Exercise Training for Chronic Heart Failure (ExTraMATCH II): Individual participant data meta-analysis of randomised controlled trials

Keywords:

Heart failure, meta-analysis, cardiac rehabilitation, randomised controlled trials, surrogate outcomes

Authors Affiliation:

Prof Rod Taylor (Chief Investigator) - University of Exeter Medical School and University of Glasgow, UK

Dr Sarah Walker - University of Exeter Medical School, UK

Dr Oriana Ciani - University of Exeter Medical School, UK and Centre for Research on Health and Social Care Management, Bocconi University, Milan, Italy

Dr Fiona C Warren - University of Exeter Medical School, UK

Professor Neil A Smart - University of New England, Australia

Professor Massimo Piepoli –Guglielmo da Saliceto Hospital, Italy

Associate Professor Constantinos H Davos – Academy of Athens, Greece

Corresponding author:

Professor Rod Taylor, Institute of Health Research, University of Exeter Medical School, College of Medicine and Health, College House, St Lukes Campus, Heavitree Road, Exeter, EX1 2LU, England, United Kingdom. Email: r.taylor@exeter.ac.uk Tel: +44 (0)7968 152537

Conflicts of interest:

Taylor is currently co-chief investigator on a National Institute for Health Research (NIHR) funded programme grant designing and evaluating the clinical and cost-effectiveness of a home-based cardiac rehabilitation intervention for heart failure patients (RP-PG-1210-12004). He is also a member of NIHR Priority Research Advisory Methodology Group (PRAMG), August 2015-present. Previous roles include: NIHR South West Research for Patient Benefit (RfPB) Committee South West, 2010-2014; Core group of Methodological

Experts for the NIHR Programme Grants for Applied Research programme, 2013-October 2017; NIHR HTA Themed Call Board, 2012-2014; NIHR HTA General Board, 2014- June 2017 and Chair of NIHR Health Services and Delivery Research Researcher-Led panel, March 2014-Feb 2018. All other authors declare no conflicts.

Abstract

(Word count 555)

Background: Current national and international guidelines on the management of heart failure (HF) recommend exercise-based cardiac rehabilitation (ExCR), but do not differentiate this recommendation according to patient subgroups.

Objective(s): (1) to obtain definitive estimates of the impact of ExCR interventions versus control (no exercise intervention) on mortality, hospitalisation, exercise capacity, and health-related quality of life (HRQoL) in HF patients; (2) to determine the differential (subgroup) effects of ExCR in HF patients according to their age, gender, ejection fraction, aetiology, New York Heart Association (NYHA) class, and baseline exercise capacity; (3) to assess whether the change in exercise capacity mediates for the impact of the ExCR on final outcomes (mortality, hospitalisation, and HRQoL) and is an acceptable surrogate endpoint.

Design: Individual participant data (IPD) meta-analysis

Setting: An international literature review

Participants: HF patients in randomised controlled trials (RCTs) of ExCR

Interventions: ExCR for at least 3 weeks compared with no exercise control with 6 months follow-up

Main outcome measures: mortality (all cause and HF-specific), hospitalisation (all-cause & HF-specific), exercise capacity, and HRQoL

Data sources: Individual participant data from eligible RCTs

Review methods: RCTs from ExTraMATCH IPD meta-analysis and 2014 Cochrane systematic review of ExCR

Results: Out of the 23 eligible RCTs (4,398 patients), 19 RCTs (3,990 patients) contributed data to this IPD meta-analysis. There was a wide variation in exercise programme prescriptions across included studies. Compared with control, there was no statistically significant difference in pooled time to event estimates in favour of ExCR although confidence intervals were wide: all-cause mortality: hazard ratio (HR) 0.83 (95% confidence interval (CI): 0.67 to 1.04), HF-related mortality: HR 0.84 (95% CI: 0.49 to 1.46), all-cause hospitalisation: HR 0.90 (95% CI: 0.76 to 1.06), and HF-related hospitalisation: HR 0.98 (95% CI: 0.72 to 1.35). There was a statistically significant difference in favour of ExCR for exercise capacity and HRQoL. Compared to control, at 12-months follow-up, improvements were seen in the six-minute walk test (6MWT) (mean: 21.0 metres, 95% CI: 1.57 to 40.4, and Minnesota Living with HF Questionnaire score (mean: -5.94, 95% CI: -1.0 to -10.9, lower scores indicate improved HRQoL). No strong evidence for differential intervention effects across patient characteristics was found for any outcomes. Moderate to good levels of correlation (R^2_{trial}>50% & ρ>0.50) between peak oxygen uptake (VO_2peak) or 6MWT with mortality and HRQoL were seen. Estimated surrogate threshold effect (STE) was an increase of 1.6 to 4.6 ml/kg/min for VO_2peak.

Limitations: Lack consistency in how included RCTs defined and collected the outcomes; we were unable to obtain IPD from all includable trials for all outcomes; and we did not seek patient level on exercise adherence. .

Conclusions: In comparison to no exercise control, participation in ExCR improves the exercise and HRQoL in HF patients but appears to have no effect on their mortality or hospitalisation. No strong evidence was found of differential intervention effects of ExCR across patient characteristics. VO_2peak and 6MWT may be suitable surrogate endpoints for the treatment effect of ExCR on mortality and HRQoL in HF.
*

Future work: Consensus on definition, collection, and reporting of core sets of outcome data future ExCR RCTs in HF; continuance of policies that encourage RCTs authors to make their datasets available.

Study registration: This study is registered as PROSPERO number CRD42014007170

Funding details: This project was funded by the NIHR Health Technology Assessment programme (HTA 15/80/30)

Table of Contents

List of tables

List of figures

List of appendices

List of abbreviations

ExCR	Exercise-based cardiac rehabilitation
HF	Heart failure
HFpEF	HF with preserved ejection fraction
HFrEF	HF with reduced ejection fraction
HRQoL	Health-related quality of life
HSCIC	Health and Social Care Information Centre
IPD	Individual participant (or patient) data
ISWT	Incremental shuttle walk test
KCCQ	Kansas City Cardiomyopathy Questionnaire
MLHFQ	Minnesota Living with Heart Failure
NICE	National Institute of Health and Care Excellence
NIH	National Institutes of Health
NIHR	National Institute of Health Research
NYHA	New York Heart Association
PPI	Patient and Public Involvement
RCT	Randomised controlled trial
VO_2peak	Peak oxygen uptake
6MWT	6-minute walk test

Plain English summary

Exercise-based cardiac rehabilitation is currently recommended in both UK and international clinical guidelines for people with heart failure. However, it is remains uncertain whether the effects of cardiac rehabilitation are consistent across patient subgroups (e.g. men versus women). We sought to review available scientific evidence using individual patient data in order to look at this issue.

We searched electronic literature databases for published studies and sought anonymised individual patient data from the researchers who conducted these research studies. We were able to bring together data from some 3,900 people with heart failure.

Although our analyses of this data show that participation in exercise-based cardiac rehabilitation does not appear to impact on the risk of death or hospitalisation, participation does offer some improvement in the physical fitness and quality of life of people with heart failure. We also found that these benefits were irrespective of patient's age, gender, ethnicity, initial level of physical fitness, or disease severity.

Scientific summary

Background

People with symptomatic heart failure (HF) are living for longer following the onset of their condition, increasing the importance of effective and accessible services for these patients. Exercise-based cardiac rehabilitation (ExCR) is recognised as integral to the comprehensive care of HF patients. ExCR is a process by which patients, in partnership with health professionals, are encouraged and supported to achieve and maintain optimal physical health. Current national and international guidelines on the management of HF recommend ExCR, but do not differentiate according to patient subgroups.

Objectives

The Exercise Training Meta-Analysis of Trials for Chronic Heart Failure (ExTraMATCH II) project aimed to determine which HF patient subgroups benefit most from ExCR using individual participant data (IPD) meta-analysis.

The project had three objectives.

1. To obtain definitive estimates of the impact of ExCR interventions versus control (no exercise intervention) on mortality, hospitalisation, exercise capacity, and health-related quality of life (HRQoL) in HF patients.
2. To determine the differential (subgroup) effects of ExCR in HF patients according to their: (i) age, (ii) gender, (iii) left ventricular ejection fraction, (iv) HF aetiology, (v) New York Heart Association (NYHA) class, and (vi) baseline exercise capacity.
3. To assess whether the change in patient exercise capacity mediates and is an acceptable surrogate endpoint for the impact of the ExCR on final outcomes (mortality, hospitalisation, and HRQoL).

The information gained from the ExTraMATCH II project will inform future UK and international clinical and policy decision-making on the use of ExCR in HF.

Methods

We conducted and reported this study in accordance the Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data (PRISMA IPD) statement. Randomised controlled trials were identified from the original ExTraMATCH IPD meta-analysis and the 2014 Cochrane systematic review of ExCR for HF; these were based on searches of the following electronic databases: Cochrane Central Register of Controlled Trials (CENTRAL) in The Cochrane Library, EMBASE, MEDLINE, CINAHL, PsycINFO, and the NHS Centre for Reviews and Dissemination. Conference proceedings and trial registers were also searched. In keeping with the original ExTraMATCH IPD meta-analysis, trials of exercise training for at least 3 weeks compared with no exercise control with 6 months' follow-up or longer were included if they provided IPD on mortality or hospitalisation (all-cause or HF-specific) time to event or exercise capacity or HRQoL. The datasets of IPD were combined into a single dataset. One-stage fixed effect meta-analyses of time-to-event endpoints were performed using Cox proportional hazards models, stratified by study. One-stage meta-analyses of continuous outcomes were performed using hierarchical linear models with adjustments for baseline values and a random effect on study. Two-stage models using fixed and random effects were also performed. Interactions terms between ExCR and participant characteristics were used to assess potential differential effects of ExCR across subgroups. Mediational analyses and meta-analytic regressions, with estimation

of R^2 at the trial level, and surrogate threshold effect (STE) were performed to assess the question of surrogate validity for exercise capacity outcomes of peak oxygen uptake (VO_2peak) and six minute walk test (6MWT).

Results

Of the 23 eligible trials (4398 patients), 19 trials contributed data to the IPD meta-analysis – 18 trials (3912 patients) to the clinical events (mortality and hospitalisation) analysis, 13 trials (3332 patients) to exercise capacity and HRQoL analysis, and 10 trials (2656 patients) to the exercise capacity mediational/surrogate endpoint analysis.

Characteristics and quality of included trials

Patient characteristics at baseline were well balanced between ExCR and control group patients. The majority of patients were male (75%), with a mean age of 61 years and predominantly with reduced ejection fraction HF (HFrEF) (mean baseline left-ventricular ejection fraction 26.7% no included trials recruited patients with preserved ejection fraction heart failure - ejection fraction >45%), and most patients were in NYHA functional class II (59%) or III (37%). Trials were from Europe and North America and were published between 1990 and 2012. Sample size ranged from 50 to 2130 patients. All trials evaluated an aerobic exercise intervention, which was most commonly delivered in either an exclusively centre-based setting or a centre-based setting in combination with some home exercise sessions. The dose of exercise training ranged widely across trials. ExCR was delivered over a period of 12 to 90 weeks, with between 2 and 7 sessions per week; median session duration was between 15 and 120 minutes (including warm-up and cool-down). The intensity of exercise ranged between 50 to 85% peak VO_2. The overall quality of included trials was judged to be moderate to good, with a median TESTEX score of 11 (range 9 to 14) out of a maximum score of 15.

Impact of ExCR on mortality and hospitalisation

Compared with control, there was no statistically significant difference in pooled time to event estimates in favour of ExCR although confidence intervals were wide: all-cause mortality: hazard ratio (HR): 0.83 (95% confidence interval (CI): 0.67 to 1.04), HF-related mortality: HR 0.84 (95% CI: 0.49 to 1.46), all-cause hospitalisation: HR 0.90 (95% CI: 0.76

to 1.06), and HF-related hospitalisation: HR 0.98 (95% CI: 0.72 to 1.35). No strong evidence for differential intervention effects across patient characteristics was found.

Impact of ExCR on exercise capacity and HRQoL

Compared with control, there was a statistically significant difference in favour of ExCR for exercise capacity and HRQoL. For example, at 12-month follow-up, improvements were seen in the 6MWT (mean: 21.0 metres, 95% CI: 1.57 to 40.4, p=0.034, $\tau 2 = 491$, $I^2 = 78\%$) and Minnesota Living with HF Questionnaire score (mean: -5.94, 95% CI -1.0 to -10.9, p=0.018, $\tau 2 = 77$, I2 =88%; lower scores indicate improved HRQoL). No strong evidence for differential intervention effects across patient characteristics was found.

Validation of exercise capacity as a surrogate endpoint

Moderate to good levels of correlation (R^2_{trial}>50% and ρ>0.50) between exercise capacity VO_2peak or 6MWT with mortality and HRQoL were seen. Estimated STE was an increase of 1.6 to 4.6 ml/kg/min for VO_2peak. Our results indicate that an increase in VO_2peak or 6MWT with ExCR to be potentially weak mediators of final outcomes.

Discussion

In HFrEF patients ExCR did not have a statistically significant effect on the risk of mortality and hospitalisation. However, uncertainty around effect estimates and lack of individual patient data on exercise adherence precludes drawing definitive conclusions in these event outcomes. ExCR significantly improves exercise capacity and HRQoL. We found no consistent differences in ExCR effects across patient subgroups. Our results provide indicative evidence that VO_2peak and 6MWT may be suitable surrogate endpoints for the treatment effect of ExCR on final outcomes in HF.

Recommendations for further research

Two central aspects of future data collection include: a consensus on the definition, collection, and reporting of core sets of outcome data, concomitant disease/comorbidities and metrics of therapy delivery/uptake plus the capture of data on patient level adherence to the amount of exercise training during the ExCR intervention period. More generally, the research community should continue to implement policies that encourage primary study

authors to make their datasets available, either by depositing in publicly available repositories or shared with IPD meta-analysis collaborations when directly requested.

Study registration

This study is registered as PROSPERO number CRD42014007170.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research (HTA 15/80/30).

Chapter 1: Background

Chronic heart failure (HF) is a burgeoning global health challenge that affects 1–2% of adults in the western world. [1] While survival after HF diagnosis has improved, prognosis is poor - 30 to 40% of patients die within a year of diagnosis. [2] Patients with HF experience limitations to their exercise capacity, activities of daily living, and health-related quality of life (HRQoL) and an increased risk of hospital admission rate and all-cause mortality. [3, 4] The cost of management of HF in the UK National Health Service (NHS) was reported to be approximately £1 billion in 2010. [5] According to the Office of National Statistics, the proportion of the UK population aged 85 and over is projected to double between 2016 and 2041. [6] Due to increases in both the incidence and prevalence in heart failure with increasing age, [7] more demands will be placed on the NHS in this time frame. An increase in the prevalence of comorbidities in an older population will lead to a greater number of hospitalisations in heart failure patients. [8]

With increasing numbers of people living longer with symptomatic HF, the effectiveness and accessibility of health services for HF patients have never been more important. Exercise-based cardiac rehabilitation (ExCR) is recognised as integral to the comprehensive care of HF patients. Cardiac rehabilitation is a process by which patients, in partnership with health professionals, are encouraged and supported to achieve and maintain optimal physical health. [9] Whilst exercise training is at the centre of cardiac rehabilitation, it is accepted that programmes should be comprehensive in nature and include education and psychological input focusing on health and life-style behaviour change and psychosocial well-being. [2-4, 9]

Previous systematic reviews and meta-analyses have shown exercise-based rehabilitation offers important health benefits for patients. [9-12] Including 33 trials across 4740 HF patients, the 2014 Cochrane review [10] shows: no difference in pooled all-cause mortality with ExCR (relative risk: 0.93; 95% CI 0.69 to 1.27), reduced risk of overall hospitalisation (relative risk: 0.75; 95% CI: 0.62 to 0.92) and HF-specific hospitalisation (relative risk: 0.61; 95% CI: 0.46 to 0.80); and a clinically important improvement in disease-specific health-related quality of life (HRQoL) on the Minnesota Living with Heart Failure (MLHFQ) questionnaire (mean difference: -5.8 points, 95% CI: -9.2 to -2.4). ExCR for HF is therefore recommended by the

National Institute of Health and Care Excellence (NICE) [3] and is a class I recommendation of the joint American College of Cardiology Foundation and the American Heart Association and the European Society of Cardiology guidelines. [13-15] These guidelines do not differentiate by patient subgroup but, rather, recommend CR to all HF patients 'who are able to participate to improve functional status'. [13]

Despite this evidence and recommendation by clinical guidelines, the uptake of ExCR for HF remains poor. Only 16% of UK CR centres have a specific rehabilitation programme for HF. [16] The recent ExtraHF survey reported that only 40% of centres from across 42 European countries implemented an exercise programme for HF.,. (17) Cardiac rehabilitation centres report lack of resources to the major barrier to providing rehabilitation services for HF, i.e., lack of finances, staff, and equipment. [16, 17] A key potential solution (if supported by evidence) could be targeting exercise-based rehabilitation services to those HF patients who might experience the greatest benefit in outcomes. Such a differential effect of treatment across HF patients could improve the overall clinical and cost-effectiveness of rehabilitation for HF and drive improvements in patient uptake of rehabilitation.

Although meta-analyses demonstrate important health benefits with ExCR, there is uncertainty whether there are differential effects across HF patient subgroups. Three data sources currently provide evidence on this issue but have weaknesses. First, in 2004, the Exercise Training Meta-Analysis of Trials in Heart Failure (ExTraMATCH) Collaborative Group published an individual participant data (IPD) meta-analysis based on 9 randomised trials in 801 HF patients, showing Ex CR reduced all-cause mortality (hazard ratio 0.65, 95% CI:, 0.46 to 0.92) and there were no subgroup (age, gender, HF aetiology, New York Heart Association (NYHA) class, ejection fraction, or exercise capacity) effects. [18] Given the small number of trials, patients, and events (193 deaths) these subgroup analyses are likely to be underpowered. Furthermore, a number of trials have been published since, including HF-ACTION, a large US National Institute of Health funded randomised trial (2331 HF patients across 82 centres). [19] Second, the original analysis of the HF-ACTION trial found no interactions between treatment allocation (ExCR or no exercise control) and patient characteristics (age, gender, HF aetiology, NHYA class, ejection fraction, or depression score) for the composite outcome of mortality or hospital admission. [19] Although the largest ExCR trial to date, the power of this study to detect small subgroup effects remains limited.

Finally, meta-regression analysis in the 2014 Cochrane review found no association between trial level patient characteristics (age, gender, ejection fraction) and the impact of ExCR. [10] However, such analysis is highly prone to study level confounding (ecological fallacy) and should be interpreted with great caution. The methodology of IPD meta-analysis allows more robust analysis of treatment effects in subgroups and consistent analysis of outcome data across trials, such as enabling time to event data analyses adjusted for baseline covariates.

Chapter 2: Aims and objectives

The Exercise Training Meta-Analysis of Trials for Chronic Heart Failure (ExTraMATCH II) project aimed to determine which HF patient subgroups benefit most from ExCR using IPD meta-analysis.

The project objectives were:

1. To obtain definitive estimates of the impact of ExCR interventions versus control (no exercise intervention) on all-cause mortality, hospitalisation, HRQoL and exercise capacity in HF patients

2. To determine the differential (sub-group) effects of exercise-based interventions in HF patients according to their (i) age, (ii) gender, (iii) left ventricular ejection fraction, (iv) HF aetiology, (v) NYHA class, and (vi) baseline exercise capacity

3. To assess whether the change in patient exercise capacity mediates and acts as a surrogate endpoint for the impact of the ExCR on all-cause mortality, all-cause hospitalisation, and disease-specific health-related quality of life.

The information gained from the ExTraMATCH II project will inform future national and international clinical and policy decision-making on the use of ExCR in HF.

Chapter 3: Methods

This project was undertaken and reported according to current reporting guidelines for individual patient data meta-analyses [20-22] and was registered with PROSPERO (CRD42014007170). [23] The project management committees are listed in Appendix 1**Error! Reference source not found.**.

Identification of trials for inclusion

Trials for were identified from ExTraMATCH IPD meta-analysis and the 2014 Cochrane systematic review of ExCR for HF. [10, 18] The Cochrane review searched (to January 2013) the following electronic databases: Cochrane Central Register of Controlled Trials (CENTRAL) in The Cochrane Library, EMBASE, MEDLINE, CINAHL, PsycINFO, and the NHS Centre for Reviews and Dissemination (CRD); search strategy is included in Appendix 2. Conference Proceedings were searched on Web of Science. Trial registers (Controlled-trials.com and Clinicaltrials.gov) and reference lists of all eligible trials and identified systematic reviews were also checked. No language limitations were imposed. Details of the search strategy used are reported elsewhere [23] and are included in Appendix 1**Error! Reference source not found.**.

Trials were included with if they met the following criteria:

1. *Study design:* Randomised controlled trials (RCTs) with a follow-up period of 6 months or more (in accord with the 2014 Cochrane review)
2. *Target population:* Adult patients, aged 18 years and over, with a diagnosis of HF with reduced ejection fraction (HFrEF) or HF with preserved ejection fraction (HFpEF), based on objective assessment of left ventricular ejection fraction and on clinical findings
3. *Setting / context:* Patients managed in any setting i.e. hospital, community facility or patient's home
4. *ExCR intervention:* An ExCR intervention that included at least an aerobic exercise training component performed by the lower limbs, lasting a minimum of 3 weeks, (24) either alone or as part of a comprehensive cardiac rehabilitation programme which may also include health education and/or a psychological intervention

5. *Comparator:* A non-exercise group receiving standard medical care or an attention placebo
6. *Sample size:* A sample size of more than 50, to ensure that the logistical effort in obtaining, cleaning and organising the data were commensurate with the contribution of the data set to the analysis. [25, 26]

Identified randomised controlled trials meeting the inclusion criteria are shown in Appendix 3**Error! Reference source not found.**. Study selection for the 2014 Cochrane review and ExTraMATCH IPD meta-analysis was performed by the original research teams who performed these studies. For the purposes of this project, a single researcher (RST) compared the included studies from these two previous studies and applied the above inclusion criteria.

Investigator requests

The principal investigators of eligible studies were invited (Collaboration Invitation**Error! Reference source not found.**) to participate in this IPD meta-analysis and share their anonymised trial data. The list of variables which principal investigators were asked to provide was reported in the study protocol [27] (see Appendix 4).

Exclusion of trials from IPD analysis

Trials were excluded if:

1. They did not respond to the invitation to provide IPD for the ExTraMATCH II analysis in spite of repeated contacts attempts being made
2. They were unable to provide IPD, either because the data had been lost or destroyed
3. There were patients included in the trial who may also have appeared in another IPD dataset

Ethical approval

The ethics of obtaining data were carefully considered and advice sought from the Health and Social Care Information Centre (HSCIC), dated April 2016. The original trials had each obtained ethical/Institutional Review Board committee approval and obtained individual

patient consent. Given the fully anonymised nature of all the trial datasets (i.e. no inclusion of data such as patient name or date of birth, that would allow individual patients to be identified), HSCIC confirmed that there was no further legal/ethical or contractual requirements for use of this data for the purpose of this project. A revision of HF-ACTION [19] data were obtained via the NIH data portal which required that we obtained a letter of approval from the University of Exeter Medical School Research Ethics Committee, dated 13th November 2017 (Ethical Approval).

Data management

Data files were received in a variety of formats, depending on the security concerns of the host institutions. In most cases data transfer was by email of password protected file with a separate email containing the password. Each raw data file was saved in its original format on receipt and then converted to a Stata file. Data cleaning was carried out in each pseudonymised dataset prior to being combined in a master dataset. Within the individual datasets, data for each variable (at the patient level) was checked for accuracy in: range; extreme values; internal consistency; missing values, and consistency with published reports. Data discrepancies or missing information was discussed with trial investigators and corrected where appropriate.

All data files were stored on a secure password protected computer server managed and in accordance with the data management standard operating procedures of the UK Clinical Research Collaboration (UKCRC) registered Exeter Clinical Trials Unit. Access to data at all stages of cleaning and analysis was restricted to core members of the research team (OC, RST, FCW, and SW).

Patient and Public Involvement (PPI)

As part of the NIHR Programme Grant (Rehabilitation Enablement in Chronic Heart Failure - REACH-HF, PGfAR RP-PG-0611-12004), a PPI Group was established in 2009, consisting of eight active members (5 with lived experience of heart failure and 3 patient carer givers). The PPI Group are familiar with our ongoing portfolio of Cochrane systematic reviews in cardiac rehabilitation.

This IPD meta-analysis was proposed to the PPI group meeting in Truro on 1st Nov 2015 where views were sought on our proposed research questions. Following receipt of funding from the National Institute of Health Research (NIHR), the ExTraMATCH II project was presented to the PPI group at a further meeting, held in March 2017. Members of the group gave views on how the results should be best presented and disseminated to patients, care-givers and clinicians in order to impact on clinical practice and patient understanding of heart failure. Kevin Paul (PPI Group Chair) was a co-applicant for the REACH-HF study and was also a member of the REACH-HF Programme Steering Committee. Kevin is a core colleague and valued member of our team. He has agreed to act as conduit between the Project Advisory Group for ExTraMATCH II and our established PPI Group. The PPI group were asked to contribute to, and give views on: (i) the ExTraMATCH II protocol (e.g. whether we have prioritised the appropriate outcomes); (ii) lay summaries of the ExTraMATCH II project; (iii) the implications for clinical practice and future research; and (iv) the planned dissemination strategy.

Kevin commented on the plain English summary of the original application and also offered advice on the Plain English Summary of this document. Based on the INVOLVE guidelines, [28] we included the cost of his time to attend Project Advisory Group meetings, plus his travel.

Statistical analysis

All analyses were carried out according to the principle of intention to treat (i.e. patients as randomised in each trial arm and with complete outcome data at follow up). Where missing data were noted within an individual trial, contact with the author was attempted and data added if available. Given the relatively small levels of missing outcome and covariate data within trials, we did not undertake data imputation. Where possible, all one-stage and two-stage analyses used random effects models as the overall dataset is likely to include a high degree of clinical heterogeneity across the individual studies (differences in population, ExCR intervention and comparator). [29] All analyses were undertaken using Stata 14.2 StataCorp LP, College Station, Texas, USA.

Main outcomes

In accordance with the study research objectives we sought individual patient data for the following outcomes from eligible trials:

1) Mortality: incidence and time-to-event data for all deaths. In addition, we also sought to obtain data on the cause of death
2) Hospital admission: incidence, time-to-event and duration of hospitalisation. We also sought to obtain data on the cause of the hospitalisation
3) Disease specific health-related quality of life, as assessed by the MLHFQ questionnaire and other validated HRQoL outcomes: value at baseline (pre-randomisation) and outcome at 6, 12, 24 and >24 months post-randomisation
4) Exercise capacity, as assessed by peak oxygen uptake (VO_2 peak) and other validated exercise capacity measures: outcome at baseline and at 6, 12, 24 and >24 months post-randomisation

Patient subgroups

We also requested individual patient demographic and clinical data, including: age, gender, ejection fraction, NYHA class, heart failure aetiology (ischemic vs. non-ischemic), race/ethnicity and exercise capacity at baseline). Details of exercise training prescription (i.e. session frequency, duration, intensity and overall programme duration) was collected as part of the 2014 Cochrane review.

Statistical Analysis Plans

A detailed Statistical Analysis Plan was produced for each of the three analyses described below:

1) Impact of ExCR on mortality and hospitalisation outcomes
2) Impact of ExCR on HRQoL and exercise capacity outcomes
3) Validation of exercise capacity as a surrogate outcome

Descriptive statistics

For each analysis, patient-level characteristics were compared for those patients in the ExCR and control groups of the included studies. A descriptive of trial-level characteristics by group are also reported.

Assessment of study quality and risk of bias

We checked for potential small study bias by visual assessing funnel plot asymmetry and using the Egger test. [30] Study quality and risk of bias was assessed using the TESTEX quality assessment tool. [31] Statistical heterogeneity was assessed using the I^2 statistic. [29]

Impact of ExCR on mortality and hospitalisation

Inclusion of trials

Trials were included in the mortality and hospitalisation analyses if IPD was provided for the one or more of the outcomes of interest detailed below.

Outcomes of interest

The final patient-relevant outcomes of interest in this study were:

1. Time to event to all-cause mortality
2. Time to event to HF-related mortality
3. Time to event to all-cause hospital admission
4. Time to event to HF-related hospital admission

Due to the inconsistency of reporting in IPD sets, we were only able to consider time to event outcomes and not incidence or duration of events. Insufficient data were made available to allow analyses on 'sudden death' to be carried out.

Each of the outcomes described above were analysed separately. Each trial contributed to between one and four analyses.

Primary analysis

In the primary analysis, a two-stage IPD meta-analysis approach was taken, with each trial first analysed using a Cox regression model and then trial-specific estimates of treatment effects (i.e. hazard ratios) or treatment-covariate interactions (i.e. differences in hazard ratios) were meta-analysed across studies. A random effects model was used to account for the high degree of clinical heterogeneity across the individual studies due to differences in population, Ex-CR intervention, and comparator. [29] An overall estimate of the effect of Ex-CR for each outcome, both by trial and as a pooled estimate, was presented as a hazard ratio (HR) and 95% confidence interval (CI). Additionally, the τ^2 and I^2 statistics were reported alongside the associated p-value for the results of the main analyses. [29, 32] The Cochrane handbook advises that using specific threshold values for the interpretation of I^2 can be misleading. [33]

Secondary analysis

Secondary analysis used a one-stage IPD meta-analysis Cox regression model, stratified by trial. Stratification allowed the baseline hazard to vary between studies, rather than forcing the hazard in individual studies to be proportionate to each other. [34] No distributional assumptions about this baseline hazard were made. Due to failure of convergence in the one-stage random effect models, likely due to the low level of heterogeneity between studies, a fixed effect approach was used.

The within-trials interaction term used here identifies any patient characteristics which influence the effectiveness of Ex-CR on an individual level, necessary for making inferences for stratified medicine as recommended by Riley et al. [35] The within-trial interaction effect is fixed across trials. Continuous covariates were centred on the mean value within each trial; binary covariates were centred on the proportion within each trial.

Sensitivity analyses

To test the robustness of primary and secondary analyses, we undertook a number of pre-specified sensitivity analyses: we excluded the largest trial (HF-ACTION[19]); truncated outcomes at 1, 2, and 5 years follow up; and included trial level outcome data for studies that could not provide IPD. [26]

Impact of ExCR on HRQoL and exercise capacity

Inclusion of trials

Trials were included in the HRQoL and exercise capacity analyses if IPD was provided for the one or more of the outcomes of interest detailed below.

Outcomes of interest

The final patient-relevant outcomes of interest in this study were:

1. HRQoL measured using the MLHFQ score
2. HRQoL measured through any validated scale
3. Exercise capacity measured using VO_2peak (ml/kg/min)
4. Exercise capacity measured using 6-minute walk test (6MWT) (metres)
5. Exercise capacity measured using a standardised exercise capacity score calculated from any of the four validated exercise capacity measures listed below

HRQoL scales of measurement

HRQoL measured using one of three validated measures was included in this analysis:

(i) Minnesota Living with Heart Failure Questionnaire (MLHFQ) [36]

(ii) Kansas City Cardiomyopathy Questionnaire (KCCQ) [37], and

(iii) Guyatt scale [38]

The first HRQoL analysis was carried out for trials providing the MLHFQ data; the second analysis used a standardised score calculated from any of the three measures above.

Exercise capacity scales of measurement

Exercise capacity measured using one of four validated measures was included in this analysis:

(i) VO_2peak in ml/kg/min

(ii) Distance (metres) walked on the 6MWT

(iii) Distance (metres) walked in an incremental shuttle walk test (ISWT) and

(iv) Workload on cycle ergometer (watts)

Exercise capacity analysis was carried out for:

(i) Trials providing VO_2peak

(ii) Trials providing 6MWT

(iii) A standardised exercise capacity score, calculated from any of the validated exercise capacity measures listed above.

One study, HF-ACTION [19], provided data on both VO_2peak and 6MWT and was included in all analyses, with the VO_2peak measure taking precedence for the standardised exercise capacity analysis.

Primary analysis

The primary analyses included one-stage and two-stage IPD meta-analyses carried out at 6 and 12 months. At each time point, we used the observation closest to and prior to the time point. All one-stage IPD models used a hierarchical random effects regression model, adjusted for the baseline value of the outcome measure. All two-stage models used random treatment effects. We performed a series of models to estimate the overall treatment effect and to investigate potential interactions between ExCR and pre-defined patient subgroups (i.e., age, gender, left ventricular ejection fraction, heart failure aetiology, NYHA class and baseline exercise capacity [23, 27]) Each model investigated one interaction effect only. The I^2 and τ^2 statistics were reported alongside the associated p-value for the results of the main analyses. [29, 32]

Secondary analysis

The secondary analyses used a random effects hierarchical model for repeated measures at multiple time points. These models utilised HRQoL and exercise capacity outcome data at all available time points. Adjustments for baseline values of the outcome measure were made; no other covariates were included in the model. This model included a time by treatment interaction term.

Sensitivity analyses

To test the robustness of the primary analyses, pre-specified sensitivity analyses were carried out:

(i) Primary analysis was repeated after exclusion of the largest trial (HF-ACTION[19])

(ii) Addition of aggregate data from studies that did not provide IPD

Surrogate analysis

Inclusion of trials

All studies in the ExTraMATCH II meta-analysis were eligible for inclusion in the surrogate analyses, dependent on the availability of data on exercise capacity and final patient-relevant outcomes, as explained below.

Outcomes of interest

The final patient-relevant outcomes of interest in this validation study were:

1. HRQoL measured by MLHFQ score
2. HRQoL measured through any validated scale
3. Time to all-cause mortality
4. Time to all-cause hospital admission

For this study, three approaches to exercise capacity definition were used:

(i) Direct assessed VO_2peak

(ii) 6MWT

(iii) Direct and indirect VO_2peak (conversion from 6MWT and ISWT. No conversion was possible for watts as it is dependent on body weight of individual patients).

