.3%		14.3% [a]	14.0%[a]		16.6%[a]	18.1%[a]	*
Atrial fibrillation/flutter	9.7%	8.9%		42.2%	46.7%	***	47.8%	54.9%	***
COPD	6.9%	9.3%	***	21.3%	18.6%	***	22.1%	20.9%	*
Depression	4.6%	2.1%	***	6.2% [a]	4.3%[a]	***	6.5%	4.6%	***
Cancer	4.0%	2.4%	***	0.0% [c]	0.0%[c]	–	32.4%	32.2%	
Concomitant medications									
ACEI or ARB	90.2%	89.8%		100.0%[c]	100.0%[c]	–	63.7%	74.2%	***
Anticoagulant	2.8%	2.6%		40.7%	48.1%	***	33.4%	41.0%	***
Antiplatelet	74.7%	79.0%	***	48.2%	46.8%	*	45.5%	47.7%	**
MRA	47.1%	41.9%	***	38.4%	39.8%	*	33.9%	32.7%	
Beta-blocker	87.3%	87.5%		100.0%[c]	100.0%[c]	–	71.6%	75.0%	***
Digitalis glycoside	14.9%	14.7%		17.1%	17.0%		15.7%	13.9%	***
Diuretic	78.9%	69.6%	***	81.9%	78.1%	***	81.0%	77.7%	***

ACE, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; BMI, body mass index; CAD, coronary artery disease; CI, confidence intervals; COPD, chronic obstructive pulmonary disease; HF, heart failure; LVEF, left ventricular ejection fraction; MRA, mineralocorticoid receptor antagonist; SBP, systolic blood pressure; SD, standard deviation; TIA, transient ischaemic attack.

Percentages may not add up to 100% due to rounding.

[a]Comparison between males and females (independent *t*-test for continuous variables and χ^2 test for categorical variables).

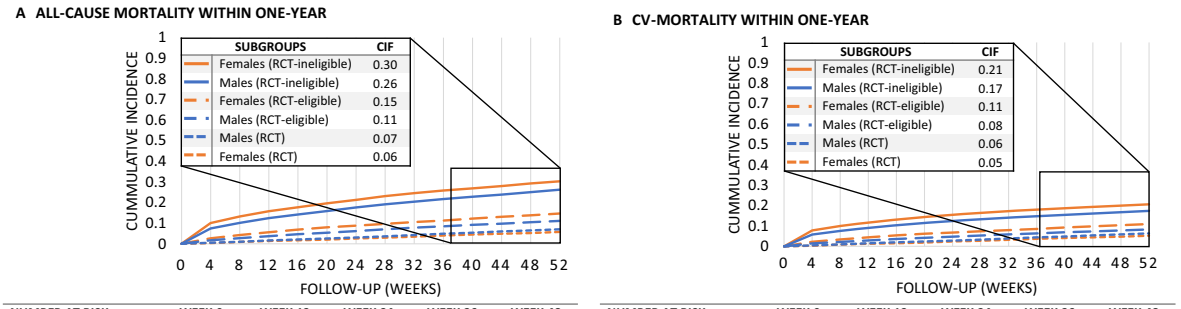
[b]Data from SwedeHF only.

[c]Statistical comparisons were not compared because they were part of the criteria for selecting RCT-eligible registry patients.

*$p < 0.05$.

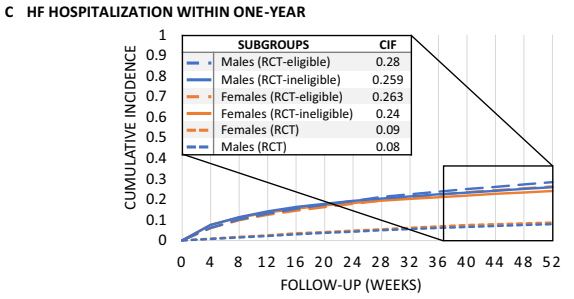
**$p < 0.01$.

