nature medicine

Article

https://doi.org/10.1038/s41591-024-03342-7

Combined endurance and resistance exercise training in heart failure with preserved ejection fraction: a randomized controlled trial

Received: 2 February 2024

Accepted: 3 October 2024

Published online: 02 January 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

Endurance exercise training (ET) is an effective treatment in heart failure with preserved ejection fraction (HFpEF), but the efficacy of resistance training in this patient population has been only scarcely evaluated. In this multicenter, randomized trial, we evaluated the effects of combined endurance and resistance training over 12 months in patients with HFpEF. The primary endpoint was a modified Packer score, including all-cause mortality, hospitalizations classified as potentially related to heart failure or exercise and changes in peak oxygen consumption ($\dot{V}O_2$), diastolic function (E/e′), New York Heart Association (NYHA) class and global self-assessment (GSA). In total, 322 patients (mean age, 70 years; 192 females (59.6%) and 130 males (40.4%)) were randomized (1:1) to ET or usual care (UC). At 12 months, the modified Packer score showed an improvement in 33 ET patients (20.5%) and in 13 UC patients (8.1%) and showed a worsening in 69 ET patients (42.9%) and in 71 UC patients (44.1%) (Kendall's tau-b = −0.073, $P = 0.17$). Although the primary endpoint was not met, clinically relevant differences favoring the ET group as compared to the UC group were observed for the following secondary endpoints: changes in peak $\dot{V}O_2$ (mean difference, 1.3 ml kg^{-1} min^{-1} (95% confidence interval (CI): 0.4–2.1)) and NYHA class (odds ratio = 7.77 (95% CI: 3.73–16.21)). No significant between-group differences were observed for other secondary endpoints, including change in E/e′, change in GSA, time to cardiovascular hospitalization or all-cause mortality. In conclusion, 1 year of combined endurance and resistance ET did not result in a significantly better modified Packer score, but it did result in improvements in important clinical parameters, such as peak $\dot{V}O_2$ and NYHA class, as compared to UC. ISRCTN registration: ISRCTN86879094.

e-mail: frank.edelmann@dhzc-charite.de

https://doi.org/10.1038/s41591-024-03342-7

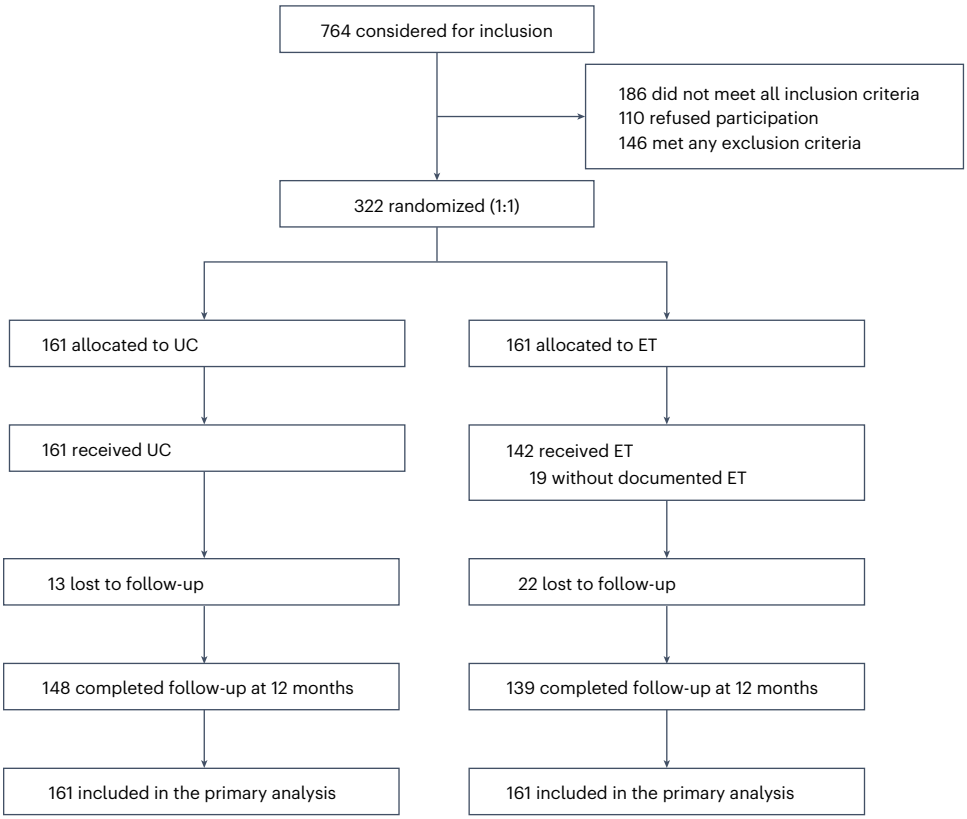


Fig. 1 | CONSORT diagram of the Ex-DHF trial. Among 764 patients who were screened for eligibility, 322 were randomized to UC or ET. All randomized patients were included in the primary analysis (intention-to-treat).

Heart failure (HF) is a major public health burden, with approximately 6.7 million individuals affected in the United States between 2017 and 2020 (ref. 1). At least half of patients with HF have preserved ejection fraction (HFpEF)[2], and the prognosis is similar to HF with reduced ejection fraction (HFrEF)[3]. However, in contrast to numerous guideline-recommended pharmacological and non-pharmacological management options in HFrEF, only sodium-glucose co-transporter 2 inhibitors, the glucagon-like peptide 1 receptor antagonist semaglutide and the non-steroidal mineralocorticoid receptor antagonist finerenone have so far improved clinical outcomes in HFpEF[4–9]. A hallmark clinical feature of HFpEF is reduced exercise tolerance, measured as decreased peak oxygen consumption (VO_2) assessed during cardiopulmonary exercise testing (CPET), secondary to impaired cardiac, vascular and skeletal muscle function[10,11]. Exercise training (ET) is an effective therapy to improve peak $\dot{V}O_2$ in clinically stable patients with HFpEF[10,12]. However, most studies have been short term (≤6 months) or limited to endurance training[10,12], and prolonged endurance training (moderate continuous or high-intensity interval training) performed over 12 months failed to result in significant improvements in peak $\dot{V}O_2$ (ref. 13). Combining endurance and resistance training (with or without additional caloric restriction) has been observed to significantly improve peak $\dot{V}O_2$ by 2.4–3.3 ml kg^{-1} min^{-1} over 12–20 weeks in individuals with HFpEF (including our Exercise training in Diastolic Heart Failure (Ex-DHF) pilot trial)[14,15]; however, long-term effects remain unknown. Moreover, because HFpEF is a systemic, multi-organ syndrome with diverse pathophysiological and clinical presentation[10,16], clinical parameters, such as HF symptoms, peak $\dot{V}O_2$, left ventricular function, health-related quality of life or morbidity in isolation, may not satisfactorily reflect the overall disease state. Therefore, Packer proposed combining symptom severity outcomes and clinical events in a hierarchical, three-level ordinal score to evaluate the efficacy of HF therapies[17,18]. Based on the results of the Ex-DHF pilot trial[14], we followed this approach and modified the Packer score for patients with HFpEF[19] to evaluate whether 12 months of supervised endurance and resistance ET will be superior to usual care (UC). This paper focuses on the primary endpoint, on safety outcomes and on secondary endpoints related to the components of the primary endpoint.

Results

Patient characteristics

Among 764 patients screened in 11 study centers in Germany and Austria, 322 patients (mean age, 70 years; 60% female, 86% New York Heart Association (NYHA) class II, 99% Caucasian) were enrolled between 21 December 2011 and 21 March 2016. Of those, 161 patients were randomized to UC and 161 to ET, of whom 142 (88%) had at least one documented ET session. At 12 months, 13 patients (8.1%) and 22 patients (13.7%) were lost to follow-up in UC and ET, respectively (Fig. 1). Last patient out was in March 2017. At baseline, patients were on stable medication for at least 4 weeks; 84% had a history of arterial hypertension; 31% had atrial fibrillation; and 38% had documented coronary artery disease (Table 1)[20].

Primary endpoint

At follow-up, the modified Packer score showed improvement in 20.5% of patients in the ET arm and in 8.1% of patients in the UC arm and worsening in 42.9% in the ET arm and 44.1% in the UC arm. The primary analysis showed no significant difference between groups (Kendall's tau-b = −0.073, $P = 0.17$; Fig. 2). Regarding the components of the primary endpoint, no significant difference was observed between groups for number of all-cause deaths ($P > 0.99$) or hospitalizations (potentially) related to exercise or HF ($P = 0.54$), whereas significant group differences favoring ET were observed for the clinical score (odds ratio (OR) = 2.43 (95% confidence interval (CI): 1.55–3.81)) and two of its components—that is, change in peak $\dot{V}O_2$ and NYHA class (Table 2).

https://doi.org/10.1038/s41591-024-03342-7

Table 1 | Demographic and clinical characteristics at baseline

	Exercise training (*n*=161)	Usual care (*n*=161)
Sex, no. (%)		
Female	100 (62.1)	92 (57.1)
Male	61 (37.9)	69 (42.9)
Age, years; mean (s.d.)	69.1 (7.4)	70.1 (7.1)
Body mass index, kg m^{-2}; mean (s.d.)	29.6 (5.4)	29.6 (5.1)
Heart rate, beats/min; mean (s.d.)	66 (11)	66 (11)
Blood pressure, mmHg; mean (s.d.)		
Systolic	128 (13)	130 (13)
Diastolic	76 (10)	77 (9)
NYHA class, no. (%)		
II (slight limitation of physical activity)	138 (85.7)	139 (86.3)
III (marked limitation of physical activity)	23 (14.3)	22 (13.7)
Smoking, no. (%)		
Ex-smoker	73 (45.3)	80 (49.7)
Current smoker	6 (3.7)	9 (5.6)
Hypertension, no. (%)	135 (83.9)	134 (83.2)
Atrial fibrillation, no. (%)	45 (28.0)	53 (32.9)
Coronary artery disease, no. (%)	67 (41.6)	56 (34.8)
Diabetes mellitus, no. (%)	39 (24.2)	29 (18)
Sleep apnea, no. (%)	23 (14.3)	17 (10.6)
Dyslipidemia, no. (%)	107 (66.5)	93 (57.8)
Medication		
ACE inhibitor/ARB, no. (%)	122 (75.8)	120 (74.5)
Beta receptor blocker, no. (%)	121 (75.2)	105 (65.2)
Mineralocorticoid receptor antagonist, no. (%)	14 (8.7)	10 (6.2)
Diuretics, no. (%)	74 (46)	80 (49.7)
Loop diuretic	25 (15.5)	35 (21.7)
Other diuretic	58 (36)	56 (34.8)
Laboratory		
Hemoglobin, mmol L^{-1}; mean (s.d.)	8.56 (0.81)	8.66 (0.83)
Anemia[a], no. (%)	19 (11.9)	16 (10.1)
Glomerular filtration rate, ml min^{-1} 1.73 m^{-2}; mean (s.d.)	55.7 (18.2)	56.2 (19.9)
Renal dysfunction[b], no. (%)	29 (18.2)	31 (19.6)
NT-proBNP, pg ml^{-1}; median (lower, upper quartile)	287 (129, 501)	253 (112, 529)
	Exercise training (*n*=161)	**Usual care (*n*=161)**
Echocardiography		
LVEF, %; no. (%)	60.5 (5.2)	62 (5.7)
LVEDD, mm; mean (s.d.)	44.9 (5.5)	43.9 (5.3)
LAVI, ml m^{-2}; mean (s.d.)	40.2 (12.8)	40.7 (12.7)
LVMI, g m^{-2}; mean (s.d.)	98.7 (22)	97.9 (23.1)
Male, mean (s.d.)	108 (24)	107 (23)
Female, mean (s.d.)	94 (19)	93 (22)
E, cm s^{-1}; mean (s.d.)	90.7 (24.9)	90.7 (27)
e' septal, cm s^{-1}; mean (s.d.)	6.01 (1.4)	5.97 (1.64)
E/e' septal, mean (s.d.)	15.7 (4.9)	16.2 (6.8)
CPET		
Peak VO_2, ml min^{-1} kg^{-1}; mean (s.d.)	18.0 (3.9)	17.8 (3.9)
Percent of reference values[c], mean (s.d.)	85.9 (16.9)	85.2 (18.4)
VO_2, anaerobic threshold, ml min^{-1} kg^{-1}; mean (s.d.)	12.2 (3.1)	12.2 (3.1)
Maximum RER, mean (s.d.)	1.09 (0.09)	1.09 (0.09)
Peak systolic blood pressure, mmHg; mean (s.d.)	179 (25)	180 (28)
Peak heart rate, beats/min; mean (s.d.)	121 (23)	120 (25)
Chronotropic incompetence[d], no. (%)	26 (16.1)	41 (25.5)
VE/VCO_2 slope, mean (s.d.)	32.5 (5.5)	32.8 (6.2)

ACE, angiotensin-converting enzyme; ARB, angiotensin II receptor blocker; LAVI, left atrial volume index; LVEDD, left ventricular end-diastolic diameter; LVEF, left ventricular ejection fraction; LVMI, left ventricular mass index; RER, respiratory exchange ratio; VCO_2, carbon dioxide production; VE, minute ventilation. [a] Anemia was defined as hemoglobin less than 13 g dl^{-1} (8.07 mmol L^{-1}) in men and less than 12 g dl^{-1} (7.45 mmol L^{-1}) in women. [b] Renal dysfunction was defined as glomerular filtration rate less than 60 ml min^{-1} 1.73 m^{-2}. [c] Percent predicted peak VO_2 is based on reference values from the Study of Health in Pomerania (SHIP)[20]. [d] Chronotropic incompetence was defined as less than 80% of predicted peak heart rate (220 − age) in patients not taking beta blockers and less than 62% of predicted peak heart rate in patients taking beta blockers.