Distances recorded as either 6MWT or ISWT at baseline were converted to VO_2peak using previously reported methods. [39-43] Details can be found in Appendix 5.

Follow-up time considerations

The following outcome follow-up times were considered: ≤ 6 months for exercise capacity outcomes; ≤12 months for HRQoL outcomes; and all available follow-up time for mortality

and hospitalisation. This approach was consistent with the assumption of temporal antecedence for a causal relationship between the surrogate endpoint and the final outcomes.

Mediation analysis

Mediation is known as the phenomenon whereby a cause affects an intermediate variable (also called mediator), and the change in the intermediate variable goes on to affect the outcome. [44, 45] The effect of the cause on the outcome that operates through the intermediate of interest is sometimes referred to as an indirect or mediated effect. Mediation analysis is usually referred to the set of techniques by which a researcher assesses the relative magnitude of these direct and indirect effects. The product method specification of this approach was used to determine whether a change in VO_2peak (ΔVO_2peak) or a change in 6MWT (Δ6MWT) mediate the relationship between treatment assignment (i.e. ExCR vs no ExCR) and each of the final outcome of interest. Linear or Cox regression analyses were conducted to evaluate the following four hypotheses:

(i) Treatment assignment (i.e. ExCR vs control) has a significant effect on ΔVO_2peak or Δ6MWT from baseline to 6 months follow-up;

(ii) ΔVO_2peak or Δ6MWT have a significant effect on ΔMLHFQ or ΔHRQL or on the hazards of developing a clinical event;

(iii) Treatment assignment (i.e. ExCR vs control) has a significant effect on ΔMLHFQ or ΔHRQL or on the hazards of developing a clinical event;

(iv) The effect of treatment assignment (i.e. ExCR vs control) on ΔMLHFQ or ΔHRQL or on the hazards of developing a clinical event is attenuated when ΔVO_2peak or Δ6MWT is added to the model.

All regression models took into account the clustering within trials to allow for study-level differences in treatment effect and unstructured covariance between random intercept and random slope. Regression models were adjusted for baseline of either exercise capacity values or baseline HRQoL values. No other adjustments were made, because patients were randomly assigned to intervention or control arm. For criterion (ii) no adjustment for made for potential confounding.

We assumed necessary to reject the null for at least the first of these hypotheses (i.e. the treatment assignment is associated with the mediator) to support the validation of ΔVO_2peak

or Δ6MWT as mediator endpoints and proceed further with the estimation of proportion of the proportion explained or proportion mediated.

Meta-analytic approach: R^2 and surrogate threshold effect

Whilst mediation analysis considers pathways by which treatment effects may arise, surrogacy principally concerns whether we are able to predict the effect treatment on the final endpoint by using the effect of treatment on the surrogate.

Given the issues described with the proportion explained and indirect effects approaches in identifying consistent surrogates, the meta-analytic approach may offer the most promise for assessing surrogate outcomes and for making policy and treatment decisions. [46, 47] This approach requires multiple studies, or at least multiple subgroups (e.g. centres within a trial), which we have through the ExTraMATCH II IPD meta-analysis. Because a true and strong association between the treatment effect on the final endpoint and the treatment effect on the surrogate is considered to be the hallmark of surrogacy, [47] this approach proceeds as follows. Let ϕ_j denote the estimate of the effect of treatment on the final outcome in the jth study, let θ_j denote the estimate of the effect of treatment on the surrogate outcome in the jth study, both derived from RCTs. For a good surrogate, a monotonic relationship would exist between ϕ_j and θ_j and, in a regression of ϕ_j on θ_j, there would be limited variability around the regression line. If the relationship between ϕ_j and θ_j is approximately linear, a reasonable measure of surrogacy is the R^2_{trial} of the regression of ϕ_j on θ_j. Another intuitive measure recommended as a surrogacy metric is the surrogate threshold effect (STE), which takes into account the variability around the regression line and represents the intercept of the prediction band of the regression line with the zero effect line on the final outcome. [46] For each trial we estimated study-level treatment effects by conducting linear regression or Cox proportional hazards regression models. Adjustment was made for baseline exercise capacity or HRQoL values. Then we conducted linear meta-regressions to relate estimated difference in exercise capacity to the estimated effect on change in HRQoL log(HR) of all-cause mortality or log(HR) of all-cause hospitalisation events. The square of the inverse standard error was used as a weight to account for uncertainty in the estimated patient-relevant outcomes effect. We calculated commonly reported indicators of surrogate validation. [48] The correlation coefficient (ρ) and the R^2 for the relationship between treatment effect difference on exercise capacity and each

of the final outcomes was estimated individually using weighting by the inverse of the variance (for the treatment effect on final outcomes). In order to estimate STE, prediction bands where calculated based on approximate prediction intervals. [48, 49]

Chapter 4: Characteristics and quality of included studies

Identification of trials for inclusion in the ExTraMATCH II master dataset

A total of 23 trials were deemed eligible for the ExTraMATCH II IPD meta-analysis. Data from six trials have been analysed previously and were available from the ExTraMATCH database.[50-55] Fourteen investigators responded positively and shared their de-identified trial data directly. [19, 56-68]

Exclusion of eligible trials from the ExTraMATCH II master dataset

We were unable to include data from three trials (355 patients); for two trials data were no longer available [69, 70] and the investigators of the other trial could not be contacted. [71] After obtaining IPD, a further trial [72] was excluded, as it was determined that it included patient data that overlapped with another trial. [62] We therefore had a total of 19 trials in the ExTraMATCH II study, [19, 50-67] with a total of 3900 patients. A flow diagram to show inclusion and exclusion of trials in the ExTraMATCH II study is shown in Figure 1**Error! Reference source not found.**. Further flow diagrams to show inclusion and exclusion of trials and participants within individual analyses are given in the appropriate results sections below.

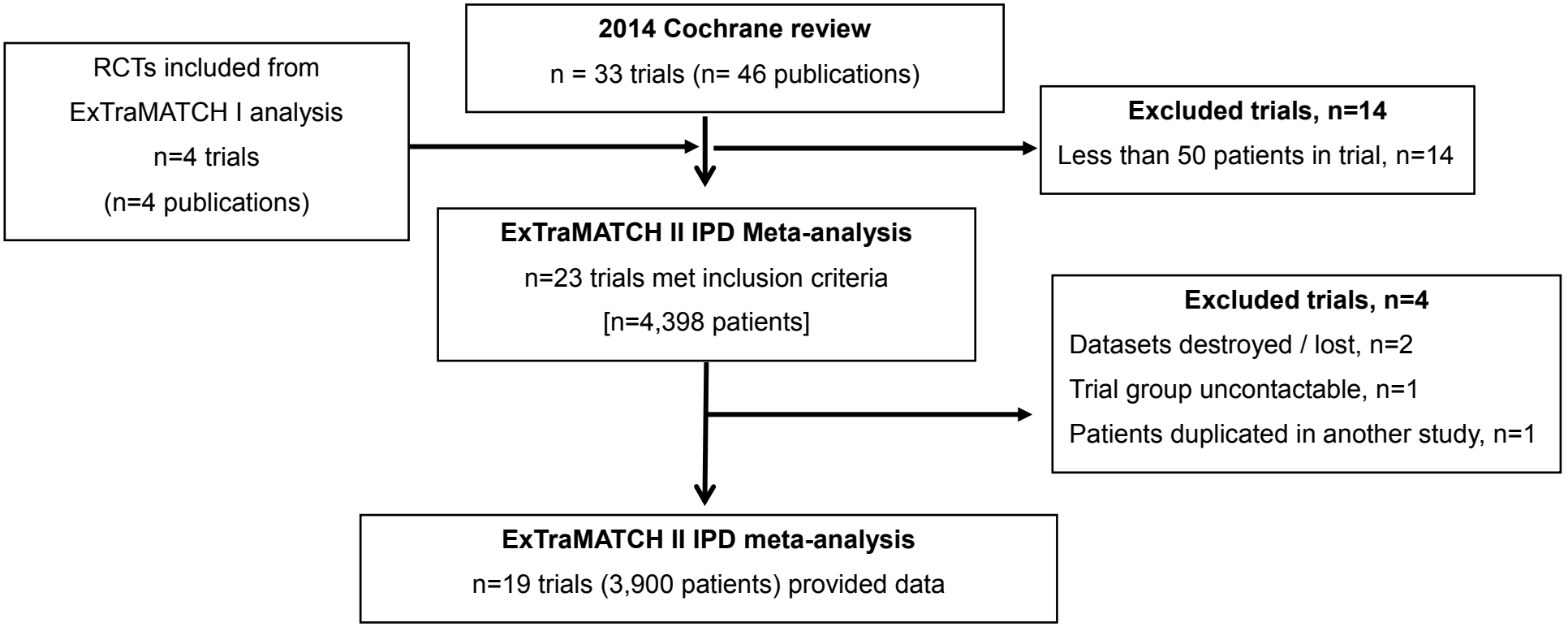


Figure 1: PRISMA flow diagram summarising selection of studies for the ExTraMATCH II study

Characteristics of included patients

Patient characteristics at baseline were well balanced between ExCR and control patients. The majority of patients were male (75%), with a mean age of 61 years (standard deviation (SD) 13). The mean baseline left-ventricular ejection fraction was 26.7% (SD 8.1%); no included trials recruited patients with preserved ejection fraction heart failure (ejection fraction >45%), and most patients were in NYHA functional class II (59%) or III (37%) (see Table 1**Error! Reference source not found.**).

Table 1: Baseline characteristics of patients in the ExTraMATCH II master dataset

Characteristic	**ExCR (n=1,986)**	**Control (n=2,003)**	**All (n=3,989)**
Age (years), mean (SD)	61.4 (12.8)	61.5 (13.1)	61.4 (13.0)
Gender			
Male	1455 (73.3)	1511 (75.4)	2966 (74.4)
Female	531 (26.7)	492 (24.6)	1023 (25.7)
Baseline ejection fraction (%); mean (SD)	27.2 (8.8)	26.9 (8.7)	26.9 (8.7)
NYHA status			
Class I	25 (1.3)	29 (1.5)	54 (1.4)
Class II	1124 (58.6)	1148 (59.5)	2272 (59.0)
Class III	721 (37.6)	728 (37.7)	1449 (37.7)
Class IV	47 (2.5)	26 (1.4)	73 (1.9)
Aetiology			
Ischaemic	1067 (57.3)	1055 (56.1)	2122 (56.7)
Non-ischemic	796 (42.7)	826 (43.9)	1622 (43.3)
Ethnicity			
White	1130 (70.2)	1163 (71.8)	2293 (71.0)
Non-white	480 (29.8)	458 (28.3)	938 (29.0)
VO_2 peak (ml/kg/min); mean (SD)	14.9 (4.3)	15.0 (4.6)	15.0 (4.4)

Data are n (%) unless otherwise indicated; percentages may not sum to 100 because of rounding; NHYA: New York Heart Association; SD: standard deviation; VO_2 peak: peak oxygen uptake.

Characteristics of included trials

Trials were from Europe and North America and were published between 1990 and 2012. Sample size ranged from 50 to 2130 patients. All trials evaluated an aerobic exercise intervention, which was most commonly delivered in either an exclusively centre-based setting or a centre-based setting in combination with some home exercise sessions. The dose of exercise training ranged widely across trials. ExCR was delivered over a period of 12 to 90 weeks, with between 2 and 7 sessions per week; median session duration was between 15 and 120 minutes (including warm-up and cool-down). The intensity of exercise ranged between 50 to 85% peak VO_2 (see Table 2**Error! Reference source not found.**).

Table 2: Characteristics of included trials in the ExTraMATCH II master dataset

Study characteristics	
Publication year	
1990 to 1999	2 (10.5)
2000 to 2009	12 (63.2)
2010 to 2012	4 (21.0)
Unpublished	1 (5.3)
Main study location	
Europe	14 (73.7)
North America[(a)]	5 (26.3)
Single study centre	
Single	13 (68.4)
Multiple	5 (26.3)
Not reported	1 (5.3)
Sample size	
0 to 99	11 (57.9)
100 to 999	7 (36.8)
1000 and over	1 (5.3)
Duration of follow-up in dataset (months), median (range)	
Mortality	29 (24)
Intervention characteristics	

Intervention type	
Exercise only programs	13 (68.4)
Comprehensive programs	5 (26.3)
Not reported	1 (5.3)
Type of exercise	
Aerobic exercise only	12 (63.2)
Aerobic plus resistance training	6 (31.6)
Not reported	1 (5.3)
Dose of intervention	
Duration of intervention (weeks), median (range)	30 (15 to 90)
Frequency (sessions per week), median (range)	2.5 (2 to 6.5)
Length of exercise session (mins), median (range)	24 (4 to 120)
Exercise intensity, range	50-85% VO2peak 11-15 BORG rating
Setting	
Centre-based	14 (73.7)
Home-based	4 (21.1)
Not reported	1 (5.3)

Assessment of study quality and risk of bias in included trials

The overall quality of included trials was judged to be moderate to good, with a median TESTEX score of 11 (range 9 to 14) out of a maximum score of 15 (see Table 3**Error! Reference source not found.**).

Table 3: Assessment of quality using TESTEX scale for trials in ExTraMATCH II

Study (publication year)	Eligibility Criteria Specified	Randomisation Specified	Allocation Concealed	Groups Similar at baseline	Blinding of Assessors	Outcome measures in >85% participant (a)	Intention to treat analysis (b)	Between-group statistical comparisons reported (c)	Point measures & measures of variability reported	Activity monitoring in control group	Relative Exercise intensity reviewed	Exercise Volume and Energy Expended	Overall TESTEX score (maximum score 15)
Belardinelli (1999)	1	0	0	1	0	3	1	1	1	0	0	1	9
Belardinelli (2012)	1	0	0	1	0	3	1	1	1	0	0	1	9
DAN-REHAB (2008)	1	1	0	1	1	3	1	2	1	0	0	0	11
Dracup (2007)	1	0	0	1	0	3	1	2	1	1	1	1	10
Gary (2010)	1	1	0	1	1	3	1	2	1	0	0	0	11
Giannuzzi (2003)	1	0	0	1	0	2	1	2	1	0	1	1	10
Hambrecht (2000)	1	1	0	1	0	3	0	2	1	0	1	1	11
HF-ACTION (2008)	1	1	1	1	1	3	1	2	1	1	0	1	14
Jolly (2009)	1	1	1	1	0	2	1	2	1	0	1	1	12
McKelvie (2002)	1	1	1	1	1	2	1	1	1	0	1	1	12
Mueller (2007)	1	0	0	1	0	2	1	2	1	0	1	1	10

Nilsson (2008)	1	1	0	1	1	2	1	2	1	0	0	1	11
Passino (2006)	1	0	0	1	0	2	1	2	1	0	1	1	10
Wielenga (1999)	1	0	0	1	0	2	1	2	1	0	0	1	9
Willenheimer (2001)	1	0	0	1	1	2	1	2	1	0	0	1	9
Witham (2005)	1	1	0	1	1	3	1	2	1	0	1	0	12
Witham (2012)	1	1	0	1	1	3	1	2	1	0	1	0	12
Yeh (2011)	1	1	0	1	1	3	1	2	1	1	0	0	12
Zanelli (unpublished)	No score***												

(a) Three points possible;

(b) If ITT was not specifically mentioned, but it was noted that no participants withdrew and all analysed 1 point was awarded;

(c) Two points possible;

Zanelli – not scored as no full publication

Chapter 5: Impact of ExCR on mortality and hospitalisation

One trial which provided IPD was not included in the mortality and hospitalisation analyses as no data were provided to allow calculation of survival time or time to hospitalisation. [59] This resulted in the inclusion of 18 trials, [19, 50, 51, 53-58, 60-67, 73] comprising 3,912 patients (1,948 ExCR, 1,964 control) with a median follow up of 19 months for mortality outcomes and 11 months for hospitalisation outcomes. Figure 2 summarises the study selection process for the mortality and hospitalisation analyses.

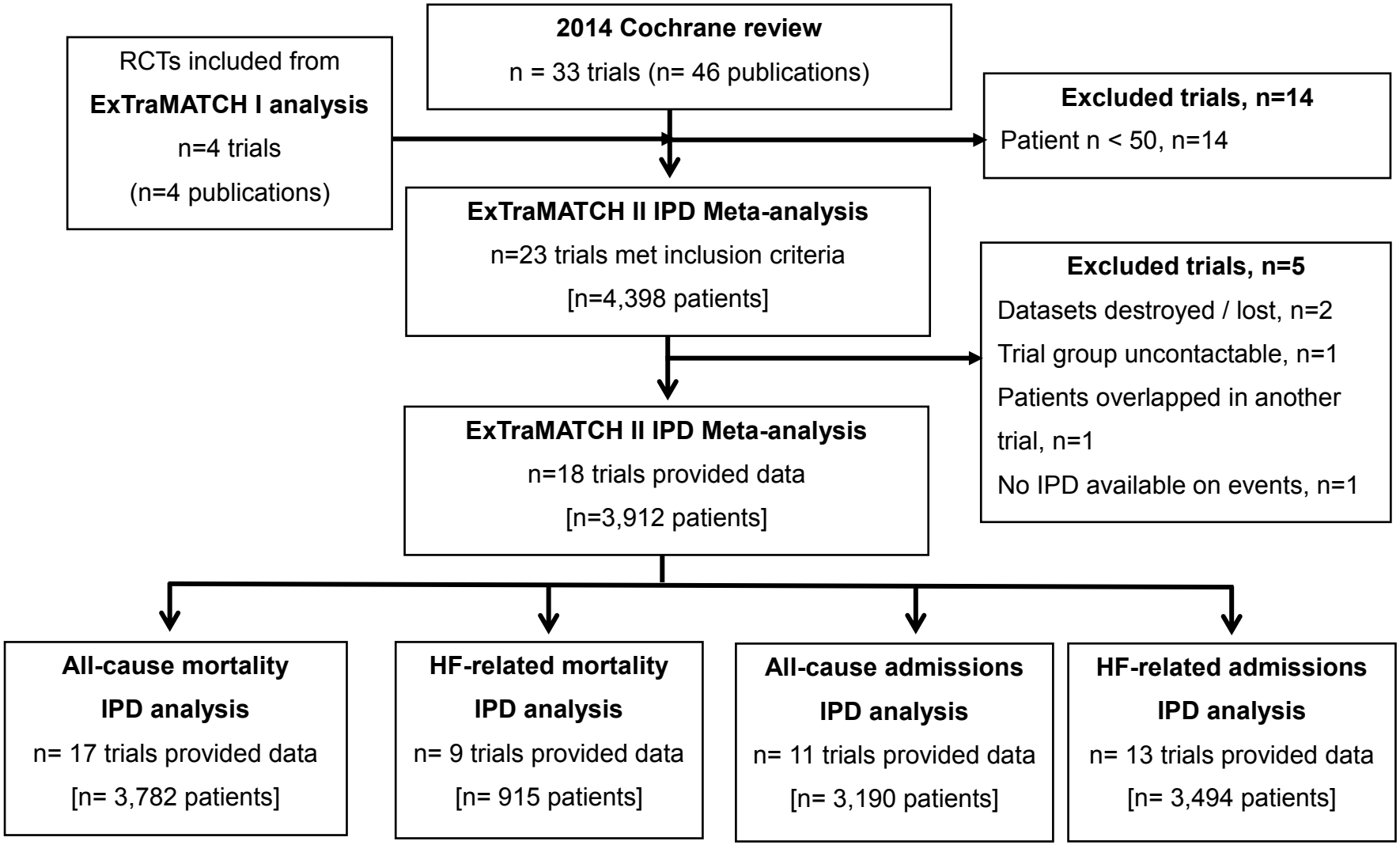


Figure 2: PRISMA flow diagram summarising selection of studies for mortality and hospitalisation analyses

Characteristics of included patients and trials

Patient baseline characteristics were well balanced between ExCR and control patients (see Table 4). The majority of patients were male (75%), with a mean age of 61 years (standard deviation (SD) 13. The mean baseline left-ventricular ejection fraction was 27% (SD 8.1%), no included studies recruited patients with preserved ejection fraction heart failure (ejection fraction >45%), and most patients were in NYHA functional class II (59%) or III (37%). Studies were published between 1999 and 2012 across a number of countries (see Table 2). Sample size ranged from 50 to 2,130 patients. All trials evaluated an aerobic exercise intervention; six also included resistance training. [52, 57, 58, 61, 64, 65] Exercise training was most commonly delivered in either an exclusively centre-based setting or a centre-based setting in combination with some home exercise sessions (see Table 5). Three trials were conducted in an exclusively home-based setting. [52, 54, 58] The dose of exercise training ranged widely across studies with an average session duration of 15 to 120 minutes (including warm-up and cool-down), 2 to 7 sessions/week, exercise intensity equivalent of 50 to 85% VO_2peak, and delivered over a duration of 12 to 90 weeks.

Table 4: Baseline characteristics of patients in the mortality and hospitalisation analyses

Characteristic	**ExCR (n=1,948)**	**Control (n=1,964)**	**All (n=3,912)**
Age (years), mean (SD)	61.3 (12.7)	61.4 (13.2)	61.3 (13.0)
Gender			
Male	1442 (74)	1489 (76)	2,931 (75)
Female	506 (26)	475 (24)	981 (25)
Baseline ejection fraction (%); mean (SD)	26.8 (8.2)	26.7 (8.1)	26.7 (8.1)
NYHA status			
Class I	25 (1)	28 (1)	53 (1)
Class II	1107 (59)	1130 (60)	2237 (59)
Class III	700 (37)	708 (37)	1408 (37)
Class IV	47 (3)	26 (1)	73 (2)
Aetiology			
Ischaemic	1094 (57)	1080 (56)	2174 (57)
Non-ischemic	809 (43)	838 (44)	1647 (43)
Ethnicity			
White	1100 (70)	1140 (72)	2240 (71)
Non-white	472 (30)	445 (28)	917 (29)
VO_2 peak (ml/kg/min); mean (SD)	14.9 (4.4)	15.0 (4.6)	14.9 (4.5)

Data are n (%) unless otherwise indicated; percentages may not sum to 100 because of rounding; NHYA: New York Heart Association; SD: standard deviation; VO_2 peak: peak oxygen uptake.

Table 5: Characteristics of included trials in the mortality and hospitalisation analyses

Study characteristics	
Publication year	
1990 to 1999	2 (11)
2000 to 2009	12 (67)
2010 to 2012	3 (17)
Unpublished	1 (6)
Main study location	
Europe	14 (78)
North America[(a)]	4 (22)
Single study centre	
Single	12 (67)
Multiple	5 (28)
Not reported	1 (6)
Sample size	
0 to 99	10 (56)
100 to 999	7 (39)
1000 and over	1 (6)
Duration of follow-up in dataset (months), median (range)	
Mortality	18.6 (11.8 to 419)
Hospitalisation	11.2 (2.6 to 98)
Intervention characteristics	
Intervention type	
Exercise only programs	5 (28)
Comprehensive programs	12 (67)
Not reported	1 (6)
Type of exercise	
Aerobic exercise only	12 (67)
Aerobic plus resistance training	6 (33)
Dose of intervention	
Duration of intervention (weeks), median (range)	30 (12 to 90)

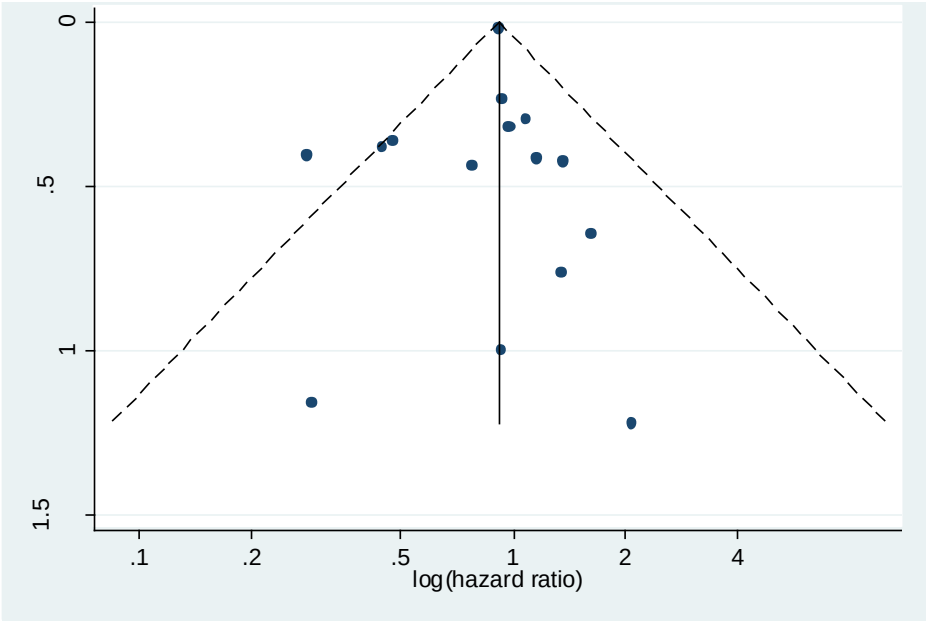
Frequency (sessions per week), median (range)	2.8 (2 to 7)
Length of exercise session (mins), median (range)	24 (15 to 120)
Exercise intensity, range	40-80% maximum heart rate 50-85% peak VO_2 12-18 Borg rating
Setting	
Centre-based only	6 (33)
Home-based only	3 (17)
Centre- and home-based	8 (44)
Not reported	1 (6)

Data are n (%) unless otherwise indicated; percentages may not sum to 100 because of rounding;

[a] HF-ACTION **(**O'Connor) study was categorised as North America but was also delivered in to a small number of patients in France

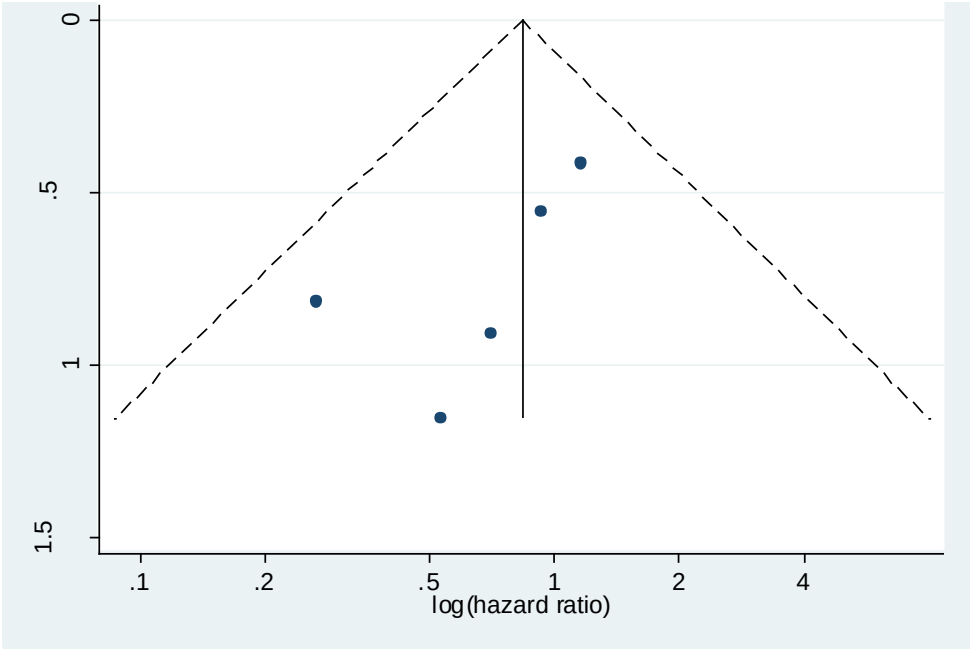
Assessment of study quality and risk of bias

There was no evidence of significant small study bias for the four outcomes (see Figure 3**Error! Reference source not found.**). The overall quality of included trials was judged to be moderate to good, with a median TESTEX [31] score of 11 (range 9 to 14) out of a maximum score of 15 (Table 6**Error! Reference source not found.**). The criteria of allocation concealment and physical activity monitoring in the control groups were met in only three studies [19, 58, 66]; the other TESTEX criteria were met in 50% or more of trials.



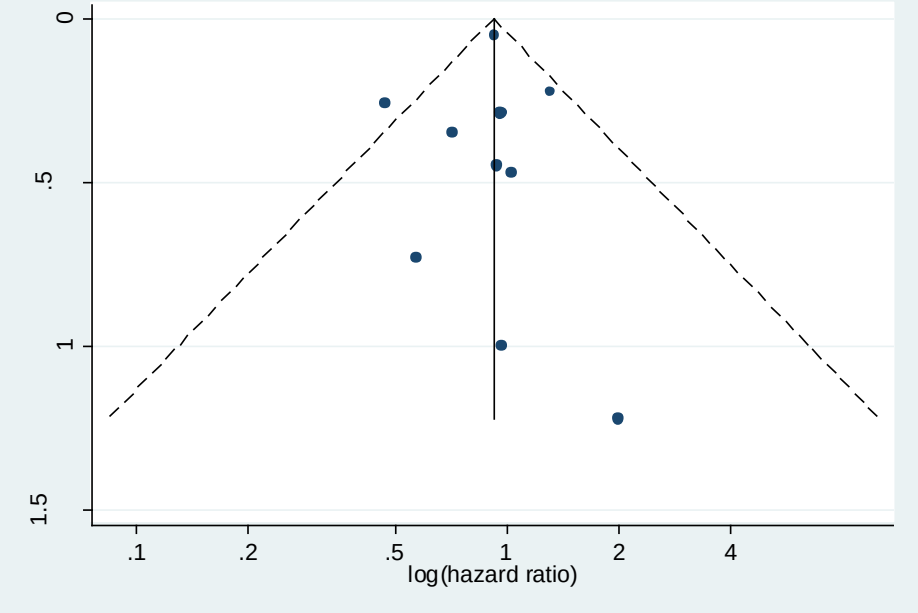
a. All-cause mortality

Egger test -0.26, p=0.458



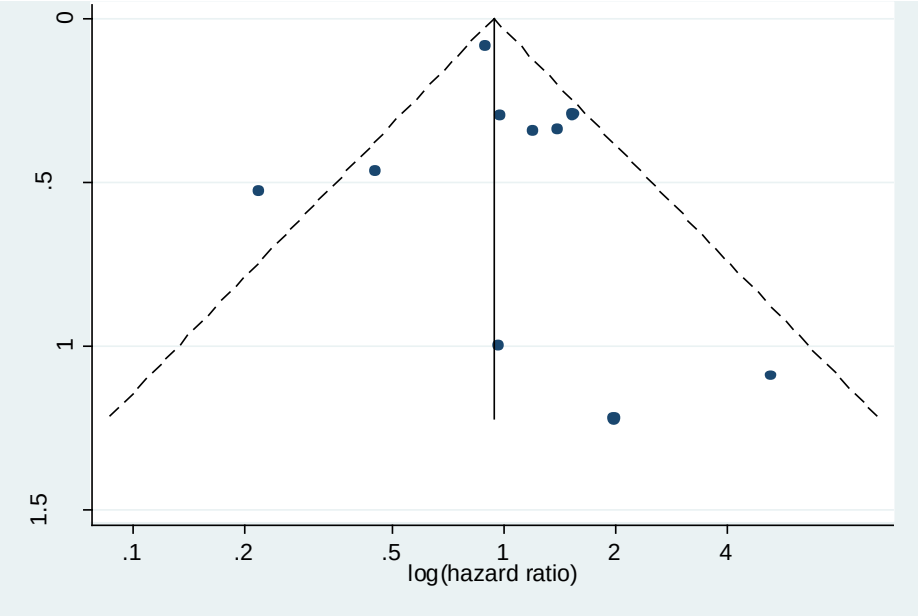
b. HF-specific mortality

Egger test -1.60, p=0.147



c. All-cause hospitalisation

Egger test 0.16, p=0.739



d. HF-specific hospitalisation

Egger test 0.32, p=0.610

Figure 3: Funnel plots for mortality and hospitalisation analyses

Table 6: Assessment of quality using TESTEX scale of included studies in mortality and hospitalisation analysis

Study (publication year)	Eligibility Criteria Specified	Randomisation Specified	Allocation Concealed	Groups Similar at baseline	Blinding of Assessors	Outcome measures in >85% participant (a)	Intention to treat analysis (b)	Between-group statistical comparisons reported (c)	Point measures & measures of variability	Activity monitoring in control group	Relative Exercise intensity reviewed	Exercise Volume and Energy Expended	Overall TESTEX score (maximum score 15)
Belardinelli (1999)	1	0	0	1	0	3	1	1	1	0	0	1	9
Belardinelli (2012)	1	0	0	1	0	3	1	1	1	0	0	1	9
DANREHAB (2008)	1	1	1	1	1	3	1	2	1	0	0	0	12
Dracup (2007)	1	0	0	1	0	3	1	2	1	1	1	1	10
Giannuzzi (2003)	1	0	0	1	0	2	1	2	1	0	1	1	10
Hambrecht (2000)	1	1	0	1	0	3	0	2	1	0	1	1	11
HF-ACTION (2008)	1	1	1	1	1	3	1	2	1	1	0	1	14
Jolly (2009)	1	1	1	1	0	2	1	2	1	0	1	1	12
McKelvie (2002)	1	1	1	1	1	2	1	1	1	0	1	1	12
Mueller (2007)	1	0	0	1	0	2	1	2	1	0	1	1	10
Nilsson (2008)	1	1	0	1	1	2	1	2	1	0	0	1	11
Passino (2006)	1	0	0	1	0	2	1	2	1	0	1	1	10

Wielenga (1999)	1	0	0	1	0	2	1	2	1	0	0	1	9
Willenheimer (2001)	1	0	0	1	1	2	1	2	1	0	0	1	9
Witham (2005)	1	1	0	1	1	3	1	2	1	0	1	0	12
Witham (2012)	1	1	0	1	1	3	1	2	1	0	1	0	12
Yeh (2011)	1	1	0	1	1	3	1	2	1	1	0	0	12
Zanelli (unpublished)	Not scored												

(a) Three points possible;

(b) If ITT was not specifically mentioned, but it was noted that no participants withdrew and all analysed 1 point was awarded;

(c) Two points possible;

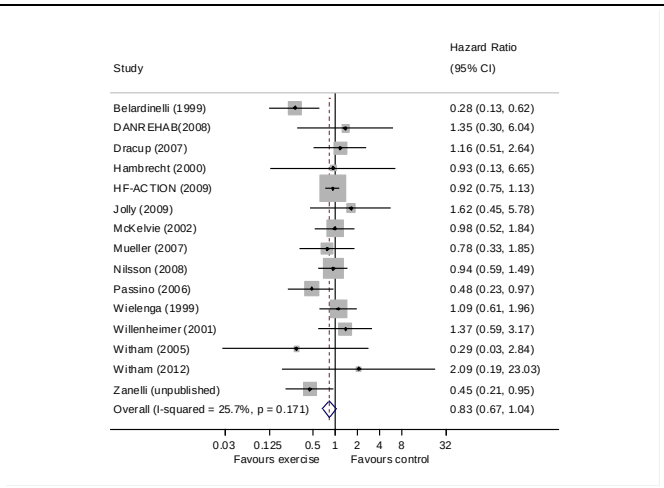
Zanelli – not scored as no full publication

Findings

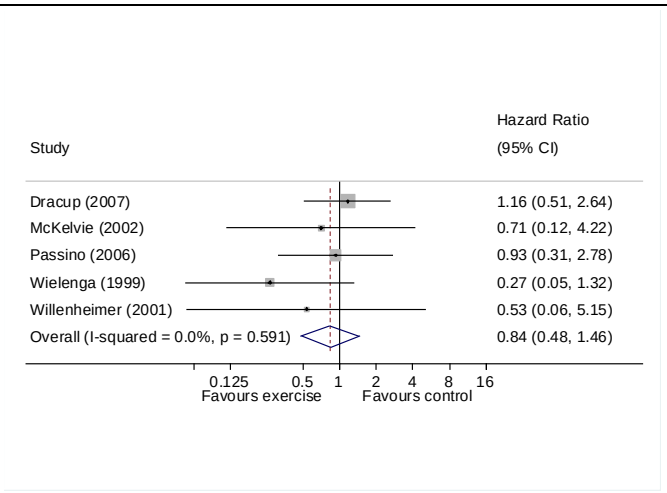
Primary analysis

Compared to control, all time to event mean treatment effects from random effects 2-stage IPD meta-analysis were in favour of ExCR but with wide confidence intervals and not statistically significant , i.e. all-cause mortality: HR: 0.83 (95% CI: 0.67 to 1.04, p=0.107, 17 studies, 3,782 patients, τ^2=0.04, I^2 =26%); HF-specific mortality (HR: 0.84, 95% CI: 0.48 to 1.46, p=0.527, 9 studies, 915 patients, τ^2=0.00, I^2 =0%); all-cause hospitalisation (HR: 0.90, 95% CI: 0.76 to 1.06, p=0.210, 11 studies, 3,190 patients, τ^2=0.01, I^2 =12.4%,), and HF-specific hospitalisation (HR: 0.98, 95% CI: 0.72 to 1.35, p=0.902, 13 studies, 3,494 patients, τ^2=0.10, I^2 =45%) (see Figure 4 **Error! Reference source not found.**and Tables 7-10)

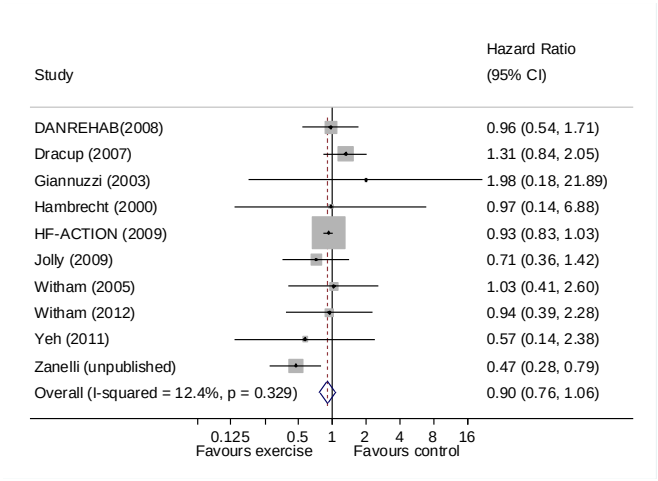
Interaction analyses for the two-stage model revealed no consistent interaction between the effect of ExCR and any of the predefined subgroups (age, gender, ejection fraction, NYHA class, HF aetiology, ethnicity or baseline exercise capacity) for all-cause mortality, HF-related mortality, all-cause hospitalisation, or HF-related hospitalisation (see Tables 7-10). In order to make further comparisons of mortality and hospitalisation rates within each subgroup, the HR and associated 95% CI from individual subgroup one-stage IPD meta-analyses are shown in Figure 5. The p-value from the interaction test in the two-stage IPD meta-analyses are presented alongside these estimates.



