ESC European Society of Cardiology

European Heart Journal - Quality of Care and Clinical Outcomes (2021) **0**, 1–9
https://doi.org/10.1093/ehjqcco/qcab070

ORIGINAL ARTICLE

Generalizability of randomized controlled trials in heart failure with reduced ejection fraction

Yvonne Mei Fong Lim [1,2], Megan Molnar[3], Ilonca Vaartjes[1], Gianluigi Savarese [4], Marinus J.C. Eijkemans[1], Alicia Uijl [1], Eleni Vradi[5], Kiliana Suzart-Woischnik[3], Jasper J. Brugts[6], Hans-Peter Brunner-La Rocca[7], Vanessa Blanc-Guillemaud[8], Fabrice Couvelard[8], Claire Baudier[8], Tomasz Dyszynski[3], Sandra Waechter[9], Lars H. Lund[4], Arno W. Hoes [1], Benoit Tyl[8], Folkert W. Asselbergs [10,11], Christoph Gerlinger[12,13], Diederick E. Grobbee[14,15], Maureen Cronin[16] and Stefan Koudstaal[1,10,17,*]

[1]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, PO Box 85500, 3508 GA Utrecht, the Netherlands; [2]Institute for Clinical Research, National Institutes of Health, 40170 Shah Alam, Malaysia; [3]Medical Affairs & Pharmacovigilance, Bayer AG, 13353 Berlin, Germany; [4]Division of Cardiology, Department of Medicine, Karolinska Insitutet, 171 76 Stockholm, Sweden; [5]Biomedical Data Science II, Bayer AG, 13353 Berlin, Germany; [6]Department of Cardiology, Erasmus MC University Medical Centre, 3015 GD Rotterdam, the Netherlands; [7]Department of Cardiology, Maastricht University Medical Center, 6229 HX Maastricht, the Netherlands; [8]Institut de Recherches Internationales SERVIER (I.R.I.S.), 92284 Suresnes, France; [9]Vifor Pharma Ltd, Glattbrugg, Switzerland; [10]Department of Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, Utrecht University, PO Box 85500, 3508 GA Utrecht, the Netherlands; [11]Institute of Cardiovascular Science and Institute of Health Informatics, Faculty of Population Health Sciences, University College London, NW1 2DA London, UK; [12]Statistics and Data Insights, Bayer AG, 13353 Berlin, Germany; [13]Gynecology, Obstetrics and Reproductive Medicine, University Medical School of Saarland, D-66421 Saar, Germany; [14]Julius Global Health, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, PO Box 85500, 3508 GA Utrecht, the Netherlands; [15]Julius Clinical, 3703 CD Zeist, the Netherlands; [16]Cronin Pharma Consulting, 6354 Lucerne, Switzerland; and [17]Department of Cardiology, Groene Hart Ziekenhuis, 2803 HH Gouda, the Netherlands

Received 20 June 2021; revised 17 September 2021; editorial decision 26 September 2021; accepted 29 September 2021; online publish-ahead-of-print 1 October 2021

Background Heart failure (HF) trials have stringent inclusion and exclusion criteria, but limited data exist regarding generalizability of trials. We compared patient characteristics and outcomes between patients with HF and reduced ejection fraction (HFrEF) in trials and observational registries.

Methods and Results Individual patient data for 16 922 patients from five randomized clinical trials and 46 914 patients from two HF registries were included. The registry patients were categorized into trial-eligible and non-eligible groups using the most commonly used inclusion and exclusion criteria. A total of 26 104 (56%) registry patients fulfilled the eligibility criteria. Unadjusted all-cause mortality rates at 1 year were lowest in the trial population (7%), followed by trial-eligible patients (12%) and trial-non-eligible registry patients (26%). After adjustment for age and sex, all-cause mortality rates were similar between trial participants and trial-eligible registry patients [standardized mortality ratio (SMR) 0.97; 95% confidence interval (CI) 0.92–1.03] but cardiovascular mortality was higher in trial participants (SMR 1.19; 1.12–1.27). After full case-mix adjustment, the SMR for cardiovascular mortality remained higher in the trials at 1.28 (1.20–1.37) compared to RCT-eligible registry patients.

Conclusion In contemporary HF registries, over half of HFrEF patients would have been eligible for trial enrolment. Crude clinical event rates were lower in the trials, but, after adjustment for case-mix, trial participants had similar rates of survival as registries. Despite this, they had about 30% higher cardiovascular mortality rates. Age and sex were the main drivers of differences in clinical outcomes between HF trials and observational HF registries.

Keywords Generalizability • External validity • Heart failure with reduced ejection fraction • Randomized clinical trials

* Corresponding author. Tel: +088 755 55 55, Email: s.koudstaal@umcutrecht.nl

Introduction

Randomized controlled trials (RCTs) are the gold standard for evaluating the efficacy and safety of investigational therapies due to their robust methodology conducted within a strict regulatory framework.[1] A well-conducted RCT has high internal validity, which ensures that the observed treatment effect is directly the result of the therapy tested.[1–4] However, high internal validity can come at the expense of external validity, defined as the degree to which the treatment effect found in the study can be generalized and replicated outside the RCT.[1] If the RCT results found in the study population are not generalizable to the target population, it is unclear which patients in routine care can receive a treatment safely and effectively.[1–5]

Physicians' uncertainty and criticism of RCTs' generalizability has been suggested as one reason for the underuse of evidence-based treatments, specifically in the field of heart failure (HF).[2,6] There is currently no consensus on how to assess generalizability, but a logical and important first step is to assess if an RCT study population is representative of the projected target population.[2–4,7] Studies comparing summary data on baseline characteristics between RCTs and observational data have already been conducted, specifically for heart failure with reduced ejection fraction (HFrEF).[5,8–10] Although these studies have shown differences in crude outcomes between trial and real-world patients, it is not known how differences in patient characteristics drive the observed differences in prognosis. In addition, some of these comparisons have been limited by the small sample sizes from single trials.

Here, we compared individual patient data of five HFrEF randomized clinical trials and two HF registries by direct data access and collaboration between academic researchers and pharmaceutical industry partners. We first determined their differences in patient characteristics, treatment, and clinical outcomes. Then, we identified the proportion of registry patients who were eligible for inclusion in the trials and compared their outcomes with trial participants while adjusting for known prognostic factors of HF at the individual patient level.