***$p < 0.001$.



NUMBER AT RISK	WEEK 0	WEEK 12	WEEK 24	WEEK 36	WEEK 48
Females (RCT-ineligible)	5 133	4 245	3 924	3 629	3 396
Males (RCT-ineligible)	12 249	10 445	9 680	9 009	8 437
Females (RCT-eligible)	7 163	6 601	6 292	5 993	5 688
Males (RCT-eligible)	15 685	14 770	14 180	13 543	12 946
Males (RCT)	13 259	13 020	12 574	12 059	11 829
Females (RCT)	3 663	3 591	3 419	3 203	3 128

NUMBER AT RISK	WEEK 0	WEEK 12	WEEK 24	WEEK 36	WEEK 48
Females (RCT-ineligible)	5 133	4 231	3 907	3 621	3 388
Males (RCT-ineligible)	12 249	10 412	9 647	8 991	8 411
Females (RCT-eligible)	7 163	6 591	6 284	5 987	5 679
Males (RCT-eligible)	15 685	14 755	14 161	13 534	12 937
Males (RCT)	13 259	13 018	12 572	12 059	11 829
Females (RCT)	3 663	3 589	3 355	3 203	3 128



NUMBER AT RISK	WEEK 0	WEEK 12	WEEK 24	WEEK 36	WEEK 48
Females (RCT-ineligible)	5 133	4 245	3 924	3 629	3 396
Males (RCT-ineligible)	12 249	10 445	9 680	9 009	8 437
Females (RCT-eligible)	7 163	6 601	6 292	5 993	5 688
Males (RCT-eligible)	15 685	14 770	14 180	13 543	12 946
Females (RCT)	3 663	3 581	3 412	3 199	3 125
Males (RCT)	13 259	13 020	12 574	12 059	11 829

Figure 2 Cumulative incidence curves and unadjusted event rates within 1 year of follow-up for (*A*) all-cause mortality, (*B*) cardiovascular (CV) mortality, and (*C*) heart failure (HF) hospitalization. Cumulative incidence functions (CIF) at 1 year for (*A*) all-cause mortality, (*B*) CV mortality, and (*C*) first HF hospitalization. CIFs calculated for males and females in the (i) randomized clinical trial (RCT) group, (ii) RCT-eligible, and (iii) RCT-non-eligible registry group. Females in the RCT-ineligible group showed the highest probability of all-cause and CV mortality at 1 year, with RCT-ineligible males having slightly lower probability of mortality. Males in the RCT-ineligible group showed highest probability of first HF hospitalization, followed by RCT-eligible males.

$p < 0.01$; RCT-ineligible: 23.7% and 25.3%, $p < 0.05$ for females and males, respectively) (*Figure 2* and *Table 2*).

Case-mix adjusted clinical outcomes

Standardized mortality ratios were calculated to compare risk of death between the RCT population and the RCT-eligible population. To interpret the SMR, a SMR ratio of 1.0 indicates that the risk of death in the RCT is identical to the risk of death in the RCT-eligible population. Therefore, SMR ratios above 1 indicates that there is higher risk of death in the RCT population than expected, and vice versa for SMR ratios below 1.0.

Unadjusted SMRs (empty model) showed that females had 55% fewer deaths in the RCT group than expected in the RCT-eligible group (SMR 0.45; 95% CI 0.39–0.52), while males had 46% fewer deaths in the RCT group (SMR 0.54; 95% CI 0.51–0.58). Model 2, which adjusted for age between the younger RCT patients (mean age 63.5 years) and RCT-eligible patients (mean age 71.1 years), showed that females still had 31% fewer observed deaths than expected (SMR 0.69; 95% CI 0.60–0.80), whereas in males there was 7% higher observed deaths in the trials than expected in the age adjusted RCT-eligible males (SMR 1.07; 95% 1.00–1.15). For cardiovascular mortality, the difference after adjusting for age was more pronounced, with 12% fewer cardiovascular deaths in females, as opposed to a 31% increased number of observed cardiovascular deaths in male trial participants than expected (SMR 0.88; 95% 0.75–1.02 vs. SMR 1.31; 95% CI 1.22–1.40). After full adjustment for all HF prognostic factors in model 4, these observed sex differences remained in place with 11% fewer cardiovascular deaths in females participating in trials than expected in RCT-eligible females, compared to 43% more observed cardiovascular deaths in male trial participants than expected from RCT-eligible males (SMR 0.89; 95% CI 0.76–1.03 vs. SMR 1.43; 95% CI 1.33–1.53) (*Figure 3*). The sensitivity analyses of SMRs calculated in subgroups of those only in sinus rhythm or those only with coronary artery disease did not meaningfully differ from the total population (online supplementary *Figures S2* and *S3*).