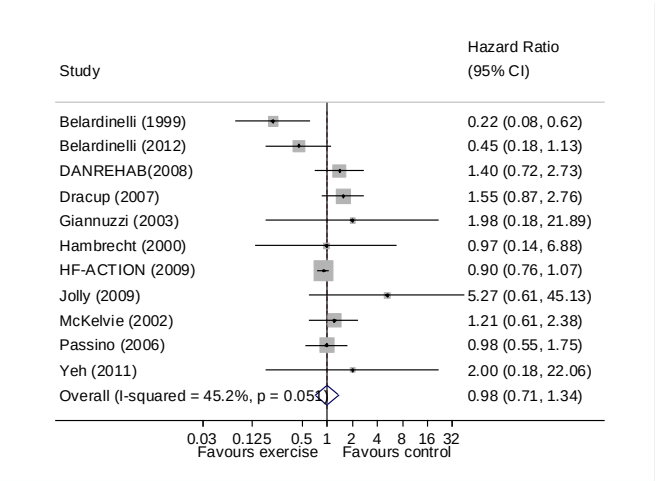
a. All-cause mortality



b. HF-specific mortality



c. All-cause hospitalisation



d. HF-specific hospitalisation

Figure 4: Effect of ExCR on mortality and hospitalisation across patient subgroups: Two –stage IPD meta-analysis

Temporary reference list:

Belardinelli (1999): Belardinelli R, Georgiou D, Cianci G, Purcaro A. Randomized, controlled trial of long-term moderate exercise training in chronic heart failure: effects on functional capacity, quality of life, and clinical outcome. Circulation. 1999;99(9):1173-82.

Belardinelli (2012): Belardinelli R, Georgiou D, Cianci G, Purcaro A. 10-year exercise training in chronic heart failure: a randomized controlled trial. J Am Coll Cardiol. 2012;60(16):1521-8.

DANREHAB (2008): Zwisler AD, Soja AM, Rasmussen S, Frederiksen M, Abedini S, Appel J, et al. Hospital-based comprehensive cardiac rehabilitation versus usual care among patients with congestive heart failure, ischemic heart disease, or high risk of ischemic heart disease: 12-month results of a randomized clinical trial. Am Heart J. 2008;155(6):1106-13.

Dracup (2007): Dracup K, Evangelista LS, Hamilton MA, Erickson V, Hage A, Moriguchi J, et al. Effects of a home-based exercise program on clinical outcomes in heart failure. Am Heart J. 2007;154(5):877-83.

Giannuzzi (2003): Giannuzzi P, Temporelli PL, Corra U, Tavazzi L. Antiremodeling effect of long-term exercise training in patients with stable chronic heart failure: results of the Exercise in Left Ventricular Dysfunction and Chronic Heart Failure (ELVD-CHF) Trial. Circulation. 2003;108(5):554-9.

Hambrecht (2000): Hambrecht R, Gielen S, Linke A, Fiehn E, Yu J, Walther C, et al. Effects of exercise training on left ventricular function and peripheral resistance in patients with chronic heart failure: A randomized trial. JAMA. 2000;283(23):3095-101.

HF-ACTION (2009): O'Connor CM, Whellan DJ, Lee KL, Keteyian SJ, Cooper LS, Ellis SJ, et al. Efficacy and Safety of Exercise Training in Patients With Chronic Heart Failure: HF-ACTION Randomized Controlled Trial. JAMA. 2009;301(14):1439-50.

Jolly (2009): Jolly K, Taylor RS, Lip GY, Davies M, Davis R, Mant J, et al. A randomized trial of the addition of home-based exercise to specialist heart failure nurse care: the Birmingham Rehabilitation Uptake Maximisation study for patients with Congestive Heart Failure (BRUM-CHF) study. Eur J Heart Fail. 2009;11(2):205-13.

McKelvie (2002): McKelvie RS, Teo KK, Roberts R, McCartney N, Humen D, Montague T, et al. Effects of exercise training in patients with heart failure: the Exercise Rehabilitation Trial (EXERT). Am Heart J. 2002;144(1):23-30.

Mueller (2007): Mueller L, Myers J, Kottman W, Oswald U, Boesch C, Arbrol N, et al. Exercise capacity, physical activity patterns and outcomes six years after cardiac rehabilitation in patients with heart failure. Clin Rehabil. 2007;21(10):923-31.

Nilsson (2008): Nilsson BB, Westheim A, Risberg MA. Long-term effects of a group-based high-intensity aerobic interval-training program in patients with chronic heart failure. Am J Cardiol. 2008;102(9):1220-4.

Passino (2006): Passino C, Severino S, Poletti R, Piepoli MF, Mammini C, Clerico A, et al. Aerobic Training Decreases B-Type Natriuretic Peptide Expression and Adrenergic Activation in Patients With Heart Failure. J Am Coll Cardiol. 2006;47(9):1835-9.

Wielenga (1999): Wielenga RP, Huisveld IA, Bol E, Dunselman PH, Erdman RA, Baselier MR, et al. Safety and effects of physical training in chronic heart failure. Results of the Chronic Heart Failure and Graded Exercise study (CHANGE). Eur Heart J. 1999;20(12):872-9.

Willenheimer (2001): Willenheimer R, Rydberg E, Cline C, Broms K, Hillberger B, Oberg L, et al. Effects on quality of life, symptoms and daily activity 6 months after termination of an exercise training programme in heart failure patients. Int J Cardiol. 2001;77(1):25-31.

Witham (2005): Witham MD, Gray JM, Argo IS, Johnston DW, Struthers AD, McMurdo ME. Effect of a seated exercise program to improve physical function and health status in frail patients > or = 70 years of age with heart failure. Am J Cardiol. 2005;95(9):1120-4.

Witham (2012): Witham MD, Fulton RL, Greig CA, Johnston DW, Lang CC, van der Pol M, et al. Efficacy and cost of an exercise program for functionally impaired older patients with heart failure: a randomized controlled trial. Circ Heart Fail. 2012;5(2):209-16.

Zanelli (unpublished): Zanelli E, Volterrani M, Scalvini S, Musmeci G, Campana M, Zappa C. Multidisciplinary non-pharmachological intervention prevents hospitalization, improves morbidity rates and functional status in patients with congestive heart failure. Eur Heart J. 1997;18(abstract suppl):647.

Yeh (2011): Yeh GY, McCarthy EP, Wayne PM, Stevenson LW, Wood MJ, Forman D, et al. Tai Chi Exercise in Patients With Chronic Heart Failure: A Randomized Clinical Trial. Arch Intern Med. 2011;171(8):750-7.

Table 7: All-cause mortality - overall treatment effect and subgroup (interaction) effects

	Primary analysis	**Secondary analysis**	**Sensitivity analyses**				
	Two-stage model, random effects HR (95% CI) p-value	**One-stage Cox model, stratified by study with fixed treatment effect** HR (95% CI) p-value	**Two-stage model, random effects excluding HF-Action** HR (95% CI) p-value	**One-stage Cox model, stratified by study with fixed effect excluding HF-Action** HR (95% CI) p-value	**Two-stage model, random effects 1 year truncation** HR (95% CI) p-value	**Two-stage model, random effects 2 year truncation** HR (95% CI) p-value	**Two-stage model, random effects 5 year truncation** HR (95% CI) p-value
Overall effect	0.83 (0.67, 1.04) p=0.107	0.85 (0.73, 0.99) p=0.034	0.81 (0.61, 1.06) p=0.129	0.79 (0.64, 0.97) p=0.027	0.87 (0.58, 1.31) p=0.507	0.86 (0.67, 1.10) p=0.217	0.84 (0.66, 1.06) p=0.140
Interaction terms							
Age (years)	0.99 (0.98, 1.00) p=0.165	0.99 (0.98, 1.01) p=0.254	0.98 (0.96, 1.01) p=0.144	0.99 (0.96, 1.01) p=0.228	0.98 (0.95, 1.00) p=0.077	0.98 (0.96, 1.00) p=0.034	0.99 (0.97, 1.00) p=0.097
Gender (male vs female)	1.10 (0.73, 1.66) p=0.660	1.06 (0.70, 1.60) p=0.783	0.71 (0.35, 1.43) p=0.341	0.70 (0.36, 1.36) p=0.300	0.76 (0.34, 1.68) p=0.490	0.96 (0.55, 1.67) p=0.872	1.17 (0.75, 1.82) p=0.481

Ejection fraction (%)	0.99 (0.97, 1.01) p=0.250	0.99 (0.97, 1.01) p=0.332	0.98 (0.95, 1.01) p=0.124	0.98 (0.96, 1.01) p=0.201	1.04 (1.00, 1.08) p=0.055	0.99 (0.97, 1.02) p=0.688	0.99 (0.97, 1.01) p=0.506
NYHA class (NYHA I/II vs NYHA III/IV)	0.80 (0.58, 1.11) p=0.182	0.79 (0.57, 1.08) p=0.134	0.82 (0.49, 1.38) p=0.459	0.75 (0.46, 1.22) p=0.244	0.50 (0.23, 1.07) p=0.073	0.84 (0.54, 1.30) p=0.431	0.83 (0.59. 1.18) p=0.297
HF aetiology (ischaemic vs non-ischaemic)	0.73 (0.38, 1.39) p=0.335	1.19 (0.86, 1.64) p=0.297	0.69 (0.36, 1.31) p=0.255	0.87 (0.54, 1.41) p=0.575	0.69 (0.19, 2.54) p=0.574	0.79 (0.38, 1.67) p=0.542	0.70 (0.33, 1.47) p=0.345
Ethnic group (white vs non-white)	1.12 (0.74, 1.69) p=0.593	1.11 (0.74, 1.68) p=0.604	(a)	1.05 (0.25, 4.31) p=0.949	0.72 (0.34, 1.53) p=0.396	0.83 (0.50, 1.38) p=0.468	1.12 (0.74, 1.69) p=0.593
Exercise capacity							
Baseline peak VO_2 directly measured	1.00 (0.95, 1.05) p=0.937	0.99 (0.95, 1.04) p=0.783	0.98 (0.90, 1.08) p=0.712	0.99 (0.91, 1.07) p=0.777	0.97 (0.88, 1.06) p=0.456	0.99 (0.93, 1.05) p=0.780	0.98 (0.91, 1.06) p=0.630
Baseline peak VO_2, directly measured and predicted	1.00 (0.95, 1.06) p=0.903	1.00 (0.96, 1.04) p=0.954	1.00 (0.91, 1.08) p=0.923	1.00 (0.93, 1.07) p=0.984	0.99 (0.90, 1.08) p=0.734	1.00 (0.94, 1.06) p=0.961	1.00 (0.93, 1.07) p=0.924
Standardised scores using baseline peak VO_2, 6MWT,	1.03 (0.83, 1.27) p=0.802	1.02 (0.85, 1.22) p=0.851	0.99 (0.71, 1.39) p=0.955	1.01 (0.75, 1.35) p=0.967	0.97 (0.66, 1.41) p=0.858	1.00 (0.78, 1.30) p=0.972	1.01 (0.76, 1.35) p=0.938

ISWT units, and watts							

(a) Study estimate not available as too few studies provide data; peak VO_2: peak oxygen uptake; ISWT: incremental shuttle walk test; 6MWT: 6-minute walk test.

Table 8: HF-specific mortality - overall treatment effect and subgroup (interaction) effects^

	Primary analysis	**Secondary analysis**	**Sensitivity analyses**		
	Two-stage model, random effects HR (95% CI) p-value	**One-stage Cox model, stratified by study with fixed treatment effect** HR (95% CI) p-value	**Two-stage model, random effects 1 year truncation** HR (95% CI) p-value	**Two-stage model, random effects 2 year truncation** HR (95% CI) p-value	**Two-stage model, random effects 5 year truncation** HR (95% CI) p-value
Overall effect	0.84 (0.48, 1.46) p=0.527	0.75 (0.44, 1.28) p=0.294	(a)	1.30 (0.59, 2.87) p=0.515	0.84 (0.49, 1.53) p=0.575
Interaction terms					
Age (years)	0.96 (0.91, 1.02) p=0.206	0.96 (0.92, 1.01) p=0.162	(a)	0.91 (0.84, 0.98) p=0.017	0.95 (0.90, 1.00) p=0.066
Gender (male vs female)	0.53 (0.08, 3.73) p=0.524	0.61 (0.11, 3.49) p=0.583	(a)	(b)	(b)
Ejection fraction (%)	0.95 (0.89, 1.02) p=0.159	0.96 (0.90, 1.02) p=0.179	(a)	1.01 (0.82, 1.24) p=0.912	0.96 (0.89, 1.04) p=0.309
NYHA class (NYHA I/II vs NYHA III/IV)	0.54 (0.07, 4.28) p=0.562	0.78 (0.23, 26.65) p=0.691	(a)	(b)	0.54 (0.07, 4.28) p=0.562

HF aetiology (ischaemic vs non-ischaemic)	Data only available for one study	3.30 (1.02, 10.7) p=0.047	(a)	(b)	(b)
Ethnic group (white vs non-white)	(b)	(b)	(a)	(b)	(b)
Exercise capacity:					
Baseline peak VO_2, directly measured	0.90 (0.76, 1.07) p=0.232	0.93 (0.78, 1.09) p=0.362	(a)	0.98 (0.73, 1.31) p=0.893	0.86 (0.69, 1.06) p=0.146
Baseline peak VO_2, directly measured and predicted	0.91 (0.77, 1.07) p=0.263	0.94 (0.80, 1.10) p=0.423	(a)	0.98 (0.76, 1.26) p=0.854	0.88 (0.72, 1.06) p=0.184
Standardised scores using baseline peak VO_2, 6MWT, ISWT units and watts score	0.69 (0.35, 1.35) p=0.276	0.82 (0.43, 1.56) p=0.545	(a)	0.86 (0.31, 2.37) p=0.773	0.61 (0.28, 1.32) p=0.210

(a) HF-Action did not provide HF-mortality so sensitivity analysis of omission not undertaken;

(b) Study estimate not available as too few studies provide data

Peak VO_2: peak oxygen uptake; ISWT: incremental shuttle walk test; 6MWT: 6-minute walk test

Table 9: All-cause hospitalisation - overall treatment effect and subgroup (interaction) effects

	Primary analysis	**Secondary analysis**	**Sensitivity analyses**				
	Two-stage model, random effects HR (95% CI) p-value	**One-stage Cox model, stratified by study with fixed treatment effect** HR (95% CI) p-value	**Two-stage model, random effects excluding HF-Action** HR (95% CI) p-value	**One-stage Cox model, stratified by study with fixed treatment effect excluding HF-Action** HR (95% CI) p-value	**Two-stage model, random effects 1 year truncation** HR (95% CI) p-value	**Two-stage model, random effects 2 year truncation** HR (95% CI) p-value	**Two-stage model, random effects 5 year truncation** HR (95% CI) p-value
Overall effect	0.90 (0.76, 1.06) p=0.210	0.91 (0.83, 1.01) p=0.072	0.86 (0.64, 1.14) p=0.293	0.85 (0.68, 1.09) p=0.210	0.94 (0.75, 1.18) p=0.583	0.91 (0.74, 1.11) p=0.330	0.90 (0.76, 1.06) p=0.210
Interaction terms							
Age (years)	1.00 (0.99, 1.01) p=0.794	1.00 (0.99, 1.01) p=0.854	1.00 (0.98, 1.03) p=0.808	1.00 (0.98, 1.02) p=-0.969	1.00 (0.99, 1.01) p=0.636	1.00 (0.99, 1.01) p=0.798	1.00 (0.99, 1.01) p=0.794
Gender (male vs female)	1.09 (0.87, 1.36) p=0.454	1.09 (0.88, 1.36) p=0.424	0.66 (0.38, 1.14) p=0.136	0.68 (0.39, 1.16) p=0.158	1.05 (0.80, 1.37) p=0.745	1.15 (0.91, 1.46) p=0.239	1.09 (0.87, 1.35) p=0.454

Ejection fraction (%)	1.00 (0.98, 1.01) p=0.629	1.00 (0.98, 1.01) p=0.646	1.00 (0.96, 1.04) p=0.857	1.00 (0.96, 1.05) p=0.831	1.00 (0.98, 1.01) p=0.632	0.99 (0.98, 1.01) p=0.343	1.00 (0.98, 1.01) p=0.629
NYHA class (NYHA I/II vs NYHA III/IV)	0.91 (0.74, 1.12) p=0.370	0.90 (0.73, 1.10) p=0.308	0.89 (0.43, 1.87) p=0.763	0.79 (0.39, 1.60) p=0.508	0.81 (0.63, 1.05) p=0.110	0.87 (0.70, 1.09) p=0.235	0.91 (0.74, 1.12) p=0.355
HF aetiology (ischaemic vs non-ischaemic)	0.96 (0.71, 1.31) p=0.810	1.00 (0.82, 1.22) p=0.988	0.73 (0.39, 1.39) p=0.340	0.73 (0.40, 1.31) p=0.284	1.08 (0.84, 1.38) p=0.562	1.04 (0.84, 1.29) p=0.723	1.01 (0.83, 1.24) p=0.910
Ethnic group (white vs non-white)	1.02 (0.83, 1.26) p=0.860	1.02 (0.83, 1.26) p=0.852	1.02 (0.47, 2.21) p=0.959	1.06 (0.49, 2.32) p=0.879	1.14 (0.88, 1.48) p=0.322	1.06 (0.85, 1.33) p=0.607	1.02 (0.83, 1.26) p=0.860
Exercise capacity:							
Baseline peak VO_2, directly measured	1.01 (0.99, 1.04) p=0.259	1.02 (0.99, 1.04) p=0.234	1.05 (0.95, 1.16) p=0.352	1.06 (0.96, 1.17) p=0.262	1.03 (0.99, 1.06) p=0.124	1.02 (0.99, 1.04) p=0.243	1.01 (0.99, 1.04) p=0.259
Baseline peak VO_2, directly measured and predicted	1.02 (0.99, 1.04) p=0.153	1.02 (0.99, 1.04) p=0.134	1.07 (0.98, 1.17) p=0.125	1.08 (0.99, 1.17) p=0.078	1.03 (1.00, 1.06) p=0.057	1.02 (0.99, 1.05) p=0.129	1.02 (0.99, 1.04) p=0.153
Standardised scores using baseline peak VO_2, 6MWT,	1.09 (0.98, 1.22) p=0.095	1.10 (0.99, 1.22) p=0.088	1.30 (0.93, 1.83) p=0.120	1.32 (0.95, 1.82) p=0.097	1.16 (1.02, 1.33) p=0.027	1.11 (0.99, 1.24) p=0.077	1.09 (0.98, 1.22) p=0.095

ISWT units and watts							

Peak VO_2: peak oxygen uptake; ISWT: incremental shuttle walk test; 6MWT: 6-minute walk test.

Table 10: HF-specific hospitalisation - overall treatment effect and subgroup (interactions) effects in studies included in IPD meta-analysis

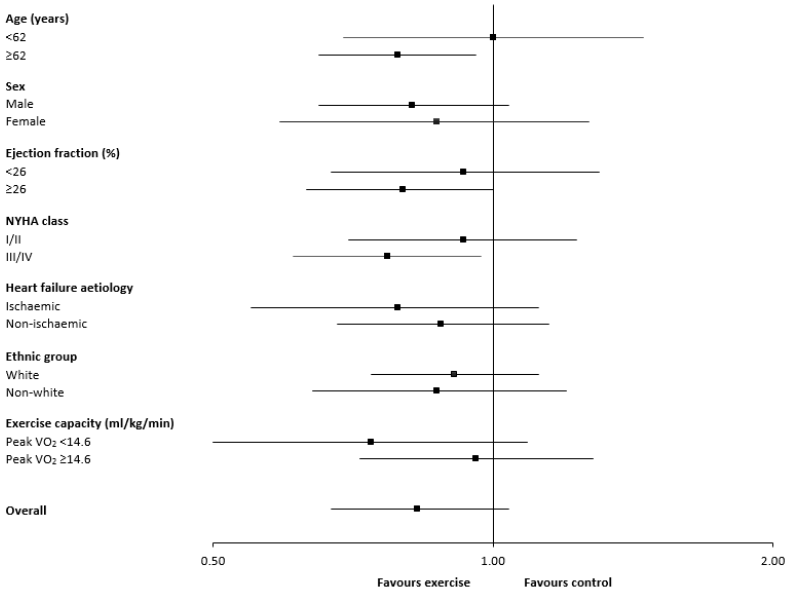
	Primary analysis	**Secondary analysis**	**Sensitivity analyses**				
	Two-stage model, random effects HR (95% CI) p-value	**One-stage Cox model, stratified by study with fixed treatment effect** HR (95% CI) p-value	**Two-stage model, random effects excluding HF-Action** HR (95% CI) p-value	**One-stage Cox model, stratified by study with fixed treatment effect excluding HF-Action** HR (95% CI) p-value	**Two-stage model, random effects 1 year truncation** HR (95% CI) p-value	**Two-stage model, random effects 2 year truncation** HR (95% CI) p-value	**Two-stage model, random effects 5 year truncation** HR (95% CI) p-value
Overall effect	0.98 (0.72, 1.35) p=0.902	0.94 (0.81, 1.08) p=0.368	1.00 (0.65, 1.54) p=0.999	1.03 (0.79, 1.35) p=0.829	1.08 (0.88, 1.33) p=0.470	1.06 (0.83, 1.34) p=0.658	0.97 (0.70, 1.34) p=0.855
Interaction terms							
Age (years)	1.00 (0.99, 1.02) p=0.603	1.00 (0.99, 1.02) p=0.632	1.00 (0.98, 1.03) p=0.958	1.00 (0.97, 1.02) p=0.906	1.00 (0.99, 1.02) p=0.640	1.00 (0.99, 1.02) p=0.611	1.00 (0.99, 1.02) p=0.580

Gender (male vs female)	1.03 (0.73, 1.46) p=0.865	0.99 (0.71, 1.39) p=0.949	0.70 (0.32, 1.53) p=0.372	0.65 (0.33, 1.29) p=0.215	0.76 (0.46, 1.24) p=0.274	1.06 (0.68, 1.66) p=0.803	0.93 (0.49, 1.75) p=0.815
Ejection fraction (%)	0.51 (0.14, 1.79) p=0.291	0.99 (0.97, 1.01) p=0.325	0.99 (0.96, 1.03) p=0.540	0.99 (0.97, 1.01) p=0.350	0.99 (0.97, 1.02) p=0.569	0.99 (0.97, 1.01) p=0.350	0.99 (0.97, 1.02) p=0.569
NYHA class (NYHA I/II vs NYHA III/IV)	1.55 (0.79, 3.02) p=0.200	1.14 (0.84, 1.54) p=0.399	2.05 (0.86, 4.92) p=0.107	1.74 (0.92, 3.29) p=0.089	0.81 (0.51, 1.29) p=0.375	1.17 (0.68, 2.03) p=0.573	1.21 (0.72, 2.04) p=0.475
HF aetiology (ischaemic vs non-ischaemic)	1.20 (0.64, 2.25) p=0.577	1.28 (0.94, 1.74) p=0.111	0.95 (0.31, 2.95) p=0.928	1.10 (0.57, 2.16) p=0.771	1.47 (0.94, 2.29) p=0.128	1.28 (0.90, 1.84) p=0.172	1.29 (0.79, 2.12) p=0.309
Ethnic group (white vs non-white)	1.18 (0.85, 1.65) p=0.318	1.19 (0.86, 1.66) p=0.291	(a)	1.79 (0.60, 5.37) p=0.301	1.25 (0.79, 1.98) p=0.334	1.20 (0.83, 1.74) p=0.327	1.18 (0.85, 1.65) p=0.318
Exercise capacity:							
Baseline peakVO_2, directly measured	0.97 (0.88, 1.07) p=0.538	0.97 (0.93, 1.01) p=0.149	0.94 (0.80, 1.11) p=0.467	0.98 (0.90, 1.07) p=0.658	0.99 (0.90, 1.10) p=0.882	0.99 (0.93, 1.06) p=0.769	0.98 (0.89,

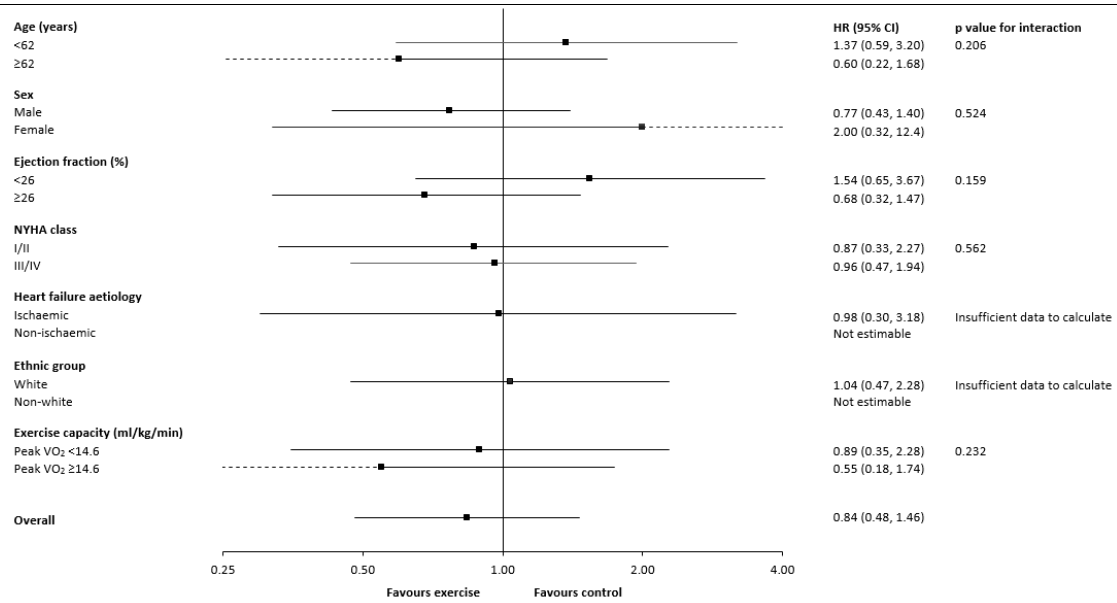
							1.08) p=0.685
Baseline peak VO_2, directly measured and predicted	0.97 (0.89, 1.07) p=0.539	0.97 (0.93, 1.01) p=0.116	0.95 (0.82, 1.10) p=0.483	0.97 (0.89, 1.05) p=0.424	0.98 (0.92, 1.05) p=0.610	0.99 (0.94, 1.03) p=0.535	0.98 (0.90, 1.07) p=0.670
Standardised scores using baseline peak VO_2, 6MWT, ISWT units and watts	0.88 (0.62, 1.26) p=0.483	0.86 (0.72, 1.03) p=0.093	0.83 (0.46, 1.49) p=0.527	0.81 (0.56, 1.16) p=0.246	0.92 (0.69, 1.23) p=0.576	0.93 (0.75, 1.16) p=0.517	0.91 (0.69, 1.20) p=0.505

(a) Study estimate not available as too few studies provide data;

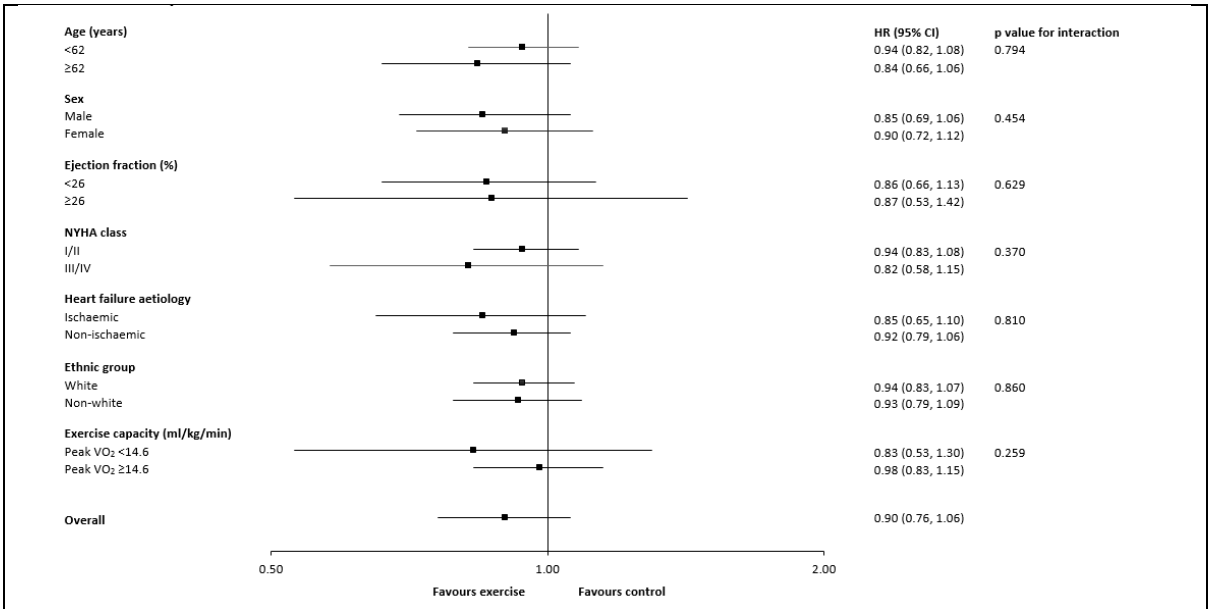
Peak VO_2: peak oxygen uptake; ISWT: incremental shuttle walk test; 6MWT: 6-minute walk test.