Methods

Data sources

Based on a collaboration with industry partners through the BigData@Heart Consortium[11], data access to patient level information was obtained for five randomized clinical trials in HFrEF patients. BEAUTIFUL and SHIFT were ivabradine trials ($n = 15\,732$),[12,13] FAIR-HF and CONFIRM were studies on intravenous iron supplementation ($n = 763$)[14,15] and PANTHEON was a trial for neladenosone bialanate ($n = 427$).[16] Of these, three were phase III trials, one was phase II and lastly, one phase IV study. All RCTs included HFrEF patients based on left ventricular ejection fraction (LVEF) values (ranging from ≤35% to ≤45%) except for the BEAUTIFUL study, which recruited coronary artery disease (CAD) patients who had left ventricular dysfunction. To maintain comparability between patients from the RCTs, only patients with New York Heart Association (NYHA) class II–IV from BEAUTIFUL ($n = 9227$) were included.

Aggregated data from both treatment and placebo arms of each RCT were pooled and compared against the HFrEF population from two observational data sources: the CHECK-HF and the SwedeHF registries.[17,18] Detailed information on the methods for both registries can be found elsewhere.[17,18] Briefly, the CHECK-HF registry included patients with chronic HF if they had an HF diagnosis based on ESC 2012 guidelines between 2013 and 2016.[17] The ongoing SwedeHF registry enrolled patients with clinician-judged HF patients in Sweden.[18] For the current analysis, outpatients registered between 2000 and 2016 ($n = 40\,230$) were included to ensure consistency with CHECK-HF.

Data from both registries were combined for describing patient characteristics and treatment but only SwedeHF data were used in the reporting on clinical outcomes because CHECK-HF did not have follow-up data. For each of the five trials, ethics approval and written informed consent were obtained by the respective study investigators.[12–16] CHECK-HF registry was granted ethics approval for anonymized analysis of existing patient data, whereas in the SwedeHF registry, enrolment was based on specific health centres' participation and patients allowed to opt-out should they wish not to participate.[17,18]

Eligibility criteria and outcomes

The inclusion and exclusion criteria listed in the study protocol of the five RCTs were tabulated (see Supplementary material online, Table S1) to identify common study entry criteria. These criteria were cross-checked for data availability within the registries and a set of most commonly used eligibility criteria was then identified to select subsets of RCT-eligible and non-eligible patients from the registries. The following inclusion criteria were used: age ≥18 years, LVEF <40%, NYHA functional class II–IV, on optimally tolerated chronic HF medications of β-blocker and angiotensin-converting enzyme inhibitor (ACEI) or angiotensin-II receptor blocker (ARB). Then, the following exclusion criteria were applied: serum haemoglobin concentrations <11 g/dL in men or <10 g/dL in women, chronic liver disease, creatinine >220 μmol/L and cancer.

Comparisons were made based on (i) patient baseline characteristics, (ii) cardiovascular medications, and (iii) mortality outcomes. For summary statistics, aggregated data were extracted from each trial and there were instances of low patient numbers in the data contingency tables. To maintain patient anonymity, all table cells with counts of 3 and below were replaced with a central number of 2.[19] For HF medications, the percentage of patients who received <50% or ≥50% target doses of the HF medications were assessed (see Supplementary material online, Table S2). Finally, the following clinical outcomes at 1 year were assessed: all-cause mortality, cardiovascular mortality (ICD-10 codes I00–I99) and first HF hospitalization (main diagnosis with codes I50, I11.0, I42.0, I42.3–I42.9, I43, I25.5, K76.1, I13.0, I32.2, or J81). Follow-up duration differed between the five trials. Three trials (BEAUTIFUL, CONFIRM-HF, and SHIFT) had follow-up data for at least 1 year, so outcome at 1 year was reported here. The remaining two trials (FAIR-HF and PANTHEON) had less than a year's follow-up and patients were censored at the end of study.

Statistical analysis

Continuous data are presented as mean with standard deviation whereas categorical variables are reported in frequencies and percentages. Mean and proportion differences between the RCT and RCT-eligible registry