The other two components of the clinical score (change in early diastolic transmitral flow velocity/early diastolic mitral annular tissue velocity (E/e') and global self-assessment (GSA)) were not significantly different between groups (Table 2). Associations between the categories of each component of the modified Packer score are shown in Supplementary Tables 1–6, and the primary reasons for a worsening of the modified Packer score are shown in Extended Data Fig. 1.

Secondary endpoints

Change in peak $\dot{V}O_2$ was significantly different between groups at 6 months and 12 months (mean differences of 0.8 ml kg^{-1} min^{-1} (95% CI: 0.0–1.6) and 1.3 ml kg^{-1} min^{-1} (95% CI: 0.4–2.1), respectively; Fig. 3). Similarly, change in NYHA functional class was also significantly different between groups at 6 months and 12 months (OR for a better outcome in ET versus UC of 2.36 (95% CI: 1.28–4.35), and 7.77 (95% CI: 3.73–16.21), respectively). In contrast, change in septal E/e' and GSA were not significantly different between groups at 6 months and 12 months (Fig. 3). The number of patients with available data for the secondary endpoints is shown in Fig. 3 or Supplementary Table 7. Time to all-cause death or cardiovascular hospitalization was also not significantly different between groups (log-rank test $P > 0.99$; Fig. 4). After multiple imputation, results were similar compared to the complete case analysis (Extended Data Table 1).

Safety

A total of 176 all-cause hospitalizations occurred in 62 patients (38.5%) in the ET group and in 54 patients (33.5%) in the UC group ($P = 0.35$). Cardiovascular hospitalizations (total of 86 events) occurred in 33 patients (20.5%) in the ET arm and in 31 patients (19.3%) in the UC arm ($P = 0.78$). The most common diagnosis during cardiovascular hospitalization was atrial fibrillation (21 events in 16 ET patients (9.9%) versus 15 events in 10 UC patients (6.2%)). Worsening HF was reported 11 times in nine ET patients (5.6%) and six times in six UC patients (3.7%). There was one death due to septic shock after diagnosis of acute leukemia in the UC group. Blinded evaluation by the endpoint

https://doi.org/10.1038/s41591-024-03342-7

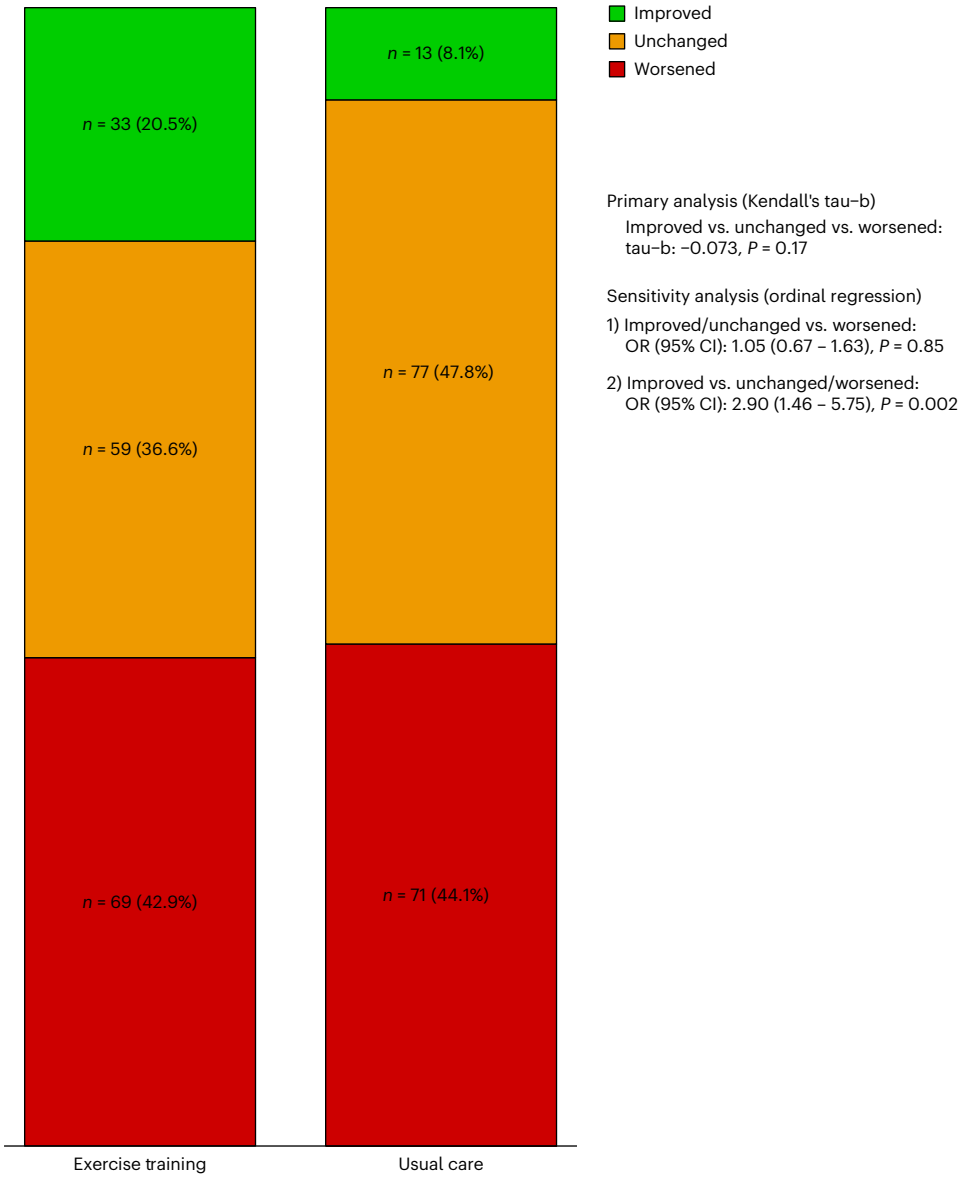


Fig. 2 | Results of the modified Packer score (primary endpoint) at 12 months. Results are shown for the ET (left) and UC (right) groups. For the primary endpoint, a two-sided test of Kendall's tau-b was performed. Values of the tau-b parameter range from −1 (all ET patients are in a better category than UC patients) to 1 (all UC patients are in a better category than ET patients). A value of 0 indicates that the distribution is identical between groups. For the pre-specified sensitivity analysis, as the assumption of proportional odds was not met, ordinal regression analyses were performed by comparing (1) 'improved/unchanged' versus 'worsened' and (2) 'improved' versus 'unchanged/worsened'. The two-sided ordinal regression analyses were adjusted for randomization strata (baseline peak VO_2 less than versus equal to or higher than 20 ml kg^{-1} min^{-1}; NYHA functional class II/III), without adjustment for multiple comparisons. ORs and 95% CIs define the odds of ET patients being in the favorable category.

committee independently identified 73 hospitalizations in 56 patients as potentially related to exercise (intervention group: 13 events in 12 patients; UC: six events in six patients) and/or HF (intervention group: 41 events in 31 patients; UC: 28 events in 22 patients) (Table 2). Based on a post hoc analysis of all hospitalizations in the ET group, 11 hospitalizations in 11 different patients were directly related to additional contacts between patients and healthcare professionals—that is, a health-related problem was identified before, during or after ET. These include five hospitalizations related to atrial fibrillation or atrial flutter, two hospitalizations related to increased blood pressure, two hospitalizations related to dizziness, one hospitalization due to angina and one hospitalization due to dyspnea (Extended Data Table 2). No acute event (such as myocardial infarction, stroke, etc.) was directly related to an ET session.

Adherence

Overall, 11,847 supervised training sessions and 1,626 home-based sessions were reported for 142 patients over 12 months in the ET group. From 139 ET patients who completed the 12-month follow-up, 53 (38.1%) and 66 (47.5%) patients performed an average of at least two supervised and overall (including home-based) sessions per week, respectively (Extended Data Fig. 2). Reasons for non-adherence were related to personal reasons ($n = 49$), orthopedic reasons ($n = 23$), HF ($n = 5$), pulmonological reasons ($n = 4$), other clinical reasons ($n = 29$) or other non-clinical reasons ($n = 23$) (multiple reasons possible).

Pre-specified sensitivity analyses

Secondary analysis of the primary endpoint (ordinal regression analysis without proportional odds assumption) revealed a significant group

https://doi.org/10.1038/s41591-024-03342-7

Table 2 | Results for the components of the modified Packer score

	No. of patients (%)		P-value
	Exercise training	Usual care	
Components of the overall score			
All-cause death	0 (0%)	1 (0.6%)	
OR (95% CI) for favored outcome in exercise training	-		>0.99
Hospitalizations potentially related to heart failure or exercise	33 (18.6%)	23 (13.7%)	
OR (95% CI) for favored outcome in exercise training	0.65 (0.36–1.16)		0.54
Dropouts			
Qualifying for 'unchanged' modified Packer score	4 (2.5%)	1 (0.6%)	
Qualifying for 'worsened' modified Packer score	18 (11.2%)	12 (7.5%)	
OR (95% CI) for better category in exercise training	0.64 (0.30–1.38)		0.25
Clinical score (based on change in septal E/e', peak $\dot{V}O_2$, NYHA class, GSA)			
Improved	39 (24.2%)	14 (8.7%)	
Unchanged	93 (57.8%)	98 (60.9%)	
Worsened	29 (18.0%)	49 (30.4%)	
OR (95% CI) for better category in exercise training	2.43 (1.55–3.81)		<0.001
Components of the clinical score			
Septal E/e'			
Improved (decrease ≥15%)	38 (23.6%)	39 (24.2%)	
Unchanged (< ±15%)	92 (57.1%)	79 (49.1%)	
Worsened (increase ≥15% and <30%)	13 (8.1%)	19 (11.8%)	
Severely worsened (increase ≥30%)	18 (11.2%)	24 (14.9%)	
OR (95% CI) for better category in exercise training	1.20 (0.79–1.82)		0.39
Peak $\dot{V}O_2$			
Improved (increase ≥15%)	53 (35.3%)	24 (15.5%)	
Unchanged (< ±15%)	87 (50.7%)	105 (63.9%)	
Worsened (decrease ≥15% and <30%)	15 (10.0%)	25 (16.1%)	
Severely worsened (decrease ≥30%)	6 (4.0%)	7 (4.5%)	
OR (95% CI) for better category in exercise training	1.09 (1.03–1.14)		<0.001
NYHA class			
Improved (decrease)	46 (33.3%)	10 (6.8%)	
Unchanged	112 (69.6%)	136 (84.5%)	
Worsened (increase)	3 (1.9%)	15 (9.5%)	
OR (95% CI) for better category in exercise training	5.89 (3.08–11.25)		<0.001
GSA			
Improved (increase ≥2 levels or to best level)	16 (9.9%)	22 (13.7%)	
Unchanged (≥ −1 level to ≤ +1 level)	128 (79.5%)	116 (72.0%)	
Worsened (decrease ≥2 levels or to worst level)	15 (9.3%)	15 (9.3%)	
Severely worsened (decrease ≥3 levels or by 2 levels to worst level)	2 (1.2%)	8 (5.0%)	
OR (95% CI) for better category in exercise training	1.02 (0.62–1.69)		0.94

The components of the clinical score are based on the difference between baseline and the last available timepoint and were assessed using ordinal regressions analyses. Exact *P* values for all-cause death, clinical score, peak VO_2 and NYHA class were 1.00, 0.0001, 0.0003 and 0.0000001, respectively.

difference favoring ET when comparing 'improved' versus 'unchanged/worsened' (OR (95% CI) = 2.90 (1.46–5.75), $P = 0.002$) but not when comparing 'improved/unchanged' versus 'worsened' (OR (95% CI) = 1.05 (0.67–1.63), $P = 0.85$) (Fig. 2). In the per-protocol analysis (excluding patients who were lost to follow-up or had major protocol deviations), the modified Packer score showed an improvement in 25% in the ET arm and in 8% in the UC arm and worsening in 35% in the ET arm and in 41% in the UC arm (Extended Data Fig. 3).