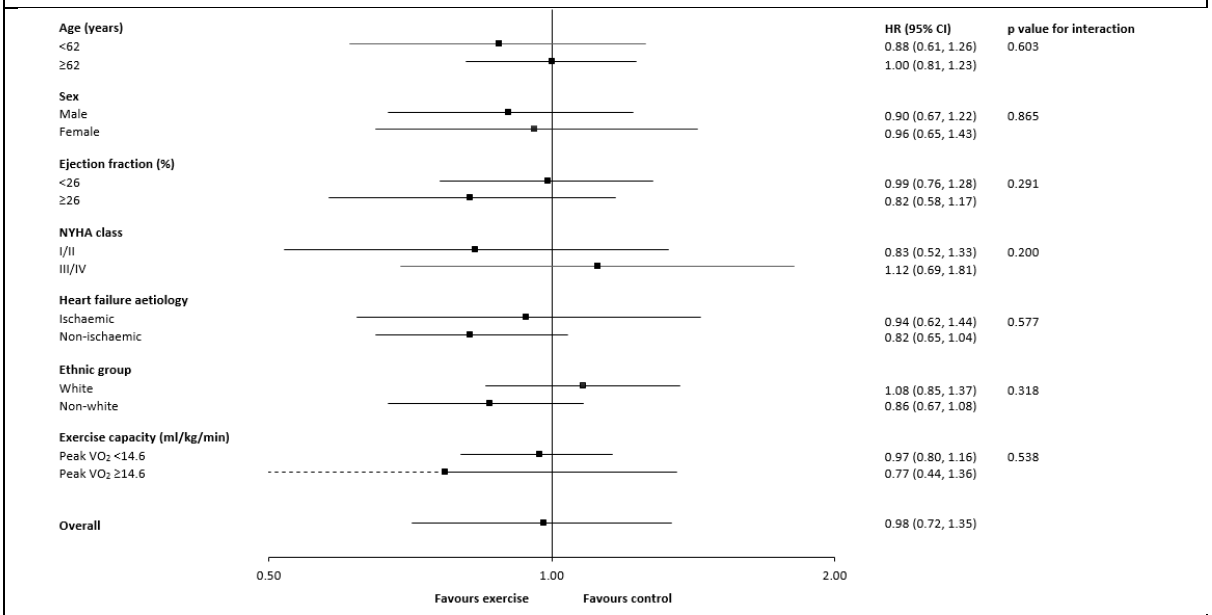
a. All-cause mortality



b. HF-specific mortality



c. All-cause hospitalisation



d. HF-specific hospitalisation

*although stratified meta-analyses are shown, the interaction P-values are calculated based on continuous distribution of age, ejection fraction, and baseline exercise capacity

Figure 5: Effect of ExCR on mortality and hospitalisation across patient subgroups: Individual subgroup one-stage IPD meta-analyses

Secondary analysis

These primary analysis results were broadly consistent across secondary analyses.

Sensitivity analyses

In sensitivity analyses, four weak interaction effects ($p<0.05$) were seen: (1) age vs. all-cause mortality ($p=0.034$) in two-stage model 2 year truncation model, i.e. larger reduction in all-cause mortality with ExCR in older patients; (2) age vs. HF-mortality ($p=0.017$) in two-stage 2-year truncation model, i.e. larger reduction in HF-mortality with ExCR in older patients; (3) ischemic status vs. HF-mortality ($p=0.047$) in one-stage model, i.e. larger reduction in HF-mortality with ExCR in ischemic patients; and (4) standardised baseline exercise capacity vs. all-cause hospitalisation ($p=0.027$) in two-stage 1-year truncation model – larger reduction in all-cause hospitalisation with ExCR in patients with lower than average baseline exercise capacity (see Tables 7-10). Inferences did not change following the addition of trial level data from trials that met our study inclusion criteria but were not able to contribute IPD (data not shown here, available from authors).

Chapter 6: Impact of ExCR on HRQoL and exercise capacity

Six trials which provided IPD but had no data on HRQoL or exercise capacity were excluded from analyses in this section. [50, 52, 54, 55, 57, 73] In addition to comparing usual care to an intervention arm of usual care plus ExCR, Gary [59] also compared the effects of cognitive behaviour therapy to cognitive behaviour therapy plus ExCR. For the purpose of analysis from this point forward, this will be described as one trial providing two comparisons. For analysis the dataset was split into two and analysed as if the data were provided by two separate trials. For the HRQoL analysis, 9 trials (10 comparisons) provided data for 3,000 patients (1,496 ExCR, 1,504 control) with a median follow-up of 33 weeks. [19, 58, 59, 61, 63-67] For the exercise capacity analysis, 13 trials (14 comparisons) provided 3,332 patients (1,662 ExCR, 1,670 control) with a median follow-up of 26 weeks. [19, 51, 56, 58-67]. **Error! Reference source not found.** Figure 6 summarises the study selection process.

Characteristics of included patients and trials

Patient baseline characteristics were well balanced between ExCR and control patients (see Table 11). The majority of patients were male (73%) with a mean age of 61 years. The mean baseline left-ventricular ejection fraction was 27%, and less than 5 percent of included patients had a preserved ejection fraction heart failure (defined as ejection fraction > 45%), and most patients were in NYHA functional class II (62%) or III (36%). Studies were published between 2000 and 2012 across a number of countries (see Table 12). Sample size ranged from 50 to 2,130 patients. All trials evaluated an aerobic exercise intervention; four also included resistance training. [58, 61, 64, 65] Exercise training was most commonly delivered in an exclusively centre-based setting. Four trials were conducted in an exclusively home-based setting. [58, 59, 61, 67] The dose of exercise training ranged across studies with an average session duration of 15 to 60 minutes (including warm-up and cool-down), 2 to 7 sessions/week, exercise intensity equivalent of 40 to 70% peak VO_2, and delivered over a duration of 4 to 120 weeks.

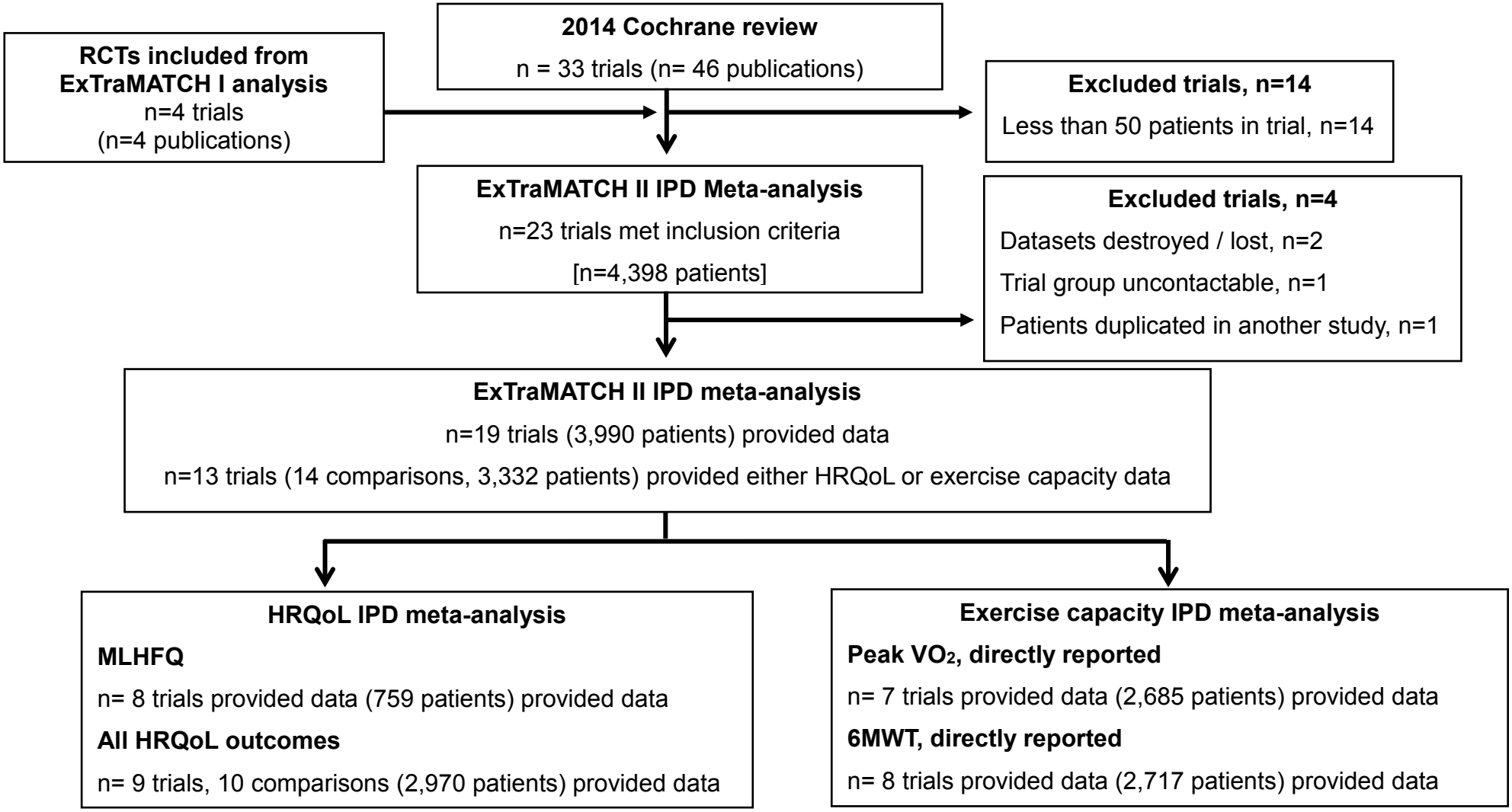


Figure 6: PRISMA flow diagram summarising selection of studies for HRQoL and exercise capacity analyses

Table 11: Baseline characteristics of patients in the HRQoL and exercise capacity analyses

Characteristic	ExCR (n=1,662)	Control (n=1,670)	All (n=3,332)
Age (years), mean (SD)	60.9 (13.2)	61.2 (13.5)	61.1 (13.4)
Gender			
Male	1,187 (71.4)	1,237 (74.1)	2,424 (72.8)
Female	475 (28.6)	433 (25.9)	908 (27.3)
Baseline ejection fraction (%); mean (SD)	27.0 (8.8)	26.9 (8.7)	26.9 (8.8)
Baseline ejection fraction:			
HFrEF (< 45%)	1,721 (96.8)	1,744 (97.5)	3,465 (97.1)
HFpEF (≥ 45%)	57 (3.2)	45 (2.5)	102 (2.9)
NYHA status			
Class I	20 (1.2)	25 (1.5)	45 (1.4)
Class II	1,002 (61)	1,032 (63)	2,034 (62.0)
Class III	597 (36)	569 (35)	1,166 (35.5)
Class IV	19 (1.2)	18 (1.1)	37 (1.1)
Aetiology			
Ischaemic	892 (54.9)	884 (54.1)	1,776 (54.5)
Non-ischemic	732 (45.1)	750 (45.9)	1,482 (45.5)
Ethnicity			
White	1,085 (69.3)	1,117 (70.9)	2,202 (70.1)
Non-white	480 (30.7)	458 (29.1)	938 (30.0)
VO_2 peak (ml/kg/min); mean (SD)	15.0 (4.5)	15.1 (4.7)	15.0 (4.6)
6MWT (metres); mean (SD)	362.6 (109.3)	362.5 (112.1)	362.6 (110.7)

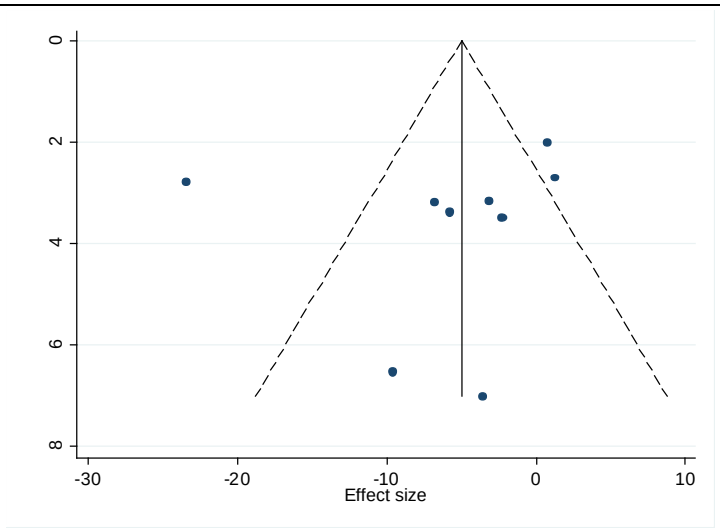
Table 12: Characteristics of included trials in the HRQoL and exercise capacity analyses

Study characteristics	
Publication year	
1990 to 1999	0 (0)
2000 to 2009	9 (64)
2010 to 2012	5 (36)
Unpublished	0 (0)
Main study location	
Europe	9 (64)
North America*	5 (36)
Single study centre	
Single	10 (71.4)
Multiple	4 (28.6)
Not reported	0 (0)
Sample size	
0 to 99	8 (57)
100 to 999	5 (36)
1000 and over	1 (7)
Duration of latest follow up (weeks); median (range)	
HRQoL outcomes	33 (26 to 104)
Exercise capacity outcomes	26 (9 to 520)
Intervention characteristics	
Intervention type	
Exercise only programs	9 (64.3)
Comprehensive programs	5 (35.7)
Type of exercise	
Aerobic exercise only	10 (71.4)
Aerobic plus resistance training	4 (28.6)
Dose of intervention	

Duration of intervention (weeks), median (range)	24 (4 to 120)
Frequency (sessions per week), median (range)	3 (2 to 6.5)
Length of exercise session (mins), median (range)	30 (15 to 60)
Exercise intensity, range	40-70% peak VO_2 11-15 Borg rating
Setting	
Centre-based only	9 (64.3)
Home-based only	5 (35.7)

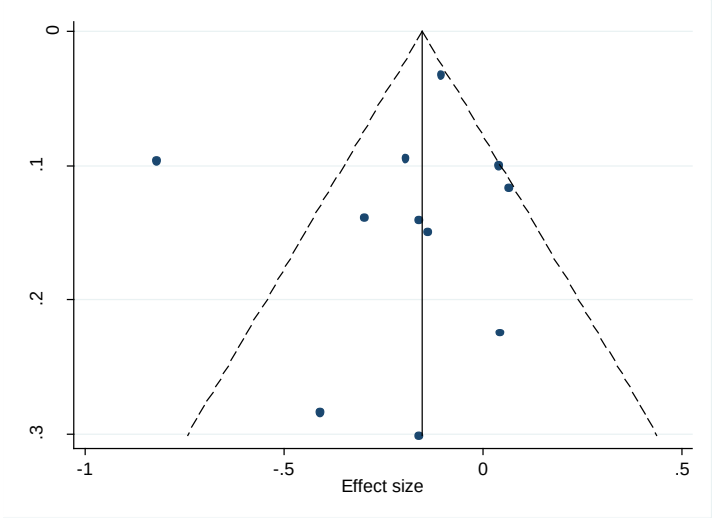
Assessment of study quality and risk of bias

There was no evidence of significant small study bias for the five outcomes studied (see Figure 7). The overall quality of included trials was judged to be moderate to good, with a median TESTEX [31] score of 11 (range 9 to 14) out of a maximum score of 15 (see Table 13**Error! Reference source not found.**). The criteria of allocation concealment and physical activity monitoring in the control groups were met in only two [19, 61] and three studies, [19, 58, 66] respectively. The other TESTEX criteria were each met in 50% or more of trials.



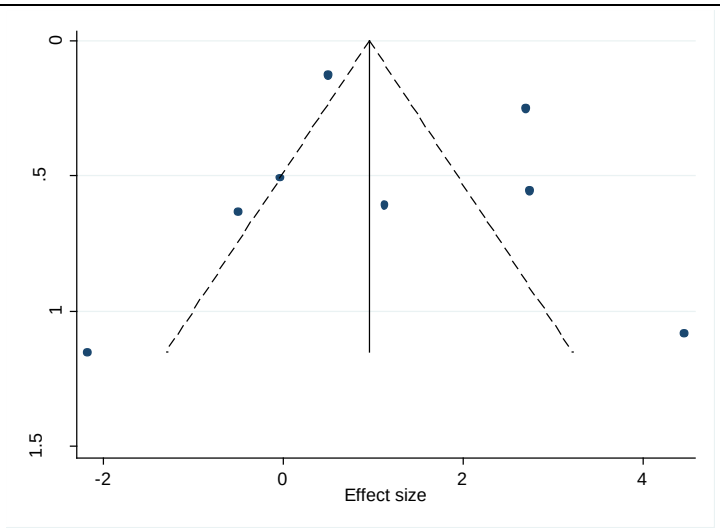
a. Minnesota Living with Heart Failure (12 months)

Egger test -1.40, p=0.656



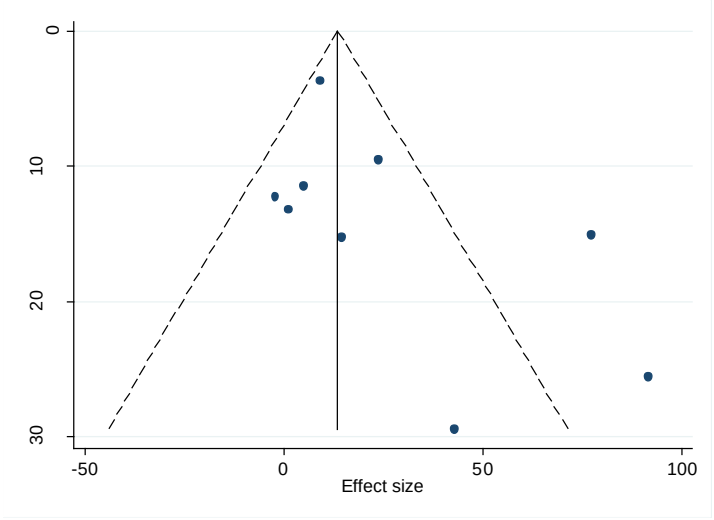
b. All HRQoL measures (12 months)

Egger test -0.72, p=0.577



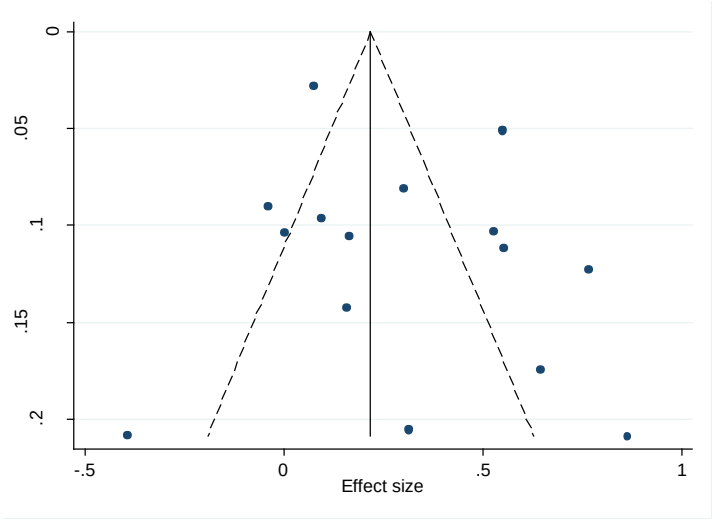
c. Peak VO_2, directly reported (12 months)

Egger test 0.99, p=0.665



d. 6MWT, directly reported (12 months)

Egger test 1.71, p=0.150



e. All exercise capacity measures (12 months)

Egger test 1.85, p=0.214

Figure 7: Funnel plots for HRQoL and exercise capacity analyses

Table 13: Assessment of quality using TESTEX scale of included studies in HRQoL and exercise capacity analysis

Study (publication year)	Eligibility Criteria Specified	Randomisation Specified	Allocation Concealed	Groups Similar at baseline	Blinding of Assessors	Outcome measures in >85% participants (a)	Intention to treat analysis (b)	Between-group statistical comparisons reported	Point measures & measures of	Activity monitoring in control group	Relative Exercise intensity reviewed	Exercise Volume and Energy Expended	Overall TESTEX score
Belardinelli (2012)	1	0	0	1	0	3	1	1	1	0	0	1	9
Dracup (2007)	1	0	0	1	0	3	1	2	1	1	1	1	10
Gary (2010)	1	1	0	1	1	3	1	2	1	0	0	0	11
Giannuzzi (2003)	1	0	0	1	0	2	1	2	1	0	1	1	10
Hambrecht (2000)	1	1	0	1	0	3	0	2	1	0	1	1	11
HF-Action (2008)	1	1	1	1	1	3	1	2	1	1	0	1	14
Jolly (2009)	1	1	1	1	0	2	1	2	1	0	1	1	12
Mueller (2007)	1	0	0	1	0	2	1	2	1	0	1	1	10
Nilsson (2008)	1	1	0	1	1	2	1	2	1	0	0	1	11
Passino (2006)	1	0	0	1	0	2	1	2	1	0	1	1	10
Witham (2005)	1	1	0	1	1	3	1	2	1	0	1	0	12
Witham (2012)	1	1	0	1	1	3	1	2	1	0	1	0	12
Yeh (2011)	1	1	0	1	1	3	1	2	1	1	0	0	12

(a) Three points possible

(b) If ITT was not specifically mentioned, but it was noted that no participants withdrew and all analysed

Findings

Primary analysis

Compared with control, treatment effects from the one-stage meta-analysis at 12 month follow-up showed a significant improvement with ExCR in exercise capacity as assessed by 6MWT (mean difference: 21.0 metres, 95% CI: 1.57 to 40.4, p=0.034, τ^2 = 491, I^2 =78%) and standardised exercise capacity score (mean difference: 0.27 standard deviation units, 95% CI 0.11 to 0.43, p=0.001, τ^2 =0.08, I^2 =91%). No significant difference in VO_2peak at 12-months was observed: 1.01 (95% CI -0.42 to 2.44, p=0.168, τ^2 = 2.17, I^2 =94%).

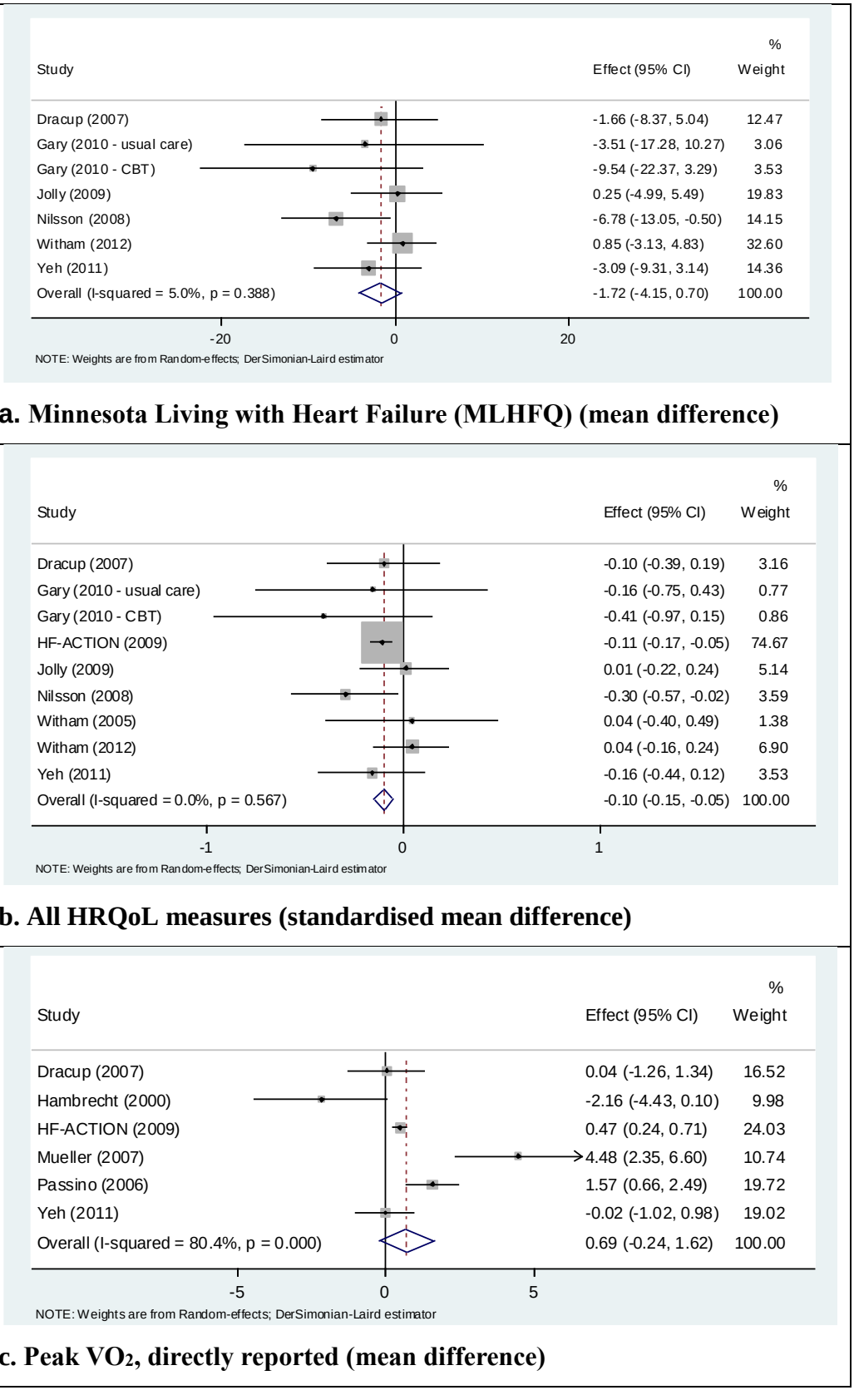
One-stage meta-analysis showed a significant improvement in HRQoL as assessed by the MLHFQ at 12-month follow-up: mean difference: -5.94, 95% CI -1.0 to -10.9, p=0.018, τ^2 =77, I^2 =88%), standardised HRQoL score (mean difference: 0.20 standard deviation units. 95% CI 0.03 to 0.37, p=0.020, τ^2 =0.07, I^2 =85%). Similar results were seen at 6-months follow up (see Forest plots for 6 month follow-up in Figure 8 and 12 month follow-up in Figure 9)). Marked statistical heterogeneity (I^2 > 70%) was seen for all exercise capacity and HRQoL outcomes.

Analyses revealed no consistent interaction between the effect of ExCR and the predefined subgroups (gender, ejection fraction, NYHA class, HF aetiology, ethnicity, and baseline exercise capacity) for either exercise or HRQoL.

A differential effect of ExCR across ages was observed in the standardised HRQoL analysis, with a reduction in HRQoL score (i.e.: an increase in standardised HRQoL score) as age increased (0.006 standard deviation units, 95% 0.002 to 0.011, p=0.006). To put this into context, based on MLHFQ standard deviation of 24, this equates to a decrease of 1.4 in the treatment effect on the MLHFQ score for every 10 years increase in patient age.

Interaction analyses for the one-stage model at 12-month follow-up showed differential effects of ExCR dependent by gender, with women showing greater benefit than men for each of VO_2peak (0.57 ml/kg/min, 95% CI: 0.04 to 1.11, p=0.036) and 6MWT (14.9 metres, 95% CI: 1.2 to 28.7, p=0.034). Differential effects of ExCR were also seen between ethnic

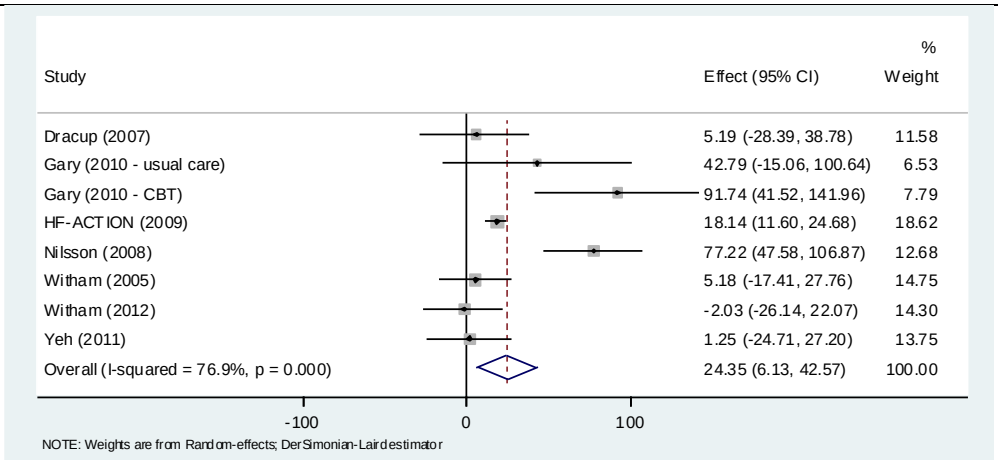
groups, with white patients showing a greater improvement in 6MWT distance compared to non-white patients:(14.2 metres, 95% CI: 0.40 to 28.0, p=0.044).



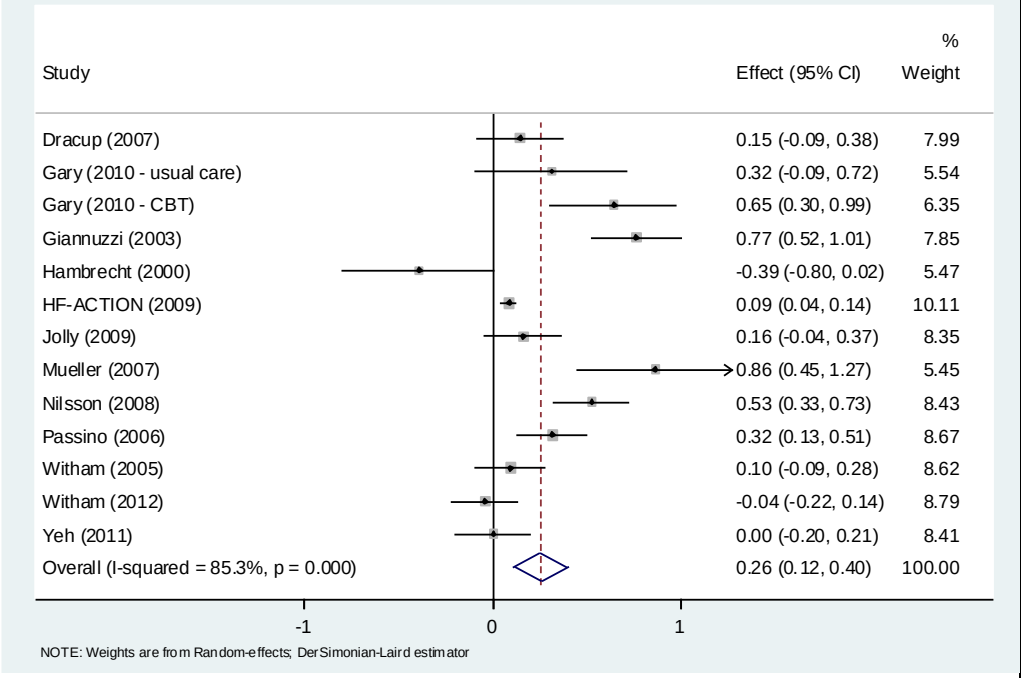
a. Minnesota Living with Heart Failure (MLHFQ) (mean difference)

b. All HRQoL measures (standardised mean difference)

c. Peak VO_2, directly reported (mean difference)



d. 6MWT (metres), directly reported (mean difference)



e. All exercise capacity measures (standardised mean difference)

Figure 8: Effect of ExCR on HRQoL and exercise capacity at 6 months: two-stage IPD meta-analysis

Temporary reference list:

Dracup (2007): Dracup K, Evangelista LS, Hamilton MA, Erickson V, Hage A, Moriguchi J, et al. Effects of a home-based exercise program on clinical outcomes in heart failure. Am Heart J. 2007;154(5):877-83.

Gary (2010): Gary RA, Dunbar SB, Higgins MK, Musselman DL, Smith AL. Combined exercise and cognitive behavioral therapy improves outcomes in patients with heart failure. J Psychosom Res. 2010;69(2):119-31.

Giannuzzi (2003): Giannuzzi P, Temporelli PL, Corra U, Tavazzi L. Antiremodeling effect of long-term exercise training in patients with stable chronic heart failure: results of the Exercise in Left Ventricular Dysfunction and Chronic Heart Failure (ELVD-CHF) Trial. Circulation. 2003;108(5):554-9.

Hambrecht (2000): Hambrecht R, Gielen S, Linke A, Fiehn E, Yu J, Walther C, et al. Effects of exercise training on left ventricular function and peripheral resistance in patients with chronic heart failure: A randomized trial. JAMA. 2000;283(23):3095-101.

HF-ACTION (2009): O'Connor CM, Whellan DJ, Lee KL, Keteyian SJ, Cooper LS, Ellis SJ, et al. Efficacy and Safety of Exercise Training in Patients With Chronic Heart Failure: HF-ACTION Randomized Controlled Trial. JAMA. 2009;301(14):1439-50.

Jolly (2009): Jolly K, Taylor RS, Lip GY, Davies M, Davis R, Mant J, et al. A randomized trial of the addition of home-based exercise to specialist heart failure nurse care: the Birmingham Rehabilitation Uptake Maximisation study for patients with Congestive Heart Failure (BRUM-CHF) study. Eur J Heart Fail. 2009;11(2):205-13.