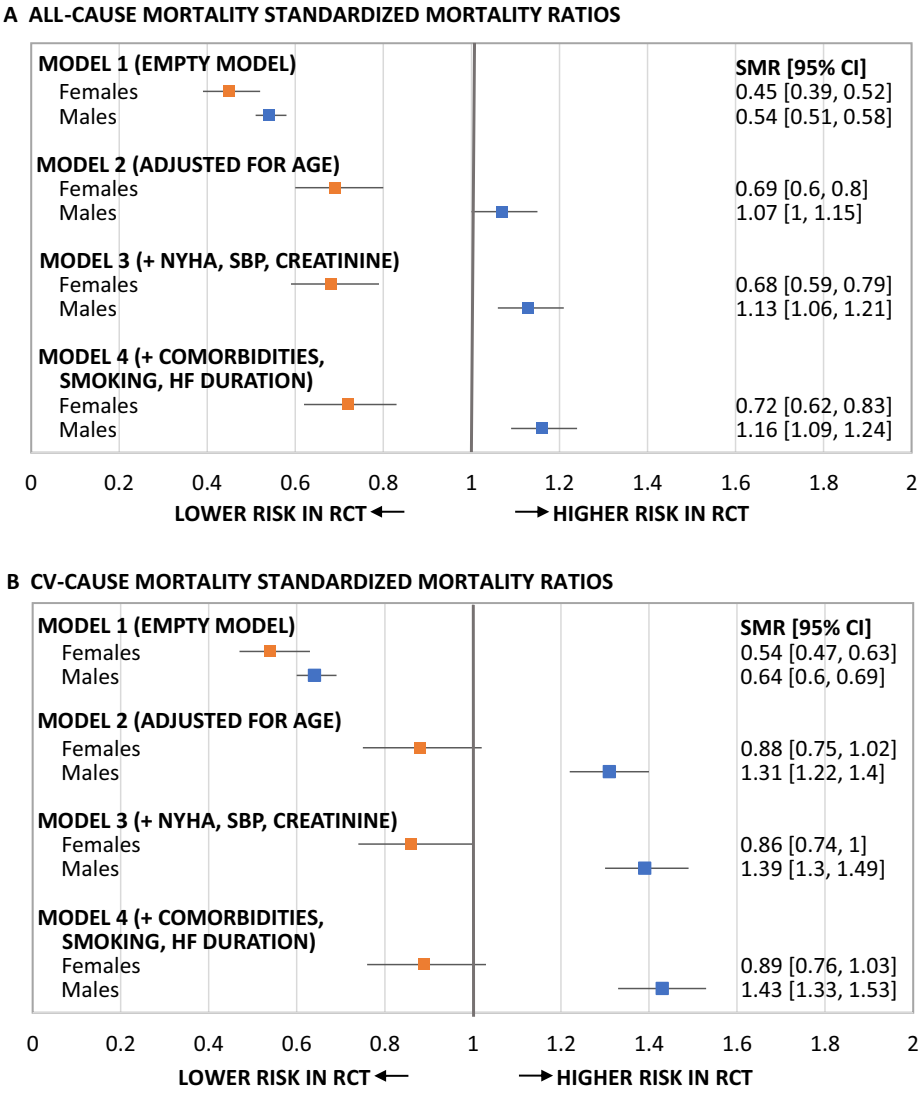


Figure 3 Standardized mortality ratios (SMR) between the randomized clinical trial (RCT) population and the RCT-eligible population stratified by sex for (*A*) all-cause mortality and (*B*) cardiovascular (CV) mortality. SMRs for (*A*) all-cause mortality and (*B*) CV mortality within 1 year and stratified by sex. Heart failure (HF) prognostic factors from the MAGGIC risk model were added stepwise to each model until the fully adjusted model 4. The models were applied to the RCT-eligible population and the derived coefficients were then applied to the RCT population to predict expected deaths. SMRs were calculated by dividing observed RCT deaths by expected RCT deaths. Pooled SMRs estimated from five trials with their 95% confidence interval (CI) were reported. Females consistently showed lower mortality rates in the RCT group compared to the registry group, while males in the RCT had higher rates of mortality. NYHA, New York Heart Association; SBP, systolic blood pressure.