Post hoc analyses

Post hoc evaluation of the components of the primary endpoint using a more widely applied hierarchical win ratio revealed significantly more wins with ET than with UC (win ratio: 1.32 (95% CI: 1.01–1.73), $P = 0.04$; Extended Data Fig. 4). Subgroup analyses did not show any significant effect modifications (for the primary endpoint and change in peak $\dot{V}O_2$) related to age, peak $\dot{V}O_2$, NYHA class or use of diuretics (interaction $P > 0.05$; Extended Data Fig. 5). Adherence to ET was significantly associated with the modified Packer score (test for trend in proportions: $P = 0.002$; Extended Data Fig. 6) and the change in peak $\dot{V}O_2$ ($P < 0.001$; Extended Data Table 3).

Discussion

Ex-DHF is, to our knowledge, the largest randomized controlled exercise intervention trial in patients with HFpEF assessing the effects of a 12-month combined thrice-weekly endurance and resistance exercise program on a clinically relevant composite endpoint score. Among 322 clinically stable patients with HFpEF and a mean age of 70 years (86% with NYHA class II), the modified Packer score, including all-cause mortality, exercise-related or HF-related hospitalizations and a clinical score consisting of investigator-reported and patient-reported symptom severity (NYHA class and GSA), myocardial diastolic function (E/e') and maximal exercise capacity (peak $\dot{V}O_2$) did not show a significant overall improvement with ET. In contrast, the results of a pre-defined secondary analysis revealed a significant difference favoring ET when comparing participants in the 'improved' versus 'unchanged/worsened' categories of the modified Packer score. Although ET was associated with an improvement in 20% of patients, this was observed in only 8% in the UC group. Irrespective of group assignment, almost half of the patients enrolled into the trial had a worsened modified Packer score at 12 months, reflecting the progressive and deteriorating nature of HFpEF. Post hoc analysis using a hierarchical win ratio significantly favored ET over UC.

Analysis of the components of the primary endpoint showed that the number of hospitalizations that were classified as potentially related to HF or exercise and dropouts associated with a worse clinical condition were not significantly different between groups (but numerically higher in the exercise group), whereas the clinical score

https://doi.org/10.1038/s41591-024-03342-7

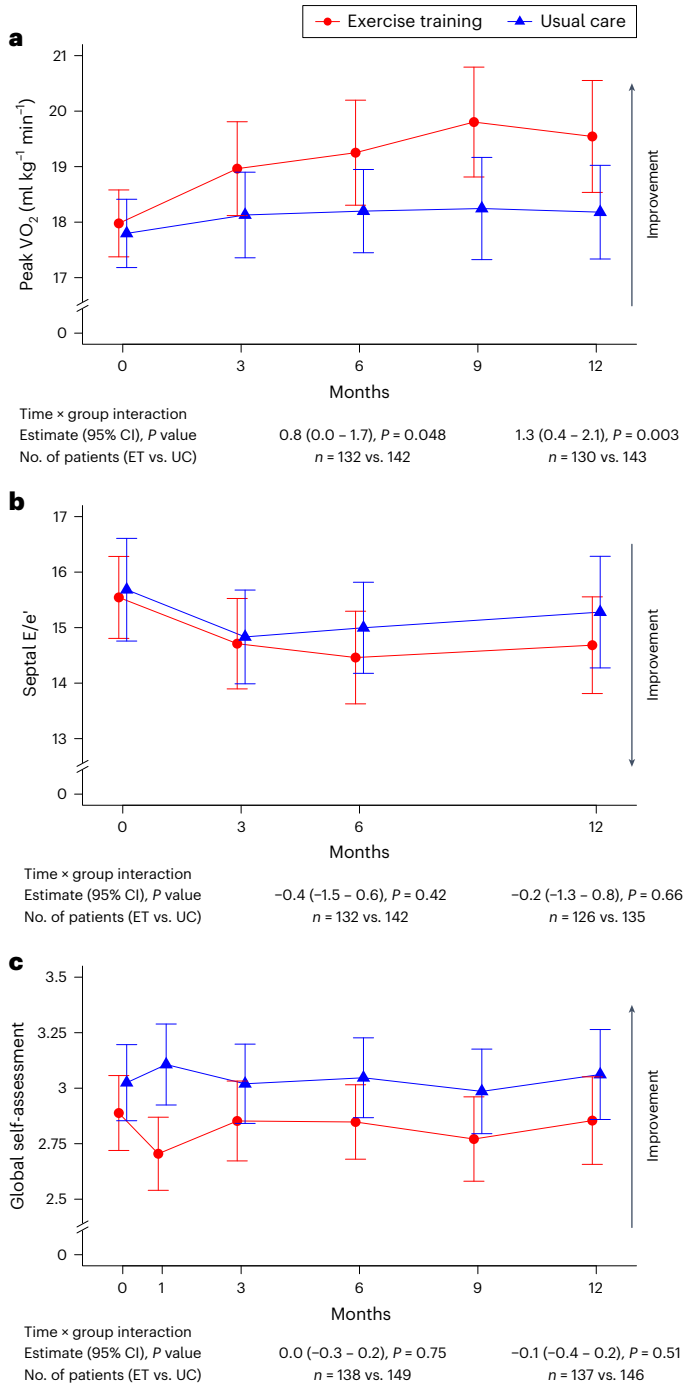


Fig. 3 | Results of secondary endpoints at 6 months and 12 months. **a**–**c**, Results from two-sided linear mixed models with the respective baseline variable as covariate and treatment as factor are shown for peak VO_2 (**a**), septal E/e′ (**b**) and GSA (**c**). The tests were based on complete cases and were not adjusted for multiple comparisons. Dots represent the mean and error bars represent the 95% CI for each group. GSA is based on a seven-point Likert scale (higher score indicates better health status).

(consisting of peak $\dot{V}O_2$, NYHA class, E/e′ and GSA) improved after ET. This improvement in the clinical score at 12 months was primarily driven by improvements in peak $\dot{V}O_2$ and NYHA class.

The difference in peak $\dot{V}O_2$ between ET and UC at 3 months was lower than observed in previous studies (0.8 ml kg^{-1} min^{-1} versus approximately 2.0 ml kg^{-1} min^{-1})[12]. However, in contrast to the OptimEx-Clin trial[13], peak $\dot{V}O_2$ continuously increased and showed a 'leveling-off' after 9 months of training with a clinically meaningful mean increase in peak $\dot{V}O_2$ of 1.3 ml kg^{-1} min^{-1} at 12 months[21]. These differences may be explained by the progressive nature of the exercise prescription with increasing duration and intensity (that is, initially three sessions per week for 30 min per session were performed at 50% peak $\dot{V}O_2$ and gradually increased to 60 min per session at 70% peak $\dot{V}O_2$ at 3 months with the addition of resistance training after the first month)[19]. This underscores the importance of prescribing a progressive load combined with different training modalities to provide sufficient exercise stimulus in the long term. Notably, resistance training may play an important role in preventing sarcopenia—a comorbidity associated with worse prognosis. Indeed, Brubaker et al.[15] recently reported that 20 weeks of combined endurance and resistance training during caloric restriction significantly improved leg muscle strength and quality. Moreover, it must be noted that, even with relatively low improvements in peak $\dot{V}O_2$, other important metrics, such as quality of life, may significantly improve[13,22].

Left ventricular diastolic function, assessed as E/e′, was not significantly different between groups. This is consistent with previous trials that reported that diastolic function does not significantly improve with ET in patients with HFpEF. However, it is in contrast with two smaller trials (including the Ex-DHF pilot trial) that found favorable effects on diastolic function[19,23]. These discrepant findings may be explained by the heterogeneity of HFpEF and the different clinical status of the patients studied. Specifically, in the present trial, patients were approximately 5 years older and had a significantly larger left atrial volume (40 ml m^{-2} as compared to 28 ml m^{-2}) compared to patients in Ex-DHF pilot trial[19]. Thus, patients in the Ex-DHF pilot trial had less severe structural changes that might be associated with a higher potential to improve diastolic function with ET. This also fits with the results of ET in patients in an early disease state with left ventricular hypertrophy and elevated cardiac biomarkers (stage B HFpEF), in which 1 year of individualized progressive ET significantly improved myocardial stiffness[24].

Taken together, it is likely that peripheral rather than central adaptations primarily drive ET-mediated improvements in peak $\dot{V}O_2$ in patients with HFpEF. Indeed, previous studies showed that changes in peak $\dot{V}O_2$ were primarily associated with increased peak exercise arterio-venous oxygen difference, whereas stroke volume and peak heart rate remained mainly unchanged[25–27]. The main mechanisms responsible for the improvement in peripheral oxygen extraction remain uncertain; however, previous studies indicated that ET does not improve peripheral vascular function[28,29], but it may induce significant molecular changes in the skeletal muscle, especially when performed at a higher intensity[30].

Hospitalizations that were classified as potentially related to HF or exercise were numerically higher in the ET group than in the UC group (33 versus 23 patients); however, not every hospitalization in the ET group should be regarded as a safety concern. Both the additional contacts between patients and healthcare professionals in the intervention arm (for example, performing heart rate and blood pressure measurements before and after every training session) as well as the training sessions themselves (for example, increased oxygen demand and continuous heart rate monitoring) may have uncovered health-related issues (for example, atrial fibrillation or flutter, angina and dizziness) that could have otherwise led to a serious clinical event, if remained undiscovered. Indeed, a post hoc evaluation of the hospitalizations in the ET group identified 11 hospitalizations that were related to an exercise session (seven of these were detected before the patients started exercising). This could be considered beneficial. However, it is unclear, whether the same event would have been detected in the UC group and whether the earlier detection prevented a deleterious event. Nevertheless, future ET trials should try to differentiate between potential beneficial observations that have been uncovered by the intervention from those that can be clearly classified as acute events (such as myocardial infarction

https://doi.org/10.1038/s41591-024-03342-7

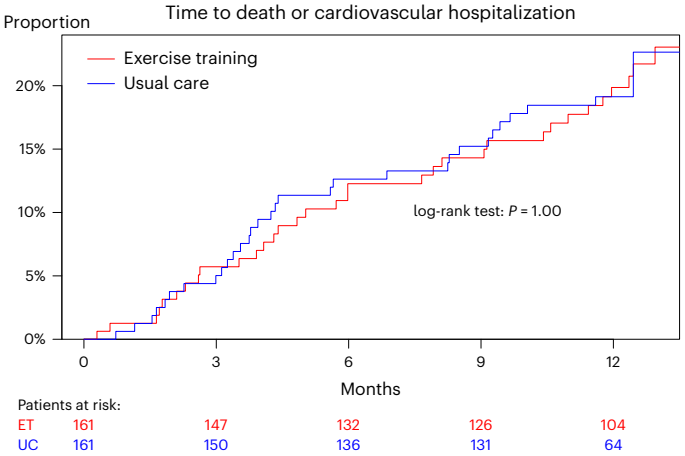


Fig. 4 | Time to all-cause death or cardiovascular hospitalizations in the ET and UC groups. The difference in time to events was evaluated using a two-sided Cox regression analysis.

or stroke). Compared to the numerically higher number of hospitalizations, time to all-cause mortality and cardiovascular hospitalizations were very similar between groups, implying that there were no relevant safety concerns. Of interest, 36 hospitalizations (42% of cardiovascular hospitalizations) were associated with atrial fibrillation. Elevated natriuretic peptides (one of our inclusion criteria) are associated with an increased risk of AF[31], and further trials in HFpEF may also focus on rhythm control strategies (for example, by early pulmonary vein isolation). Overall, the number of all-cause hospitalizations was high (176 events in 116 patients (36%), 49% cardiovascular), which is similar to the OptimEx-Clin trial (88 events in 52 patients (30%), 60% cardiovascular)[13] but higher than in other ET trials in HFpEF[15,19,23,28,32–34]. The high number of hospitalizations is in line with the results of epidemiological studies and may likely be explained by the advanced age and high burden of (non-cardiovascular) comorbidities in HFpEF[35,36].

Adherence to the intervention was relatively low, a finding in accordance with other long-term exercise studies in patients with HFpEF, HFrEF or other conditions, such as chronic kidney disease[13,37,38]. Similar to previous studies[39], we found a strong association between adherence to the intervention and the effects of ET. Although these results are prone to bias and must be interpreted as observational (as they break randomization), they nevertheless reinforce the need to evaluate ways to improve adherence to ET. This could include psychological components or the use of new telemedical approaches to increase intrinsic motivation. However, because low adherence rates are also largely due to the high number of hospitalizations and clinical events in patients with HFpEF, another way to improve long-term adherence may be a faster reintegration into regular ET after an acute clinical event, as elegantly shown in the REHAB-HF trial[40].

Similar to previous ET trials in HFpEF[13,33], the participants included in the present trial appeared to have had relatively mild disease (for example, 85% with NYHA II, peak $\dot{V}O_2$ of approximately 85% predicted). This is probably related to the inclusion criteria that patients had to be clinically stable, able to perform ET and able to travel to the training facility. However, subgroup analyses of the present trial did not show significant interaction effects between groups and age, peak $\dot{V}O_2$, NYHA class or diuretic use for either the primary endpoint or the change in peak $\dot{V}O_2$, and similar results were also previously observed in patients with HFpEF or HFrEF[13,41,42]. This indicates that, for clinically stable patients with HF, age and disease severity do not significantly influence treatment effects and the results of ET programs. In contrast, very sick, recently hospitalized patients with acute decompensated HF require a novel, tailored and personalized intervention, as prescribed in the REHAB-HF trial[40].