Mueller (2007): Mueller L, Myers J, Kottman W, Oswald U, Boesch C, Arbrol N, et al. Exercise capacity, physical activity patterns and outcomes six years after cardiac rehabilitation in patients with heart failure. Clin Rehabil. 2007;21(10):923-31.

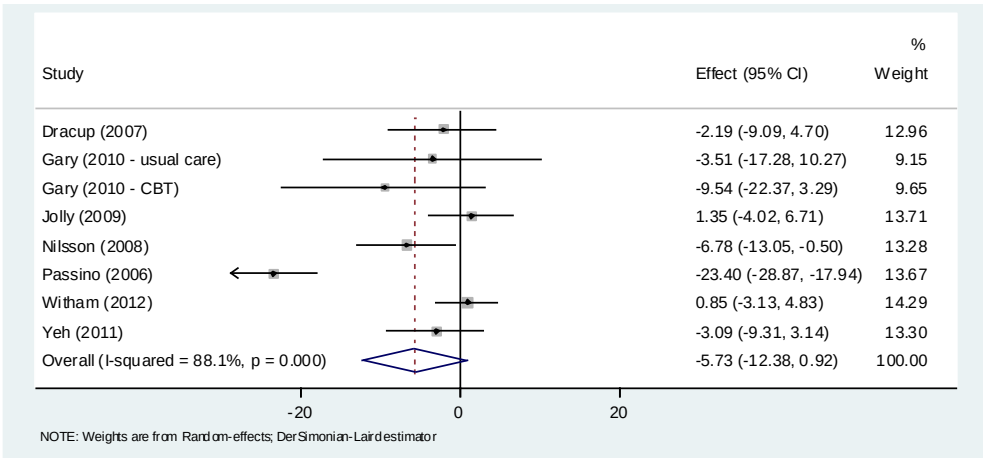
Nilsson (2008): Nilsson BB, Westheim A, Risberg MA. Long-term effects of a group-based high-intensity aerobic interval-training program in patients with chronic heart failure. Am J Cardiol. 2008;102(9):1220-4.

Passino (2006): Passino C, Severino S, Poletti R, Piepoli MF, Mammini C, Clerico A, et al. Aerobic Training Decreases B-Type Natriuretic Peptide Expression and Adrenergic Activation in Patients With Heart Failure. J Am Coll Cardiol. 2006;47(9):1835-9.

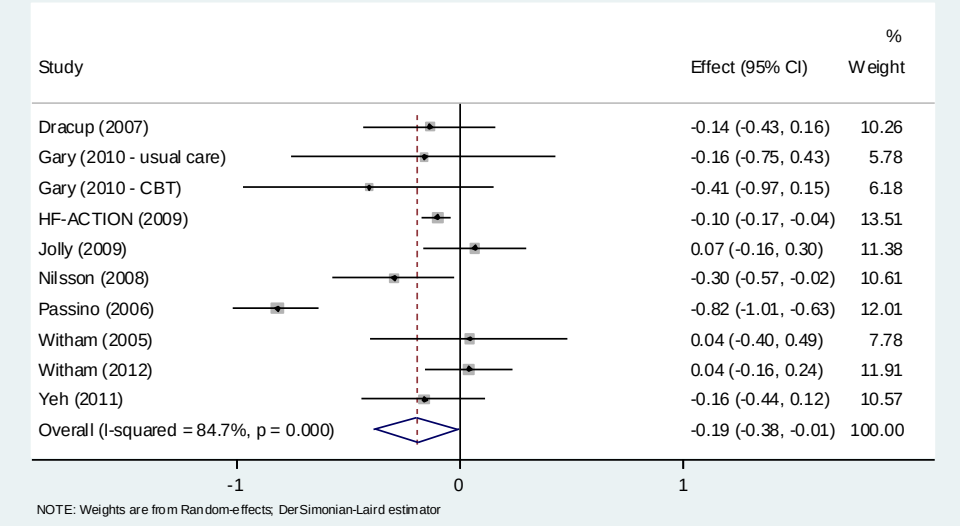
Witham (2005): Witham MD, Gray JM, Argo IS, Johnston DW, Struthers AD, McMurdo ME. Effect of a seated exercise program to improve physical function and health status in frail patients > or = 70 years of age with heart failure. Am J Cardiol. 2005;95(9):1120-4.

Witham (2012): Witham MD, Fulton RL, Greig CA, Johnston DW, Lang CC, van der Pol M, et al. Efficacy and cost of an exercise program for functionally impaired older patients with heart failure: a randomized controlled trial. Circ Heart Fail. 2012;5(2):209-16.

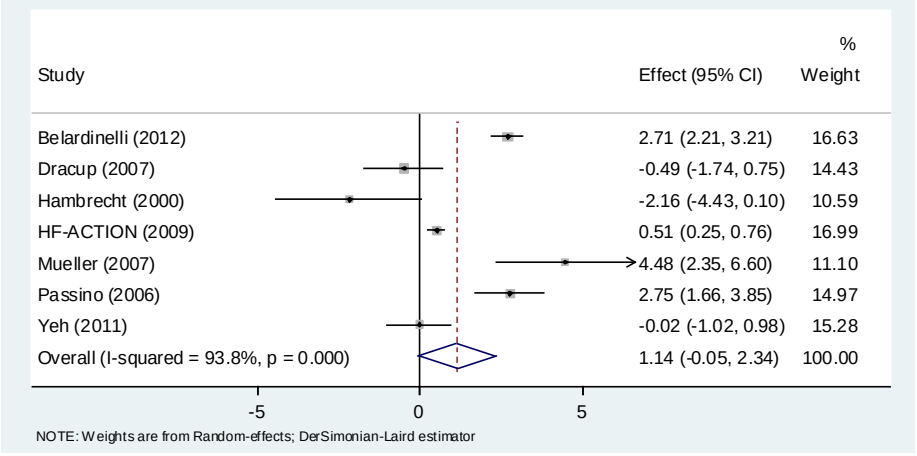
Yeh (2011): Yeh GY, McCarthy EP, Wayne PM, Stevenson LW, Wood MJ, Forman D, et al. Tai Chi Exercise in Patients With Chronic Heart Failure: A Randomized Clinical Trial. Arch Intern Med. 2011;171(8):750-7.



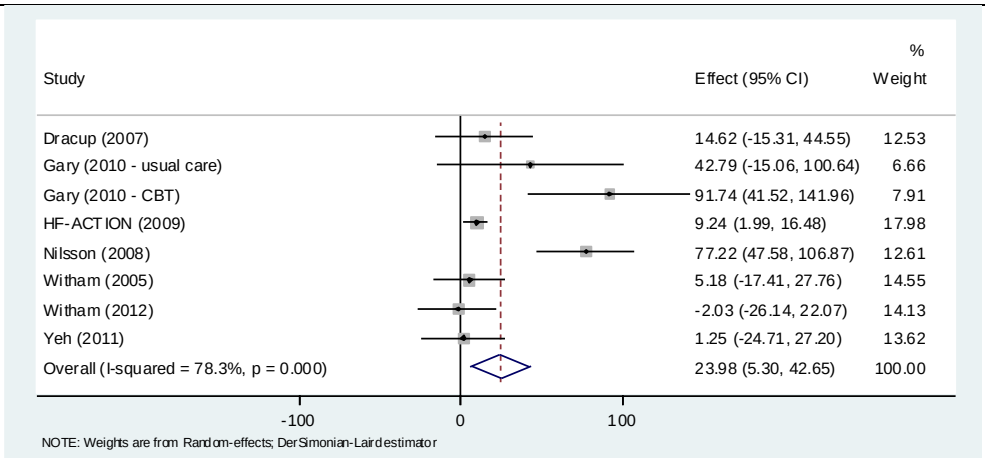
a. Minnesota Living with Heart Failure (MLHFQ) (mean difference)



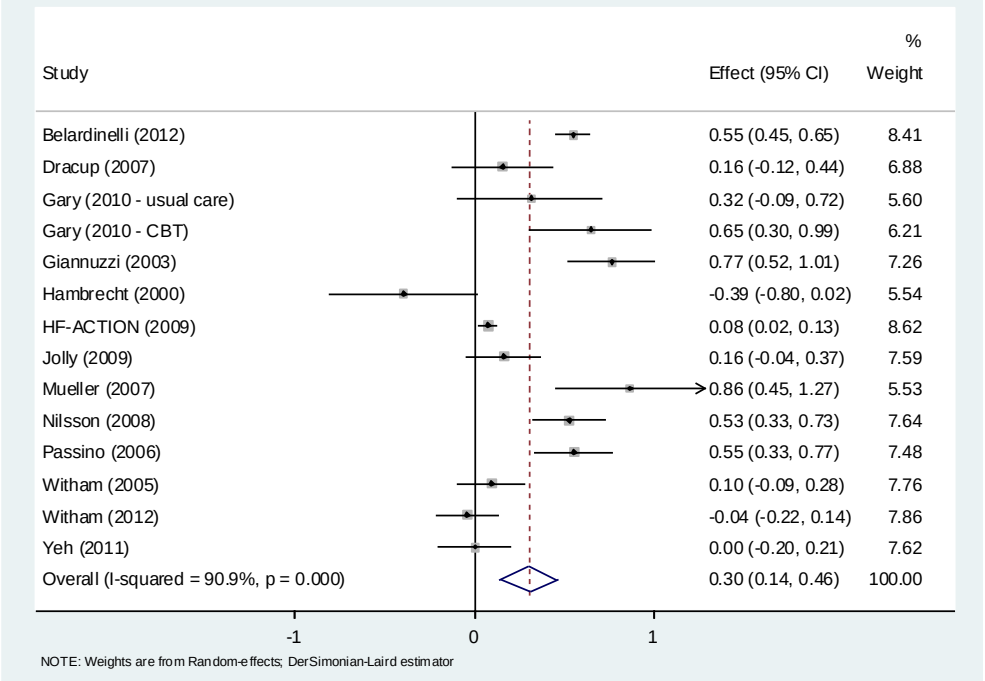
b. All HRQoL measures (standardised mean differences)



c. Peak VO_2, directly reported (mean difference)



d. 6MWT (metres), directly reported (mean difference)



e. All exercise capacity measures (standardised mean difference)

Figure 9: Effect of ExCR on HRQoL and exercise capacity at 12 months: two-stage IPD meta-analysis

Temporary reference list:

Dracup (2007): Dracup K, Evangelista LS, Hamilton MA, Erickson V, Hage A, Moriguchi J, et al. Effects of a home-based exercise program on clinical outcomes in heart failure. Am Heart J. 2007;154(5):877-83.

Gary (2010): Gary RA, Dunbar SB, Higgins MK, Musselman DL, Smith AL. Combined exercise and cognitive behavioral therapy improves outcomes in patients with heart failure. J Psychosom Res. 2010;69(2):119-31.

Giannuzzi (2003): Giannuzzi P, Temporelli PL, Corra U, Tavazzi L. Antiremodeling effect of long-term exercise training in patients with stable chronic heart failure: results of the Exercise in Left Ventricular Dysfunction and Chronic Heart Failure (ELVD-CHF) Trial. Circulation. 2003;108(5):554-9.

Hambrecht (2000): Hambrecht R, Gielen S, Linke A, Fiehn E, Yu J, Walther C, et al. Effects of exercise training on left ventricular function and peripheral resistance in patients with chronic heart failure: A randomized trial. JAMA. 2000;283(23):3095-101.

HF-ACTION (2009): O'Connor CM, Whellan DJ, Lee KL, Keteyian SJ, Cooper LS, Ellis SJ, et al. Efficacy and Safety of Exercise Training in Patients With Chronic Heart Failure: HF-ACTION Randomized Controlled Trial. JAMA. 2009;301(14):1439-50.

Jolly (2009): Jolly K, Taylor RS, Lip GY, Davies M, Davis R, Mant J, et al. A randomized trial of the addition of home-based exercise to specialist heart failure nurse care: the Birmingham Rehabilitation Uptake Maximisation study for patients with Congestive Heart Failure (BRUM-CHF) study. Eur J Heart Fail. 2009;11(2):205-13.

Mueller (2007): Mueller L, Myers J, Kottman W, Oswald U, Boesch C, Arbrol N, et al. Exercise capacity, physical activity patterns and outcomes six years after cardiac rehabilitation in patients with heart failure. Clin Rehabil. 2007;21(10):923-31.

Nilsson (2008): Nilsson BB, Westheim A, Risberg MA. Long-term effects of a group-based high-intensity aerobic interval-training program in patients with chronic heart failure. Am J Cardiol. 2008;102(9):1220-4.

Passino (2006): Passino C, Severino S, Poletti R, Piepoli MF, Mammini C, Clerico A, et al. Aerobic Training Decreases B-Type Natriuretic Peptide Expression and Adrenergic Activation in Patients With Heart Failure. J Am Coll Cardiol. 2006;47(9):1835-9.

Witham (2005): Witham MD, Gray JM, Argo IS, Johnston DW, Struthers AD, McMurdo ME. Effect of a seated exercise program to improve physical function and health status in frail patients > or = 70 years of age with heart failure. Am J Cardiol. 2005;95(9):1120-4.

Witham (2012): Witham MD, Fulton RL, Greig CA, Johnston DW, Lang CC, van der Pol M, et al. Efficacy and cost of an exercise program for functionally impaired older patients with heart failure: a randomized controlled trial. Circ Heart Fail. 2012;5(2):209-16.

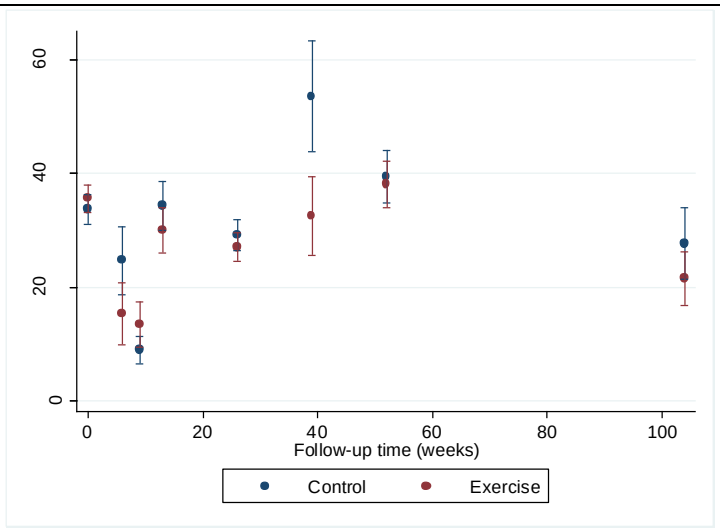
Yeh (2011): Yeh GY, McCarthy EP, Wayne PM, Stevenson LW, Wood MJ, Forman D, et al. Tai Chi Exercise in Patients With Chronic Heart Failure: A Randomized Clinical Trial. Arch Intern Med. 2011;171(8):750-7.

Secondary analysis

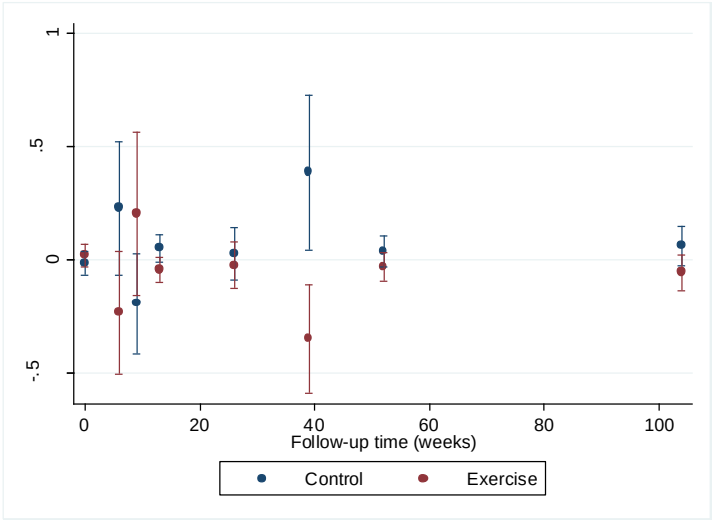
In the repeated measures analyses for each HRQoL and exercise capacity outcome, a significant interaction between ExCR and time was observed (see Figure 10**Error! Reference source not found.**). In order to visualise comparisons of changes in HRQoL and exercise capacity within each subgroup, the effect estimates and associated 95% CI from individual subgroup one-stage IPD meta-analyses are shown in Figure 11. The p-value from the interaction test in the two-stage IPD meta-analyses are presented alongside these estimates.

Sensitivity analyses

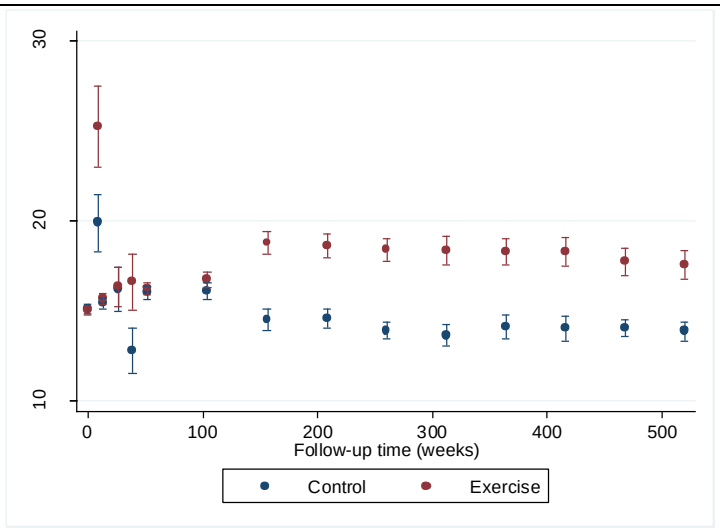
In sensitivity analyses, the results of the analyses excluding HF-ACTION, were broadly consistent with the overall results. (see Tables 14-18). Similar results were found with the addition of the study-level aggregate data to the two-stage model at 12 months follow-up.



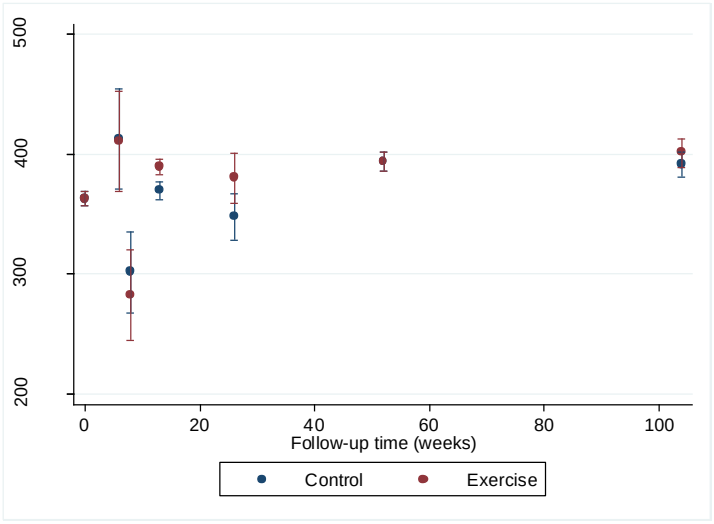
a. Minnesota Living with Heart Failure (mean difference)



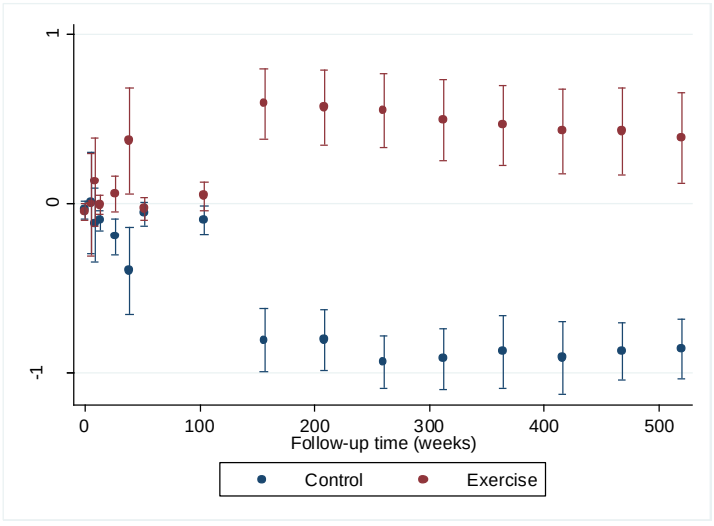
b. All HRQoL measures (standardised mean difference)



c. Peak VO_2, directly reported (mean difference)

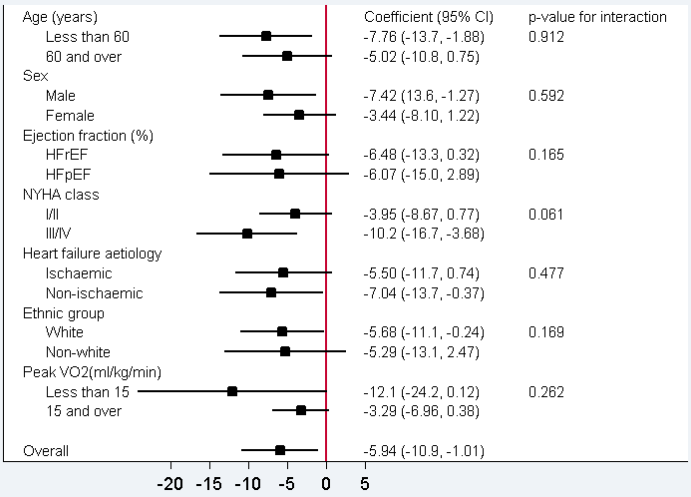


d. 6MWT, directly reported (mean difference)

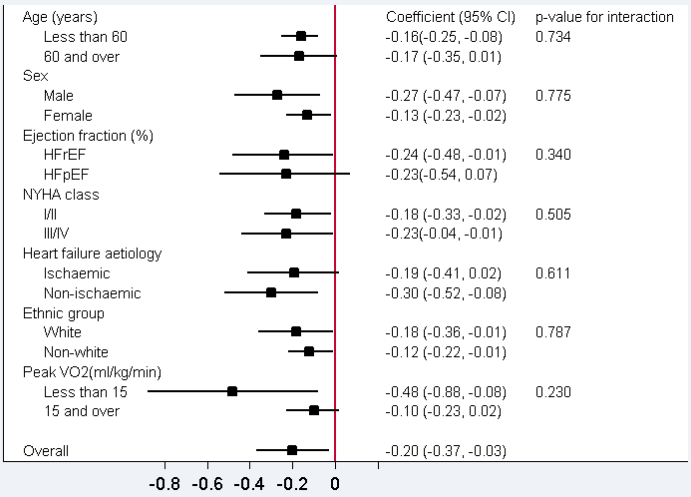


e. All exercise capacity measures (standardised mean difference)

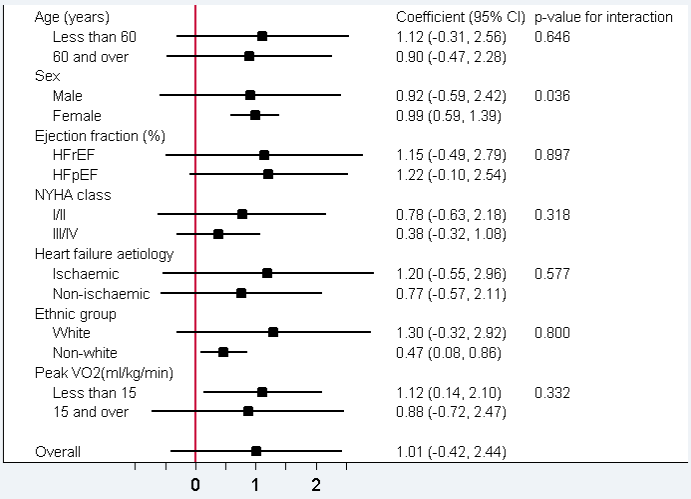
Figure 10: Effect of ExCR on HRQoL and exercise capacity



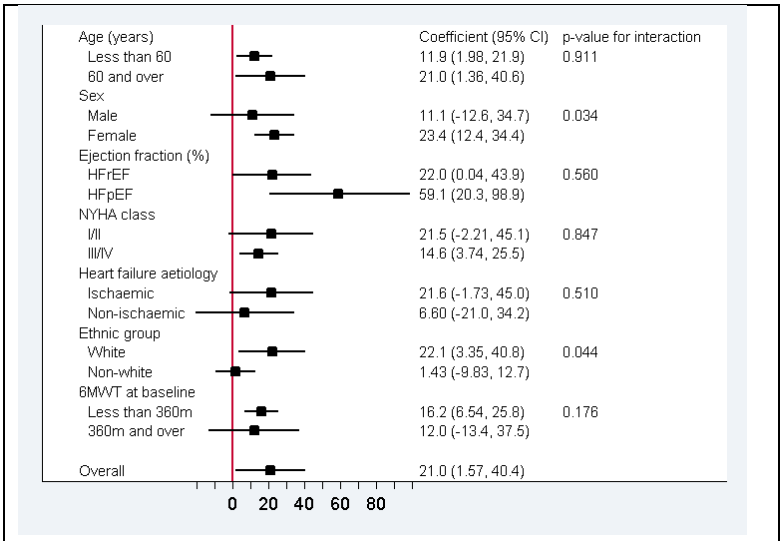
a. Minnesota Living with Heart Failure (MLHFQ)



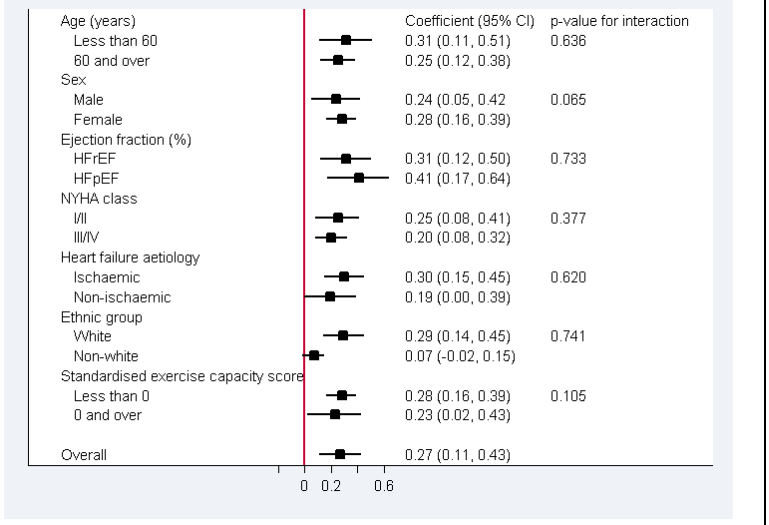
b. All HRQoL measures (standardised score)



c. Peak VO_2, directly reported



d. 6MWT (metres), directly reported



e. All exercise capacity measures (standardised score)

*although stratified meta-analyses are shown, the interaction P-values are calculated based on continuous distribution of age, ejection fraction, and baseline exercise capacity

Figure 11: Effect of ExCR on HRQoL and exercise capacity across patient subgroups (individual subgroup one-stage IPD meta-analyses)

Table 14: MLHFQ - overall treatment effect and subgroup (interaction) effects

	Primary analyses				**Sensitivity analyses, excluding HF-Action**			
	One-stage model, 6 months FU, with random treatment effect Mean difference (95% CI) p-value	**Two-stage model, 6 months FU , with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value
Overall effect	-2.85 (-5.85, 0.14), p=0.062	-1.73 (-4.15, 0.70), p=0.163	-5.94 (-10.87, -1.01), p=0.018	-5.73 (-12.38, 0.93), p=0.091	Not applicable to MLHFQ analyses as HF-Action only supplied Kansas scores			
Interaction terms								

Age (years)	0.12 (-0.10, 0.35), p=0.280		0.01 (-0.20, 0.22), p=0.912					
Gender (male vs female)	-5.31 (-11.01, 0.39), p=0.068		-1.49 (-6.95, 3.96), p=0.592					
Ejection fraction (%)	0.22 (-0.14, 0.58), p=0.227		0.24 (-0.07, 0.56), p=0.127					
Ejection Fraction (HFrEF vs HFpEF)	4.06 (-11.0, 19.1), p=0.597		8.02 (-3.29, 19.3), p=0.165					
NYHA class (NYHA I/II vs NYHA III/IV)	-6.38 (-12.31, -0.45), p=0.035		-5.30 (-10.9, 0.24), p=0.061					
HF aetiology (ischaemic vs non-ischaemic)	4.67 (-1.65, 11.0), p=0.147		2.08 (-3.64, 7.80), p=0.477					

Ethnic group (white vs non-white)	3.15 (-4.31, 10.6), p=0.408		5.17 (-2.19, 12.5), p=0.169					
Exercise capacity								
Baseline peak VO_2 directly measured	0.24 (-0.82, 1.31), p=0.654		0.47 (-0.35, 1.29), p=0.262					
Baseline peak VO_2, directly measured and predicted	0.72 (-0.01, 1.45), p=0.053		0.62 (-0.02, 1.26), p=0.058					

Peak VO_2: peak oxygen uptake

Table 15: Standardised HRQoL measure- overall treatment effect and subgroup (interaction) effects

	Primary analyses				**Sensitivity analyses, excluding HF-Action**			
	One-stage model, 6 months FU, with random treatment effect Mean difference (95% CI) p-value	**Two-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value
Overall effect	-0.11 (-0.16, -0.06), p<0.001	-0.10 (-0.15, -0.05), p<0.001	-0.20 (-0.37, -0.03), p=0.020	-0.19 (-0.38, -0.01), p=0.043	-0.11 (-0.24, 0.01), p=0.069	-0.08 (-0.18, 0.02), p=0.131	-0.17 (-0.28, -0.07), p=0.001 *	-0.21 (-0.45, 0.04), p=0.106
Interaction terms								

Age (years)	0.006 (0.002, 0.011), p=0.006		0.001 (-0.004, 0.005), p=0.734		0.003 (-0.007, 0.014), p=0.536		-0.001 (-0.011, 0.008), p=0.788	
Gender (male vs female)	0.050 (-0.068, 0.168), p=0.407		0.018 (-0.105, 0.140), p=0.775		-0.223 (-0.469, 0.024), p=0.077		-0.106 (-0.335, 0.123, p=0.365	
Ejection fraction (%)	-0.000 (-0.007, 0.007), p=0.963		-0.004 (-0.011, 0.004), p=0.340		0.010 (-0.006, 0.025), p=0.225		0.010 (-0.003, 0.023), p=0.150	
Ejection Fraction (HFrEF vs HFpEF)	-0.03 (-0.46, 0.41), p=0.902		0.13 (-0.26, 0.53), p=0.505		0.16 (-0.47, 0.84), p=0.581		0.34 (-0.14, 0.81), p=0.163	
NYHA class (NYHA I/II vs NYHA III/IV)	-0.013 (-0.126, 0.100), p=0.824		0.031 (-0.086, 0.149), p=0.599		-0.126 (-0.380, 0.129), p=0.334		-0.082 (-0.314, 0.151), p=0.491	

HF aetiology (ischaemic vs non-ischaemic)	0.076 (-0.036, 0.187), p=0.182		0.030 (-0.085, 0.145), p=0.611		0.220 (-0.055, 0.494), p=0.117		0.080 (-0.162, 0.322), p=0.517	
Ethnic group (white vs non-white)	0.041 (-0.079, 0.161), p=0.506		0.017 (-0.108, 0.142), p=0.787		0.173 (-0.172, 0.519), p=0.325		0.243 (-0.086, 0.573), p=0.147	
Exercise capacity								
Baseline peak VO_2 directly measured	-0.002 (-0.014, -0.011), p=0.775		0.008 (-0.005, 0.021), p=0.230		0.012 (-0.035, 0.059), 0.612		0.021 (-0.012, 0.055), p=0.216	
Baseline peak VO_2, directly measured and predicted	0.000 (-0.012, 0.013), p=0.956		0.008 (-0.004, 0.021), p=0.208		0.023 (-0.010, 0.056), p=0.171		0.020 (-0.008, 0.048), p=0.172	
Standardised scores using baseline peak	N/A as no further data available				N/A as no further data available			

VO_2, 6MWT, ISWT units, and watts	over analysis in row above				over analysis in row above			

* Fixed effect on treatment with a random effect on study, due to non-convergence of the random treatment effect model; peak VO_2: peak oxygen uptake; ISWT: incremental shuttle walk test; 6MWT: 6-minute walk test.

Table 16: Peak VO2 directly measured - overall treatment effect and subgroup (interaction) effects

	Primary analyses				**Sensitivity analyses, excluding HF-Action**			
	One-stage model, 6 months FU, with random treatment effect Mean difference (95% CI) p-value	**Two-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value
Overall effect	0.62 (-0.82, 2.07), p=0.397	0.69 (-0.24, 1.62), p=0.145	1.01 (-0.42, 2.44), p=0.168	1.14 (-0.05, 2.34), p=0.061	0.71 (-1.10, 2.52), p=0.444	0.77 (-0.73, 2.28), p=0.315	1.15 (-0.60, 2.90), p=0.196	1.26 (-0.31, 2.82), p=0.115
Interactions terms								

Age (years)	0.00 (-0.02, 0.02), p=0.980		-0.00 (-0.02, 0.14), p=0.646		-0.01 (-0.07, 0.04), p=0.628		-0.02 (-0.06, 0.03), p=0.415	
Gender (male vs female)	-0.25 (-0.78, 0.27), p=0.345		-0.57 (-1.11, -0.04), p=0.036		-0.67 (-2.47, 1.14), p=0.468		-0.42 (-1.80, 0.95), p=0.549	
Ejection fraction (%)	0.03 (0.00, 0.06), p=0.034		0.02 (-0.01, 0.05), p=0.157		0.05 (-0.04, 0.13), p=0.280		0.03 (-0.04, 0.11), p=0.349	
Ejection Fraction (HFrEF vs HFpEF)	0.07 (-1.88, 2.01), p=0.947		-0.13 (-2.07, 1.81), p=0.897		-1.34 (-2.42, 5.09), p=0.485		-0.19 (-3.34, 2.97), p=0.907	
NYHA class (NYHA I/II vs NYHA III/IV)	-0.10 (-0.58, 0.38), p=0.687		-0.25 (-0.75, 0.24), p=0.318		-0.50 (-2.13, 1.13), p=0.549		-0.75 (-1.95, 0.46), p=0.224	
HF aetiology (ischaemic vs non-ischaemic)	0.02 (-0.44, 0.47), p=0.945		-0.13 (-0.60, 0.34), p=0.577		-0.63 (-2.04, 0.79), p=0.386		-0.24 (-1.39, 0.91), p=0.683	

Ethnic group (white vs non-white)	-0.19 (-0.66, 0.29), p=0.447		-0.07 (-0.58, 0.45), p=0.800		-0.47 (-2.36, 1.43), p=0.628		0.16 (-1.71, 2.03), p=0.870	
Exercise capacity								
Baseline peak VO_2 directly measured	0.01 (-0.04, 0.06), p=0.719		0.03 (-0.03, 0.08), p=0.332		-0.06 (-0.21, 0.09), p=0.435		-0.04 (-0.17, 0.10), p=0.602	
Baseline peak VO_2, directly measured and predicted	0.01 (-0.04, 0.06), p=0.702		0.03 (-0.02, 0.08), p=0.299		-0.06 (-0.21, 0.09), p=0.452		-0.03 (-0.16, 0.10), p=0.660	

Peak VO_2: peak oxygen uptake; ISWT: incremental shuttle walk test; 6MWT: 6-minute walk test.