Discussion

Using individual patient data of over 62 000 patients from five HFrEF RCTs and two HF registries, we found several sex differences that impacted the efficacy of enrichment strategies in the clinical trials itself and influenced the generalizability of their results into daily clinical practice. Thirty-one percent of patients in the registries were females, whereas 22% trial participants were females. Contrary to males, females in trials had a significantly better survival than expected from the registries, even after extensive adjustments for HF prognostic factors. HF hospitalizations were much more frequent in the observational registry compared to the trials, but here there were no relevant differences between the sexes (*Graphical Abstract*). Taken together, these data show that although in- and exclusion criteria are similar, the populations of males and females enrolled in the RCTs show substantial differences in comparison with HF patients in the general population, and the magnitude and direction of these differences were unique to both sexes.

Table 2 Unadjusted outcome rates between males and females, stratified by randomized clinical trial (RCT) population, RCT-eligible registry population, and RCT-ineligible population at 1 year

	RCT			RCT-eligible			RCT-ineligible		
	Females	Males	Proportion difference (99% CI)	Females	Males	Proportion difference (99% CI)	Females	Males	Proportion difference (99% CI)
All-cause mortality	5.6% (201)	6.9% (910)	−1.4% (−2.5%, −0.3%)**	14.0% (1005)	10.7% (1671)	3.4% (2.1%, 4.6%)***	28.6% (1468)	24.6% (3017)	4.0% (2.1%, 5.9%)***
CV mortality	5.0% (183)	6.2% (822)	−1.2% (−2.3%, −0.1%)**	10.6% (761)	8.1% (1267)	2.5% (1.5%, 3.6%)***	20.2% (1038)	17.0% (2078)	3.3% (1.6%, 4.9%)***
First HF hospitalization	8.4% (308)	7.8% (1034)	0.6% (−0.7%, 1.9%)	23.2% (1660)	24.8% (3885)	−1.6% (−3.2%, 0.0%)**	23.7% (1214)	25.3% (3097)	−1.6% (−3.5%, 0.2%)*

CI, confidence interval; CV, cardiovascular; HF, heart failure.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ based on χ^2 test, 99% CI calculated.

We confirm that there are sex-related differences in clinical profile, comorbidities, medication use, and outcomes in HFrEF.[2,31–33] Females in all three groups were older, less often smokers, had higher LVEF, less ischaemic-related disease, more often diagnosed with hypertension, and had higher NYHA class III/IV proportions across all populations.[2,4,5,33–35] Females typically have shorter HF duration due to later onset of HFrEF which was only confirmed here in the RCT population, but not in the registry populations.[8] Depression rates were more than doubled compared to males.[36] These sex differences were consistent across the three groups, however the proportion differences between the sexes were much more striking in the registry populations. Females and males in the RCTs were more similar. Target dosing did not meaningfully differ between the sexes in any group, which emphasizes the impact of male-derived treatment guidelines and the need for this topic to be explored further.

Data on prognostic differences between males and females with HFrEF are conflicting, although females most often seem to fare better than males.[4,7,8,37–39] In the present study, females in both registry populations, i.e., RCT-eligible and -ineligible, experienced higher unadjusted mortality rates due to all causes and cardiovascular causes compared to males in the registries, whereas the mortality rates were roughly similar between males and females in the trials. However, after adjusting for known prognostic factors in HF, males in the RCTs had consistently higher mortality risk in comparison to males in the RCT-eligible population, with cardiovascular mortality risk 43% higher in the RCTs than expected in the registry. The higher percentage of cardiovascular death in the RCTs is consistent with use of enrichment strategies in inclusion/exclusion criteria. However, despite the same inclusion/exclusion criteria for males and females, females in trials showed no evidence of enrichment. On the contrary, there seemed a trend towards lower-than-expected cardiovascular mortality risk for females enrolled in trials compared to eligible females from the registries.