Our study has several limitations, including the asymmetric definition of the modified Packer score. Specifically, one hard criterion of clinical worsening (for example, increase in E/e′ >30%) or a hospitalization related to exercise or HF led to a worsened modified Packer score, irrespective of changes in all other score components. Indeed, among 56 patients in whom a hospitalization was classified as (potentially) related to HF or exercise, seven (12.5%) had an improved clinical score. Moreover, 30% of patients with a worsened modified Packer score had an increase in E/e′ of more than 30% (42 of 140 patients), which is likely explained by the fact that even relatively small changes in E and/or e′ may lead to a significant increase in the ratio of E/e′. The win ratio seems to be a better interpretable hierarchical composite endpoint and effect measure than the Packer score. However, at the time of designing and starting the present study, the win ratio approach had not yet been published[43]. Another limitation was that the study did not allow differentiation between the effects of endurance and resistance training and did not include specific strength outcomes, such as changes in body composition, muscle strength or physical function. Because resistance training may have a positive effect on these parameters, this gap in evidence should be specifically addressed in future research. Moreover, as discussed above, adherence to the intervention was relatively low, and the association between adherence and the effects of ET must be interpreted with caution. There has been some delay between last patient–last visit and time of publication. The main reason for this has been the repetitions of echo core laboratory measurements, because of a discrepancy between unblinded site echocardiography measurements and blinded echocardiography measurements assessed in the echo core laboratory (for details, see Methods and Supplementary Tables 8 and 9). Despite the delay, to our knowledge, the present trial is, to date, still the largest randomized controlled exercise intervention trial in patients with HFpEF. Moreover, the same long-term intervention period of 12 months has been investigated only in a smaller randomized controlled trial in HFpEF[13]. Finally, the fact that study participants cannot be blinded to treatment arm assignment constitutes a general limitation in ET trials. In contrast to objective parameters, such as clinical events, peak VO_2 or E/e′, this could have influenced the results of subjective parameters, such as NYHA class.

In conclusion, 1 year of combined supervised endurance and resistance ET in clinically stable patients with HFpEF did not result in a significantly better modified Packer score compared to UC (primary endpoint). In contrast, secondary analysis of the primary endpoint revealed significant differences favoring ET when comparing ‘improved’ versus ‘unchanged/worsened’ modified Packer score, which was primarily driven by significant and clinically meaningful improvements in peak VO_2 and NYHA functional class, with no significant differences in change in E/e′, change in GSA, hospitalization rate and mortality between groups. However, the numerically higher number of hospitalizations in the ET group merits further investigation.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-024-03342-7.

References

1. Tsao, C. W. et al. Heart disease and stroke statistics—2023 update: a report from the American Heart Association. *Circulation* **147**, e93–e621 (2023).
2. Gerber, Y. et al. A contemporary appraisal of the heart failure epidemic in Olmsted County, Minnesota, 2000 to 2010. *JAMA Intern. Med.* **175**, 996–1004 (2015).
3. Shah, K. S. et al. Heart failure with preserved, borderline, and reduced ejection fraction: 5-year outcomes. *J. Am. Coll. Cardiol.* **70**, 2476–2486 (2017).

4. Anker, S. D. et al. Empagliflozin in heart failure with a preserved ejection fraction. *N. Engl. J. Med.* **385**, 1451–1461 (2021).
5. Packer, M. et al. Effect of empagliflozin on worsening heart failure events in patients with heart failure and preserved ejection fraction: EMPEROR-Preserved trial. *Circulation* **144**, 1284–1294 (2021).
6. Solomon, S. D. et al. Dapagliflozin in heart failure with mildly reduced or preserved ejection fraction. *N. Engl. J. Med.* **387**, 1089–1098 (2022).
7. Meifang, W. et al. Advance in the pharmacological and comorbidities management of heart failure with preserved ejection fraction: evidence from clinical trials. *Heart Fail. Rev.* **29**, 305–320 (2024).
8. Kosiborod M. N. et al. Semaglutide versus placebo in patients with heart failure and mildly reduced or preserved ejection fraction: a pooled analysis of the SELECT, FLOW, STEP-HFpEF, and STEP-HFpEF DM randomised trials. *Lancet* **404**, 949–961 (2024).
9. Solomon S. D. et al. Finerenone in heart failure with mildly reduced or preserved ejection fraction. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa2407107 (2024).
10. Pandey, A. et al. Exercise intolerance in older adults with heart failure with preserved ejection fraction: *JACC* State-of-the-Art Review. *J. Am. Coll. Cardiol.* **78**, 1166–1187 (2021).
11. Del Buono, M. G. et al. Exercise intolerance in patients with heart failure: *JACC* State-of-the-Art Review. *J. Am. Coll. Cardiol.* **73**, 2209–2225 (2019).
12. Sachdev, V. et al. Supervised exercise training for chronic heart failure with preserved ejection fraction: a scientific statement from the American Heart Association and American College of Cardiology. *J. Am. Coll. Cardiol.* **81**, 1524–1542 (2023).
13. Mueller, S. et al. Effect of high-intensity interval training, moderate continuous training, or guideline-based physical activity advice on peak oxygen consumption in patients with heart failure with preserved ejection fraction: a randomized clinical trial. *JAMA* **325**, 542–551 (2021).
14. Edelmann, F. et al. Exercise training improves exercise capacity and diastolic function in patients with heart failure with preserved ejection fraction: results of the Ex-DHF (Exercise training in Diastolic Heart Failure) pilot study. *J. Am. Coll. Cardiol.* **58**, 1780–1791 (2011).
15. Brubaker, P. H. et al. A randomized, controlled trial of resistance training added to caloric restriction plus aerobic exercise training in obese heart failure with preserved ejection fraction. *Circ. Heart Fail.* **16**, e010161 (2023).
16. Redfield, M. M. & Borlaug, B. A. Heart failure with preserved ejection fraction: a review. *JAMA* **329**, 827–838 (2023).
17. Packer, M. Proposal for a new clinical end point to evaluate the efficacy of drugs and devices in the treatment of chronic heart failure. *J. Card. Fail.* **7**, 176–182 (2001).
18. Packer, M. Development and evolution of a hierarchical clinical composite end point for the evaluation of drugs and devices for acute and chronic heart failure: a 20-year perspective. *Circulation* **134**, 1664–1678 (2016).
19. Edelmann, F. et al. Exercise training in Diastolic Heart Failure (Ex-DHF): rationale and design of a multicentre, prospective, randomized, controlled, parallel group trial. *Eur. J. Heart Fail.* **19**, 1067–1074 (2017).
20. Gläser, S. et al. [The Study of Health in Pomerania (SHIP) reference values for cardiopulmonary exercise testing]. *Pneumologie* **67**, 58–63 (2013).
21. Tucker, W. J. et al. Impact of exercise training on peak oxygen uptake and its determinants in heart failure with preserved ejection fraction. *Card. Fail. Rev.* **2**, 95–101 (2016).
22. Borlaug, B. A. et al. Inorganic nitrite to amplify the benefits and tolerability of exercise training in heart failure with preserved ejection fraction: the INABLE-Training Trial. *Mayo Clin. Proc.* **99**, 206–217 (2024).
23. Alves, A. J. et al. Exercise training improves diastolic function in heart failure patients. *Med. Sci. Sports Exerc.* **44**, 776–785 (2012).
24. Hieda, M. et al. One-year committed exercise training reverses abnormal left ventricular myocardial stiffness in patients with stage b heart failure with preserved ejection fraction. *Circulation* **144**, 934–946 (2021).
25. Haykowsky, M. J. et al. Effect of endurance training on the determinants of peak exercise oxygen consumption in elderly patients with stable compensated heart failure and preserved ejection fraction. *J. Am. Coll. Cardiol.* **60**, 120–128 (2012).
26. Fu, T. C. et al. Aerobic interval training elicits different hemodynamic adaptations between heart failure patients with preserved and reduced ejection fraction. *Am. J. Phys. Med. Rehabil.* **95**, 15–27 (2016).
27. Mueller, S. et al. Peak O_2-pulse predicts exercise training-induced changes in peak $\dot{V}O_2$ in heart failure with preserved ejection fraction. *ESC Heart Fail.* **9**, 3393–3406 (2022).
28. Kitzman, D. W. et al. Effect of endurance exercise training on endothelial function and arterial stiffness in older patients with heart failure and preserved ejection fraction: a randomized, controlled, single-blind trial. *J. Am. Coll. Cardiol.* **62**, 584–592 (2013).
29. Gevaert, A. B. et al. Effect of training on vascular function and repair in heart failure with preserved ejection fraction. *J. Am. Coll. Cardiol. HF* **11**, 454–464 (2023).
30. Winzer, E. B. et al. Impact of different training modalities on molecular alterations in skeletal muscle of patients with heart failure with preserved ejection fraction: a substudy of the optimex trial. *Circ. Heart Fail.* **15**, e009124 (2022).
31. Uhe, T. et al. Signs of left atrial disease and 10-year risk of atrial fibrillation. *PLoS ONE* **17**, e0266848 (2022).
32. Kitzman, D. W. et al. Effect of caloric restriction or aerobic exercise training on peak oxygen consumption and quality of life in obese older patients with heart failure with preserved ejection fraction: a randomized clinical trial. *JAMA* **315**, 36–46 (2016).
33. Kitzman, D. W., Brubaker, P. H., Morgan, T. M., Stewart, K. P. & Little, W. C. Exercise training in older patients with heart failure and preserved ejection fraction. *Circ. Heart Fail.* **3**, 659–667 (2010).
34. Gary, R. A. et al. Home-based exercise improves functional performance and quality of life in women with diastolic heart failure. *Heart Lung* **33**, 210–218 (2004).
35. Streng, K. W. et al. Non-cardiac comorbidities in heart failure with reduced, mid-range and preserved ejection fraction. *Int. J. Cardiol.* **271**, 132–139 (2018).
36. Pandey, A. et al. Temporal trends in prevalence and prognostic implications of comorbidities among patients with acute decompensated heart failure. *Circulation* **142**, 230–243 (2020).
37. O'Connor, C. M. et al. Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial. *JAMA* **301**, 1439–1450 (2009).
38. Anding-Rost K. et al. Exercise during hemodialysis in patients with chronic kidney failure. *NEJM Evid*. EVIDoa2300057 (2023).
39. Alonso, W. W., Keteyian, S. J., Leifer, E. S., Kitzman, D. W. & Sachdev, V. Adherence to exercise in adults with heart failure. *J. Cardiopulm. Rehabil. Prev.* **44**, E17–E18 (2024).
40. Kitzman, D. W. et al. Physical rehabilitation for older patients hospitalized for heart failure. *N. Engl. J. Med.* **385**, 203–216 (2021).
41. Taylor, R. S. et al. Impact of exercise rehabilitation on exercise capacity and quality-of-life in heart failure: individual participant meta-analysis. *J. Am. Coll. Cardiol.* **73**, 1430–1443 (2019).

42. Taylor, R. S. et al. Impact of exercise-based cardiac rehabilitation in patients with heart failure (ExTraMATCH II) on mortality and hospitalisation: an individual patient data meta-analysis of randomised trials. *Eur. J. Heart Fail.* **20**, 1735–1743 (2018).
43. Pocock, S. J., Ariti, C. A., Collier, T. J. & Wang, D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur. Heart J.* **33**, 176–182 (2012).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025

Frank Edelmann[1,2,3,19] ✉, Rolf Wachter[4,5,6,19], André Duvinage[7,19], Stephan Mueller[7,8], Isabel Fegers-Wustrow[7,8], Silja Schwarz[7], Jeffrey W. Christle[7,9], Elisabeth Pieske-Kraigher[1,2], Melchior Seyfarth[10], Markus Knapp[11], Marcus Dörr[12,13], Kathleen Nolte[14], Hans-Dirk Düngen[1,2], Christoph Herrmann-Lingen[5,15], Katrin Esefeld[7,8], Andreas Hagendorff[6], Mark J. Haykowsky[16], Gerd Hasenfuss[4,5], Volker Holzendorf[17], Christiane Prettin[17], Meinhard Mende[17], Burkert Pieske[18,19] & Martin Halle[7,8,19]

[1]Department of Cardiology, Angiology and Intensive Care Medicine, Campus Virchow Klinikum, Deutsches Herzzentrum der Charité, Berlin, Germany. [2]Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany. [3]DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, Berlin, Germany. [4]Department of Cardiology and Pneumology, University Medical Center Göttingen, Göttingen, Germany. [5]DZHK (German Centre for Cardiovascular Research), Partner Site Göttingen, Göttingen, Germany. [6]Department of Cardiology, University of Leipzig, Leipzig, Germany. [7]Department for Preventive Sports Medicine and Sports Cardiology, Technical University of Munich, School of Medicine and Health, TUM University Hospital, Munich, Germany. [8]DZHK (German Centre for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany. [9]Department of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, CA, USA. [10]Witten Herdecke University, Helios Heart Center Wuppertal, Wuppertal, Germany. [11]Practice for Cardiology, Schwäbisch Hall, Germany. [12]Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany. [13]DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany. [14]Department of Cardiology, Evangelic Hospital Göttingen-Weende, Göttingen, Germany. [15]Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany. [16]College of Health Sciences, University of Alberta, Edmonton, AB, Canada. [17]Clinical Trial Centre Leipzig, University of Leipzig, Leipzig, Germany. [18]Division of Cardiology, Department of Internal Medicine, University Medicine Rostock, Rostock, Germany. [19]These authors contributed equally: Frank Edelmann, Rolf Wachter, André Duvinage, Burkert Pieske, Martin Halle. ✉e-mail: frank.edelmann@dhzc-charite.de