Table 17: 6MWT directly measured - overall treatment effect and subgroup (interaction) effects

	Primary analyses				**Sensitivity analyses, excluding HF-Action**			
	One-stage model, 6 months FU, with random treatment effect Mean difference (95% CI) p-value	**Two-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value
Overall effect	22.1 (1.87, 42.3), p=0.032	24.4 (6.13, 42.6), p=0.009	21.0 (1.57, 40.4), p=0.034	24.0 (5.30, 42.7), p=0.012	22.1 (-1.64, 45.8), p=0.068	27.9 (1.25, 54.6), p=0.040	24.0 (1.25, 46.7), p=0.039	29.0 (3.05, 55.0), p=0.029
Interaction terms								

Age (years)	0.01 (-0.49, 0.50), p=0.973		-0.03 (-0.56, 0.50), p=0.911		0.45 (-0.81, 1.72), p=0.482		0.97 (-0.23, 2.17), p=0.115	
Gender (male vs female)	-10.7 (-23.6, 2.26), p=0.106		-14.9 (-28.7, -1.16), p=0.034		-19.7 (-47.3, 7.92), p=0.162		-13.5 (-39.9, 12.9), p=0.317	
Ejection fraction (%)	0.34 (-0.46, 1.14), p=0.399		0.21 (-0.64, 1.06), p=0.634		1.05 (-0.78, 2.88), p=0.262		0.04 (-1.69, 1.77), p=0.963	
Ejection fraction (HFrEF vs HFpEF)	0.68 (-47.8, 49.2), p=0.978		15.4 (-36.3, 67.0), p=0.560		13.8 (-6.09, 88.6), p=0.717		14.7 (-56.1, 85.4), p=0.685	
NYHA class (NYHA I/II vs NYHA III/IV)	-1.81 (-14.3, 10.6), p=0.776		1.31 (-12.0, 14.6), p=0.847		-5.90 (-34.6, 22.8), p=0.687		-8.14 (-35.7, 19.4), p=0.563	
HF aetiology (ischaemic vs non-ischaemic)	3.73 (-8.26, 15.7), p=0.542		-4.30 (-17.1, 8.51), p=0.510		37.9 (9.34, 66.4), p=0.009		26.9 (-0.13, 54.0), p=0.051	

Ethnic group (white vs non-white)	10.46 (-2.55, 23.5), p=0.115		14.2 (0.40, 28.0), p=0.044		-20.7 (-60.5, 19.0), p=0.307		8.34 (-29.5, 46.1), p=0.665	
Exercise capacity								
Baseline 6MWT directly measured	-0.05 (-0.11, 0.01), p=0.079		0.19 (-0.08, 0.46), p=0.176		-0.06 (-0.18, 0.06), p=0.321		-0.05 (-0.16, 0.07), p=0.421	

6MWT: 6-minute walk test.

Table 18: Standardised exercise capacity score - overall treatment effect and subgroup (interaction) effects

	Primary analyses				**Sensitivity analyses, excluding HF-Action**			
	One-stage model, 6 months FU, with random treatment effect Mean difference (95% CI) p-value	**Two-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 6 months FU, with random treatment effect** Mean difference (95% CI) p-value	**One-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value	**Two-stage model, 12 months FU, with random treatment effect** Mean difference (95% CI) p-value
Overall effect	0.230 (0.067, 0.392), p=0.006	0.256 (0.116, 0.396), p<0.001	0.268 (0.110, 0.426), p=0.001	0.302 (0.142, 0.462), p<0.001	0.256 (0.079, 0.433), p=0.005	0.278 (0.105, 0.451), p=0.002	0.298 (0.125, 0.471), p=0.001	0.324 (0.150, 0.497), p<0.001
Interaction terms								

Age (years)	0.001 (-0.003, 0.004), p=0.758		-0.001 (-0.005, 0.003), p=0.636		0.003 (-0.008, 0.014), p=0.565		-0.000 (-0.010, 0.009), p=0.948	
Gender (male vs female)	-0.063 (-0.157, 0.319), p=0.194		-0.096 (-0.197, 0.006), p=0.065		-0.066 (-0.250, 0.118), p=0.484		-0.065 (-0.240, 0.110), p=0.464	
Ejection fraction (%)	0.007 (0.001, 0.012), p=0.021		0.005 (-0.001, 0.011), p=0.108		0.008 (-0.003, 0.019), p=0.131		0.008 (-0.003, 0.018), p=0.169	
Ejection Fraction (HFrEF vs HFpEF)	0.11 (-0.20, 0.43), p=0.487		0.06 (-0.28, 0.40), p=0.733		0.21 (-0.23, 0.65), p=0.348		0.06 (-0.36, 0.49), p=0.766	
NYHA class (NYHA I/II vs NYHA III/IV)	-0.010 (-0.098, 0.079), p=0.826		-0.043 (-0.138, 0.052), p=0.377		-0.011 (-0.184, 0.162), p=0.900		-0.061 (-0.224, 0.101), p=0.459	

HF aetiology (ischaemic vs non-ischaemic)	0.012 (-0.074, 0.098), p=0.783		0.024 (-0.070, 0.117), p=0.620		0.035 (-0.143, 0.213), p=0.701		0.049 (-0.121, 0.219), p=0.573	
Ethnic group (white vs non-white)	-0.064 (-0.159, 0.031), p=0.187		0.018 (-0.088, 0.124), p=0.741		-0.096 (-0.352, 0.160), p=0.461		0.078 (-0.195, 0.351), p=0.577	
Exercise capacity								
Standardised scores using baseline peak VO_2, 6MWT, ISWT units, and watts	-0.025 (-0.066, 0.017), p=0.240		-0.017 (-0.048, 0.508), p=0.105		-0.070 (-0.147, 0.007), p=0.077		-0.052 (-0.129, 0.026), p=0.191	

Peak VO_2: peak oxygen uptake; ISWT: incremental shuttle walk test; 6MWT: 6-minute walk test.

Chapter 7: Results from surrogate analysis

Inclusion of trials in the ExTraMATCH II Surrogate analyses

All 19 trials from the ExTraMATCH II study were eligible for analysis here if they provided the required data (as detailed in the Methods section above). Only 10 trials [19, 51, 58, 61-67] provided data for the surrogate analyses. Figure 12**Error! Reference source not found.** summarises the availability of studies and patient data for exercise capacity and the patient-relevant outcomes of mortality, hospitalisation, and HRQoL.

Characteristics of included patients and trials

Across ExCR and control groups, patient baseline characteristics were well balanced (see Table 19**Error! Reference source not found.**). Patients had mean age of 62 years and the majority were male (73%). The mean baseline left-ventricular ejection fraction was 26%; most patients were in NYHA functional class II (63%) or III (34%). Studies were published between 2000 and 2012 from a range of geographical locations (see Table 20**Error! Reference source not found.**). Sample size was typically small and ranged from 50 to 2130 patients. All trials included ExCR based on an aerobic exercise intervention. The dose of ExCR ranged widely across studies: average session duration of 15 to 60 minutes; 2 to 7 sessions/week; exercise intensity equivalent of 40 to 70% VO_2peak, delivered over a duration of 4 to 120 weeks. The change in exercise capacity and final patient-relevant outcomes for each included studies is shown in Table 21**Error! Reference source not found.**.

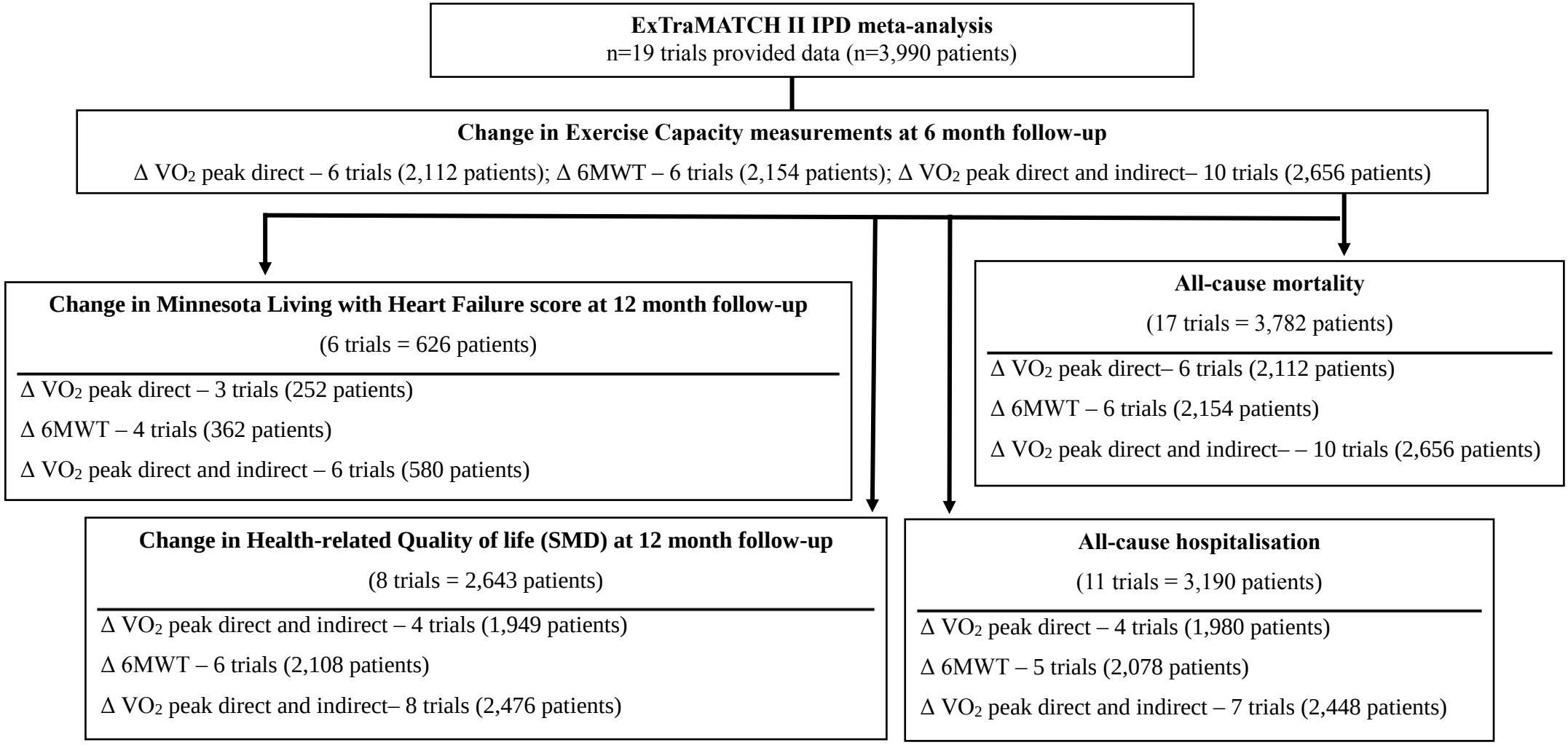


Figure 12: PRISMA flow diagram summarising selection of studies for ExTraMATCH II surrogate analysis

Table 19: Baseline characteristics of patients in surrogate analyses

	ExCR group (n=1,345)	**Control group (n= 1,311)**	**All patients (n= 2,656)**
Age (years), mean (SD)	61.2 (13.0)	61.6 (13.4)	61.39 (13.19)
Gender (male), n (%)	970 (72.1)	973 (74.2)	1,943 (73.2)
Baseline ejection fraction; mean (SD)	26.0 (7.9)	26.2 (7.6)	26.1 (7.8)
NYHA class, n (%)			
Class I	13 (1)	27 (2)	27 (2)
Class II	834 (62)	861 (64)	848 (63)
Class III	485 (36)	444 (33)	457 (34)
Class IV	13 (1)	13 (1)	13 (1)
Aetiology, ischemic, n (%)	713 (53)	708 (54%)	1,421 (54)
Ethnicity, white (%)	914 (70)	908 (71)	1,822 (70)
VO_2peak (ml/kg/min), mean (SD)	15.1 (4.6)	15.2 (4.8)	15.1 (4.7)
6MWT (metres), mean (SD)	368 (108)	366 (110)	367 (109)

6MWT: six minute walk test; NYHA: New York Heart Association;

VO_2peak: peak oxygen uptake

Table 20: Characteristics of included studies and interventions in surrogate analyses

Study characteristics	
Publication year, n (%)	
2000 to 2009	8 (80)
2010 to 2012	2 (20)
Main study location, n (%)	
Europe	6 (60)
North America [(1)]	4 (40)
Single study centre, n (%)	
Single	7 (70)
Multiple	3 (30)
Sample size, n (%)	
0 to 99	5 (50)
100 to 999	4 (40)
1000 and over	1 (10)
Duration of latest follow up (weeks); median (range)	10.5 (6 to 30)
Intervention characteristics	
Type of exercise n (%)	
Aerobic exercise only	6 (60)
Aerobic plus resistance training	4 (40)
Dose of intervention	
Duration of intervention (weeks), median (range)	24 (4 to 120)
Frequency (sessions per week), median (range)	2.75 (2.5 to 6.5)
Length of exercise session (mins), median (range)	30 (15 to 60)
Exercise intensity, range	40 to 70% VO_2peak 11 to 15 Borg rating
Setting, n (%)	
Centre-based only	3 (30)
Home-based only	2 (20)
Both home and centre-based	5 (50)

[(1)] HF-ACTION trial also includes French patients

HRQoL: health-related quality of life; VO_2peak: peak oxygen uptake

Table 21: Change in exercise capacity and final patient-relevant outcomes for each included study

Study (Year)	ΔVO2peak direct (ml/kg/min) mean difference	Δ6MWT (metres) mean difference	ΔVO2peak direct and indirect (ml/kg/min) mean difference	ΔMLHFQ mean difference	ΔHRQoL any validated measure mean difference	All-cause mortality HR	All-cause hospital admission HR
Dracup (2007)	0.04 (-1.26 to 1.34)	5.19 (-28.39 to 38.78)	0.15 (-0.91 to 1.21)	-2.19 (-9.09 to 4.70)	-0.15 (-0.44 to 0.15)	1.16 (0.51 to 2.64)	1.31 (0.84 to 2.05
Hambrecht (2000)	-2.16 (-4.43 to 0.10)		-2.16 (-4.43 to 0.10)			0.93 (0.13 to 6.65)	0.97 (0.14 to 6.88)
HF-ACTION (2009)	0.47 (0.24 to 0.71)	18.14 (11.60 to 24.68)	0.43 (0.20 to 0.66)		-0.10 (-0.17 to -0.04)	0.92 (0.75 to 1.13)	0.93 (0.83 to 1.03)
Jolly (2009)			0.57 (-0.15 to 1.29)	1.35 (-4.02 to 6.71)	0.07 (-0.16 to 0.30)	1.62 (0.45 to 5.78)	0.72 (0.36 to 1.42)
Mueller (2007)	4.47 (2.35 to 6.60)		4.48 (2.35 to 6.60)			0.78 (0.33 to 1.85)	
Nilsson (2008)		77.22 (47.58 to 106.87)	1.78 (1.09 to 2.46)	-6.78 (-13.05 to -0.50)	-0.30 (-0.57 to -0.02)		

Passino (2006)	1.57 (0.66 to 2.49)		1.57 (0.66 to 2.49)	-23.41 (-28.87 to -17.94)	-0.82 (-1.01 to -0.63)	0.48 (0.23 to 0.97)	
Witham (2005)		5.18 (-17.41 to 27.76)	0.12 (-0.40 to 0.64)		0.04 (-0.40 to 0.49)	0.29 (0.03 to 2.84)	1.03 (0.41 to 2.60)
Witham (2012)		-2.03 (-26.14 to 22.08)	-0.05 (-0.60 to 0.51)	0.86 (-3.13 to 4.84)	0.04 (-0.16 to 0.24)	2.09 (0.19 to 23.03)	0.94 (0.39 to 2.28)
Yeh (2011)	-0.02 (-1.02 to 0.98)	1.25 (-24.71 to 27.20)	-0.17 (-1.16 to 0.82)	-3.09 (-9.31 to 3.14)	-0.16 (-0.43 to 0.12)		0.57 (0.14 to 2.38)
Pooled results	**0.69 (-0.24 to 1.62) p=0.145 I^2 = 80.4%**	**16.69 (-1.08 to 34.36) p=0.066 I^2 = 76.5%**	**0.61 (0.10 to 1.11) p=0.019 I^2 = 80.3%**	**-5.53 (-13.27 to 2.21) p=0.162 I^2 = 91.5%**	**-0.18 (-0.39 to 0.02) p=0.084 I^2 = 87.9%**	**0.83 (0.67 to 1.04) p=0.107 I^2 = 25.7%**	**0.90 (0.76 to 1.06) p=0.210 I^2 = 12.4%**

ΔMLHFQ: change in Minnesota Living with Heart Failure; ΔHRQoL: change in health-related quality of life;

ΔVO2peak: change in peak oxygen uptake; Δ6MWT: change in 6-minute walk test

Assessment of study quality and risk of bias

The overall quality of included trials was judged to be moderate to good, with a median TESTEX [31] score of 11 (range 10 to 14) out of a maximum score of 15 (see Table 22**Error! Reference source not found.**).

Table 22: Assessment of quality using TESTEX scale of included studies in surrogate analyses

Study (publication year)	Eligibility Criteria Specified	Randomisation Specified	Allocation Concealed	Groups Similar at baseline	Blinding of Assessors	Outcome measures in >85% participants [1]	Intention to treat analysis [2]	Between-group statistical comparisons reported†	Point measures & measures of variability reported	Activity monitoring in control group	Relative Exercise intensity reviewed	Exercise Volume and Energy Expended	Overall TESTEX score (maximum score 15)
Dracup (2007)	1	0	0	1	0	3	1	2	1	1	1	1	10
Hambrecht (2000)	1	1	0	1	0	3	0	2	1	0	1	1	11
HF-ACTION (2008)	1	1	1	1	1	3	1	2	1	1	0	1	14
Jolly (2009)	1	1	1	1	0	2	1	2	1	0	1	1	12
Mueller (2007)	1	0	0	1	0	2	1	2	1	0	1	1	10
Nilsson (2008)	1	1	0	1	1	2	1	2	1	0	0	1	11
Passino (2006)	1	0	0	1	0	2	1	2	1	0	1	1	10
Witham (2005)	1	1	0	1	1	3	1	2	1	0	1	0	12
Witham (2012)	1	1	0	1	1	3	1	2	1	0	1	0	12
Yeh (2011)	1	1	0	1	1	3	1	2	1	1	0	0	12

[1] Three points possible; [2] If ITT was not specifically mentioned, but it was noted that no participants withdrew and all analysed

Findings

Mediation analysis

The four criteria which must be satisfied in order to establish that change in exercise capacity is a mediator of mortality, hospitalisation and change in HRQoL are listed in Table 23**Error! Reference source not found.**. First, mean improvements were seen in all exercise capacity metrics with ExCR compared to control, although none reached statistical significance at $p<0.05$. Second, greater differences in exercise capacity significantly reduced the risk of mortality and hospitalisation and was associated with a larger gain in HRQoL. Third, although ExCR decreased both the risk of mortality and hospitalisation and was also associated with a larger gain in HRQoL, there was no statistically significant difference compared to control. Finally, the effect of ExCR versus control on final patient-relevant outcomes was attenuated by adding Δ6MWT and ΔVO_2peak (directly and indirectly measured) into the model. No attenuation was seen with the addition of ΔVO_2peak when measured directly.

Meta-analytic regression: R^2 and surrogate threshold effect

Regression coefficients of determination (R^2) and correlation coefficients (ρ) between the change in exercise capacity and hospitalisation were poor (R^2_{trial} <50% and $\rho<0.50$). Moderate to good levels of correlation (R^2_{trial} >50% and $\rho>0.50$) between exercise capacity VO_2peak and 6MWT with mortality and HRQoL were seen (Table 24**Error! Reference source not found.**). STE for MLHFQ ranged from an increase of 1.6 to 4.6 ml/kg/min for VO_2peak. STE was not estimable for 6MWT. Negative correlation coefficients indicate that larger ExCR effects on exercise capacity are associated with larger ExCR effects on mortality and HRQoL. Figure 13**Error! Reference source not found.**, Figure 14 **Error! Reference source not found.**and Figure 15**Error! Reference source not found.** illustrate the results of our meta-regression and STE analysis.

Small study bias

There was no evidence of significant small study bias as shown by the funnel plots (see Figure 16**Error! Reference source not found.**) or Egger's test p-values for any of the exercise capacity outcomes (ΔVO_2peak direct, $p = 0.699$; Δ6MWT, $p = 0.93$; ΔVO_2peak

direct and indirect, p = 0.553) or the four patient-relevant final outcomes (ΔMLHFQ, p = 0.607; ΔHRQoL outcomes, p = 0.659; mortality, p = 0.745; hospitalisation, p = 0.733).

Table 23: Criteria to establish change in exercise capacity as a mediator in the relationship between treatment effect on patient-relevant final outcomes

	ΔVO_2 peak direct (ml/kg/min)	Δ6MWT (metres)	ΔVO_2 peak direct and indirect (ml/kg/min)
Criteria 1			
1. Treatment assignment has a significant effect on exercise capacity	0.61 (95%CI, -0.89 to 2.11)	14.61 (95% CI, -6.16 to 35.37)	0.58 (95% CI, -0.35 to 1.51)
Criteria 2			
2. Exercise capacity has a significant effect on ΔMLHFQ	-1.64 (95% CI, -2.57 to -0.71)	-0.06 (95% CI, -0.08 to -0.03)	-1.80 (95% CI, -2.77 to -0.83)
2. Exercise capacity has a significant effect on ΔHRQoL all measures (standard deviation units)	-0.06 (95% CI, -0.08 to -0.04)	-0.002 (95% CI, -0.003 to -0.001)	-0.07 (95% CI, -0.08 to -0.05)
2. Exercise capacity has a significant effect on all-cause mortality HR	0.88 (95% CI, 0.84 to 0.92)	0.997 (95% CI, 0.995 to 0.998)	0.88 (95% CI, 0.84 to 0.92)
2. Exercise capacity has a significant effect on all-cause hospital admission HR	0.93 (95% CI, 0.91 to 0.96)	0.998 (95% CI, 0.997 to 0.999)	0.94 (95% CI, 0.92 to 0.96)
Criteria 3			
3. Treatment assignment has a significant effect on patient-relevant final outcomes: ΔMLHFQ : -5.84 (95%CI, -11.96 to 0.77) ΔHRQoL: all outcomes (standard deviation units):-0.22 (95% CI, -0.38 to -0.07)			

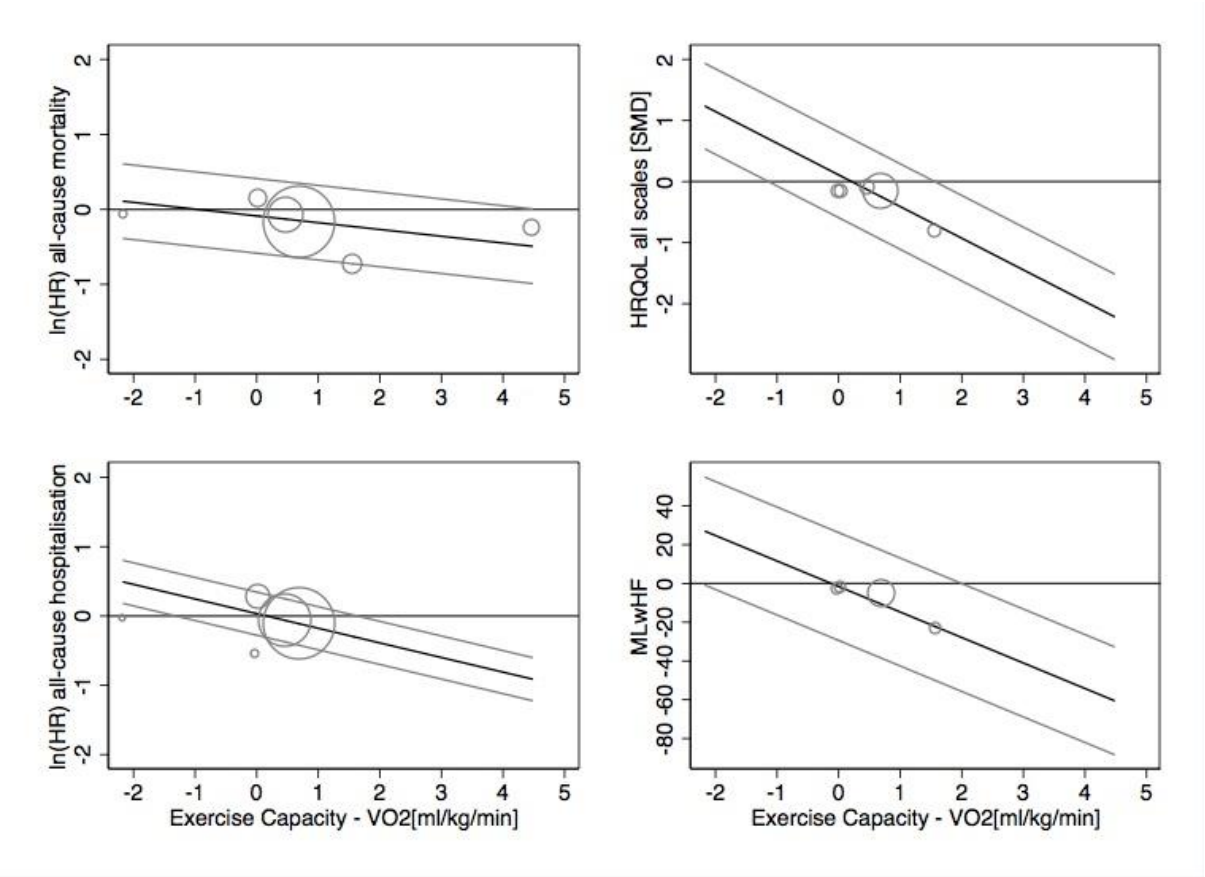
All-cause mortality HR: 0.85 (95% CI, 0.73 to 0.99) All-cause hospital admission HR: 0.91 (95% CI, 0.83 to 1.00)			
Criteria 4*			
4. The effect of treatment assignment on ΔMLHFQ is attenuated when the change in exercise capacity is added to the model	-8.28 (95% CI, -18.56 to 2.01)	-1.77 (95% CI, -4.76 to 1.23)	-4.70 (95% CI, -10.81 to 1.40)
4. The effect of treatment assignment on ΔHRQoL all outcomes is attenuated when the change in exercise capacity is added to the model	-0.28 (95% CI, -0.56 to -0.01)	-0.05 (95% CI, -0.12 to 0.01)	-0.17 (95% CI, -0.31 to -0.02)
4. The effect of treatment assignment on all-cause mortality HR is attenuated when the change in exercise capacity is added to the model	0.99 (95% CI, 0.79 to 1.24)	1.00 (95% CI, 0.81 to 1.24)	1.01 (95% CI, 0.83 to 1.22)
4. The effect of treatment assignment on all-cause hospital admission HR is attenuated when the change exercise capacity is added to the model	0.93 (95% CI, 0.82 to 1.04)	0.97 (95% CI, 0.86 to 1.09)	0.95 (95% CI, 0.85 to 1.06)

HR: hazard ratio; ΔMLHFQ: change in Minnesota Living with Heart Failure Questionnaire; ΔHRQoL: change in health-related quality of life; ΔVO_2 peak: change in peak oxygen uptake; Δ6MWT: change in 6-minute walk test; * mediator-adjusted coefficient

Table 24: Surrogacy metrics for change in exercise capacity and final outcomes

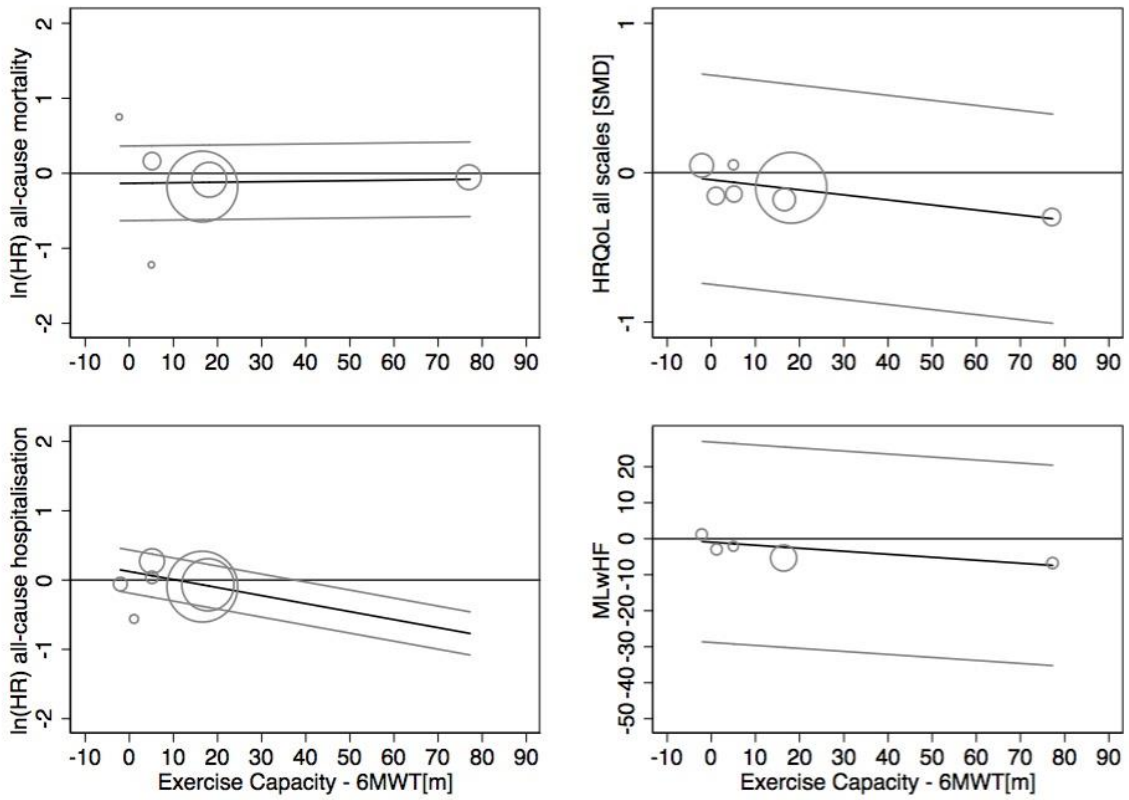
	ΔVO2 peak direct (ml/kg/min)	**Δ6MWT (metres)**	**ΔVO2 peak direct and indirect (ml/kg/min)**
ΔMLHFQ	R^2_{trial} 94% ρ -0.80 STE 2 ml/kg/min	R^2_{trial} 65% ρ -0.90** STE not estimable	R^2_{trial} 54% ρ -0.64 STE 3.2 ml/kg/min
ΔHRQoL all outcomes (standard deviation units)	R^2_{trial} 81% ρ -0.60 STE 1.6 ml/kg/min	R^2_{trial} 54% ρ -0.57 STE not estimable	R^2_{trial} 62% ρ -0.53 STE 2 ml/kg/min
All-cause mortality HR	R^2_{trial} 21% ρ -0.89** STE 4.6 ml/kg/min	R^2_{trial} 1% ρ -0.20 STE not estimable	R^2_{trial} 7% ρ -0.31 STE not estimable
All-cause hospital admission HR	R^2_{trial} 26% ρ -0.20 STE 1.8 ml/kg/min	R^2_{trial} 9% ρ -0.03 STE 38 m	R^2_{trial} 14% ρ -0.21 STE 1.8 ml/kg/min

HR: hazard ratio; ΔMLHFQ: change in Minnesota Living with Heart Failure Questionnaire; ΔHRQoL: change in health-related quality of life; ΔVO2peak: change in peak oxygen uptake; Δ6MWT: change in 6-minute walk test; STE: surrogate threshold effect



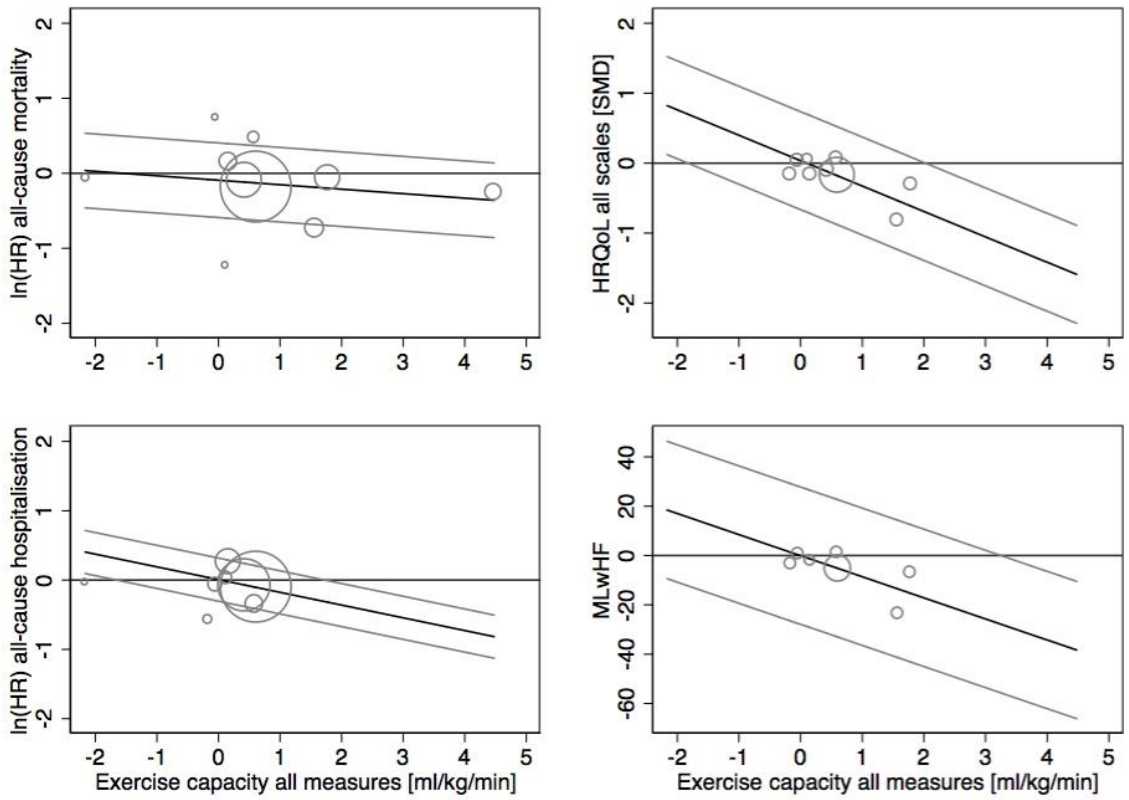
Circles represent trial-level treatment effects, with sizes proportionate to study weights (based on inverse variance weighting). Solid grey lines correspond to the bounds of the 95% prediction interval for the true effect on the final outcomes in a new study.
HR = hazard ratio; HRQoL = health-related quality of life; MLHFQ = Minnesota Living with Heart Failure Questionnaire; SMD = standardised mean difference.