Enrichment strategies are often used in RCTs to identify patients who will experience cardiovascular events sooner than non-cardiovascular events in order to decrease time to target endpoint and reduce costs of the RCT.[40] It is unclear what could explain this opposing response to enrichment. One explanation could be that there are some unmeasured sex-specific factors affecting patient selection at the time of recruitment. There are numerous reports that point out that females can be underrepresented due to significant patient-oriented biopsychosocial barriers which results in the exclusion of females who are elderly, obese, depressed, non-white, with greater comorbidity, and who have less social support.[7,31–33] This could hold true for the studied population here, as baseline characteristic differences between the RCT and registry groups were larger for females than males. In addition, although females and males in the RCTs were prescribed medication similarly, females in the registries were less often prescribed anticoagulants and ACEI or ARBs, which is consistent with previous literature.[2,4,15] This is concerning because the use of ACEI or ARBs was a significant driver for RCT ineligibility in registry females and is possibly an additional barrier for female recruitment in RCTs. Patients in RCTs are also known to receive better care, and gender-related differences in clinical management have been

shown to negatively affect females in the real world.[37,39] Taken together, these barriers could lead to a healthier female RCT population that is less representative of their real-world counterparts, especially in comparison to males.

Lastly, it is also conceivable that risk factors used to calculate the SMRs have sex-specific impact. Although the MAGGIC risk model was chosen for this study due to its validity in predicting mortality for both sexes,[41,42] there are strong arguments that testing the interaction with sex in the models or that sex disaggregation of results should be the norm in cardiovascular research.[43] In addition, sex is currently used as a surrogate for underlying variables that determine gender such as caregiver strain, work strain, emotional intelligence, social support and discrimination. Although these variables are known to have a significant effect on health outcomes, current research remains sex-specific until reporting of gender variables increase.[44]

Study limitations

The strength of this study lies in large sample sizes and access to individual patient data from both trial and observational datasets. There are also potential limitations of the study. The harmonized criteria selected to define the RCT-eligible population were chosen based on data availability and commonality between trials. Therefore, the percentage of patients eligible for trial inclusion is likely overestimated but still considered appropriate for mortality analysis to make fairer comparisons between the RCT population and the real world. The RCTs involved in this study were selected based on the availability of data from industry partners. However, a comparison of baseline characteristics with contemporary trials does not show meaningful differences (online supplementary *Table S4*). Combining RCTs can always present a source of heterogeneity in participant characteristics due to different investigational drugs being studied, trial phases and study countries. However, our sensitivity analyses in the coronary artery disease and sinus rhythm subgroups support that differences in exclusion criteria between the RCTs do not affect the main conclusion. Pooling of the placebo and treatment arms does not allow extrapolation of the mortality rates and risk from this study; however, pooling does not explain the sex differences seen in these results which was the main research question and conclusion of these results. Lastly, registry patients are a fair representation of real-world patients, however there are likely to be some differences in characteristics and treatment practices between patients who were and were not enrolled in the registries. We also acknowledge that the trial and real-world populations differed in geographical location, healthcare systems and time of data collection.

Conclusion

In conclusion, the efficacy of enrichment in RCTs and the generalizability of RCTs towards the HFrEF population in the community differed substantially between the sexes, with females having lower trial participation and females who are enrolled in trials having lower than expected mortality rates compared to similar females in the registries, while males had higher than expected cardiovascular mortality rates in trials compared to similar males in registries. Failure to account for these differences and not stratifying future analysis by sex may influence appropriate translation of clinical trial results towards daily clinical practice or lead to under-powered RCTs because of ineffective enrichment.

Supplementary Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Acknowledgement

The authors are grateful for the assistance of Vasuki Kumpeson in the statistical analyses for the clinical trial data. Open Access funding enabled and organized by Projekt DEAL.

Funding

This study is part of the BigData@Heart project and is funded by the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement number 116074. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations. The funders had no role in the design, analysis, interpretation of data or writing of the manuscript.