Methods

Study design

Ex-DHF (ISRCTN registry number ISRCTN86879094) was a randomized, multicenter parallel trial with two groups conducted at 11 trial sites (Göttingen, Munich, Berlin, Greifswald, Schwäbisch Hall, Wuppertal, Magdeburg, Tübingen, Bremen and Ulm, Germany, and Graz, Austria). The study protocol is provided in the Supplementary Information and was previously published[19]. In brief, the Ex-DHF trial recruited patients with symptomatic but clinically stable HFpEF (NYHA classes II and III, left ventricular ejection fraction ≥50%, diastolic dysfunction defined as either E/e′ >15 or E/e′ 8–15 in combination with N-terminal pro-brain natriuretic peptide (NT-proBNP) >220 ng L^{-1} or the presence of atrial fibrillation)[44] and peak VO_2 less than 25 ml kg^{-1} min^{-1} (a full description of inclusion and exclusion criteria is shown in Supplementary Table 10). Patients were randomized (1:1) to combined endurance and resistance ET or UC. Randomization was stratified by center, NYHA class and peak VO_2 (less than versus equal to or greater than 20 ml kg^{-1} min^{-1}) and performed by the local investigators in a web-based system using Pocock's minimization algorithm with random component[45]. The study was performed in accordance with the Declaration of Helsinki and was approved by the ethics committee of the University Medical Center Göttingen and the local ethics committees at all participating sites. All participants provided written informed consent. Patients were treated according to HF guidelines at time of recruitment, and medication had to be stable for at least 4 weeks before patients were eligible to enter the trial. Follow-up visits were scheduled at 1, 3, 6, 9 and 12 months after randomization. All procedures were performed according to standard operating procedures and supervised by core laboratories for echocardiography (Berlin, Germany) and CPET (Munich, Germany). Local assessments of left ventricular ejection fraction and E/e′ were used to assess inclusion criteria, whereas only data analyzed in the blinded echocardiography core laboratory were used for statistical analyses. Due to differences between local measurements (unblinded) and core laboratory measurements (blinded for clinical data and group assignment; Supplementary Table 8), analyses of E and e′ (both septal and lateral) were repeated by the echo core laboratory in Berlin (blinded for clinical data and group assignment and study visit) and additionally validated (~70% randomly selected E/e′ measurements) by the echo laboratory in Leipzig. There was a high agreement between the second echo core laboratory measurements of E, e′ and E/e′ and the external validation (Krippendorff's alpha[46] >0.8 for all comparisons; Supplementary Table 9). In general, an alpha ≥ 0.8 is accepted as good agreement[47]. Echo data from this second core laboratory analysis were used for the analysis of the primary endpoint. CPET was performed on a bicycle ergometer with an initial power output of 20 W, which was increased by 20 W every 2 min until symptom-limited exhaustion. Peak VO_2 was defined as the highest 10-s average.

Intervention

Patients randomized to the intervention group were prescribed supervised ET on 3 d per week for 12 months and were encouraged to increase their leisure time physical activity. As previously described in detail[19], the exercise intervention consisted of a combination of endurance and resistance training based on our Ex-DHF pilot study[14]. Endurance training was performed on a bicycle ergometer starting with 3 × 30 min at 50% peak VO_2 in the first 2 weeks and gradually increased to 3 × 60 min at 70% peak VO_2. After 4 weeks, resistance training was added for 2–3 training sessions per week (seven exercises on weight training machines for major muscle groups; 1–2 sets and 12–15 repetitions at 60% of one-repetition-maximum (1-RM)). Individual exercise intensities were based on repeated CPET and 1-RM testing at baseline (only CPET) as well as 4 weeks (only 1-RM), 3 months, 6 months and 9 months after randomization. Patients were considered adherent if they completed at least 66.6% of their scheduled exercise sessions[19]. Patients randomized to UC were not offered any tailored exercise prescription, supervised training or structured advice to increase their daily physical activities. They were instructed at the time of enrollment to continue all therapies.

Endpoints

The primary endpoint was the difference in the modified Packer score between groups at 12 months. The modified Packer score combined six HF outcome parameters, including all-cause mortality, hospitalizations that were potentially related to HF or exercise and a clinical score consisting of change in peak VO_2, change in septal E/e′, change in NYHA class and change in GSA (seven-point Likert scale from 'very bad' to 'very good'—higher values indicating better health status) in a hierarchical manner and was graded as 'worsened', 'unchanged' or 'improved' (for more details, see Supplementary Fig. 1, the study protocol (provided in the supplement) or the study design manuscript)[19]. As the algorithm also allowed the assessment of patients who were lost to follow-up, the primary endpoint was available for all study participants. The pre-specified secondary endpoints included in this paper are related to the components of the primary endpoint—that is, time to all-cause death or cardiovascular hospitalization as well as the change in peak VO_2, E/e′, NYHA class and GSA at 6 months and 12 months. All other pre-specified secondary endpoints will be published separately. An independent clinical endpoint committee adjudicated all deaths and hospitalizations that occurred during the study period. Potential associations with exercise or HF were evaluated blinded to treatment group assignment based on the medical reports. Notably, although patients randomized to UC did not receive supervised ET, hospitalizations could have been classified as potentially related to exercise based on the nature of the clinical event. To evaluate whether the additional contacts between patients and healthcare professionals and the supervised ET may have uncovered health-related problems that could have otherwise remained undiscovered, we performed an additional post hoc evaluation of the hospitalizations in the ET group to identify hospitalizations that were detected in the context of an ET session (but not related to ET per se).

Statistical analysis

The sample size calculation was based on the Ex-DHF pilot study[19] and the I-PRESERVE trial[48] and was previously described in detail[19]. Assuming an improved modified Packer score in 20% of ET patients and in 5% of UC patients and a worsened modified Packer score in 15% of ET patients and in 25% of UC patients, a power of 98% and a significance level of $\alpha = 0.05$ results in 160 patients per treatment group. The main analysis of the primary endpoint was performed using the test of Kendall's tau[19]. Moreover, in a pre-specified secondary analysis, the primary endpoint was also evaluated using an ordinal regression analysis including the randomization strata NYHA class and peak VO_2 (less than versus equal to or greater than 20 ml min^{-1} kg^{-1}) as covariates and center as random effect. Because the assumption of proportional odds was violated, the ordinal regression was performed without this assumption. Therefore, the three categories were combined as follows: (1) 'improved/unchanged' versus 'worsened' and (2) 'improved' versus 'unchanged/worsened'. The categories 'worsened', 'unchanged' and 'improved' are also presented for each component of the primary endpoint, and, although this was not pre-specified in the trial protocol, we evaluated the OR for improvement with ET for each of these components, as this improves the understanding of the contribution of each component to the results of the primary endpoint. In a pre-specified per-protocol analysis, patients who had at least one major protocol deviation or were lost to follow-up at 12 months were excluded. The pre-specified secondary endpoints reported in this paper are limited to the components of the primary endpoint and were analyzed by Cox regression analysis (time to all-cause death or cardiovascular hospitalizations), separate ordinal regression analyses (change in NYHA class from baseline to 6 months and 12 months) or

linear mixed models (change in septal E/e', peak VO_2 and GSA) with the respective baseline variable as covariate and treatment as factor (complete cases). Analyses of secondary endpoints included in the clinical score (peak VO_2, E/e', NYHA class and GSA) were repeated after performing a multiple imputation approach to account for missing values. To use a maximum of information and to minimize a potential bias, demographic and clinical data, parameters of echocardiography, cardiopulmonary exercise testing and quality of life were included to generate 50 complete datasets applying multiple imputation with 20 simulations by chained equations[49]. In each imputed dataset, the components of the primary endpoint were combined into the ordinal primary endpoint. Imputed data were analyzed by linear mixed models including random intercept for center and important covariates, such as age, sex, training compliance and daily activity. The results were pooled following the rules from Rubin[50]. In post hoc analyses, we also calculated a win ratio with 95% CI[51] using a similar hierarchy and the same cutoffs as for the primary endpoint (all-cause death > number of hospitalizations due to worsening heart failure > change in peak $\dot{V}O_2$ > change of NYHA class > change in GSA > change in E/e'; see Supplementary Fig. 1 for cutoffs). The win ratio is now commonly used in clinical trials as an easily interpretable hierarchical composite endpoint but was introduced only after the beginning of the present trial[43]. To evaluate the association with adherence to ET, we performed another post hoc analysis comparing the results of the modified Packer score as well as the change in peak VO_2 between UC and three adherence subgroups of ET with cutoffs at 33.3% and 66.7% adherence to the scheduled training sessions. This was done using the Mantel–Haenszel test of trend in proportions for the modified Packer score and a single-factor ANOVA with post hoc tests following the Dunnett method for the change in peak VO_2. Moreover, to examine whether the results of the modified Packer score or the change in peak VO_2 were dependent on age or parameters of disease severity (peak VO_2, NYHA class and use of diuretics), we built ordinal (for modified Packer score) and mixed linear (for change in peak VO_2) regression models with tests for interaction between these variables and study group.

All patients were analyzed according to their randomization group. Data were extracted from the database and prepared for analysis with SPSS software version 24 and later (IBM). Descriptive statistics and the primary analysis were performed with SPSS. Linear mixed models and ordinal regression analyses were performed with R (versions 3 and 4) including the packages 'VGAM', 'mice', lme4, 'WinRatio' and 'survival'. To ensure correct coding of the composite primary endpoint, the calculation was independently performed by two statisticians on two software platforms (SPSS and R). All statistical analyses were performed using two-tailed tests with a significance level of $\alpha = 5\%$. Results of the secondary endpoints were not adjusted for multiple testing and should be interpreted as exploratory. In this paper, we followed the CONSORT guidelines for reporting randomized controlled trials.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

As patients have not explicitly consented to sharing pseudonymized data, we are not allowed to share the individual participant data for legal reasons. However, upon reasonable request to the corresponding author (F.E., frank.edelmann@dhzc-charite.de), aggregated data (data on demographics, clinical history, pharmacological treatment, electrocardiography, laboratory parameters, echocardiography, cardiopulmonary exercise testing and global self-assessment) that do not allow identification of individual patients may be shared after consultation with the data protection officers and legal representatives of the participating institutions and after signing a data sharing agreement. A response to requests for data access can be expected within 4 weeks.

Code availability

No computer code was used to collect the data in this study.

References

44. Paulus, W. J. et al. How to diagnose diastolic heart failure: a consensus statement on the diagnosis of heart failure with normal left ventricular ejection fraction by the Heart Failure and Echocardiography Associations of the European Society of Cardiology. *Eur. Heart J.* **28**, 2539–2550 (2007).
45. Pocock, S. J. & Simon, R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* **31**, 103–115 (1975).
46. Hayes, A. F. & Krippendorff, K. Answering the call for a standard reliability measure for coding data. *Commun. Methods Meas.* **1**, 77–89 (2007).
47. Krippendorff K. *Content Analysis: An Introduction to Its Methodology* 2nd edn (Sage, 2004).
48. Massie, B. M. et al. Irbesartan in patients with heart failure and preserved ejection fraction. *N. Engl. J. Med.* **359**, 2456–2467 (2008).
49. van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).
50. Rubin, D. B. Multiple imputation after 18+ years. *J. Am. Stat. Assoc.* **91**, 473–489 (1996).
51. Bebu, I. & Lachin, J. M. Large sample inference for a win ratio analysis of a composite outcome based on prioritized components. *Biostatistics* **17**, 178–187 (2016).

Acknowledgements

This study was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft (DFG)) (project numbers ED 196/2-1 (F.E.), HA 5812/4-1 (M.H.), GE 2048/2-1, ED 196/2-2 (F.E.), HA 5812/4-2 (M.H.), PR 1454/1-2 (C.P.) and SFB-1470 – Z02 (F.E.)) and the German Centre for Cardiovascular Research (Deutsches Zentrum für Herz-Kreislauf-Forschung (DZHK)) (grant number: 81Z0600603) (S.M., I.F.-W. and M.H.). The funders had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; or decision to submit the manuscript for publication.

Author contributions

F.E., B.P. and M.H. made substantial contributions to the conception or design of the work. All authors made substantial contributions to the acquisition, analysis or interpretation of the data. F.E., R.W., S.M., I.F.-W., J.W.C., M.J.H., M.M., B.P. and M.H. drafted or substantially revised the manuscript. All authors approved the submitted version of the manuscript. All authors agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature.