Figure 13: Regression analyses showing the relationship between ΔVO2peak direct at 6 month follow-up versus log(HR) of all-cause mortality and all-cause hospitalisation, and ΔVO2peak direct versus ΔMLHFQ and ΔHRQoL all outcomes



Circles represent trial-level treatment effects, with sizes proportionate to study weights (based on inverse variance weighting). Solid grey lines correspond to the bounds of the 95% prediction interval for the true effect on the final outcomes in a new study.
HR = hazard ratio; HRQoL = health-related quality of life; MLHFQ = Minnesota Living with Heart Failure Questionnaire; SMD = standardised mean difference.

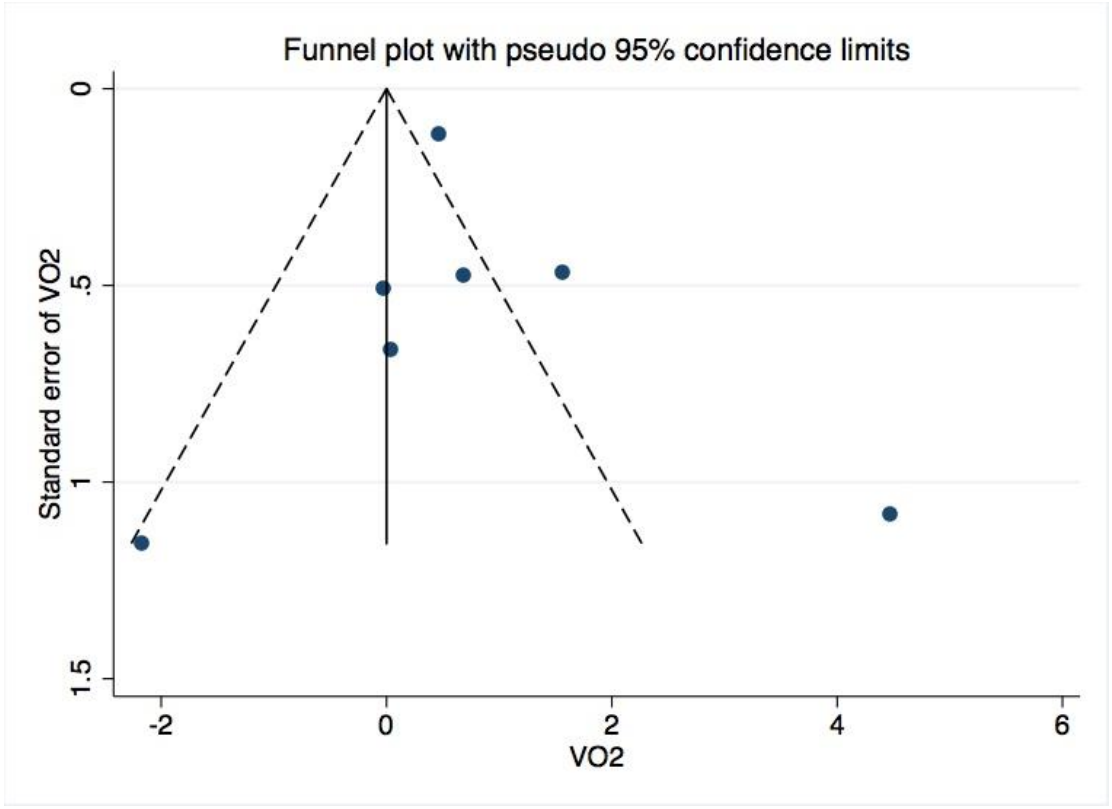
Figure 14: Regression analyses showing the relationship between Δ6MWT at 6 months follow-up versus log(HR) of all-cause mortality and all-cause hospitalisation, and Δ6MWT versus ΔMLHFQ and ΔHRQoL all outcomes



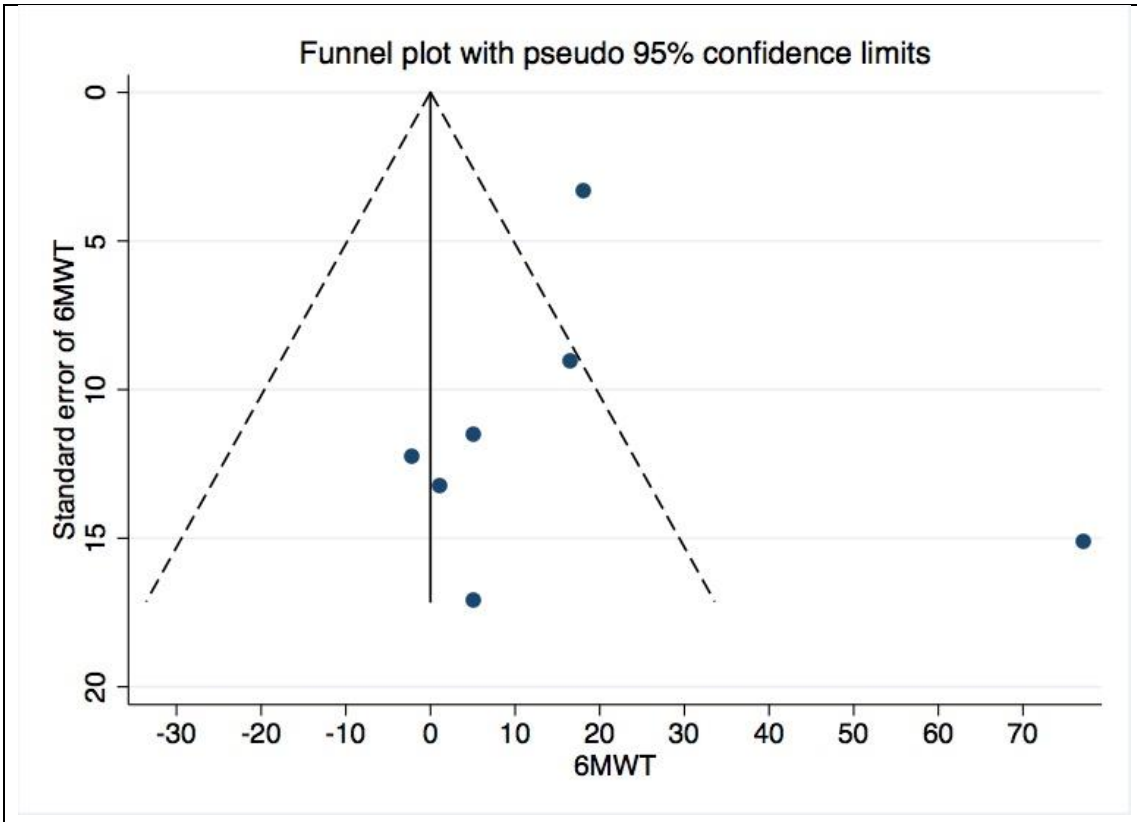
Circles represent trial-level treatment effects, with sizes proportionate to study weights (based on inverse variance weighting). Solid grey lines correspond to the bounds of the 95% prediction interval for the true effect on the final outcomes in a new study.

HR = hazard ratio; HRQoL = health-related quality of life; MLHFQ = Minnesota Living with Heart Failure Questionnaire; SMD = standardised mean difference.

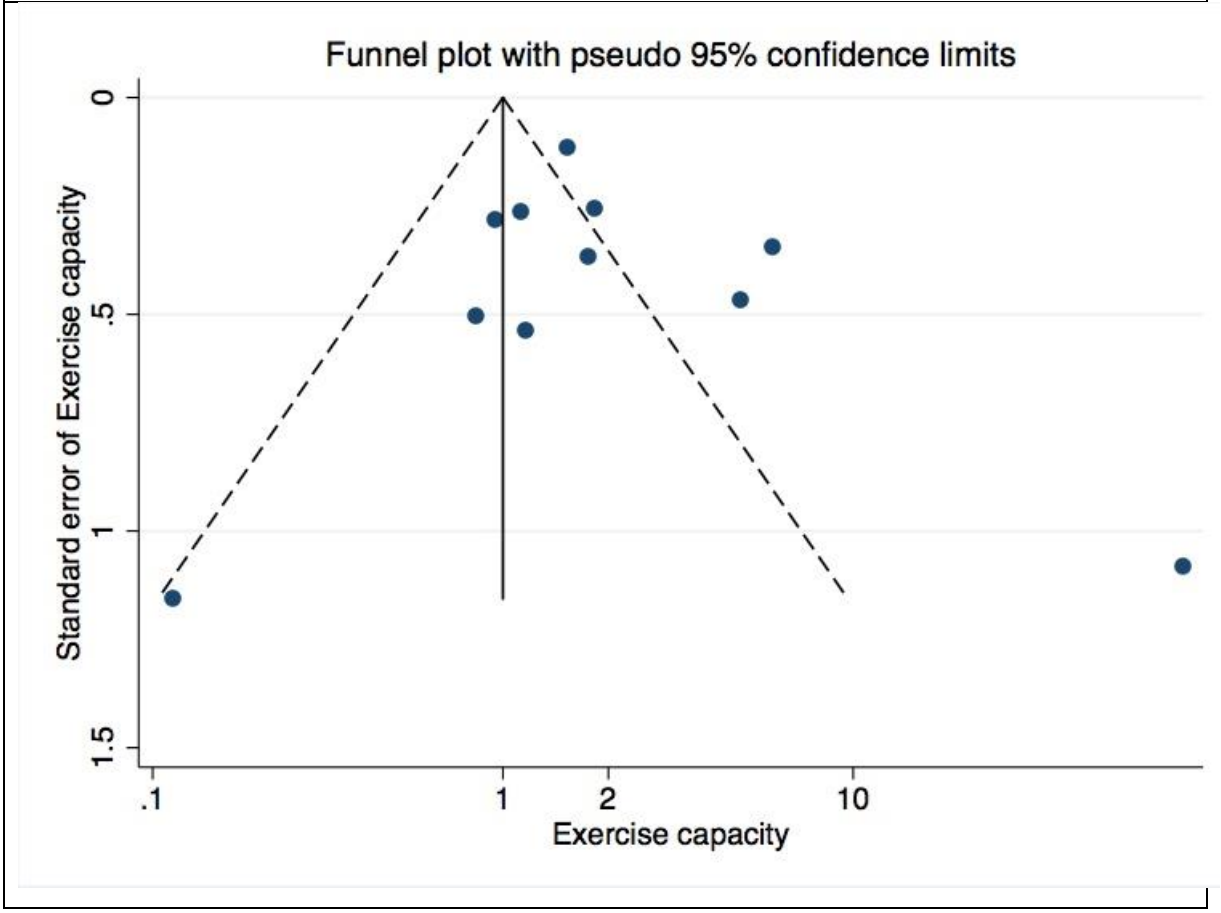
Figure 15: Regression analyses showing the relationship between ΔVO2peak direct and indirect versus log(HR) of all-cause mortality and all-cause hospitalisation, and ΔVO2peak direct and indirect versus ΔMLHFQ and ΔHRQoL all outcomes



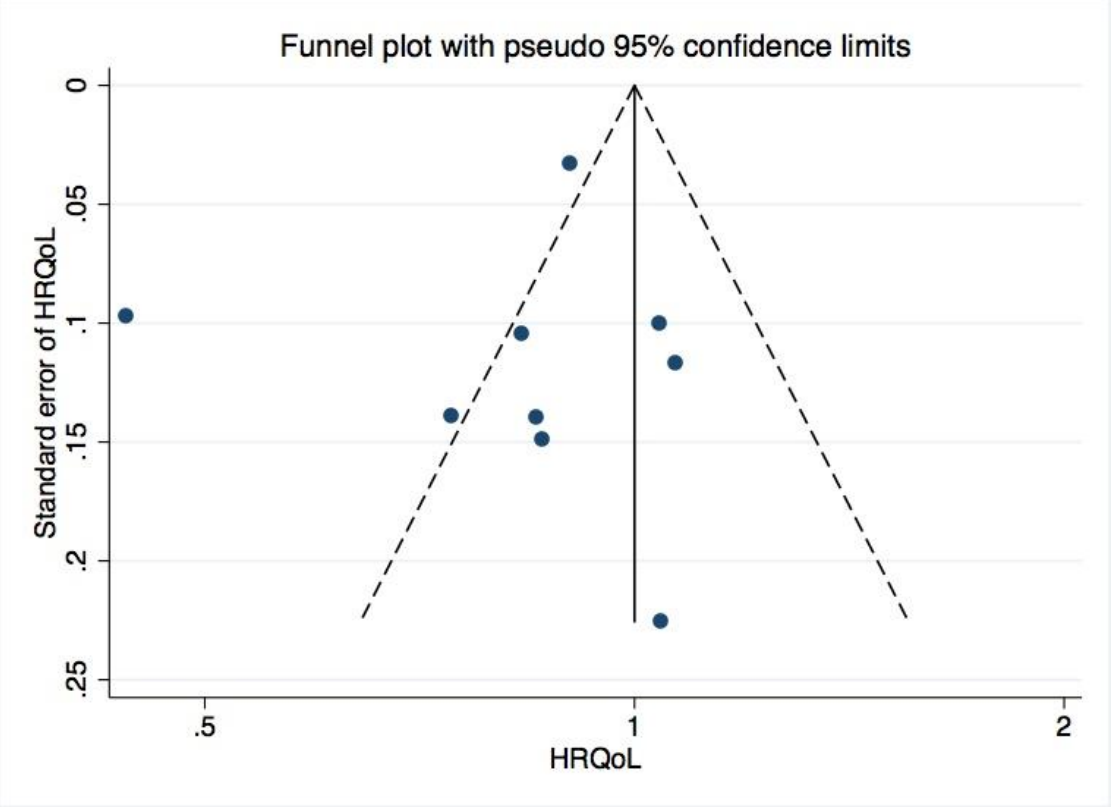
a. VO_2peak


Funnel plot with pseudo 95% confidence limits
Standard error of 6MWT
0
5
10
15
20
-30
-20
-10
0
10
20
30
40
50
60
70
6MWT

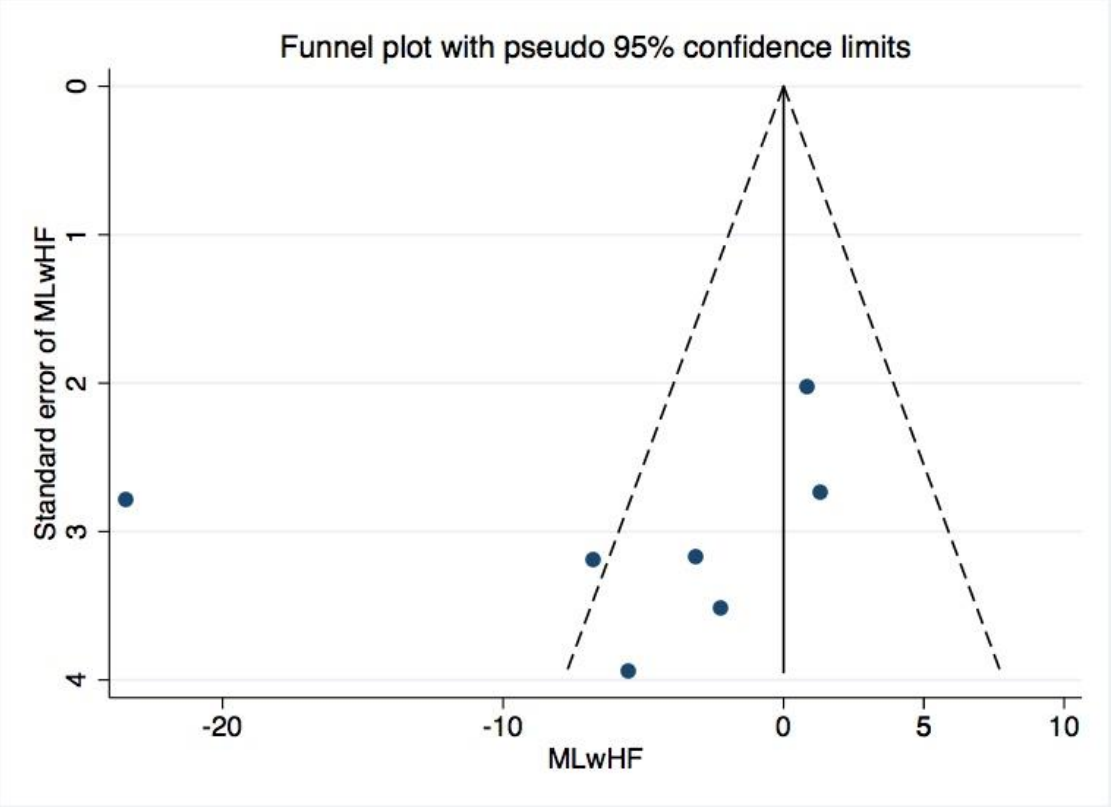
b. 6MWT


Funnel plot with pseudo 95% confidence limits
Standard error of Exercise capacity
0
.5
1
1.5
.1
1
2
10
Exercise capacity

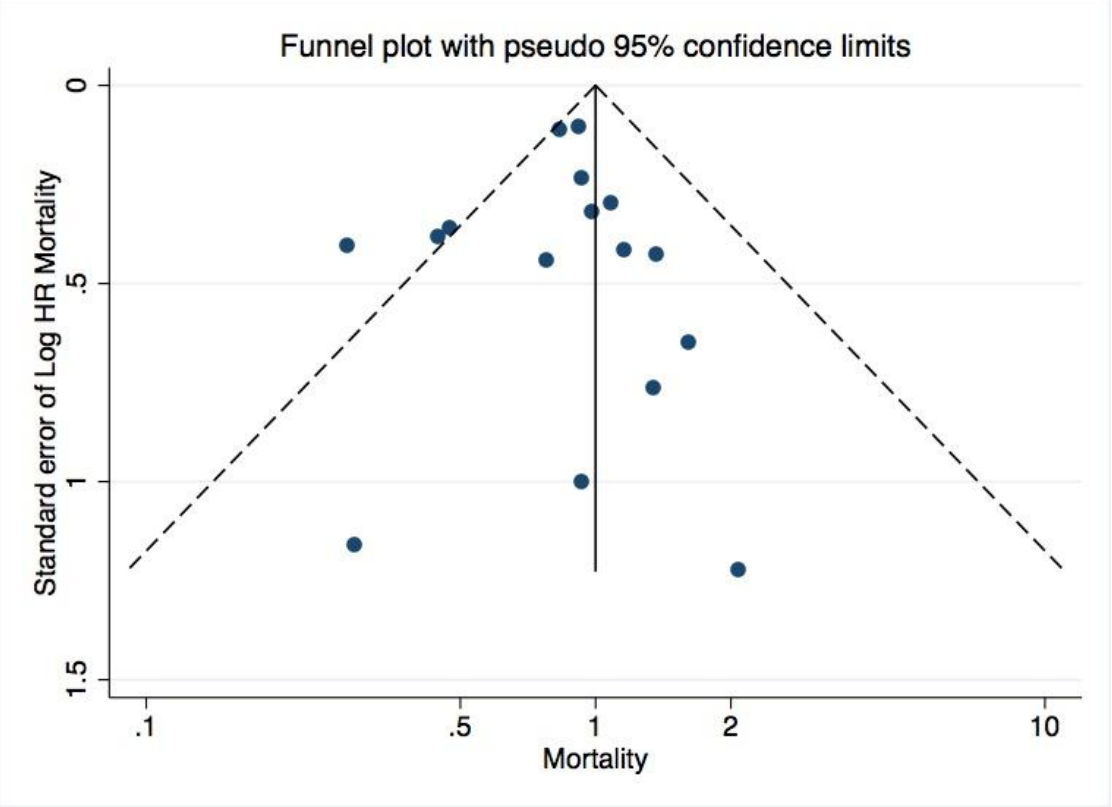
c. Converted exercise capacity score



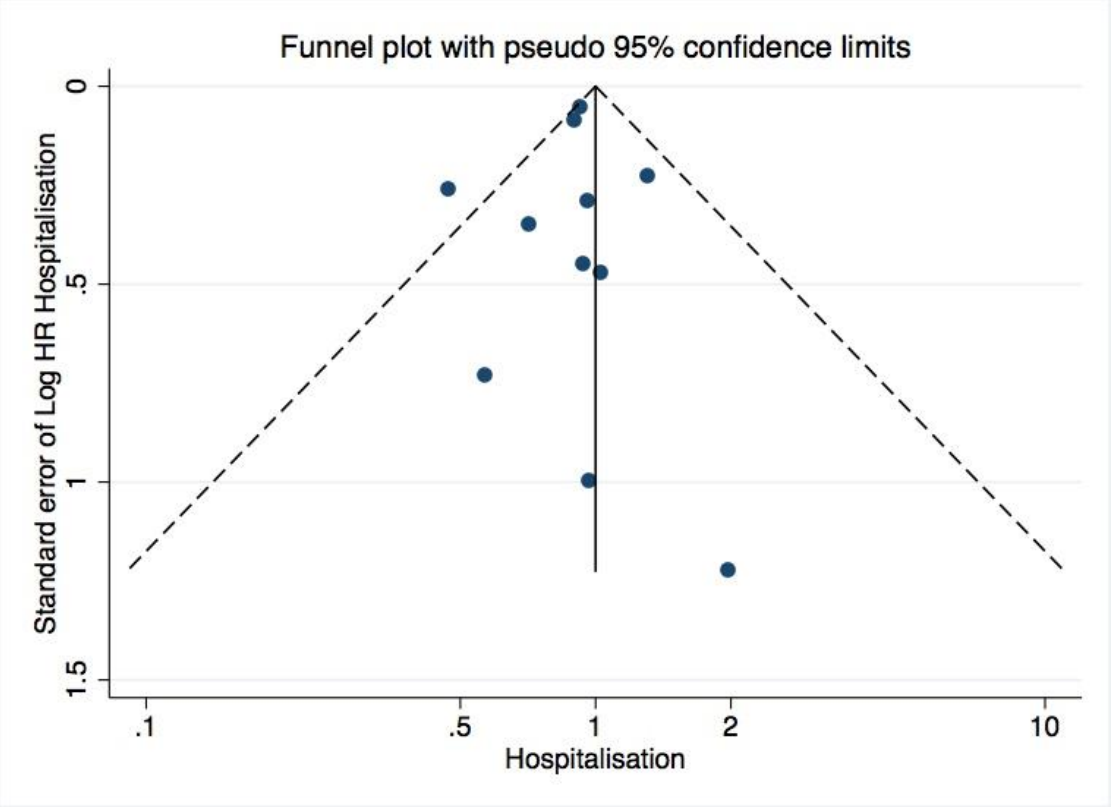
d. HRQoL



e. MLHFQ



f. Mortality



g. Hospitalisation

Figure 16: Funnel Plots for surrogate analysis

Chapter 8: Discussion

The ExTraMATCH II project is a meta-analysis of IPD from HF patients recruited to RCTs conducted worldwide that sought to determine which HF patient subgroups benefit most from ExCR and assess the suitability of exercise capacity as surrogate endpoint.

Summary of findings

Of the 37 eligible trials, 19 contributed data to the IPD meta-analysis – 18 trials (3912 patients) to the clinical events (mortality and hospitalisation) analysis, 13 trials (3332 patients) to exercise capacity and HRQoL analysis, and 10 trials (2,656 patients) to the exercise capacity mediational/surrogate endpoint analysis.

Patient characteristics at baseline were well balanced between ExCR and control patients. The majority of patients were male (74%), with a mean age of 61 years and with HFrEF (mean left-ventricular ejection fraction: 26.9%), and in NYHA functional class II (59%) or III (38%). Trials from Europe and North America were published between 1990 and 2012. Sample size ranged from 50 to 2130 patients. All trials evaluated an aerobic exercise intervention, which was most commonly delivered in either an exclusively centre-based setting or a centre-based setting in combination with some home exercise sessions. The dose of exercise training ranged widely across trials. ExCR was delivered over a period of 15 to 90 weeks, with between 2 and 7 sessions per week; median session duration was between 4 and 120 minutes (including warm-up and cool-down). The intensity of exercise ranged between 50 to 85% peak VO_2. The overall quality of included trials was judged to be moderate to good, with a median TESTEX score of 11 (range 9 to 14) out of a maximum score of 15. Compared to no exercise control, ExCR did not have a statistically significant effect on the risk of mortality and hospitalisation. However, uncertainty around effect estimates precludes drawing definitive conclusions for these event outcomes. In contrast, ExCR was found to significantly improve both exercise capacity and HRQoL, the improvement in MLHFQ being also clinically important (i.e. a mean reduction ≥5 points). (74) We found no consistent differences in ExCR effects across patient subgroups (age, sex, ethnicity, NYHA functional class, ischaemic aetiology, ejection fraction, and baseline exercise capacity) across of mortality, hospitalisation, exercise capacity or HRQoL. Our validation of exercise capacity as a putative surrogate endpoint for patient-relevant outcomes (mortality, hospitalisation, and

HRQoL) was limited by access to a small number of trials that were able to contribute suitable patient level data. Although, subject to considerably statistical uncertainty, our results provide indicative evidence that VO_2peak and 6MWT may be suitable surrogate endpoints for the treatment effect of ExCR on final outcomes in HF.

Comparison to existing evidence

Our finding of a lack of consistent evidence for HF patient subgroup effects of ExCR agrees with both the previous ExTraMATCH and Cochrane 2014 analyses. [10, 18] However, these two previous studies had major limitations that are likely to have limited their ability to detect subgroup effects. ExTraMATCH included data on only 801 HF patients and observed 88 deaths and 300 patients with a composite outcome of death or hospitalisation, and therefore lacked statistical power. Using meta-regression analysis, the 2014 Cochrane review found no association between trial level patient characteristics and ExCR. However, meta-regression analysis is highly prone to study level confounding (ecological fallacy) and should be interpreted with great caution. [75]

Our findings are also consistent with the IPD subgroup analyses from the multicentre HF-ACTION trial. The HF-ACTION investigators reported no significant interaction effect of exercise training intervention on their composite primary outcome (all-cause mortality or hospitalisation) and subgroups of age (≤70 vs. > 70 years), gender, race (white vs. non-white), HF aetiology (ischaemic vs. non ischaemic), ejection fraction (≤25% vs. >25%), or NHYA class (II vs. III/IV).[19] A post hoc analysis by HF ACTION investigators, found a significant (adjusted p = 0.02) interaction between ExCR and the change in 6MWT with ExCR and ethnicity (+26 metres in black patients vs +11 metres in white patients), consistent with the current study. [76]

Our validation study of the suitability of exercise as surrogate outcome, albeit uncertain, are broadly in agreement with our recent study based on a trial level meta-analyses. [77]

Strengths and limitations

The ExTraMATCH-II project has a number of strengths. Our IPD meta-analysis is the largest to date and has greater power to detect any differential treatment effect across groups than

single trials or aggregate meta-analysis. We were able to standardise the handling and analysis of time to event outcomes and continuous outcomes across trials. We found no evidence of publication bias. The project was conducted and reported in accordance with current IPD guidance and Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data (PRISMA IPD) statement. [21, 78]

Whilst systematic reviews and meta-analyses of IPD from randomised trials are recognised as the gold standard for assessing intervention effects [79] our study has a number of limitations. First, there was a lack consistency in how included trials with IPD in our analyses defined and collected the outcomes of interest, i.e. time to event for death and hospitalisation, exercise capacity and HRQoL. We made considerable efforts to contact study authors in order to clarify issues around the definition of outcomes, especially HF-related mortality and hospitalisations. Although we were able to resolve data issues in many cases, we recognise that a lack of consistency in outcome definition across included trials may exist, weakening the strength of our conclusions. Second, we were not able to obtain IPD from all includable trials for all outcomes - not all investigator for the trials that met our inclusion criteria were able to provide IPD and of the trials that did provide IPD, not all collected the outcomes of interest. For example, the large National Institutes of Health (NIH) funded US multicentre HF-ACTION trial did not collect HF-specific hospitalisation [19], thus reducing our statistical power for this outcome. Third, we did not seek patient level data on 'ExCR dose', i.e. adherence according to exercise training duration, frequency and intensity undertaken by an individual patient. Using IPD from HF-ACTION, Keteyian et al found exercise volume (defined as metabolic equivalent of task (MET)-hour per week) to be a predictor for the composite outcome of all-cause mortality or hospitalisation ($p=0.03$) [80] Fourth, there were high levels of statistical heterogeneity for both exercise capacity and HRQoL outcomes. This heterogeneity may well have reflected the variation in ExCR interventions across the included trials. Fifth, our analysis is based on randomised trials identified by literature searches up to 2013 and therefore did not include IPD from more recent trials that may have met the inclusion criteria of this study.

Finally, in terms of our surrogate validation analysis a particular limitation was the proportion of included trials that provided patient level data on both exercise capacity and patient-relevant outcomes. Of the 19 trials (3,990 patients) that met our inclusion criteria, only a

maximum of 10 trials (2656 patients) provided paired data on exercise capacity and either mortality, hospitalisation, or HRQoL. This has a number of implications for the interpretation of our findings: (1) the statistical power of our analysis was low; evidenced by the wide confidence intervals in pooled analysis and whilst all outcomes were in direction of benefit of ExCR, none reached a level of formal statistical significance at 5% level, (2) and relatedly, we had limited statistical power to detect an association between changes in exercise capacity and the final patient-related outcomes, and (3) the results are likely to be subject to selection bias and therefore may not be representative of all RCT evidence.

Relevance to clinical practice

The observed improvements in patient exercise capacity and HRQoL with ExCR participation support the Class I recommendation of current international clinical guidelines that ExCR should be offered to HF patients. [3, 13, 15] Our findings do not endorse limiting ExCR interventions to subgroups of HF patients.

Research recommendations

1. In spite of the comprehensiveness of this IPD meta-analysis, findings of this study demonstrate that further evidence is still required to definitively assess the impact of ExCR on mortality and hospitalisation in patients with reduced ejection fraction HF; in particular, to increase the power to examine whether the effect of ExCR varies according to patient characteristics. To more reliably quantify the impact of ExCR on clinical outcomes and examine how these effects may vary across HF patients, there is an urgent need for trial investigators to more consistently collect, report, and share patient-level data in the future.

   Two central aspects of future data collection include a consensus on the definition, collection, and reporting of clinical event data, especially hospitalisation, plus the capture of data on patient level adherence to the amount of exercise training during the ExCR intervention period. More generally, the research community should continue to implement policies that encourage primary study authors to make their datasets available, either by depositing in publicly available repositories or shared with IPD meta-analysis collaborations when directly requested. [81]

2. Given the vast majority of IPD in this study was from HFrEF patients, future trials including HFpEF patients are needed to assess the effectiveness of ExCR and whether there are differential effects of ExCR in this patient group.
3. Future IPD meta-analyses of RCTs for interventions in HF are needed to confirm the tentative conclusion that VO_2peak and 6MWT may be suitable surrogate endpoints for the final patient-related outcomes. Such future IPD meta-analyses also need to consider individual patient adherence to exercise training.

