Title: Objective assessment of fatigue or dyspnoea as the mechanism of exercise limitation in heart failure: *Implications for individualised therapy*

1. Introduction

This Statistical Analysis Plan (SAP) predefines statistical analysis, population, variables, and analysis methods of this trial. It is based on the study protocol (ClinicalTrials.gov ID:NCT04332536). The trial has been designed to demonstrate 1) whether a novel exercise test can identify the primary mechanism of exercise limitation in patients with CHF and 2) whether a specific intervention targeting dyspnoea influences exercise tolerance in CHF patients.

2. Rationale

Chronic heart failure (CHF) is a complex multifaceted disease that has a rising prevalence and poor prognosis, imposing an ever-growing demand on limited health care resources: annual CHF costs are ~2 % of the total NHS budget.[1] Optimal contemporary CHF therapy has reduced hospitalisations and extended the life-span of CHF patients, but these treatments are less effective at alleviating symptoms. Thus, there is a large and growing CHF patient population living with persistent and debilitating symptoms.

The primary symptom and a defining characteristic of CHF is exercise intolerance, and it is by the degree of exercise intolerance that CHF severity is clinically defined.[2,3] Impaired exercise tolerance limits the ability to undertake activities of daily living, reduces health-related quality of life, and perpetuates a cycle of inactivity and

deconditioning that further accelerates the decline in functional capacity and disease progression.

The gold-standard measure of exercise tolerance is $\dot{V}O_{2peak}$ measured from a treadmill or cycle ergometer based cardiopulmonary exercise (CPX) test. $\dot{V}O_{2peak}$ is the strongest predictor of survival.[4] and interventions that increase $\dot{V}O_{2peak}$ convey protection against adverse events (i.e. hospitalisations and death).[5] The clinical usefulness of $\dot{V}O_{2peak}$, and other CPX test parameters (e.g. $\dot{V}_E/\ \dot{V}CO_2$) particularly in relation to prognosis, is established.[4] However, the *mechanism* limiting $\dot{V}O_{2peak}$ is not identified by current CPX tests. At the point of exercise limitation patients report intolerable fatigue (i.e. muscle limited) or dyspnoea (i.e. symptom limited by breathlessness). Objectively identifying the mechanism of exercise limitation with a CPX test could be an essential first step to target treatments and personalise therapy to ameliorate the exercise limitation across a heterogeneous population of CHF patients.

We developed an innovative CPX technique that distinguishes between fatigue and dyspnoea as the operant mechanism that determines $\dot{V}O_{2peak}$. We have successfully used this CPX-test in COPD patients.[6,7,8]

This project will use our CPX technique in CHF to provide a step-change in the information provided by traditional CPX tests and understand how fatigue or dyspnoea limits $\dot{V}O_{2peak}$ in this heterogeneous population (*Objective 1*). An opioid intervention will then be used to demonstrate that our novel CPX technique can: (1) identify the mechanism of exercise limitation; (2) be used to target mechanism-specific therapies

to the aetiology of the exercise limitation. Specifically, our hypothesis is that symptom-alleviation with opioid use will increase exercise tolerance in symptom-limited (breathlessness), but not muscle-limited CHF patients (*Objective 2*).

Limitations to $\dot{V}O_{2peak}$ in health: $\dot{V}O_{2peak}$ measured during dynamic whole-body exercise (e.g. walking or cycling) quantifies the capacity of the O_2 cascade to deliver *and* utilise O_2. The effectiveness of the O_2 cascade determines the extent to which oxidative phosphorylation can meet the demands for ATP turnover, minimising the requirement for substrate-level phosphorylation and associated accumulation of fatigue-related metabolites. This is particularly important as these fatigue-related metabolites are directly associated with muscle fatigue and indirectly increase ventilation, thus contribute to dyspnoea. Therefore, task failure at $\dot{V}O_{2peak}$ could be due to (a) fatigue-induced reductions in the capacity of the neuromuscular system to voluntarily generate power such that the required power for the task can no longer be produced (i.e. $\dot{V}O_{2peak}$ muscle-limited), or (b) overwhelming symptoms of dyspnoea that result in exercise cessation prior to muscular limits for power production being attained. The latter situation implies a reserve in the capacity of the exercising muscle to generate power above that required by the task (i.e. $\dot{V}O_{2peak}$ symptom-limited). These opposing mechanisms could be considered as 'unable' to continue the exercise (muscle-limited) vs. 'unwilling' to continue due to overwhelming dyspnoea (symptom-limited).[9] Identifying the aetiology of the exercise limitation offers the intriguing potential to personalise the approach to interventions designed to ameliorate exercise intolerance and thereby symptoms.