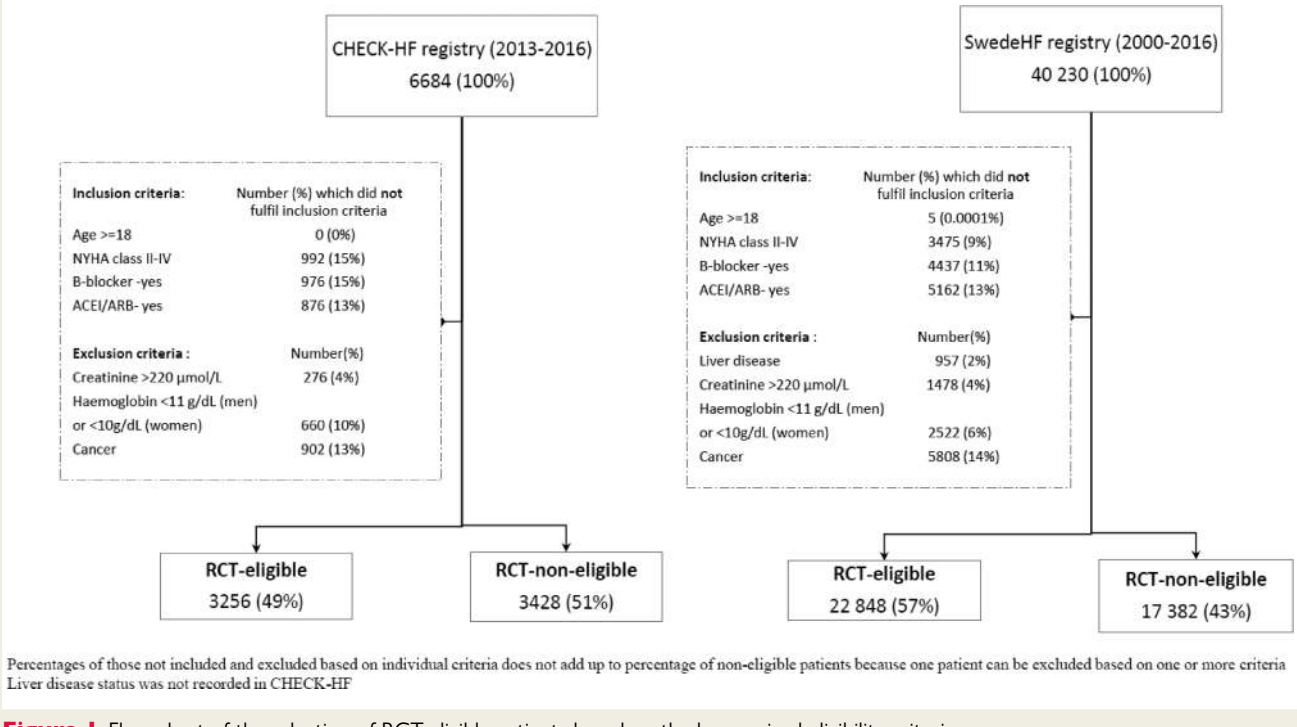


Figure 1 Flow chart of the selection of RCT-eligible patients based on the harmonized eligibility criteria.

patients were calculated and reported with their corresponding 99% confidence intervals (CI). Data are presented by three groups: (i) RCT participants, (ii) RCT-eligible, and (iii) RCT-non-eligible registry patients. Cumulative incidence curves were used to compare unadjusted outcomes between study groups. For cardiovascular mortality, deaths due to other causes were treated as competing events. For first HF hospitalization, all-cause deaths were treated as competing events. Then, standardized mortality ratios (SMRs) were used to compare adjusted mortality rates between the trials and the SwedeHF registry population. First, we fitted a Poisson model with 11 prognostic indicators from a validated MAGGIC HF risk score [age, sex, LVEF, NYHA class, serum creatinine, chronic obstructive pulmonary disease (COPD), diabetes, systolic blood pressure, body mass index (BMI), HF duration, smoking status] in a stepwise manner to the trial-eligible SwedeHF patients' data.[20,21] Next, the model with the derived β-coefficients was applied to each trial to estimate each individual's expected mortality, which was then summed across all participants to derive total expected mortality counts. The observed mortality count for each trial was divided by the expected mortality count to give the SMRs. An SMR value >1 indicated that the observed risk of mortality in a trial was higher than the risk predicted based on SwedeHF patients as the reference population. The SMR was risk-adjusted for 11 prognostic factors to address heterogeneity between the trials. This was considered sufficient adjustment to pool the trials using fixed effect meta-analysis without introducing partial pooling. The corresponding 95% CI was determined using methods described by Breslow and Day.[22] SMRs were not estimated for HF hospitalization because its existing risk prediction models do not have adequate discriminative performance compared to those designed to predict mortality.[23]

For cardiovascular causes of mortality, the Poisson model has taken into account competing risk from other causes of death as every patient's follow-up duration was included in the estimation of the number of events. Rather than predicting cumulative probabilities, the Poisson model gives a prediction of the number of events for each individual which can be summed to obtain the total expected number of events in a trial. Missing data were multiply imputed by chained equations using the mice package in R.[24] The number of imputations was set at 20.[25] Statistical significance was set at 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).[26,27]

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 CAD; therefore, sensitivity analyses were conducted in subsets of registry patients who were (i) in sinus rhythm or (ii) diagnosed with CAD. The fully adjusted SMRs from each subset were then compared to the original estimates. A third sensitivity analysis was performed to determine the effects of time period differences between trial and registry data on HF medication prescription.

Results

Study population

Majority of registry patients (56%) were eligible for inclusion in the trials (*Figure 1*). Compared to the overall registry group, RCT patients were younger (mean 63.6 years vs. 72.7 years), less frequently women (22% vs. 31%), had longer duration of HF, were more often in LVEF category of 30–39% as opposed to

