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

Megan Schroeder¹,²*, Yvonne Mei Fong Lim³,⁴, Gianluigi Savarese², Kiliana Suzart-Woischnik⁵, Claire Baudier⁶, Tomasz Dyszynski⁵, Ilonca Vaartjes³, Marinus J.C. Eijkemans³, Alicia Uijl²,⁷, Ronald Herrera⁵, Eleni Vradi⁸, Jasper J. Brugts⁷, Hans-Peter Brunner-La Rocca⁹, Vanessa Blanc-Guillemaud⁹, Sandra Waechter⁶, Fabrice Couvelard⁹, Benoit Tyl⁹, Samuel Fatoba⁵, Arno W. Hoes³, Lars H. Lund², Christoph Gerlinger¹¹,¹², Folkert W. Asselbergs⁷,¹³, Diederick E. Grobbee³,¹⁴, Maureen Cronin¹⁵, and Stefan Koudstaal¹⁶

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

© 2023 The Authors. European Journal of Heart Failure published by John Wiley & Sons Ltd on behalf of European Society of Cardiology. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
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). Females and males differ in HF aetiology, age, risk factors, biomarkers, pathophysiology, comorbidities, and clinical presentation. 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.

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%, whereas the percentage of enrolled females in HFrEF trials is on average 24%. As a consequence, contemporary treatment guidelines are still predominantly based on male-derived data. Post-hoc analyses stratified for sex from trials, and observational data, currently suggest that females may need lower dosages. 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. Briefly, five HFrEF RCTs and two HF registries were included in this study. BEAUTIFUL and SHIFT were phase III ivabradine trials (n = 15732). FAIR-HF and CONFIRM were phase III and phase IV studies on intravenous iron supplementation (n = 763) and PANTHEON was a phase II trial for neladenosone bialanate (n = 427). For final analysis, aggregated data from both treatment and placebo arms of each RCT were pooled to represent one RCT population (n = 16917).

The Dutch CHECK-HF and SwedeHF registries enrol patients with clinician-judged HF and detailed information on the methods can be found elsewhere. 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 = 40230) 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). 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). However, one limitation in SMR analysis is the inability to account for case-mix between populations. 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 DerSimonian and Laird.

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. 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).

Results

Eligibility for potential trial enrolment

Out of 46914 patients from the registries, 14584 were females (31.1%). After applying the harmonized set of eligibility and ineligibility criteria, 8294 out of 14584 (56.9%) females and 17818 out of 32330 (55.1%) males in the registries were considered eligible for RCT recruitment for a final RCT-eligible group of 26104 (31.8% females). Cancer was the most restricting criterion with 27.4% of females and 28.1% of males excluded. The exclusion criteria...
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 16917 (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. 26.4%, 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%).
Sex differences in HFrEF RCTs