Competing interests

F.E. reports personal fees from AstraZeneca, Bayer, Berlin Chemie, Boehringer Ingelheim, CVRx, Medtronic, Merck, Merck Sharp & Dohme, Novartis, Pfizer, PharmaCosmos, Resmed, Servier and Vifor Pharma; non-financial support from Novartis; and grants from AstraZeneca, Boehringer Ingelheim, Servier and Thermo Fisher Scientific, outside the submitted work. R.W. reports receiving personal fees from AstraZeneca, Bayer, Bristol Myers Squibb, Boehringer Ingelheim, CVRx, Daiichi Sankyo, Medtronic, Novartis, Pfizer, Pharmacosmos and Servier and research support from Boehringer Ingelheim, Bundesministerium für Bildung und Forschung, Deutsche Forschungsgemeinschaft, the European Union and Medtronic, outside

the submitted work. S.M. reports personal fees from Bristol Myers Squibb (consulting services), outside the submitted work. E.P.-K. holds minor shares in ICTS GmbH (Imaging in Clinical Trials Services) and is a part-time employee of this company and reports receiving minor personal fees from AstraZeneca. H.-D.D. reports receiving personal fees from Bayer, Bristol Myers Squibb, Boehringer Ingelheim, Merck, Merck Sharp & Dohme, Novartis and Servier, outside the submitted work. H.-D.D. also holds shares in SCIRENT GmbH and d.o.o. (Clinical Trials Services). C.H.-L. reports a personal fee from Novartis; royalties from Hogrefe Publishing Group; and research grants from the German Ministry of Education and Research, the German Research Foundation and the EU Commission. K.E. reports personal fees from Bristol Myers Squibb and Boehringer Ingelheim (honoraria for lectures), outside the submitted work. A.H. reports personal fees from AstraZeneca, Boehringer Ingelheim, Bristol Myers Squibb, Cardiac Dimensions, Daiichi Sankyo, GE Healthcare and Novartis. G.H. reports personal fees from AstraZeneca, Bayer, Boehringer Ingelheim, Corvia, Impulse Dynamics, Novartis, Servier and Vifor Pharma; being a co-principal investigator for Impulse Dynamics; and business ownership of Avocet Bio GmbH. B.P. reports institutional grants from AstraZeneca, Bayer Healthcare and Boston Scientific and personal fees for steering committee, consulting and speaker services from Bayer Healthcare, Merck Sharp & Dohme, AstraZeneca, Boehringer Ingelheim, Novartis, Boston Scientific and Abbott, outside the submitted work. B.P. also holds minor shares in ICTS GmbH (Imaging in Clinical Trials Services). M.H. reports receiving personal fees from Abbott, Amgen, Boehringer Ingelheim, Bristol Myers Squibb, Daiichi Sankyo, Sanofi-Aventis, Novartis and Medical Park (consulting fees and honoraria for lectures) and being the past president of the European Association of Preventive Cardiology (2020–2022), outside the submitted work. No other potential conflicts of interest were reported.

Additional collaborators

Stephan Gielen, Stephan von Haehling, Stefan Störk, Tobias D. Trippel, Stefan Anker, Hugo Saner, Mitja Lainscak, Andrea Berghold, Daniel Morris, Evgeny Belyavski, Martin Kropf and Aravind Kumar Radhakrishnan.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-024-03342-7.

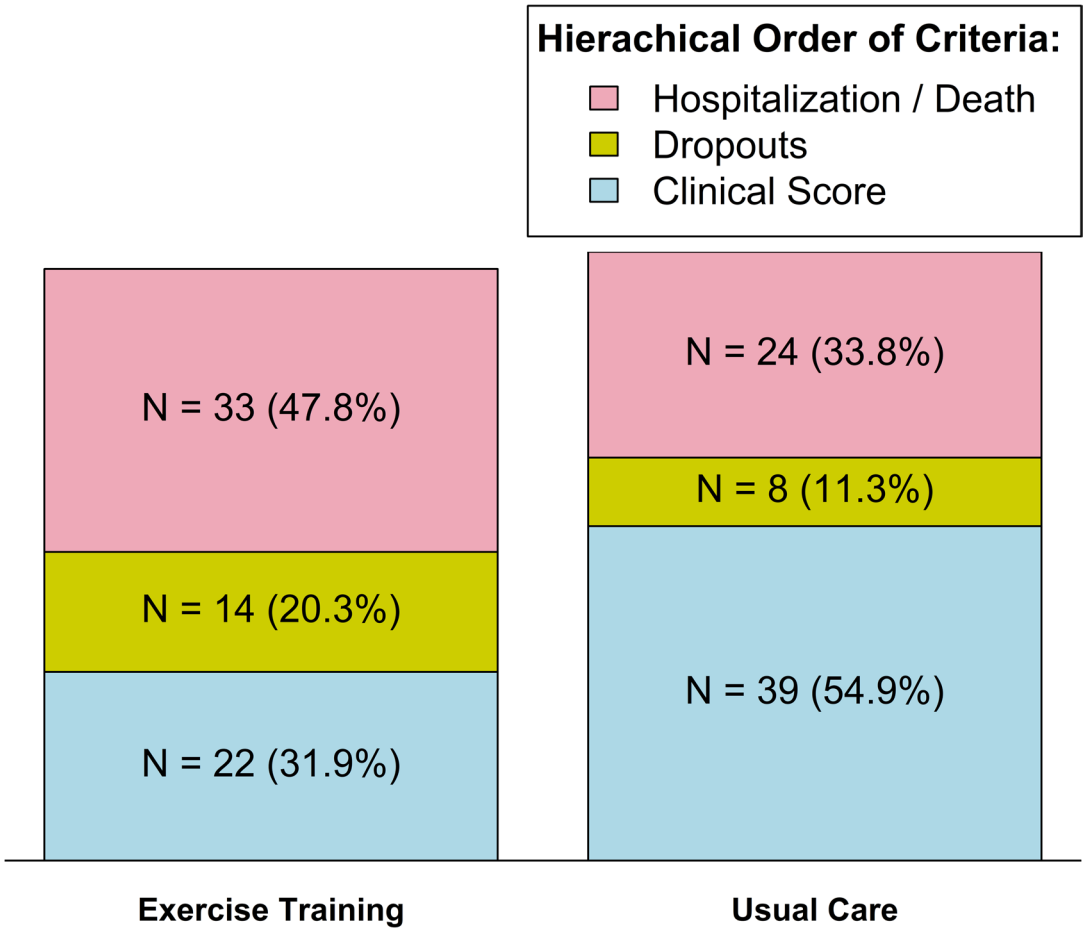
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-024-03342-7.

Correspondence and requests for materials should be addressed to Frank Edelmann.

Peer review information *Nature Medicine* thanks Dalane Kitzman, Rod Taylor and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Michael Basson, in collaboration with the *Nature Medicine* team.

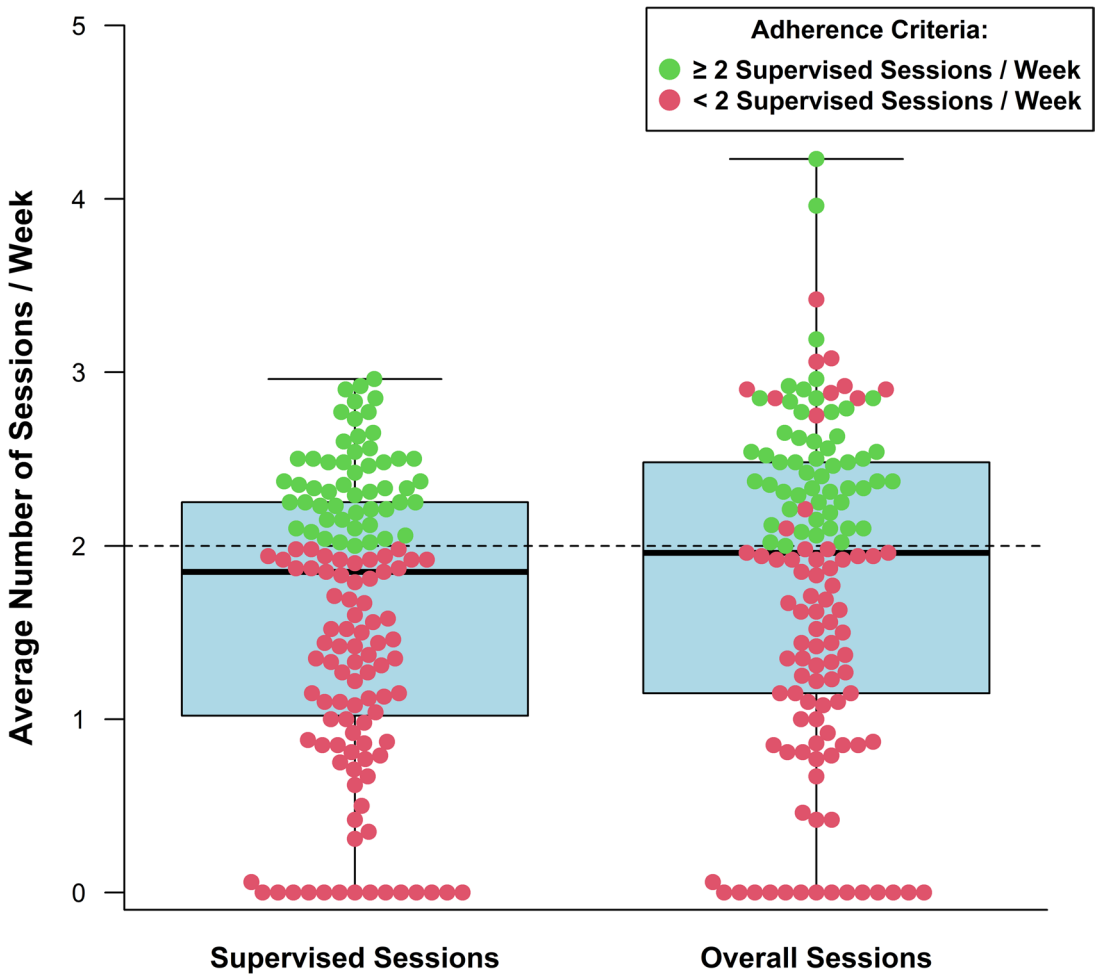
Reprints and permissions information is available at www.nature.com/reprints.

https://doi.org/10.1038/s41591-024-03342-7



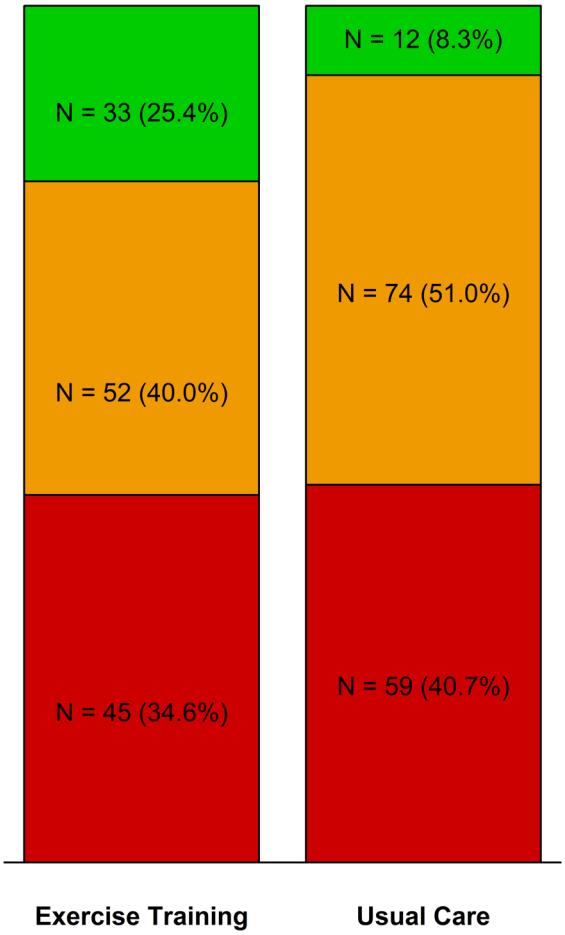
Extended Data Fig. 1 | Reasons for Worsening of the Modified Packer Score (MPS). Reasons are shown separately for exercise training (left) and usual care (right) as absolute and relative frequency.

https://doi.org/10.1038/s41591-024-03342-7



Extended Data Fig. 2 | Adherence to exercise training. Green dots represent individuals who performed an average of at least 2 supervised sessions/week between baseline and 12-months follow-up (N = 53). Red dots represent individuals who performed an average of less than 2 supervised (left) or overall (right) sessions/week between baseline and 12-months follow-up (N = 108). Medians are shown by thick vertical lines and box boundaries were determined by the first and third quartile. The whiskers were defined as the highest/smallest value that was within 1.5× interquartile range ± upper/lower box boundary. For supervised sessions, minimum and maximum adherence were 0 sessions/week and 2.96 sessions/week, respectively. For overall sessions, minimum and maximum were 0 sessions/week and 4.23 sessions/week.