Table 1 Characteristics of HFrEF patients by RCT and registry groups

		Registry population		RCT vs. RCT-eligible	
	RCT study population N = 16 922	RCT-eligible N = 26 104 (56%)	RCT-non-eligible N = 20 810 (44%)	Difference in mean or proportion (99% CI)	P-value[b]
Patient characteristics					
Age (years)	63.6 ± 10.0	71.1 ± 12.6	74.7 ± 13.1	−7.5 (−7.8, −7.2)	***
Women	3663 (22%)	8294 (32%)	6290 (30%)	−10.1% (−11.2%, −9.0%)	***
Body mass index (kg/m^2)	28.3 ± 4.4	27.1 ± 5.7	25.9 ± 6.1	−1.2 (−1.3, −1.1)	***
Systolic blood pressure—mmHg	125.2 ± 13.4	124.6 ± 21.0	124.5 ± 22.7	0.6 (0.1, 1.1)	***
Diastolic blood pressure—mmHg	76.6 ± 8.4	73.7 ± 12.3	72.3 ± 13.4	2.9 (2.6, 3.2)	***
Heart rate—beats per minute	74.4 ± 9.5	74.5 ± 15.9	75.7 ± 17.4	−0.1 (−0.4, 0.2)	
Serum creatinine—μmol/L	99.2 ± 29.5	99.4 ± 37.9	123.4 ± 91.2	−0.2 (−1.1, 0.6)	
Haemoglobin—g/dL	14.1 ± 1.3	13.7 ± 2.1	12.9 ± 2.6	0.4 (0.4, 0.4)	***
Current smoker	2667 (16%)	3370 (15%)[a]	2301 (13%)[a]	1.1% (0.1%, 2.0%)	**
Heart failure severity measures					
Duration of heart failure—months	42.0 ± 56.4	29.8 ± 61.7	32.9 ± 61.9	12.2 (10.7, 13.7)	***
LVEF categories—no (%)					
<30	5338 (32%)	13 936 (53%)	9751 (47%)	-	***
30–39	11 225 (66%)	12 168 (47%)	11 059 (53%)		
≥40	247 (1%)	-	-		
Missing	112(1%)	-	-		
Mean LVEF (%)	31	-	-	-	
NYHA Functional Class—no (%)					
I	3 (0.02%)	0 (0%)	4459 (21%)	-	***
II	10 394 (61%)	14 478 (55%)	7231 (35%)		
III	6422 (38%)	10 623 (41%)	7673 (37%)		
IV	113 (1%)	1003 (4%)	1447 (7%)		
Medical history—no (%)					
Hypertension	11 517(68%)	14 654 (56%)	11 505 (55%)	11.9% (10.7%, 13.2%)	***
Diabetes mellitus	5711 (34%)	7083 (27%)	5649 (27%)	6.6% (5.4%, 7.8%)	***
Coronary artery disease	14 541 (86%)	11 916 (52%)[a]	9497 (55%)[a]	33.8% (32.7%, 34.9%)	***
History of MI	5721 (34%)	8120 (31%)	6776 (33%)	2.7% (1.5%, 3.9%)	***
Atrial fibrillation	449 (38%)[c]	12 563 (48%)	10 014 (48%)	−10.4% (−14.1% −6.7%)	***
Valvular disease	2009 (12%)	5616 (22%)	5280 (25%)	−10.7% (−11.7%, −9.8%)	***
Stroke/TIA	1564 (9%)	3220 (14%)[a]	3067 (18%)[a]	−4.8% (−5.7%, −4.0%)	***
Anaemia	588 (3%)	6611 (25%)	9064 (44%)	−21.8% (−22.6%, −21.1%)	***
COPD	1482 (9%)	5084 (19%)	4417 (21%)	−10.7% (−11.6%, −9.9%)	***
Depression	451 (3%)	1117 (5%)[a]	894 (5%)[a]	−2.2% (−2.7%, −1.7%)	***
Cancer	462 (3%)	0 (0%)	6710 (32%)	2.7% (2.4%, 3.1%)	***
Coronary revascularization					
PCI	1538 (9%)	1994 (8%)	1528 (7%)	-	***
CABG	1029 (6%)	3038 (12%)	2609 (13%)		
PCI + CABG	236 (1%)	2913 (11%)	2020 (10%)		
Clinical outcomes at 1 year					
All-cause mortality	1112 (7%)	2674 (12%)[a]	4482 (26%)[a]	−5.1% (−5.9%, −4.4%)	***
Cardiovascular mortality	1005 (6%)	2026 (9%)[a]	3114 (18%)[a]	−2.9% (−3.6%, −2.3%)	***
First HF hospitalization	1399 (8%)	5544 (24%) [a]	4310 (25%)[a]	−16.0% (−16.9%, −15.1%)	***
Cardiovascular medications at baseline					
ACEI/ARB[d]	15 251 (90%)	26 104 (100%)	14 773 (71%)	-	-
Mineralocorticoid receptor antagonist	7294 (43%)	10 275 (40%)	6880 (33%)	3% (2%, 5%)	***
Antiplatelet	13 208 (78%)	12 329 (47%)	9788 (47%)	31% (30%, 32%)	***
β-Blocker[d]	14 808 (88%)	26 104 (100%)	15 392 (75%)	-	-

Table 1 Continued.

		Registry population		RCT vs. RCT-eligible	
	RCT study population *N* = 16 922	RCT-eligible *N* = 26 104 (56%)	RCT-non-eligible *N* = 20 810 (44%)	Difference in mean or proportion (99% CI)	*P*-value[b]
Diuretic	12 120 (72%)	20 697 (79%)	16 379 (79%)	−8% (−9%, −7%)	***
Digitalis	2500 (15%)	4447 (17%)	3002 (14%)	−2% (−3%, −1%)	***
Statins	11 231 (66%)	13 995 (54%)	9674 (47%)	13% (11%, 14%)	***

Values are expressed as mean ± standard deviation or number (%).
Percentages may not add up to 100% due to rounding.
*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.
ACEI, angiotensin-converting enzyme inhibitor; ARB, angiotensin-II receptor blocker; CABG, coronary artery bypass graft; CI, confidence intervals; COPD, chronic obstructive pulmonary disease; HF, heart failure; LVEF, left ventricular ejection fraction; MI, myocardial infarction; PCI, percutaneous coronary intervention; TIA, transient ischaemic attack.
[a] Data from SwedeHF only.
[b] Comparison between RCT and registry (RCT-eligible) population (independent *t*-test for continuous and ^{2}-test for categorical variables).
[c] RCT data only from CONFIRM, FAIR HF, and PANTHEON.
[d] Statistical comparisons were not done for ACEI/ARB and β-blocker because these treatments were part of the criteria for selecting RCT-eligible registry patients.

<30% and predominantly in NYHA class II rather than class III–IV (*Table 1*). The baseline characteristics of each registry are provided in Supplementary material online, Table S3.

Hypertension, diabetes, and CAD were more common in the RCT group compared to the overall registry group. However, the proportion of patients with valve disease, stroke, anaemia, COPD, cancer, and coronary revascularization were markedly lower in the RCT patients. After restricting the registry group to those who would be eligible for inclusion in the RCTs, this RCT-eligible registry group was more similar to the RCT group in NYHA class, serum creatinine, and haemoglobin, but differences in comorbidities largely remained (*Table 1*). In the selection of trial-eligible patients, the most restrictive inclusion criteria were NYHA class II–IV and the use of ACEI/ARB and ß-blockers while the most restrictive exclusion criterion was cancer (*Figure 1*).

Use and target doses of cardiovascular medication

Prescription of medications was higher for antiplatelets, mineralocorticoid receptor antagonists, and statins in the RCTs compared with registry patients. Despite similar proportions in use of ACEI/ARB (87% vs. 90%), more registry than RCT patients received higher doses (≥50% of target doses) of these medications (see Supplementary material online, Table S4). We then restricted the comparison to the same time periods (2005–09) between the two largest trials and SwedeHF registry patients and found that the proportion of patients who were given target doses did not differ much from the main findings, which used data from 2001 to 2016 (see Supplementary material online, Table S5).