Conflict of interest: M.S. is a full-time employee at Boehringer Ingelheim and received grants from Bayer AG for this work. G.S. reports grants and personal fees from Vifor, AstraZeneca, grants and non-financial support from Boehringer Ingelheim, personal fees from Società Prodotti Antibiotici, Roche, Servier, GENESIS, Cytokinetics, Medtronic, grants from Novartis, Boston Scientific, outside the submitted work. K.S.W. reports personal fees from Bayer AG, during the conduct of the study; personal fees from Bayer AG, outside the submitted work. C.B. reports personal fees from Institut de Recherches Internationales Servier during the conduct of the study. T.D. is a full-time employee of Bayer AG, Berlin, Germany. E.V. is a fulltime employee of Bayer AG, Berlin, Germany. J.J.B reports grants from Abbott, outside the submitted work. H.P.B.LR. reports grants and personal fees from Novartis, Vifor, Roche Diagnostics, personal fees from AstraZeneca, Boehringer Ingelheim, outside the submitted work. V.B.G. reports personal fees from Institut de Recherches Internationales Servier during the conduct of the study. S.W. is an employee of Vifor Pharma Int. B.T. reports personal fees from Institut de Recherches Internationales Servier, during the conduct of the study. L.H.L. reports personal fees from Merck, Sanofi, Bayer, Pharmacosmos, Abbott, Medscape, Myokardia, grants and personal fees from Vifor-Fresenius, AstraZeneca, Relypsa, Mundipharma, Boehringer Ingelheim, Novartis, grants from Boston Scientific, outside the submitted work. C.G. is a fulltime employee of Bayer AG, Berlin, Germany. M.C. reports personal fees from Vifor Pharma, during the conduct of the study; personal fees from Ava AG, outside the submitted work. All other authors have nothing to disclose.

References

1. Vogel B, Acevedo M, Appelman Y, Bairey Merz CN, Chieffo A, Figtree GA, et al. The Lancet women and cardiovascular disease Commission: reducing the global burden by 2030. *Lancet*. 2021;**397**:2385–438.
2. Lam CSP, Arnott C, Beale AL, Chandramouli C, Hilfiker-Kleiner D, Kaye DM, et al. Sex differences in heart failure. *Eur Heart J*. 2019;**40**:3859–3868c.