https://doi.org/10.1038/s41591-024-03342-7



Kendall's tau-b

Improved vs. Unchanged vs. Worsened:
tau-b: -0.14, P = 0.015

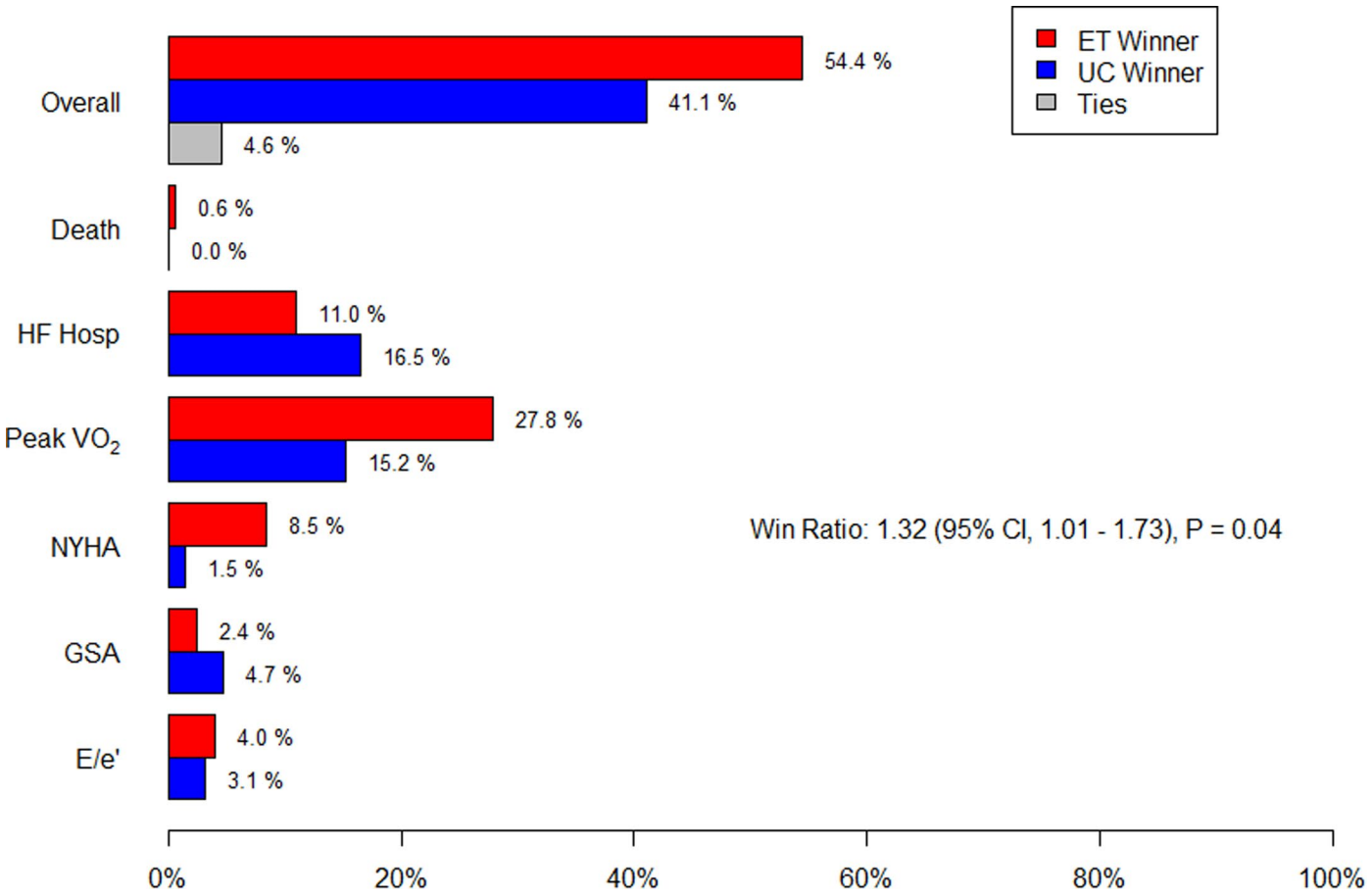
Ordinal Regression Analysis

1) Improved / Unchanged vs. Worsened:
OR (95% CI): 1.30 (0.79 to 2.12), P = 0.30

2) Improved vs. Unchanged / Worsened:
OR (95% CI): 3.77 (1.85 to 7.67), P < 0.001

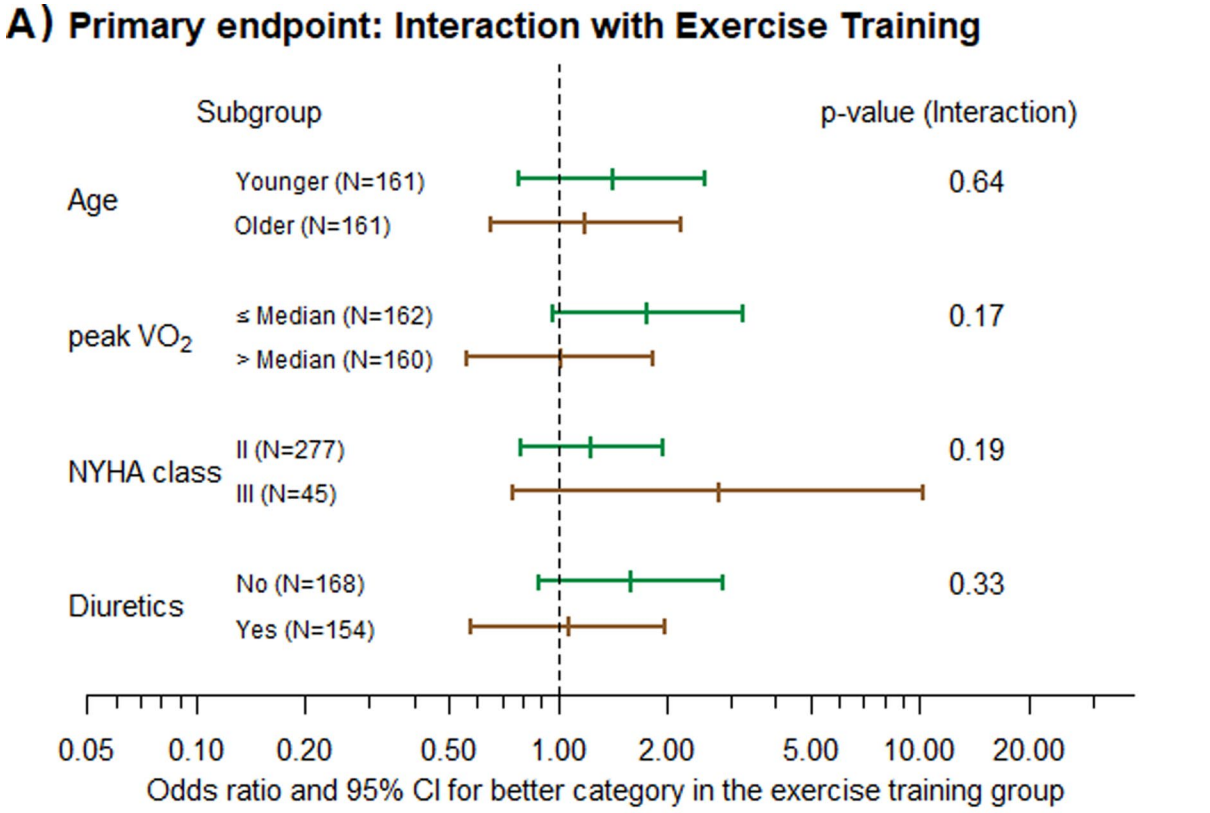
Extended Data Fig. 3 | Per-Protocol analysis of the primary endpoint (modified Packer score). Patients who were lost to 12-months follow-up or who had a major protocol deviation were excluded. Results are shown for a two-sided test of Kendall's tau-b and based on two-sided ordinal regression analyses, which were adjusted for randomization strata (baseline peak VO_2 less vs. equal or higher than 20 mL/kg/min; New York Heart Association functional class II / III). P-values were not adjusted for multiple comparisons. The exact P-value for the comparison of 'improved vs. unchanged / worsened' modified Packer score is P = 0.0003.

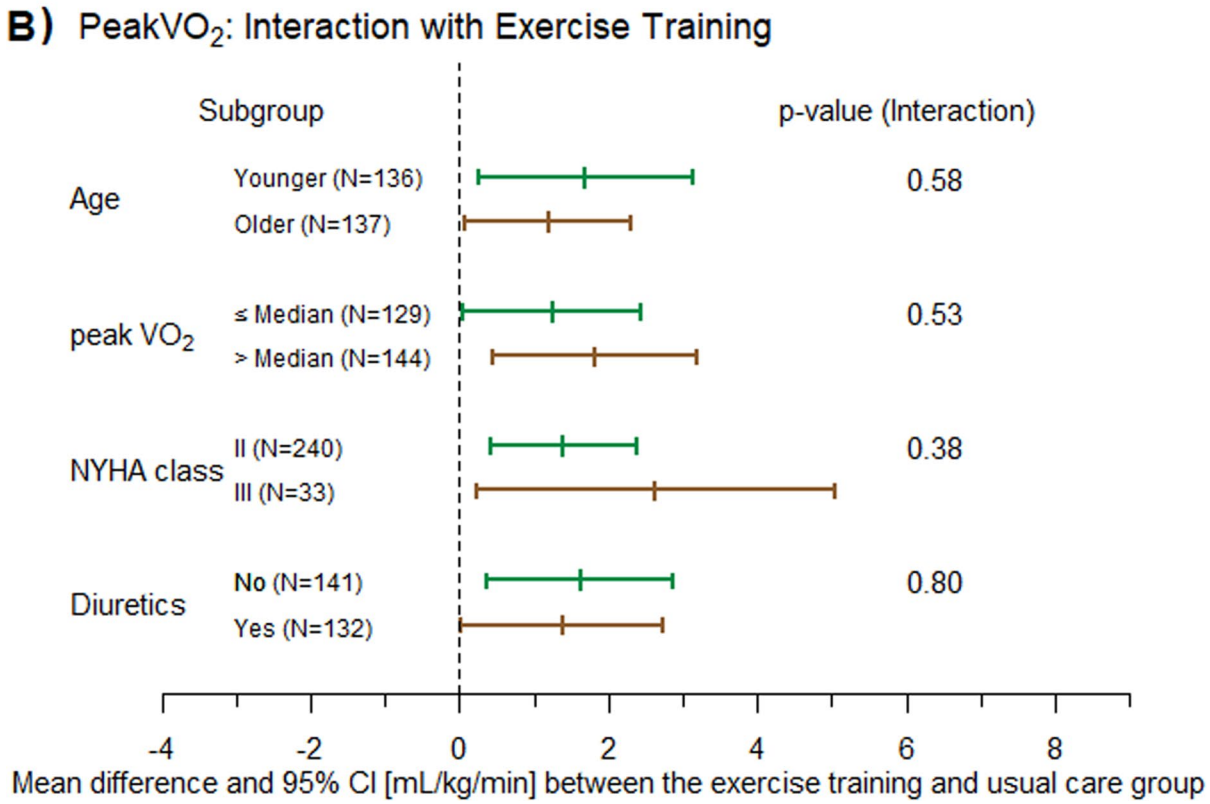
https://doi.org/10.1038/s41591-024-03342-7



Extended Data Fig. 4 | Win Ratio for the components of the primary endpoint. In this post-hoc analysis, a two-sided win ratio (based on all 322 randomized patients) was calculated using a similar hierarchy as for the primary endpoint (all-cause death > number of hospitalizations due to worsening heart failure > change in peak $\dot{V}O_2$ > change of NYHA class > change in GSA > change in E/e'). The cutoffs used to categorize peak $\dot{V}O_2$, NYHA class, GSA and E/e' are the same as the predefined cutoffs used for the primary endpoint (-2 = hard criterion for clinical worsening, -1 = mild criterion for clinical worsening, 0 = unchanged, +1 = criterion for clinical improvement; for details see Supplementary Fig. 1).