Clinical outcomes at one year

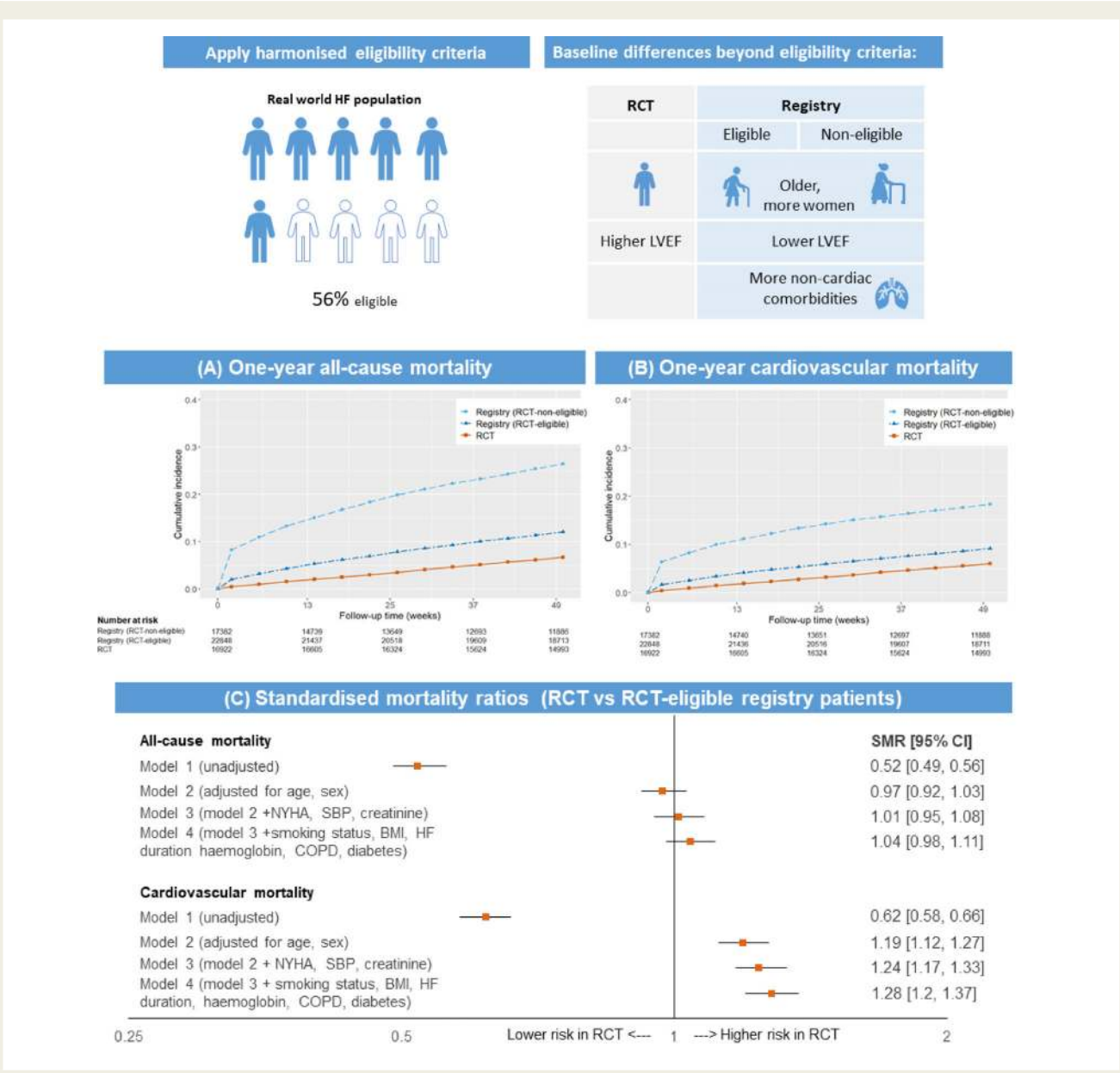
Cumulative incidence curves are shown in the central illustration and *Figure 2*. All-cause mortality, cardiovascular mortality, and first HF hospitalization at 1 year were lower in the RCTs than in trial-eligible and trial non-eligible registry groups.

There was no remaining difference in all-cause mortality risk between trial and registry patients after adjusting for known HF prognostic factors [fully adjusted (Model 4) SMR 1.04; 95%CI 0.98–1.11)] (Central Illustration). However, higher cardiovascular mortality risk persisted in the RCT group compared to trial-eligible registry patients [fully adjusted (Model 4) SMR 1.28; 95%CI 1.20–1.37)]. Age and sex explained most of the mortality difference between patient groups, as reflected in the large shift of SMR between Model 1 (empty model) to Model 2 (with age and sex). Stepwise addition of prognostic factors changed SMR in the same direction but to a lesser degree, as seen in the shift of SMR in Model 2 (with age and sex) to Model 4 (fully adjusted) for all-cause and cardiovascular mortality.

Sensitivity analyses were conducted by estimating SMRs in a subset of patients who were in sinus rhythm and estimates were similar to those obtained in the main results (see Supplementary material online, Figures S1 and S2).

Discussion

The present study has individual patient data of over 62 000 patients from five clinical trials and two observational HF registries, which allowed direct and adjusted comparisons on patient characteristics for both all-cause and cause-specific mortality. Overall, we found that over half of patients in the registries met the most commonly used inclusion and exclusion criteria for trial enrolment. Unadjusted survival was markedly lower in registries than trials. However, after adjusting for case-mix, all-cause mortality rates were comparable between the trials and registries whereas cardiovascular mortality occurred more frequently in the trial participants compared to registry patients.



Central illustration. Cumulative incidence and case-mix adjusted SMRs for all-cause and cardiovascular mortality at 1 year. (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 and cardiovascular mortality with stepwise adjustment for HF prognostic factors. Pooled SMRs estimated from five trials with their 95% CI were reported.