3. Beale AL, Meyer P, Marwick TH, Lam CSP, Kaye DM. Sex differences in cardiovascular pathophysiology: why women are overrepresented in heart failure with preserved ejection fraction. *Circulation*. 2018;**138**:198–205.
4. Lainščak M, Milinković I, Polovina M, Crespo-Leiro MG, Lund LH, Anker SD, et al. Sex- and age-related differences in the management and outcomes of chronic heart failure: an analysis of patients from the ESC HFA EORP Heart Failure Long-Term Registry. *Eur J Heart Fail*. 2020;**22**:92–102.
5. Lawson CA, Zaccardi F, Squire I, Ling S, Davies MJ, Lam CSP, et al. 20-year trends in cause-specific heart failure outcomes by sex, socioeconomic status, and place of diagnosis: a population-based study. *Lancet Public Health*. 2019;**4**: e406–20.
6. Meyer S, van der Meer P, van Deursen VM, Jaarsma T, van Veldhuisen DJ, van der Wal MHL, et al. Neurohormonal and clinical sex differences in heart failure. *Eur Heart J*. 2013;**34**:2358–47.
7. O'Meara E, Clayton T, McEntegart MB, McMurray JJV, Piña IL, Granger CB, et al. Sex differences in clinical characteristics and prognosis in a broad spectrum of patients with heart failure: results of the Candesartan in Heart Failure: Assessment of Reduction in Mortality and morbidity (CHARM) program. *Circulation*. 2007;**115**:3111–20.
8. Stolfo D, Uijl A, Vedin O, Strömberg A, Faxén UL, Rosano GMC, et al. Sex-based differences in heart failure across the ejection fraction spectrum: phenotyping, and prognostic and therapeutic implications. *JACC Heart Fail*. 2019;**7**: 505–15.
9. Greene SJ, DeVore AD, Sheng S, Fonarow GC, Butler J, Califf RM, et al. Representativeness of a heart failure trial by race and sex. *JACC Heart Fail*. 2019;**7**:980–92.
10. Tahhan AS, Vaduganathan M, Greene SJ, Fonarow GC, Fiuzat M, Jessup M, et al. Enrollment of older patients, women, and racial and ethnic minorities in contemporary heart failure clinical trials: a systematic review. *JAMA Cardiol*. 2018;**3**:1011–9.
11. Mentzer G, Hsich EM. Heart failure with reduced ejection fraction in women: epidemiology, outcomes, and treatment. *Heart Fail Clin*. 2019;**15**:19–27.
12. Norberg H. Clinical trial enrolment favours men: fewer women meet eligibility criteria for trials of heart failure medication is discussed by Helena Norberg from Sweden. *Eur Heart J*. 2019;**40**:1104–5.
13. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). With the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur J Heart Fail*. 2022;**24**:4–131.
14. Veenis JF, Brunner La Rocca HP, Linssen GCM, Erol-Yilmaz A, Pronk ACB, Engelen DJM, et al.; CHECK-HF Investigators. Impact of sex-specific target dose in chronic heart failure patients with reduced ejection fraction. *Eur J Prev Cardiol*. 2021;**28**:957–65.
15. Santema BT, Ouwerkerk W, Tromp J, Sama IE, Ravera A, Regitz-Zagrosek V, et al. Identifying optimal doses of heart failure medications in men compared with women: a prospective, observational, cohort study. *Lancet*. 2019;**394**: 1254–63.
16. Lim YMF, Molnar M, Vaartjes I, Savarese G, Eijkemans MJC, Uijl A, et al. Generalizability of randomized controlled trials in heart failure with reduced ejection fraction. *Eur Heart J Qual Care Clin Outcomes*. 2022;**8**:761–9.
17. Swedberg K, Komajda M, Böhm M, Borer JS, Ford I, Dubost-Brama A, et al.; SHIFT Investigators. Ivabradine and outcomes in chronic heart failure (SHIFT): a randomised placebo-controlled study. *Lancet*. 2010;**376**:875–85.
18. Fox K, Ferrari R, Tendera M, Steg PG, Ford I; BEAUTIFUL Steering Committee. Rationale and design of a randomized, double-blind, placebo-controlled trial of ivabradine in patients with stable coronary artery disease and left ventricular systolic dysfunction: the morBidity-mortality EvAlUaTion of the I(f) inhibitor ivabradine in patients with coronary disease and left ventricULar dysfunction (BEAUTIFUL) study. *Am Heart J*. 2006;**152**:860–6.
19. Ponikowski P, van Veldhuisen DJ, Comin-Colet J, Ertl G, Komajda M, Mareev V, et al.; CONFIRM-HF Investigators. Beneficial effects of long-term intravenous iron therapy with ferric carboxymaltose in patients with symptomatic heart failure and iron deficiency. *Eur Heart J*. 2015;**36**:657–68.
20. Anker SD, Colet JC, Filippatos G, Willenheimer R, Dickstein K, Drexler H, et al.; FAIR-HF Committees and Investigators. Rationale and design of Ferinject® Assessment in patients with IRon deficiency and chronic Heart Failure (FAIR-HF) study: a randomized, placebo-controlled study of intravenous iron supplementation in patients with and without anaemia. *Eur J Heart Fail*. 2009;**11**:1084–91.