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

<table>
<thead>
<tr>
<th></th>
<th>RCT Females</th>
<th>RCT Males</th>
<th>p-value&lt;sup&gt;a&lt;/sup&gt;</th>
<th>RCT-eligible registry Females</th>
<th>RCT-eligible registry Males</th>
<th>p-value&lt;sup&gt;a&lt;/sup&gt;</th>
<th>RCT-ineligible registry Females</th>
<th>RCT-ineligible registry Males</th>
<th>p-value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>3663</td>
<td>13254</td>
<td></td>
<td>8294</td>
<td>17810</td>
<td></td>
<td>6290</td>
<td>14520</td>
<td></td>
</tr>
<tr>
<td>Demographics and lifestyle, mean (SD) or %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>66.3 (9.9)</td>
<td>62.8 (9.9)</td>
<td></td>
<td>73.9 (9.1)</td>
<td>69.8 (12.2)</td>
<td></td>
<td>76.7 (9.7)</td>
<td>73.8 (9.3)</td>
<td></td>
</tr>
<tr>
<td>Smoking history</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>72.4%</td>
<td>29.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous/current</td>
<td>27.6%</td>
<td>71.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical parameters, mean (SD) or %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF duration (months)</td>
<td>42.4 (58.4)</td>
<td>41.6 (58.7)</td>
<td></td>
<td>21.0 (58.6)</td>
<td>30.8 (58.8)</td>
<td></td>
<td>21.1 (32.2)</td>
<td>24.4 (31.2)</td>
<td></td>
</tr>
<tr>
<td>SBP (mmHg)</td>
<td>126.5 (14.9)</td>
<td>124.9 (14.9)</td>
<td></td>
<td>126.4 (16.0)</td>
<td>123.7 (20.5)</td>
<td></td>
<td>125.7 (17.1)</td>
<td>124.0 (15.2)</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>28.6 (5.5)</td>
<td>28.2 (4.6)</td>
<td></td>
<td>26.6 (6.5)</td>
<td>27.2 (6.0)</td>
<td></td>
<td>25.6 (4.7)</td>
<td>26.0 (4.3)</td>
<td></td>
</tr>
<tr>
<td>Creatinine (µmol/L)</td>
<td>90.3 (33.9)</td>
<td>101.6 (41.3)</td>
<td></td>
<td>90.6 (37.6)</td>
<td>103.4 (38.2)</td>
<td></td>
<td>108.9 (0.3)</td>
<td>129.7 (60.6)</td>
<td></td>
</tr>
<tr>
<td>LVEF (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–29</td>
<td>28.7%</td>
<td>32.6%</td>
<td></td>
<td>47.0%</td>
<td>56.1%</td>
<td></td>
<td>43.0%</td>
<td>48.6%</td>
<td></td>
</tr>
<tr>
<td>30–39</td>
<td>67.4%</td>
<td>66.6%</td>
<td></td>
<td>53.0%</td>
<td>43.9%</td>
<td></td>
<td>57.0%</td>
<td>51.4%</td>
<td></td>
</tr>
<tr>
<td>≥40</td>
<td>3.9%</td>
<td>0.8%</td>
<td></td>
<td>0.0%</td>
<td>0.0%</td>
<td></td>
<td>0.0%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>NYHA functional class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>0.0%</td>
<td>0.1%</td>
<td></td>
<td>0.0%</td>
<td>0.0%</td>
<td></td>
<td>17.9%</td>
<td>22.9%</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>53.5%</td>
<td>63.5%</td>
<td></td>
<td>51.9%</td>
<td>57.1%</td>
<td></td>
<td>35.4%</td>
<td>34.5%</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>45.8%</td>
<td>35.7%</td>
<td></td>
<td>43.5%</td>
<td>39.4%</td>
<td></td>
<td>38.5%</td>
<td>36.2%</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>0.7%</td>
<td>0.7%</td>
<td></td>
<td>4.6%</td>
<td>3.5%</td>
<td></td>
<td>8.2%</td>
<td>6.4%</td>
<td></td>
</tr>
<tr>
<td>III/IV</td>
<td>46.5%</td>
<td>36.4%</td>
<td></td>
<td>48.1%</td>
<td>42.9%</td>
<td></td>
<td>46.6%</td>
<td>42.6%</td>
<td></td>
</tr>
<tr>
<td>Comorbidities %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>72.0%</td>
<td>67.0%</td>
<td></td>
<td>59.1%</td>
<td>54.7%</td>
<td></td>
<td>56.6%</td>
<td>54.7%</td>
<td></td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>34.7%</td>
<td>33.5%</td>
<td></td>
<td>25.8%</td>
<td>27.8%</td>
<td></td>
<td>25.7%</td>
<td>27.8%</td>
<td></td>
</tr>
<tr>
<td>CAD</td>
<td>79.3%</td>
<td>87.8%</td>
<td></td>
<td>49.0%</td>
<td>48.6%</td>
<td></td>
<td>48.1%</td>
<td>57.4%</td>
<td></td>
</tr>
<tr>
<td>Valvular heart disease</td>
<td>14.9%</td>
<td>11.0%</td>
<td></td>
<td>24.2%</td>
<td>20.2%</td>
<td></td>
<td>26.5%</td>
<td>25.0%</td>
<td></td>
</tr>
<tr>
<td>Stroke or TIA</td>
<td>9.1%</td>
<td>9.3%</td>
<td></td>
<td>14.3%</td>
<td>14.0%</td>
<td></td>
<td>16.6%</td>
<td>18.1%</td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation/flutter</td>
<td>9.7%</td>
<td>8.9%</td>
<td></td>
<td>42.2%</td>
<td>46.7%</td>
<td></td>
<td>47.8%</td>
<td>54.9%</td>
<td></td>
</tr>
<tr>
<td>COPD</td>
<td>6.9%</td>
<td>9.3%</td>
<td></td>
<td>21.3%</td>
<td>18.6%</td>
<td></td>
<td>22.1%</td>
<td>20.9%</td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>4.6%</td>
<td>2.1%</td>
<td></td>
<td>6.2%</td>
<td>4.3%</td>
<td></td>
<td>6.5%</td>
<td>4.6%</td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>4.0%</td>
<td>2.4%</td>
<td></td>
<td>0.0%</td>
<td>0.0%</td>
<td></td>
<td>–</td>
<td>32.4%</td>
<td>32.2%</td>
</tr>
<tr>
<td>Concomitant medications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACEI or ARB</td>
<td>90.2%</td>
<td>89.8%</td>
<td></td>
<td>100.0%</td>
<td>100.0%</td>
<td></td>
<td>–</td>
<td>63.7%</td>
<td>74.2%</td>
</tr>
<tr>
<td>Anticoagulant</td>
<td>2.8%</td>
<td>2.6%</td>
<td></td>
<td>40.7%</td>
<td>48.1%</td>
<td></td>
<td>33.4%</td>
<td>41.0%</td>
<td></td>
</tr>
<tr>
<td>Antiplatelet</td>
<td>74.7%</td>
<td>79.0%</td>
<td></td>
<td>48.2%</td>
<td>46.8%</td>
<td></td>
<td>*</td>
<td>45.5%</td>
<td>47.7%</td>
</tr>
<tr>
<td>MRA</td>
<td>47.1%</td>
<td>41.9%</td>
<td></td>
<td>38.4%</td>
<td>39.8%</td>
<td></td>
<td>*</td>
<td>33.9%</td>
<td>32.7%</td>
</tr>
<tr>
<td>Beta-blocker</td>
<td>87.3%</td>
<td>87.5%</td>
<td></td>
<td>100.0%</td>
<td>100.0%</td>
<td></td>
<td>–</td>
<td>71.6%</td>
<td>75.0%</td>
</tr>
<tr>
<td>Digitalis glycoside</td>
<td>14.9%</td>
<td>14.7%</td>
<td></td>
<td>17.1%</td>
<td>17.0%</td>
<td></td>
<td>15.7%</td>
<td>13.9%</td>
<td></td>
</tr>
<tr>
<td>Diuretic</td>
<td>78.9%</td>
<td>69.6%</td>
<td></td>
<td>81.9%</td>
<td>78.1%</td>
<td></td>
<td>81.0%</td>
<td>77.7%</td>
<td></td>
</tr>
</tbody>
</table>