We identified a higher proportion of trial-eligible patients compared to previous studies on patients with acute decompensated HF and HF with reduced and preserved ejection fraction: 56% vs. 13–42%.[8,28,29] Furthermore, the percentage of trial-eligible registry patients who were given at least 50% target doses of HF medications were slightly higher than in RCTs. This higher proportion compared to previous reports could be explained at least in part by extensive HF programmes and nurse-led up-titration of disease-modifying therapies in the Netherlands and Sweden. Also, data in the registries were from more recent years than the trials, thus reflecting more contemporary prescribing practices. Accordingly, we would expect background therapies in newer HF trials to be at a higher rate than the ones described here. Therefore, our findings, along with other recent studies in acute HF suggest that the gap in HF guideline-adherent treatment between trial and real-world patients is narrowing.[6,30]

The differences observed between trial participants and trial-eligible registry patients highlight other factors besides eligibility criteria that influence patient selection in RCTs. Physicians intuitively recruit patients who are deemed less likely to drop out to ensure low attrition rates which retain high internal validity.[31–33] Older patients and those with comorbidities are not always

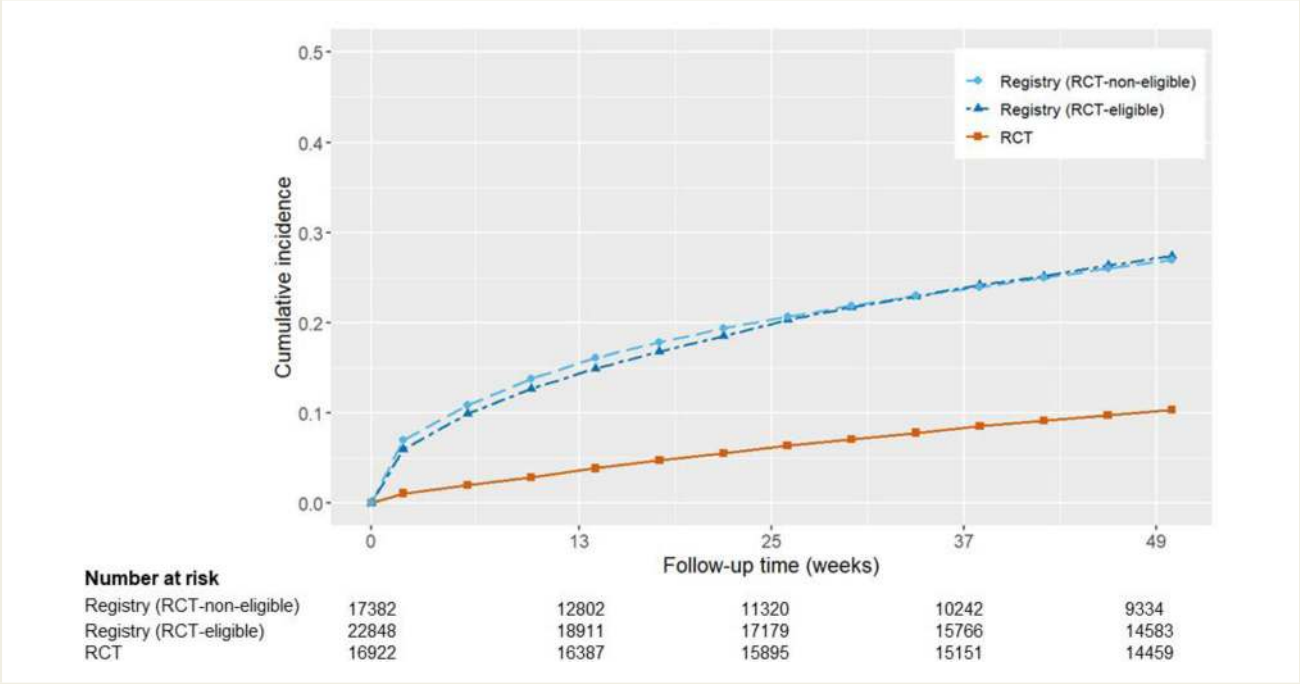


Figure 2 Cumulative incidence curves for first HF hospitalization at 1 year by (i) RCT participants, (ii) RCT-eligible, and (iii) RCT-non-eligible registry patients.

physically or mentally able to comply and finish the treatment protocol due to frailty, low mobility and increased risk for adverse events.[7,34] Women with HF tend to be older and are less likely to participate due to perceived harm from clinical studies, transportation difficulties, or constraints from a caregiving role.[33,35,36] Consequently, the additional criteria introduced by investigators alongside the eligibility criteria consistently cause underrepresentation of older patients, those with comorbidities and women in CV trials.[37] However, expanding the study population to include these groups would increase the cost of already expensive HF trials, and other solutions to improve generalizability that have been proposed include individual participant data meta-analysis, proper reporting of subgroup analysis, registry-based trials, or comparative effectiveness studies.[38–40] The growing trend to conduct RCTs as site-less or direct-to-patient studies may reduce this bias in the future.

We have shown, by direct comparisons between study groups that the risk of mortality and HF hospitalization was lowest in the trial population. However, after accounting for known prognostic factors for survival in HF, differences in survival between trial and registry patients disappeared. In fact, age and sex combined explained the largest variation in SMRs between trials and registries. This observation is evident for both all-cause and cardiovascular mortality and highlights their important contribution on the generalizability of HF trials.

Taken together, it seems that differences in overall survival between HF trials and registries behave predictably, and could be addressed by clinical variables which are readily available in daily clinical practice. Although well accepted, we have demonstrated for the first time that there are increased cardiovascular mortality rates in the HF trial participants compared to trial-eligible registry patients, as high up to 30% even after adjustment for prognostic factors. From a drug developer and/or regulatory perspective, prognostic enrichment strategies were advocated and used in many cardiovascular trials to identify patients who have higher likelihood of cardiovascular events.[32] In addition, excluding patients with other comorbidities in these trials could lead to lower competing risks of death from non-cardiovascular causes. On a broadly similar note, trial-eligible registry patients selected for the PARADIGM-HF trial criteria had higher risk of both cardiovascular and non-cardiovascular mortality compared to non-eligible registry patients.[41] From the clinicians' perspective, it is important to be aware that half of patients were ineligible, and that even among trial-eligible patients, residual differences between cardiovascular and non-cardiovascular outcomes persists.