21. Voors AA, Shah SJ, Bax JJ, Butler J, Gheorghiade M, Hernandez AF, et al. Rationale and design of the phase 2b clinical trials to study the effects of the partial adenosine A1-receptor agonist neladenoson bialanate in patients with chronic heart failure with reduced (PANTHEON) and preserved (PANACHE) ejection fraction. *Eur J Heart Fail*. 2018;**20**:1601–10.
22. Savarese G, Vasko P, Jonsson Å, Edner M, Dahlström U, Lund LH. The Swedish Heart Failure Registry: a living, ongoing quality assurance and research in heart failure. *Ups J Med Sci*. 2019;**124**:65–9.
23. Brugts JJ, Linssen GCM, Hoes AW, Brunner-La Rocca HP; CHECK-HF Investigators. Real-world heart failure management in 10,910 patients with chronic heart failure in The Netherlands: design and rationale of the Chronic Heart failure ESC guideline-based Cardiology practice Quality project (CHECK-HF) registry. *Neth Heart J*. 2018;**26**:272–9.
24. Taylor P. Standardized mortality ratios. *Int J Epidemiol*. 2013;**42**:1882–90.
25. Roessler M, Schmitt J, Schoffer O. Can we trust the standardized mortality ratio? A formal analysis and evaluation based on axiomatic requirements. *PLoS One*. 2021;**16**:e0257003.
26. Breslow NE, Day NE. Statistical methods in cancer research. Volume II – The design and analysis of cohort studies. *IARC Sci Publ*. 1987;1–406.
27. Rahimi K, Bennett D, Conrad N, Williams TM, Basu J, Dwight J, et al. Risk prediction in patients with heart failure: a systematic review and analysis. *JACC Heart Fail*. 2014;**2**:440–6.
28. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;**30**:377–99.
29. R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2023. Available from: https://www.R-project.org [last accessed 6 Jan 2023].
30. StataCorp. *Stata statistical software: release 15*. College Station, TX, USA. 2023. Available from: https://www.stata.com [last accessed 6 Jan 2023].
31. Eisenberg E, Di Palo KE, Piña IL. Sex differences in heart failure. *Clin Cardiol*. 2018;**41**:211–6.
32. Hsich EM, Grau-Sepulveda MV, Hernandez AF, Peterson ED, Schwamm LH, Bhatt DL, et al. Sex differences in in-hospital mortality in acute decompensated heart failure with reduced and preserved ejection fraction. *Am Heart J*. 2012;**163**:430–437.e3.
33. Sillars A, Ho FK, Pell GP, Gill JMR, Sattar N, Gray S, et al. Sex differences in the association of risk factors for heart failure incidence and mortality. *Heart*. 2020;**106**:203–12.
34. Azad N, Kathiravelu A, Minoosepeher S, Hebert P, Fergusson D. Gender differences in the etiology of heart failure: a systematic review. *J Geriatr Cardiol*. 2011;**8**:15–23.
35. Lawson CA, Zaccardi F, Squire I, Okhai H, Davies M, Huang W, et al. Risk factors for heart failure. *Circ Heart Fail*. 2020;**13**:e006472.
36. Waraich P, Goldner EM, Somers JM, Hsu L. Prevalence and incidence studies of mood disorders: a systematic review of the literature. *Can J Psychiatry*. 2004;**49**:124–38.
37. Aimo A, Vergaro G, Barison A, Maffei S, Borrelli C, Morrone D, et al. Sex-related differences in chronic heart failure. *Int J Cardiol*. 2018;**255**:145–51.
38. Scrutinio D, Guida P, Passantino A, Lagioia R, Raimondo R, Venezia M, et al. Female gender and mortality risk in decompensated heart failure. *Eur J Intern Med*. 2018;**51**:34–40.
39. Dewan P, Rørth R, Jhund PS, Shen L, Raparelli V, Petrie MC, et al. Differential impact of heart failure with reduced ejection fraction on men and women. *J Am Coll Cardiol*. 2019;**73**:29–40.
40. U.S. Food and Drug Administration. Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and biological Products. March 2019. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enrichment-strategies-clinical-trials-support-approval-human-drugs-and-biological-products [last accessed 6 Jan 2023].
41. Uijl A, Koudstaal S, Direk K, Denaxas S, Groenwold RHH, Banerjee A, et al. Risk factors for incident heart failure in age- and sex-specific strata: a population-based cohort using linked electronic health records. *Eur J Heart Fail*. 2019;**21**: 1197–206.
42. Vishram-Nielsen JKK, Foroutan F, Ross HJ, Gustafsson F, Alba AC. Performance of prognostic risk scores in heart failure patients: do sex differences exist? *Can J Cardiol*. 2020;**36**:45–53.
43. Woodward M. Rationale and tutorial for analysing and reporting sex differences in cardiovascular associations. *Heart*. 2019;**105**:1701–8.
44. Nielsen MW, Stefanick ML, Peragine D, Neilands TB, Ioannidis JPA, Pilote L, et al. Gender-related variables for health research. *Biol Sex Differ*. 2021;**12**:23.