**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.

<sup>a</sup>Percentages may not add up to 100% due to rounding.
<sup>b</sup>Comparison between males and females (independent t-test for continuous variables and χ² test for categorical variables).
<sup>c</sup>Data from SwedeHF only.
<sup>d</sup>Statistical comparisons were not compared because they were part of the criteria for selecting RCT-eligible registry patients.

<sup>p</sup> < 0.05.
<sup>p</sup> < 0.01.
<sup>p</sup> < 0.001.
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 1.0 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 from RCT-eligible females, compared to 43% more observed cardiovascular deaths in male trial participants than expected from RCT-eligible males (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% CI 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 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.

### 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 1.0 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% CI 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).

---

Table 2

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Week 0</th>
<th>Week 12</th>
<th>Week 24</th>
<th>Week 36</th>
<th>Week 48</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males (RCT)</td>
<td>13,259</td>
<td>13,020</td>
<td>12,574</td>
<td>12,059</td>
<td>11,829</td>
</tr>
<tr>
<td>Males (RCT-eligible)</td>
<td>15,685</td>
<td>14,770</td>
<td>14,180</td>
<td>13,543</td>
<td>12,946</td>
</tr>
<tr>
<td>Males (RCT-ineligible)</td>
<td>12,249</td>
<td>10,445</td>
<td>9,680</td>
<td>9,009</td>
<td>8,437</td>
</tr>
<tr>
<td>Females (RCT)</td>
<td>7,163</td>
<td>6,601</td>
<td>6,292</td>
<td>5,993</td>
<td>5,688</td>
</tr>
<tr>
<td>Females (RCT-eligible)</td>
<td>7,163</td>
<td>6,601</td>
<td>6,292</td>
<td>5,993</td>
<td>5,688</td>
</tr>
<tr>
<td>Females (RCT-ineligible)</td>
<td>5,133</td>
<td>4,245</td>
<td>3,915</td>
<td>3,609</td>
<td>3,316</td>
</tr>
</tbody>
</table>

---

© 2023 The Authors. European Journal of Heart Failure published by John Wiley & Sons Ltd on behalf of European Society of Cardiology.
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.
We confirm that there are sex-related differences in clinical profiles and gender-related differences in clinical management have been reported in RCTs. Patients in RCTs are also known to receive better care, 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 prescribed anticoagulants and ACEI or ARBs, which is consistent with previously literature.35 This is concerning because the use of ACEI or ARBs was a significant driver for RCT ineligibility in registry populations, i.e., RCT-eligible and -ineligible, experienced 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 lower-than-expected cardiovascular mortality risk for females.34,35

We confirm that there are sex-related differences in clinical profiles and gender-related differences in clinical management have been reported in RCTs. Patients in RCTs are also known to receive better care, 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 prescribed anticoagulants and ACEI or ARBs, which is consistent with previously literature.35 This is concerning because the use of ACEI or ARBs was a significant driver for RCT ineligibility in registry populations, i.e., RCT-eligible and -ineligible, experienced 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 lower-than-expected cardiovascular mortality risk for females.34,35
shown to negatively affect females in the real world. \cite{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, \cite{31,42} 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. \cite{40} 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. \cite{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 underestimated 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 Project 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 full-time employee of Bayer AG, Berlin, Germany. J.J.B reports grants from Abbott, outside the submitted work. H.P.B.L. 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 full-time 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


© 2023 The Authors. *European Journal of Heart Failure* published by John Wiley & Sons Ltd on behalf of European Society of Cardiology.