Strengths and limitations

The strength of this study lies in the large sample sizes from both trial and observational datasets. Direct access to individual patient data also enabled the reporting of case-mix-adjusted differences in outcomes between trials and registry. There are also several limitations to this study. First, we applied a harmonized set of criteria which were common across the trials based only on data that were also available from the registries. There was not sufficient depth in the data from the registries to assess many of the eligibility criteria such as worsening HF in the past 12 months, scheduled coronary revascularization within 3 months or severe valve disease. Also, not all criteria per RCT have been considered but only the most common ones. For these reasons, the percentage of patients eligible for trial inclusion is likely overestimated. The trials included in this study

were a convenient sample based on data accessibility; thus, it can be difficult to infer these findings to other HF trials. Second, a large proportion of trials patients came from two RCTs which excluded patients with atrial fibrillation (BEAUTIFUL and SHIFT), which might have impacted the results. However, we believe that this impact is not substantial, as supported by sensitivity analyses (see Supplementary material online, Figures S1 and S2). Although the trials evaluated here were not the most recent HFrEF trials, we do not expect large changes in patient and clinical characteristics among those enrolled in trials then and now. This is supported by a baseline characteristics comparison with DAPA-HF and PARADIGM-HF, which showed comparable patient characteristics in terms of mean age, percentage of women, percentage in NYHA class III/IV and mean LVEF, except for percentage with atrial fibrillation which was lower in this study.[42] It is also necessary to note that, although registry patients are a fair representation of real-world patients, 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 on geographical location, healthcare systems and time of data collection.[43]

Conclusion

In summary, over half of patients in registries met the most commonly used inclusion and exclusion criteria for potential trial enrolment. In terms of generalizability, age and sex were the main drivers of differences in clinical outcomes between HF trials and observational HF registries. As expected, HF trial participants showed higher residual cardiovascular mortality rates after correction for case mix.

Supplementary material

Supplementary material is available at *European Heart Journal—Quality of Care and Clinical Outcomes* online.

Acknowledgements

The authors are grateful for the assistance of Vasuki Kumpeson and Loïs Ravel-Sibillot in the statistical analyses for the clinical trial data.

Funding

This study is part of the BigData@Heart project and is funded by the Innovative Medicines Initiative 2 Joint Undertaking (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.

Conflict of interest M.M. reports grants from Bayer AG, outside the submitted work. G.S. reports grant and personal fees from Vifor, grants and non-financial support from Boehringer Ingelheim, personal fees from Societa´ Prodotti Antibiotici, grants and personal fees from AstraZeneca, personal fees from Roche, personal fees from Servier, grants from Novartis, personal fees from GENESIS, personal fees from Cytokinetics, personal fees from Medtronic, grants from Boston Scientific, outside the submitted work. E.V. is a fulltime employee of Bayer AG, Berlin, Germany. K.S.-W. reports personal fees from Bayer AG, during the conduct of the study; personal fees from Bayer AG, outside the submitted work. J.J.B reports grants from Abbot, outside the submitted work. H.-P.B.-LR reports grants and personal fees from Novartis, grants and personal fees from Vifor, grants and personal fees from Roche Diagnostics, personal fees from AstraZeneca, personal fees from Boehringer-Ingelheim outside the submitted work. V.B.-G. is a fulltime employee of Institut de Recherches Internationales Servier during the conduct of the study. F.C. is a fulltime employee of Institut de Recherches Internationales Servier. C.B. is a fulltime employee of Institut de Recherches Internationales Servier during the conduct of the study. T.D. is a full-time employee of Bayer AG, Berlin, Germany. S.W. is an employee of Vifor Pharma Int. L.H.L. reports personal fees from Merck, personal fees from Sanofi, grants and personal fees from Vifor-Fresenius, grants and personal fees from AstraZeneca, grants and personal fees from Relypsa, personal fees from Bayer, grants from Boston Scientific, personal fees from Pharmacosmos, personal fees from Abbott, grants and personal fees from Mundipharma, personal fees from Medscape, personal fees from Myokardia, grants and personal fees from Boehringer Ingelheim, grants and personal fees from Novartis, outside the submitted work. B.T. is a fulltime employee of Institut de Recherches Internationales Servier, during the conduct of the study. 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. The other authors have no conflicts of interest to disclose.

Data availability

The data underlying this article cannot be shared publicly due to privacy and ethical restrictions.

References

1. Spieth PM, Kubasch AS, Penzlin AI, Illigens BM-W, Barlinn K, Siepmann T. Randomized controlled trials – a matter of design. *Neuropsychiatr Dis Treat* 2016;**12**:1341–1349.
2. Rothwell PM. External validity of randomised controlled trials: 'to whom do the results of this trial apply?' *Lancet Lond Engl* 2005;**365**:82–93.
3. Dekkers OM, Elm E von, Algra A, Romijn JA, Vandenbroucke JP. How to assess the external validity of therapeutic trials: a conceptual approach. *Int J Epidemiol* 2010;**39**:89–94.
4. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results: a potential outcomes perspective. *Epidemiol Camb Mass* 2017;**28**:553–561.
5. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* 2015;**16**:495.
6. Lenzen MJ, Boersma E, Reimer WJMSO, Balk AHMM, Komajda M, Swedberg K *et al.* Under-utilization of evidence-based drug treatment in patients with heart failure is only partially explained by dissimilarity to patients enrolled in landmark trials: a report from the Euro Heart Survey on Heart Failure. *Eur Heart J* 2005;**26**:2706–2713.
7. Hanlon P, Hannigan L, Rodriguez-Perez J, Fischbacher C, Welton NJ, Dias S *et al.* Representation of people with comorbidity and multimorbidity in clinical trials of novel drug therapies: an individual-level participant data analysis. *BMC Med* 2019;**17**:201.
8. Miró Ò, Gil V, Müller C, Mebazaa A, Bueno H, Martín-Sánchez FJ *et al.* How does a clinical trial fit into the real world? The RELAX-AHF study population into the EAHFE registry. *Clin Res Cardiol Off J Ger Card Soc* 2015;**104**:850–860.
9. Patel HC, Hayward C, Dungu JN, Papadopoulou S, Saidmeerasah A, Ray R *et al.* Assessing the eligibility criteria in phase III randomized controlled trials of drug therapy in heart failure with preserved ejection fraction: the critical play-off