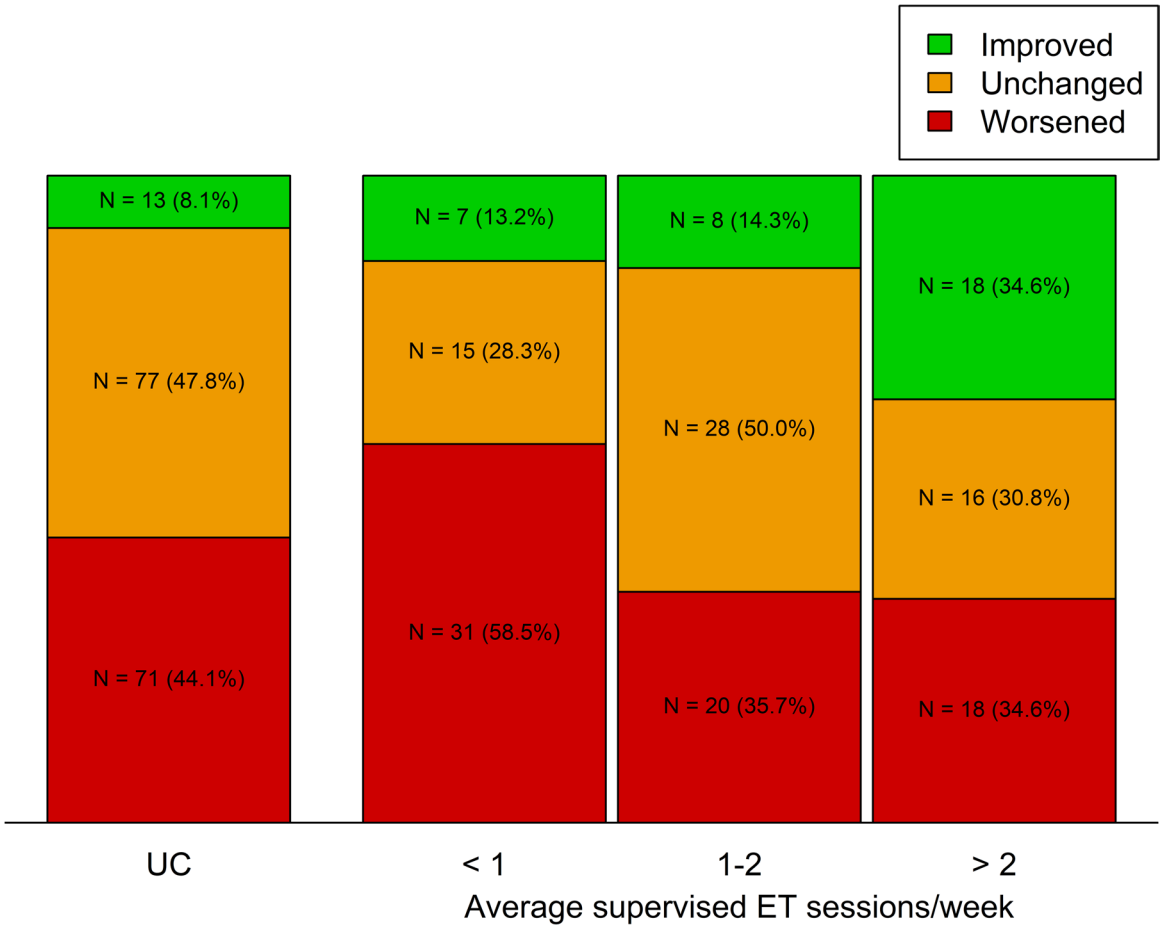
https://doi.org/10.1038/s41591-024-03342-7





Extended Data Fig. 5 | Subgroup analysis for the modified Packer score (A) and the change in peak oxygen consumption $\dot{V}O_2$ (B). Subgroup analyses for the modified Packer score were based on two-sided ordinal regression analyses with tests for interaction between the respective variables and study group. Subgroup analyses for the change in peak VO_2 were based on two-sided mixed-linear regression models with tests for interaction between the respective variables and study group. P-values have not been adjusted for multiple comparisons.

https://doi.org/10.1038/s41591-024-03342-7



Extended Data Fig. 6 | Association between adherence to exercise training (ET) and the modified Packer score compared to usual care (UC). Based on a two-sided Mantel-Haenszel test of trend in proportions, there was a significant association between exercise training frequency and the modified Packer score (P = 0.002). (UC = no exercise, ET = divided into three groups based on the adherence to the supervised exercise training).

https://doi.org/10.1038/s41591-024-03342-7

Extended Data Table 1 | Results of the secondary endpoints after multiple imputation

End point	Timepoint	Exercise Training	Usual Care	Time × Group Interaction [§]
Change in peak $\dot{V}O_2$ [$]	Baseline	18.0 (4.8)	17.8 (4.8)	- - -
	6 months	19.0 (4.9)	18.0 (5.5)	0.8 (0.49), P = 0.10
	12 months	19.4 (5.1)	18.0 (5.2)	1.1 (0.47), P = 0.02
Change in septal E/e' [$]	Baseline	15.7 (6.0)	16.2 (6.1)	- - -
	6 months	15.1 (6.5)	15.9 (6.5)	-0.3 (0.66), P = 0.62
	12 months	15.5 (6.6)	16.5 (6.9)	-0.5 (0.69), P = 0.45
Change in GSA [$]	Baseline	2.9 (1.1)	3.0 (1.1)	- - -
	6 months	2.9 (1.1)	3.1 (1.3)	-0.1 (0.15), P = 0.74
	12 months	2.9 (1.1)	3.1 (1.3)	-0.1 (0.15), P = 0.73
NYHA class [#]	Baseline	0 / 138 / 23	0 / 139 / 22	- - -
	Change at 6 months	33 / 126 / 2	17 / 136 / 8	2.40 (1.35), P = 0.004
	Change at 12 months	50 / 107 / 4	14 / 133 / 14	4.50 (1.34), P < 0.001

Demographic and clinical data, parameters of echocardiography, CPET and quality of life were included to generate 50 complete datasets applying multiple imputation with 20 simulations by chained equations. In each imputed dataset, the components of the primary endpoint were combined into the ordinal primary endpoint. Imputed data were analyzed by linear mixed models including random intercept for center and important covariates, such as age, sex, training compliance and daily activity. The results were pooled following the rules from Rubin[50]. The exact *P* value for the change in NYHA class at 12 months was *P*=0.000000002. [$] Values are shown as mean (s.d.) for each visit. [#] Values are shown as number of patients in NYHA I/II/III at baseline and number of patients with improved/unchanged/worsened NYHA class at 6 months and 12 months. [§] Results for the time×group interaction are shown as effect measure (standard error) for peak VO_2, septal E/e' and GSA and as OR (standard error) for NYHA class.

https://doi.org/10.1038/s41591-024-03342-7

Extended Data Table 2 | Hospitalizations that were related to a health-related problem that was identified before, during or after a supervised ET session

Diagnosis / Symptoms	Consequence / Procedure
Health issue that were detected before the start of an exercise training session	
• Atrial fibrillation	Catheter ablation
• Atrial flutter	Cardioversion
• Atrial flutter	Isthmus ablation
• Atrial flutter	Electrical cardioversion and beta blocker therapy
• Hypertensive blood pressure	Intensified drug treatment
• Hypertensive blood pressure	Intensified drug treatment, pacemaker implantation for III° AV block
• Atrial flutter	Catheter ablation and antiarrhythmic medical therapy
Health issue that occurred during exercise training	
• Dyspnea	Coronary artery disease suspected, ruled out by coronary angiography, no other cause identified
• Unstable angina	Coronary angiography, no intervention necessary
• Dizziness	Severe carotid artery stenosis, successfully treated by surgery
• Dizziness	No pathophysiology explaining dizziness was identified

Based on a post hoc evaluation of all hospitalizations in patients randomized to the ET group.

https://doi.org/10.1038/s41591-024-03342-7

Extended Data Table 3 | Association between adherence to ET and change in peak VO_2

Adherence to supervised ET (%)	Number of patients (N [%])	Mean difference in peak $\dot{V}O_2$ between ET subgroups and UC (mL/kg/min)	Standard error	adj. P-value	95% confidence interval	
≤ 33.3%	27 [20.8 %]	0.17	0.95	1.00	-2.10	2.43
33.3 - 66.7%	40 [30.8%	0.89	0.61	0.37	-0.57	2.35
66.7 - 100%	63 [48.5%]	3.12	0.60	<0.001	1.68	4.56

Adherence to exercise training was significantly associated with the change in peak VO_2 (single-factor two-sided ANOVA $P < 0.001$). *P* values for post hoc tests were adjusted for multiple testing using the Dunnett method. The exact *P* value for the 66.7–100% subgroup was 0.000001. The analysis is based on complete cases (peak VO_2 values at baseline and at 12 months).

nature portfolio

Corresponding author(s): Frank Edelmann

Last updated by author(s): Sep 25, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We did not use computer code to collect the data in this study
Data analysis	Data analysis was performed with IBM SPSS (version 24 and later) and R (versions 3-4), including the packages VGAM, mice, lme4, WinRatio and survival

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

As patients have not explicitly consented to sharing pseudonymized data, we are not allowed to share the individual participant data for legal reasons. However, upon request to the corresponding author (Frank Edelmann, frank.edelmann@dhzc-charite.de), aggregated data (data on demographics, clinical history, pharmacological treatment, electrocardiography, laboratory parameters, echocardiography, cardiopulmonary exercise testing, global self-assessment) that does not

allow identification of individual patients may be shared after consultation with the data protection officers and legal representatives of the participating institutions and after signing a data sharing agreement. A response to requests for data access can be expected within four weeks.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Self-reported sex is reported in the manuscript. We did not yet perform sex-based analyses. In order not to overload the manuscript, results reported in this manuscript have been limited to the composite primary endpoint and the components of the primary endpoints (as secondary endpoints). Additional secondary endpoints and subgroup analyses are planned to be published separately.
Reporting on race, ethnicity, or other socially relevant groupings	Self-reported ethnicity is reported in the results section. Of the 322 randomized patients, 319 stated that they were Caucasian (99%).
Population characteristics	All patients fullfilled the diagnostic criteria for heart failure with preserved ejection fraction according to Paulus et al. (2007). Among the 322 patients, 60% were female and 99% stated they were Caucasian. Mean age at randomization was 70 years and 86% of patients were classified as having symptoms according to NYHA II .At baseline, patients were on stable medication for at least four weeks, 84% had a history of arterial hypertension, 31% had atrial fibrillation and 38% had documented coronary artery disease. Additional baseline characteristics are shown in Table 1 of the manuscript.
Recruitment	Patients were recruited in cardiological or sport-cardiological outpatient clinics or cardiological practices. There might be a natural self-selection of patients who are able and willing to participate in an exercise training study. This could lead to the fact that the average study patient may be less sick than the average patient in the entire HFpEF population. Nevertheless, subgroup analyses of previous trials in stable patients with HFpEF (OptimEx-Clin; JAMA. 2021;325(6):542–551) and HFrEF (ExTraMATCH II; Eur J Heart Fail. 2018;20(12):1735–1743; J Am Coll Cardiol. 2019;73(12):1430-1443) did not show any significant effect modifications related to age, peak V O2 or NYHA class. Similarly, post-hoc subgroup analyses of the present trial did not show significant interaction effects between groups and age, peak V O2, NYHA class or diuretics for either the primary endpoint or the change in peak V O2. This indicates that for stable patients with HF, these parameters do not significantly influence treatment effects and the results of the present and previous exercise intervention trials may likely be also applied to sicker yet stable patients with HF.
Ethics oversight	The study protocol was approved by the ethics committee of the University Medical Center Göttingen (Germany) and the ethics committees at all participating sites.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size calculation was based on the Ex-DHF pilot study and the I-PRESERVE trial. Assuming an improved modified Packer score in 20% of ET and 5% of UC patients, and a worsened modified Packer score in 15% of ET and 25% of UC patients, a power of 98% and a significance level of $\alpha = 0.05$ resulted in 160 patients per treatment group. More details can be found in the trial protocol included in the supplementary material.
Data exclusions	No data was excluded from the analyses
Replication	To ensure correct coding of the composite primary endpoint, the calculation was independently performed by two statisticians on two software platforms (SPSS and R). Analyses of the primary and secondary endpoints were repeated in sensitivity analyses.
Randomization	Randomization was stratified by center, NYHA class, and peak V O2 (less vs. equal or greater than 20 mL/kg/min) and performed in a web-based system using Pocock's minimization algorithm with random component.
Blinding	Due to the nature of the intervention (lifestyle internvetion), blinding of participants and local investigators was not feasible. However, several outcomes were evaluated blinded to treatment group assignment, including echocardiographic parameters and the evaluation of adverse events by the safety committee. Calculation of the hierarchical composite endpoint and the statistical analyses were also performed in a blinded manner.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	ISRCTN86879094 at https://www.isrctn.com/
Study protocol	the study protocol has been uploaded as supplementary material
Data collection	The data was collected in cardiological or sport-cardiological outpatient clinics or cardiological practices in [Göttingen, Munich, Berlin, Greifswald, Schwäbisch Hall, Wuppertal, Magdeburg, Tübingen, Bremen, Ulm, Germany; and Graz, Austria] from December 2011 (first-patient-first-visit) until March 2017 (last-patient-last-visit).
Outcomes	The primary endpoint was the difference in the modified Packer score between groups at 12 months. The modified Packer score combined six heart failure outcome parameters including all-cause mortality, hospitalizations that were potentially related to heart failure or exercise training and a clinical score consisting of change in peak V O2, change in septal E/e', change in NYHA class and change in global self assessment (7-point Likert scale from "very bad" to "very good"; higher values indicating better health status) in a hierarchical manner and was graded as "worsened", "unchanged" or "improved". More details can be found in Supplementary Figure 1 or the trial protocol). As the algorithm also allowed to assess patients who were lost to follow-up, the primary endpoint was available for all study participants. Secondary endpoints reported in this manuscript are limited to the components of the primary endpoint including time to all-cause death or cardiovascuflar hospitalization as well as the change in peak V O2, E/e', NYHA class and GSA at 6 and 12 months.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A