MLHFQMLHFQMLHFQMLHFQMLHFQMLHFQ

Acknowledgements

We thank Mr Tim Eames, Exeter Clinical Trials Unit, Exeter, UK for his advice and support on data management for this study. We also thank the investigator teams who provided individual patient data: Professor K Dracup, University of California San Francisco, School of Nursing, USA; Prof LS Evangelista, University of California Irvine, California, USA; Prof K Jolly, Institute of Applied Health Research, University of Birmingham, UK; Prof RS McKelvie, St. Joseph's Health Care Centre, London, Ontario, Canada; Prof J Myers, Palo Alto Health Care System/Stanford University Division of Medicine, USA; Associate Prof BB Nilsson, Oslo University Hospital and Faculty of Health Sciences, Oslo Metropolitan University, Norway; Prof C O'Connor, Duke Clinical Research Institute, North Carolina; Prof C Passino, Fondazione G. Monasterio and Scuola Superiore Sant'Anna, Pisa, Italy; Dr MD Witham, School of Medicine, University of Dundee, Dundee, UK; Associate Prof GY Yeh, Beth Israel Deaconess Medical Center, Department of Medicine, Boston, Massachusetts, USA; Prof AD Zwisler, National Knowledge Centre for Rehabilitation and Palliative Care, University Hospital Odense and University of Southern Denmark, Denmark.

Our thanks to the International Steering Committee (see Appendix 1**Error! Reference source not found.**). We acknowledge the contribution of patient data from Dr Rebecca Gary, Dr Rainer Hambrecht, and the late Dr Romualdo Belardinelli, and the late Dr Pantaleo Giannuzzi.

We extend our thanks to Kevin Paul and the PPI group of the NIHR Programme Grant (Rehabilitation Enablement in Chronic Heart Failure - REACH-HF, PGfAR RP-PG-0611-12004) who provided patient involvement in this project and also reviewed the Plain English summary of this report.

We are conscious that this research would not have been possible the data of the heart failure patients included in the trials and we are indebted to them.

Contributions of authors

Rod Taylor (Professor of Health Services Researcher) was project chief investigator with overall responsibility for the project and guarantor, designed the project, obtained the data,

contributed to the design of the data analysis, and led the report write up and edited the final report

Sarah Walker (Research Fellow in Medical Statistics) performed the data checking and cleaning process undertook the statistical analysis for the impact of ExCR on clinical events and exercise capacity/HRQoL outcomes, and contributed to drafting of the paper

Oriana Ciani (Postdoctoral Research Fellow) designed the project, undertook the statistical analysis for the mediation and surrogate validation of exercise capacity and contributed to drafting of the paper

Fiona Warren (Senior Lecturer in Medical Statistics) designed the project, advised on the statistical analysis, and contributed to the report write up

Neil Smart (Professor in Exercise and Sports Science) designed the project, provided advice and content-specific expertise, contributed to the data checking and cleaning and the report write up

Massimo Piepoli (Professor of Cardiology) designed the project, provided advice and content-specific expertise, contributed to obtaining the data, and contributed to the report write up

Costantinos Davos (Associate Professor Cardiovascular Diseases) contributed to obtaining the data, provided advice and content-specific expertise, and contributed to the report write up

Publications

Taylor RS, Piepoli MF, Smart N, Coats AJS, Ellis S, Dalal H, O'Connor CM, Warren FC, Whellan D, Ciani O, and ExTraMATCH II Collaborators Exercise training for chronic heart failure (ExTraMATCH II): Protocol for an individual participant data meta-analysis. *Int J Cardiol.* 2014;174:683-7

Taylor RS, Walker S, Smart NA, Piepoli MF, Warren FC, et al. Impact of exercise-based rehabilitation in patients with heart failure (ExTraMATCH II) on mortality and hospitalisation: an individual-patient data meta-analysis of randomised trials – manuscript submitted for publication. *Eur J Heart Failure* 2018;20:1735-43

Taylor SR, Walker S, Smart NA, Piepoli MF, Warren FC, et al. Impact of exercise-based rehabilitation in patients with heart failure (ExTraMATCH II) on exercise capacity and health-related quality of life: a systematic review and meta-analysis of individual participant data from randomised trials. *J Am Coll Cardiol.* 2019 – accepted for publication.

Ciani O, Taylor RS, Walker S, Warren FC, et al. Validation of Exercise capacity as a surrogate endpoint in exercise-based rehabilitation for heart failure: an individual patient data meta-analysis of randomised controlled trials. Manuscript in preparation.

Data sharing statement

All data requests should be submitted to the corresponding author for consideration. Access to anonymised data may be granted following review.

References

1. Braunwald E. The war against heart failure: the Lancet lecture. The Lancet. 2015;385(9970):812-24.
2. Ponikowski P, Anker SD, AlHabib KF, Cowie MR, Force TL, Hu S, et al. Heart failure: preventing disease and death worldwide. ESC Heart Failure. 2014;1(1):4-25.
3. National Institute for Health and Care Excellence. Chronic heart failure in adults: management. Clinical guideline [CG108] 2010 (Last accessed: 25th June 2018). Available from: https://www.nice.org.uk/guidance/CG108.
4. McKelvie RS. Heart failure. BMJ Clin Evid. 2010;0204.
5. Editorial (no authors listed). On the horizon in heart failure. Lancet. 2011;378.
6. Office of National Statistics. National Population Projections: 2016-based statistical bulletin2017 (Last accesses: 16th January 2019). Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/nationalpopulationprojections/2016basedstatisticalbulletin.
7. Cowie MR, Wood DA, Coats AJ, Thompson SG, Poole-Wilson PA, Suresh V, et al. Incidence and aetiology of heart failure; a population-based study. Eur Heart J. 1999;20(6):421-8.
8. Braunstein JB, Anderson GF, Gerstenblith G, Weller W, Niefeld M, Herbert R, et al. Noncardiac comorbidity increases preventable hospitalizations and mortality among medicare beneficiaries with chronic heart failure. J Am Coll Cardiol. 2003;42(7):1226-33.
9. Lewinter C, Doherty P, Gale CP, Crouch S, Stirk L, Lewin RJ, et al. Exercise-based cardiac rehabilitation in patients with heart failure: a meta-analysis of randomised controlled trials between 1999 and 2013. Eur J Prev Cardiol.. 2015;22(12):1504-12.
10. Taylor RS, Sagar VA, Davies EJ, Briscoe S, Coats AJS, Dalal H, et al. Exercise-based rehabilitation for heart failure. Cochrane Database Syst Rev. 2014(4).
11. Haykowsky MJ, Timmons MP, Kruger C, McNeely M, Taylor DA, Clark AM. Meta-analysis of aerobic interval training on exercise capacity and systolic function in patients with heart failure and reduced ejection fractions. Am J Cardiol. 2013;111(10):1466-9.
12. Taylor RS, Davies EJ, Dalal HM, Davis R, Doherty P, Cooper C, et al. Effects of exercise training for heart failure with preserved ejection fraction: a systematic review and meta-analysis of comparative studies. Int J Cardiol. 2012;162(1):6-13.
13. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Jr., Drazner MH, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. Circulation. 2013;128(16):e240-327.
14. Working Group on Cardiac Rehabilitation & Exercise Physiology and Working Group on Heart Failure of the European Society of Cardiology. Recommendations for exercise training in chronic heart failure patients. Eur Heart J. 2001;22(2):125-35.
15. McMurray JJV, Adamopoulos S, Anker SD, Auricchio A, Böhm M, Dickstein K, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012. The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. Eur Heart J. 2012;33(14):1787-847.
16. Dalal HM, Wingham J, Palmer J, Taylor R, Petre C, Lewin R. Why do so few patients with heart failure participate in cardiac rehabilitation? A cross-sectional survey from England, Wales and Northern Ireland. BMJ Open. 2012;2(2).

17. Piepoli MF, Binno S, Corra U, Seferovic P, Conraads V, Jaarsma T, et al. ExtraHF survey: the first European survey on implementation of exercise training in heart failure patients. Eur J Heart Fail. 2015;17(6):631-8.
18. ExTraMATCH Collaborative. Exercise training meta-analysis of trials in patients with chronic heart failure (ExTraMATCH). BMJ. 2004;328(7433):189.
19. O'Connor CM, Whellan DJ, Lee KL, Keteyian SJ, Cooper LS, Ellis SJ, et al. Efficacy and Safety of Exercise Training in Patients With Chronic Heart Failure: HF-ACTION Randomized Controlled Trial. JAMA. 2009;301(14):1439-50.
20. Sagar VA, Davies EJ, Briscoe S, Coats AJS, Dalal HM, Lough F, et al. Exercise-based rehabilitation for heart failure: systematic review and meta-analysis. Open Heart. 2015;2(1).
21. Tierney J, Vale C, Riley R, Tudur Smith C, Stewart L, Clarke M, et al. Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: Guidance on Their Use. PLoS Med. 2015;12(7):e1001855.
22. Stewart LA, Clarke M, Rovers M, et al. Preferred reporting items for a systematic review and meta-analysis of individual participant data: The prisma-ipd statement. JAMA. 2015;313(16):1657-65.
23. Taylor R, Piepoli M, Smart N, Coats A, Ellis S, Dalal H, et al. Exercise training for chronic heart failure (ExTraMATCH II): protocol for an individual participant data meta-analysis PROSPERO [Internet]. 2014 (Last accessed: 23rd October 2017). Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42014007170.
24. Smart N, Marwick TH. Exercise training for patients with heart failure: a systematic review of factors that improve mortality and morbidity. Am J Med. 2004;116(10):693-706.
25. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340.
26. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. BMJ. 2012;344.
27. Taylor RS, Piepoli MF, Smart N, Coats AJS, Ellis S, Dalal H, et al. Exercise training for chronic heart failure (ExTraMATCH II): Protocol for an individual participant data meta-analysis. Int J Cardiol. 2014;174(3):683-7.
28. Cartwright J, Kabir T, Simons L. Budgeting for involvement: Practical advice on budgeting for actively involving the public in research studies 2013 (Last accessed 25th June 2018.) Available from: http://www.invo.org.uk/wp-content/uploads/2014/11/10002-INVOLVE-Budgeting-Tool-Publication-WEB.pdf.
29. Higgins J, Thompson S. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539-58.
30. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-34.
31. Smart NA WM, Ismail H, Giallauria F, Vigorito C, Cornelissen V, Dieberg G. Validation of a new tool for the assessment of study quality and reporting in exercise training studies: TESTEX. Int J Evid Based Healthc . 2015;13(1):9-18.
32. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-60.
33. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 2011 [Available from: http://handbook.cochrane.org/.
34. Cleves M, Gould W, Marchenko Y. An Introduction to Survival Analysis Using Stata. Third ed: Stata Press; 2016.

35. Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Stat Med. 2008;27(11):1870-93.
36. Bilbao A, Escobar A, García-Perez L, Navarro G, Quirós R. The Minnesota Living with Heart Failure Questionnaire: comparison of different factor structures. Health Qual Life Outcomes. 2016;14:23.
37. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. J Am Coll Cardiol. 2000;35(5):1245-55.
38. Guyatt GH, Nogradi S, Halcrow S, Singer J, Sullivan MJJ, Fallen EL. Development and testing of a new measure of health status for clinical trials in heart failure. J Gen Intern Med. 1989;4(2):101-7.
39. Ross RM, Murthy JN, Wollak ID, Jackson AS. The six minute walk test accurately estimates mean peak oxygen uptake. BMC Pulm Med. 2010;10(1):31.
40. Keell SD, Chambers JS, Francis DP, Edwards DF, Stables RH. Shuttle-walk test to assess chronic heart failure. The Lancet.352(9129):705.
41. Fowler SJ, Singh SJ, Revill S. Reproducibility and validity of the incremental shuttle walking test in patients following coronary artery bypass surgery. Physiotherapy.91(1):22-7.
42. Du H, Wonggom P, Tongpeth J, Clark RA. Six-Minute Walk Test for Assessing Physical Functional Capacity in Chronic Heart Failure. Curr Heart Fail Rep. 2017;14(3):158-66.
43. Gabler NB, French B, Strom BL, Palevsky HI, Taichman DB, Kawut SM, et al. Validation of 6-Minute Walk Distance as a Surrogate End Point in Pulmonary Arterial Hypertension Trials. Circulation. 2012;126(3):349-56.
44. Emsley R, Dunn G, White IR. Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. Stat Methods Med Res. 2010;19(3):237-70.
45. VanderWeele TJ. Explanation in causal inference : methods for mediation and interaction. 2015.
46. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat. 2006;5(3):173-86.
47. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points—the challenge of statistical validation. Nature Reviews Clinical Oncology. 2010;7:309.
48. Ciani O, Davis S, Tappenden P, Garside R, Stein K, Cantrell A, et al. Validation of surrogate endpoints in advanced solid tumours: ststematic review of statistical methods, results, and implications for policy makers. Int J Technol Assess Health Care. 2014;30(3):312-24.
49. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ. 2011;342:d549.
50. Belardinelli R, Georgiou D, Cianci G, Purcaro A. Randomized, controlled trial of long-term moderate exercise training in chronic heart failure: effects on functional capacity, quality of life, and clinical outcome. Circulation. 1999;99(9):1173-82.
51. Hambrecht R, Gielen S, Linke A, Fiehn E, Yu J, Walther C, et al. Effects of exercise training on left ventricular function and peripheral resistance in patients with chronic heart failure: A randomized trial. JAMA. 2000;283(23):3095-101.
52. McKelvie RS, Teo KK, Roberts R, McCartney N, Humen D, Montague T, et al. Effects of exercise training in patients with heart failure: the Exercise Rehabilitation Trial (EXERT). Am Heart J. 2002;144(1):23-30.

53. Wielenga RP, Huisveld IA, Bol E, Dunselman PH, Erdman RA, Baselier MR, et al. Safety and effects of physical training in chronic heart failure. Results of the Chronic Heart Failure and Graded Exercise study (CHANGE). Eur Heart J. 1999;20(12):872-9.
54. Willenheimer R, Rydberg E, Cline C, Broms K, Hillberger B, Oberg L, et al. Effects on quality of life, symptoms and daily activity 6 months after termination of an exercise training programme in heart failure patients. Int J Cardiol. 2001;77(1):25-31.
55. Zanelli E, Volterrani M, Scalvini S, Musmeci G, Campana M, Zappa C. Multidisciplinary non-pharmachological intervention prevents hospitalization, improves morbidity rates and functional status in patients with congestive heart failure. Eur Heart J. 1997;18(abstract suppl):647.
56. Belardinelli R, Georgiou D, Cianci G, Purcaro A. 10-year exercise training in chronic heart failure: a randomized controlled trial. J Am Coll Cardiol. 2012;60(16):1521-8.
57. Zwisler AD, Soja AM, Rasmussen S, Frederiksen M, Abedini S, Appel J, et al. Hospital-based comprehensive cardiac rehabilitation versus usual care among patients with congestive heart failure, ischemic heart disease, or high risk of ischemic heart disease: 12-month results of a randomized clinical trial. Am Heart J. 2008;155(6):1106-13.
58. Dracup K, Evangelista LS, Hamilton MA, Erickson V, Hage A, Moriguchi J, et al. Effects of a home-based exercise program on clinical outcomes in heart failure. Am Heart J. 2007;154(5):877-83.
59. Gary RA, Dunbar SB, Higgins MK, Musselman DL, Smith AL. Combined exercise and cognitive behavioral therapy improves outcomes in patients with heart failure. J Psychosom Res. 2010;69(2):119-31.
60. Giannuzzi P, Temporelli PL, Corra U, Tavazzi L. Antiremodeling effect of long-term exercise training in patients with stable chronic heart failure: results of the Exercise in Left Ventricular Dysfunction and Chronic Heart Failure (ELVD-CHF) Trial. Circulation. 2003;108(5):554-9.
61. Jolly K, Taylor RS, Lip GY, Davies M, Davis R, Mant J, et al. A randomized trial of the addition of home-based exercise to specialist heart failure nurse care: the Birmingham Rehabilitation Uptake Maximisation study for patients with Congestive Heart Failure (BRUM-CHF) study. Eur J Heart Fail. 2009;11(2):205-13.
62. Mueller L, Myers J, Kottman W, Oswald U, Boesch C, Arbrol N, et al. Exercise capacity, physical activity patterns and outcomes six years after cardiac rehabilitation in patients with heart failure. Clin Rehabil. 2007;21(10):923-31.
63. Nilsson BB, Westheim A, Risberg MA. Long-term effects of a group-based high-intensity aerobic interval-training program in patients with chronic heart failure. Am J Cardiol. 2008;102(9):1220-4.
64. Witham MD, Gray JM, Argo IS, Johnston DW, Struthers AD, McMurdo ME. Effect of a seated exercise program to improve physical function and health status in frail patients > or = 70 years of age with heart failure. Am J Cardiol. 2005;95(9):1120-4.
65. Witham MD, Fulton RL, Greig CA, Johnston DW, Lang CC, van der Pol M, et al. Efficacy and cost of an exercise program for functionally impaired older patients with heart failure: a randomized controlled trial. Circ Heart Fail. 2012;5(2):209-16.
66. Yeh GY, McCarthy EP, Wayne PM, Stevenson LW, Wood MJ, Forman D, et al. Tai Chi Exercise in Patients With Chronic Heart Failure: A Randomized Clinical Trial. Arch Intern Med. 2011;171(8):750-7.
67. Passino C, Severino S, Poletti R, Piepoli MF, Mammini C, Clerico A, et al. Aerobic Training Decreases B-Type Natriuretic Peptide Expression and Adrenergic Activation in Patients With Heart Failure. J Am Coll Cardiol. 2006;47(9):1835-9.

68. Myers J, Goebbels U, Dzeikan G, Froelicher V, Bremerich J, Mueller P, et al. Exercise training and myocardial remodeling in patients with reduced ventricular function: One-year follow-up with magnetic resonance imaging. Am Heart J. 2000;139(2, Part 1):252-61.
69. Davidson PM, Cockburn J, Newton PJ, Webster JK, Betihavas V, Howes L, et al. Can a heart failure-specific cardiac rehabilitation program decrease hospitalizations and improve outcomes in high-risk patients? Eur J Cardiovasc Prev Rehabil. 2010;17(4):393-402.
70. Austin J, Williams R, Ross L, Moseley L, Hutchison S. Randomised controlled trial of cardiac rehabilitation in elderly patients with heart failure. Eur J Heart Fail. 2005;7(3):411-7.
71. Klecha A, Kawecka-Jaszcz K, Bacior B, Kubinyi A, Pasowicz M, Klimeczek P, et al. Physical training in patients with chronic heart failure of ischemic origin: effect on exercise capacity and left ventricular remodeling. Eur J Cardiovasc Prev Rehabil. 2007;14(1):85-91.
72. Dubach P, Myers J, Dziekan G, Goebbels U, Reinhart W, Muller P, et al. Effect of High Intensity Exercise Training on Central Hemodynamic Responses to Exercise in Men With Reduced Left Ventricular Function. J Am Coll Cardiol. 1997;29(7):1591-8.
73. Wielenga RP, Erdman RAM, Huisveld IA, Bol E, Dunselman PHJM, Baselier MRP, et al. Effect of exercise training on quality of life in patients with chronic heart failure. J Psychosom Res. 1998;45(5):459-64.
74. American Thoracic Society. Minnesota Living with Heart Failure Questionnaire 2004; (Last accessed: 18th June 2018). Available from: http://qol.thoracic.org/sections/instruments/ko/pages/mlwhfq.html.
75. Thompson S, Higgins J. How should meta-regression analyses be undertaken and interpreted? Stat Med. 2002;21(11):1559-73.
76. Mentz RJ, Babyak MA, Bittner V, Fleg JL, Keteyian SJ, Swank AM, et al. Prognostic significance of depression in blacks with heart failure. Insights from heart failure: a controlled trial investigating outcomes of exercise training. Circ Heart Fail. 2015;8(3):497-503.
77. Ciani O, Piepoli M, Smart N, Uddin J, Walker S, Warren F, et al. Validation of exercise capacity as a surrogate endpoint in exercise-based rehabilitation for chronic heart failure: a meta-analysis of randomised controlled trials. JACC HF. 2018;6(7):596-604.
78. Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. Stat Med. 1995;14(19):2057-79.
79. Chalmers I. The Cochrane Collaboration: Preparing, Maintaining, and Disseminating Systematic Reviews of the Effects of Health Care. Ann N Y Acad Sci. 1993;703(1):156-65.
80. Keteyian SJ, Levine AB, Brawner CA, et al. EXercise training in patients with heart failure a randomized, controlled trial. Ann Intern Med. 1996;124(12):1051-7.
81. Veroniki AA, Straus SE, Ashoor H, Stewart LA, Clarke M, Tricco AC. Contacting authors to retrieve individual patient data: study protocol for a randomized controlled trial. Trials. 2016;17(1):138.

Appendices

Appendix 1: Project committees

Project Management Group

Professor Rod S Taylor (Chair), Institute of Health Research, University of Exeter Medical School, Exeter, UK

Dr Oriana Ciani, Institute of Health Research, University of Exeter Medical School, Exeter, UK and Centre for Research on Health and Social Care Management, Bocconi University, Milan, Italy

Associate Professor Constantinos H Davos, Cardiovascular Research Laboratory, Biomedical Research Foundation, Academy of Athens, Athens, Greece

Professor Massimo Piepoli, Cardiology Unit, Guglielmo da Saliceto Hospital, Piacenza, Italy

Professor Neil A Smart, University of New England, Armidale, Australia

Dr Sarah Walker, Institute of Health Research, University of Exeter Medical School, Exeter, UK

Dr Fiona C Warren, Institute of Health Research, University of Exeter Medical School, Exeter, UK

International Steering Committee

All members of the Project Management Group details above, plus:

Professor Andrew Coats (Chair), IRCCS, San Raffaele, Pisana, Italy and University of Warwick, UK

Professor Stephen Ellis, Duke Clinical Research Institute, North Carolina, USA

Associate Professor Hasnain M Dalal, University of Exeter Medical School, Exeter, UK and Research, Development & Innovation, Royal Cornwall Hospital, UK

Professor Steven Keteyian, Department of Medicine, Henry Ford Hospital, Detroit, Michigan, USA

Professor Christopher O'Connor, Duke Clinical Research Institute, North Carolina, USA

Professor David Whellan, Department of Medicine, Sidney Kimmel Medical College, Philadelphia, Pennsylvania, USA

Appendix 2: Example database search strategy from Cochrane 2014 review

MEDLINE(R) Ovid 1946 to January Week 4 2013

1. exp Myocardial Ischemia/
2. (myocard$4 adj5 (ischaemi$2 or ischemi$2)).ti,ab.
3. ((ischaemi$2 or ischemi$2) adj5 heart).ti,ab.
4. exp Coronary Artery Bypass/
5. coronary.ti,ab.
6. exp Coronary Disease/
7. exp Myocardial Revascularization/
8. Myocardial Infarction/
9. (myocard$5 adj5 infarct$5).ti,ab.
10. (heart adj5 infarct$5).ti,ab.
11. exp Angina Pectoris/
12. angina.ti,ab.
13. exp Heart Failure/
14. (heart adj5 failure).ti,ab.
15. (HFNEF or HFPEF or HFREF or "HF NEF" or "HF PEF" or "HF REF").ti,ab.
16. or/1-15
17. exp Heart Diseases/
18. (heart adj5 disease$2).ti,ab.
19. myocard$5.ti,ab.
20. cardiac$2.ti,ab.
21. CABG.ti,ab.
22. PTCA.ti,ab.
23. (stent$4 and (heart or cardiac$4)).ti,ab.
24. Heart Bypass, Left/ or exp Heart Bypass, Right/
25. or/17-24
26. *Rehabilitation Centers/
27. exp Exercise Therapy/
28. *Rehabilitation/
29. exp Sports/
30. Physical Exertion/ or exertion.ti,ab.

31. exp Exercise/
32. rehabilitat$5.ti,ab.
33. (physical$4 adj5 (fit or fitness or train$5 or therap$5 or activit$5)).ti,ab.
34. (train$5 adj5 (strength$3 or aerobic or exercise$4)).ti,ab.
35. ((exercise$4 or fitness) adj5 (treatment or intervent$4 or programs$2 or therapy)).ti,ab.
36. Patient Education as Topic/
37. (patient$2 adj5 educat$4).ti,ab.
38. ((lifestyle or life-style) adj5 (intervent$5 or program$2 or treatment$2)).ti,ab.
39. *Self Care/
40. (self adj5 (manage$5 or care or motivate$5)).ti,ab.
41. *Ambulatory Care/
42. exp Psychotherapy/
43. psychotherap$2.ti,ab.
44. (psycholog$5 adj5 intervent$5).ti,ab.
45. relax$6.ti,ab.
46. exp Relaxation Therapy/ or exp Mind-Body Therapies/
47. exp Counseling/
48. (counselling or counseling).ti,ab.
49. exp Cognitive Therapy/
50. exp Behavior Therapy/
51. ((behavior$4 or behaviour$4) adj5 (modify or modificat$4 or therap$2 or change)).ti,ab.
52. *Stress, Psychological/
53. (stress adj5 management).ti,ab.
54. (cognitive adj5 therap$2).ti,ab.
55. meditat$4.ti,ab.
56. *Meditation/
57. exp Anxiety/
58. (manage$5 adj5 (anxiety or depress$5)).ti,ab.
59. CBT.ti,ab.
60. hypnotherap$5.ti,ab.
61. (goal adj5 setting).ti,ab.

62. (goal$2 adj5 setting).ti,ab.
63. (psycho-educat$5 or psychoeducat$5).ti,ab.
64. (motivat$5 adj5 (intervention or interv$3)).ti,ab.
65. Psychopathology/
66. psychopathol$4.ti,ab.
67. psychosocial$4.ti,ab.
68. distress$4.ti,ab.
69. exp Health Education/
70. (health adj5 education).ti,ab.
71. (heart adj5 manual).ti,ab.
72. Autogenic Training/
73. autogenic$5.ti,ab.
74. or/26-39
75. or/40-73
76. 16 or 25
77. 74 or 75
78. 76 and 77
79. randomized controlled trial/
80. randomized controlled trial.pt.
81. controlled clinical trial.pt.
82. controlled clinical trial/
83. Random Allocation/
84. Double-Blind Method/
85. single-blind method/
86. (random$ or placebo$).ti,ab.
87. ((singl$3 or doubl$3 or tripl$3 or trebl$3) adj5 (blind$3 or mask$3)).ti,ab.
88. exp Research Design/
89. Clinical Trial.pt.
90. exp clinical trial/
91. (clinic$3 adj trial$2).ti,ab.
92. or/79-91
93. 78 and 92

94. (Animals not Humans).sh.

95. 93 not 94

96. limit 95 to yr="2008 -Current"

Appendix 3: Identified randomised controlled trials meeting inclusion criteria

First author (year)	Total patients (N) [1]	Trial setting (single or multi-centre)	NYHA class	Mean ejection fractio(%)	Mean age (years)	Male (%)	Exercise type [2]	Overall exercise duration (minutes)	Exercise frequency (sessions/ week)	Mean programme duration (weeks)	Exercise setting [3]	Longest follow- up (months)
Cochrane 2014 review												
Austin (2005/8)	200	Single	II/III	NR	72	43	Mix	120	2.5	24	Both	60
Belardinelli (1999)	99	Single	II/IV	28	55	89	Aerobic	40	2.5	56	Centre	26
Belardinelli (2012)	123	Single	II/III	37	59	78	Aerobic	40	2.5	56	Centre	120
Davidson (2010)	105	Single	I/II/III/IV	.NR	72.3	67	Mix	40	1	12	Centre	12
Dracup (2007)	173	Single	II/IV	26	54	72	Mix	28	4	52	Home	12
DANREHAB (2008)	91	Single	I/II/III	NR	66	90	Mix	90	3	12	Both	12
Gary (2010)	65	Single	II/III	NR	65.8	42	Aerobic	37.5	3	12	Home	6
Giannuzzi (2003)	90	Multi	II/III	25	60.5	.	Aerobic	30	4	24	Both	6
Hambrecht (2000)	73	Single	I/II/III	29	54	100	Aerobic	15	6.5	24	Both	6
HF-ACTION (2009)	2331	Multi	II/III/IV	25	59	72	Aerobic	30	2.5	120	Both	48
Jolly (2009)	169	Multi	I/II/IV	NR	66	75	Mix	25	5	48	Home	12
Klecha (2007)	50	Single	II/III	28	61	100	Aerobic	20	3	24	Centre	6
McKelvie (2002)	181	Multi	I/II/III	NR	65.5	81	Mix	30	2	36	Both	12
Mueller (2007)	50	Single	NR	NR	55	100	Aerobic	120	5	4	Centre	74
Nilsson (2008)	80	Single	II/III	31	70	79	Aerobic	50	2	16	Centre	12
Passino (2006)	95	Single	I/II/III	34	60.5	87	Aerobic	30	3	36	Home	9
Willenheimer (2000)	54	Single	NR	36.5	64	71.5	Aerobic	30	2.5	16	Centre	10
Witham (2005)	82	Single	II/III	NR	80.5	55	Mix	20	2.5	24	Both	6
Witham (2012)	107	Single	II/III	NR	81	100	Mix	60	2	24	Both	6
Yeh (2011)	100	Multi	I/II/III	29	67.5	64	Aerobic	30	2.5	12	Both	6
ExTraMATCH I (2004)												
Dubach (1997)/ Meyers (2002)	51	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	8.5
Zanelli (1997)	155	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	10
Wielenga (1999)	80	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	47.3

[1] Total number of patients randomised; [2] 'Mix' includes aerobic and resistance training; [3] Whether exercise setting is home or centre or both; NR: not reported in either Cochrane (2014) or ExTraMATCH I (2004) reports; NHYA: New York Heart Association

Table reproduced from ExTraMATCH II Study Protocol: Taylor RS, Piepoli MF, Smart N, Coats AJS, Ellis S, Dalal H, O'Connor CM, Warren FC, Whellan D, Ciani O, and ExTraMATCH II Collaborators Exercise training for chronic heart failure (ExTraMATCH II): Protocol for an individual participant data meta-analysis. Int J Cardiol. 2014;174:683-7

Appendix 4: ExTraMATCH II core data fields

Variable	Description
Study level data	
Centre ID	Centre name
Randomised control patients (N)	
Randomised exercise patients (N)	
Patient level data – descriptive	
Patient ID	
Date of randomisation	dd/mm/yyyy
Allocated treatment	1 Exercise 2 Control
Date of birth	dd/mm/yyyy
Gender	1 Male 2 Female 9 Data unavailable
Race	1 White/Caucasian 2 African/African-American 3 Asian 4 Other 9 Data unavailable
Aetiology of heart failure	1 Ischaemic heart disease 2 Idiopathic dilated cardiomyopathy 3 Other/Unknown 9 Data unavailable
Year of heart failure diagnosis	yyyy
New York Heart Association class at entry/baseline	1 NYHA Class I 2 NYHA Class II 3 NYHA Class III 4 NYHA Class IV 9 Unknown/Unavailable
Ejection fraction at entry/baseline (%)	

Patient level data - Outcomes	
Method of exercise capacity assessment	1 6-minute walk test 2 Bicycle ergometer test 3 Treadmill test 4 Other [state]
Exercise capacity score at entry (units)	
Follow-up 1 exercise capacity score	Follow-up time (months)
Follow-up 2 exercise capacity score	Follow up time (months)
Follow-up 3 exercise capacity score	Follow up time([months)
Health related quality of life	1 Minnesota Living With Heart Failure 2 Other measure (state)
HRQoL at entry	Total & subscores
Follow-up 1 HRQoL score	Total & subscores Follow up time (months)
Follow-up 2 HRQoL score	Total & subscores Follow up time (months)
Follow-up 3 HRQoL score	Total & subscores Follow up time (months)
Date of death	dd/mm/yyyy
Cause of death	1 Acute myocardial infarction 2 Sudden death 3 Heart failure 4 Other cardiac 5 Stroke 6 Other vascular/thrombo-embolic 7 Non-cardiovascular 8 Unknown [1–4, cardiac; 1–6, cardiovascular]
Date of first all-cause hospital admission	dd/mm/yyyy 1 de novo hospitalisation 2 rehospitalisation
Date of first HF hospital admission	dd/mm/yy 1 de novo hospitalisation 2 rehospitalisation
Number of all-cause hospitalisations	

Number of all HF hospitalisations	
Drop-out	
Date of study discontinuation	dd/mm/yyyy
Reason for study discontinuation	
Exercise training (only applies to exercise group patients)	
Study level data	
Prescribed exercise training Overall duration Session duration Frequency of sessions Intensity	--- weeks (ranges if appropriate) ---- minutes (range if appropriate) --- sessions/week (range if appropriate) ----% units (range if appropriate)
Setting	1 Centre only 2 Home only 3 Both centre and home (define proportion of sessions at each location) 4 Other (state)
Patient level data	
Attended first exercise training	1 Yes 2 No 3 Not reported
Are details available at patient level on exercise dose received?	1 Yes 2 No

Appendix 5: Prediction of VO_2peak in HF from submaximal exercise tests

1: 6-minute walk test (6MWT)

A number of studies have examined the relationship of 6MWT and VO_2peak in HF patients and reported variable levels of association/correlation. Many studies failed to report a prediction equation or reported a multivariate equation that incorporated clinical parameters not available in the EMII IPD set. A recent discussion paper on the use of the 6MWT in HF, has questioned the reliability of prediction of VO_2peak (42). However, a review in 2010 by Ross et al of 11 studies in 1,083 patients with cardiopulmonary disease (many with HF) found generally high level of association of VO_2peak and 6MWT (average correlation coefficient of 0.59). (39) Using a study level random effects linear regression approach the authors derived the following overall prediction model with standard error of estimate (SEE) of 1.1 ml/kg/min.

$$VO_2peak\ (ml/kg/min) = 4.948 + 0.023 \times 6\text{-}MWD\ distance\ (metres)$$

2: Incremental shuttle walk test (ISWT)

Keell et al (40) tested the safety and acceptability of the SWT in patients with chronic heart failure and examined the relationship between SWT performance and VO_2peak.

$$VO_2peak\ (ml/kg/min) = (0{\cdot}27 \times number\ of\ 10m\ shuttles) + 7{\cdot}77$$

Similarly, Fowler et al (41) (41) proposed the following formula in patients following coronary artery bypass surgery:

$$VO_2peak\ (ml/kg/min) = 7.81 + [0.03 \times ISWT\ distance\ (m)]$$