between a 'pure' patient phenotype and the generalizability of trial findings. *J Card Fail* 2017;**23**:517–524.
10. Kapelios CJ, Lainscak M, Savarese G, Laroche C, Seferovic P, Ruschitzka F *et al.* Heart Failure Long-Term Registry Investigators. Sacubitril/valsartan eligibility and outcomes in the ESC-EORP-HFA Heart Failure Long-Term Registry: bridging between European Medicines Agency/Food and Drug Administration label, the PARADIGM-HF trial, ESC guidelines, and real world. *Eur J Heart Fail* 2019;**21**:1383–1397.
11. Innovative Medicines Initiative. BigData@Heart. IMI Innov. Med. Initiat. 2021. http://www.imi.europa.eu/projects-results/project-factsheets/bigdataheart (7 September 2021).
12. 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**:860–866.
13. Swedberg K, Komajda M, Böhm M, Borer JS, Ford I, Tavazzi L. Rationale and design of a randomized, double-blind, placebo-controlled outcome trial of ivabradine in chronic heart failure: the Systolic Heart Failure Treatment with the I(f) Inhibitor Ivabradine Trial (SHIFT). *Eur J Heart Fail* 2010;**12**:75–81.
14. 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**:1084–1091.
15. Ponikowski P, Veldhuisen DJ van, Comin-Colet J, Ertl G, Komajda M, Mareev V *et al.* Rationale and design of the CONFIRM-HF study: a double-blind, randomized, placebo-controlled study to assess the effects of intravenous ferric carboxymaltose on functional capacity in patients with chronic heart failure and iron deficiency. *ESC Heart Fail* 2014;**1**:52–58.
16. 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**:1601–1610.
17. 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 Mon J Neth Soc Cardiol Neth Heart Found* 2018;**26**:272–279.
18. 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* 2018;**0**:1–5.
19. Schoffer O. Statistik in Sachsen. *SAS im Forschungsdatenzentrum der Statistischen Landesämter* Sachsen; 2008. p. 123–126.
20. Sartipy U, Dahlström U, Edner M, Lund LH. Predicting survival in heart failure: validation of the MAGGIC heart failure risk score in 51,043 patients from the Swedish heart failure registry. *Eur J Heart Fail* 2014;**16**:173–179.
21. Loomis D, Richardson D, Elliott L. Poisson regression analysis of ungrouped data. *Occup Environ Med* 2005;**62**:325–329.
22. Breslow N, Day N. *Statistical Methods in Cancer Research. Volume II: The Design and Analysis of Cohort Studies.* IARC Scientific Publication No. 82.
23. 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**:440–446.
24. Buuren S van, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *Journal of Statistical Software* 2011;**45**:1–67.
25. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;**30**:377–399.
26. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019.
27. StataCorp. *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LP; 2017.
28. Wang TS, Hellkamp AS, Patel CB, Ezekowitz JA, Fonarow GC, Hernandez AF. Representativeness of RELAX-AHF clinical trial population in acute heart failure. *Circ Cardiovasc Qual Outcomes* 2014;**7**:259–268.
29. Rodrigues G, Tralhão A, Aguiar C, Freitas P, Ventosa A, Mendes M. Is the PARADIGM-HF cohort representative of the real-world heart failure patient population? *Rev Port Cardiol (Engl Ed)*2018;**37**:491–496.
30. Ezekowitz JA, Hu J, Delgado D, Hernandez AF, Kaul P, Leader R *et al.* Acute heart failure: perspectives from a randomized trial and a simultaneous registry. *Circ Heart Fail* 2012;**5**:735–741.
31. Schulz K, Grimes DA. *Essential Concepts in Clinical Research: Randomised Controlled Trials and Observational Epidemiology*. Elsevier Health Sciences; 2018.
32. U.S. Food & Drug Administration. *Enrichment Strategies for Clinical Trials to Support Determination of Effectiveness of Human Drugs and Biological Products Guidance for Industry*. Maryland: U.S Food & Drug Administration; 2019.
33. Martin SS, Ou F-S, Newby LK, Sutton V, Adams P, Felker GM *et al.* Patient- and trial-specific barriers to participation in cardiovascular randomized clinical trials. *J Am Coll Cardiol* 2013;**61**:762–769.
34. Vitale C, Spoletini I, Rosano G. Frailty in heart failure: implications for management. *Card Fail Rev* 2018;**4**:104–106.
35. Beckie TM, Mendonca MA, Fletcher GF, Schocken DD, Evans ME, Banks SM. Examining the challenges of recruiting women into a cardiac rehabilitation clinical trial. *J Cardiopulm Rehabil Prev* 2009;**29**:13–23.
36. Cheung AM, Lee Y, Kapral M, Scher J, Ho I, Lui-Yee D *et al.* Barriers and motivations for women to participate in cardiovascular trials. *J Obstet Gynaecol Can* 2008;**30**:332–337.
37. Van Spall HGC, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA* 2007;**297**:1233–1240.
38. Lund LH, Oldgren J, James S. Registry-based pragmatic trials in heart failure: current experience and future directions. *Curr Heart Fail Rep* 2017;**14**:59–70.
39. Zuidgeest MGP, Goetz I, Groenwold RHH, Irving E, Thiel GJMW van, Grobbee DE. Series: pragmatic trials and real world evidence: paper 1. Introduction. *J Clin Epidemiol* 2017;**88**:7–13.
40. Shah RU, Chang TI, Fonarow GC. Comparative effectiveness research in heart failure therapies: women, elderly patients, and patients with kidney disease. *Heart Fail Clin* 2013;**9**:79–92.
41. Savarese G, Hage C, Benson L, Schrage B, Thorvaldsen T, Lundberg A *et al.* Eligibility for sacubitril/valsartan in heart failure across the ejection fraction spectrum: real-world data from the Swedish Heart Failure Registry. *J Intern Med* 2020.
42. McMurray JJV, DeMets DL, Inzucchi SE, Køber L, Kosiborod MN, Langkilde AM *et al.* The Dapagliflozin and prevention of adverse-outcomes in heart failure (DAPA-HF) trial: baseline characteristics. *Eur J Heart Fail* 2019;**21**:1402–1411.
43. World Health Organization. *The World Health Report 2000. Health Systems: Improving Performance*. Geneva, Switzerland: World Health Organization; 2000.