The novel exercise test: To investigate this, we have developed an innovative CPX test to distinguish between these opposing mechanisms of task failure at $\dot{V}O_{2peak}$. This CPX test uses a commercially available cycle ergometer that has two key features; (1) strain gauges in the crank that directly measure the force produced by the leg muscles during cycling every 2° of pedal rotation, allowing calculation of instantaneous power production; and (2) the ability to switch instantaneously from standard hyperbolic, cadence-independent cycling, to isokinetic cycling (experimentally-controlled, constant cadence). Since muscle torque is velocity dependent, isokinetic (cadence-controlled) measurements are essential to appropriately measure maximal voluntary neuromuscular power producing capacity in both the fresh and fatigued state.[10]

Using our CPX test we can therefore measure the power-producing capacity of the leg muscles at any point during an exercise task from a brief (~5 s) maximal isokinetic effort. By comparing this maximal voluntary isokinetic power (MVIP) with the instantaneous demands of the exercise task (power set on the cycle ergometer), we can delineate between muscle and symptom limitations. In a single test we make three power measurements that might be clinically important: 1. Fresh MVIP at baseline ($MVIP_{FRESH}$); 2. the standard cycle ergometer measure of power at $\dot{V}O_{2peak}$ (P-$\dot{V}O_{2peak}$) 3. MVIP in the fatigued state, instantaneously at the point of $\dot{V}O_{2peak}$ at task failure ($MVIP_{FATIGUE}$; Figure 1).

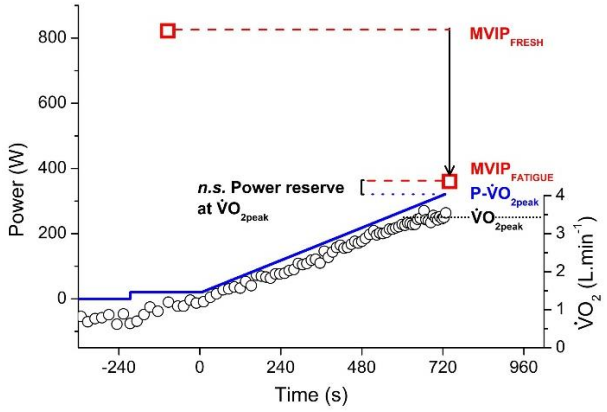


Figure 1: Healthy participant responses to our innovative ramp-incremental CPX test in which fresh maximal voluntary isokinetic power (□ **MVIP$_{FRESH}$**) decreases as the ramp-incremental test progresses to the extent that at $\dot{V}O_{2peak}$ (o) fatigued maximal voluntary isokinetic power (□ **MVIP$_{FATIGUE}$**) is not different from ramp-incremental power (**P-$\dot{V}O_{2peak}$**); i.e. there is no power reserve at $\dot{V}O_{2peak}$[7]

These power measurements allow us to assess skeletal muscle power generating capacity (MVIP$_{FRESH}$), and the mechanism of task failure at $\dot{V}O_{2peak}$, delineating fatigue (muscle-limited) from dyspnoea (symptom-limited) through comparison of P-$\dot{V}O_{2peak}$ and MVIP$_{FATIGUE}$ – the so-called 'power reserve' (Figure 1) – allowing us to provide the following important differentiation at the end of the test:

- *No power reserve*: The absence of a meaningful difference between P-$\dot{V}O_{2peak}$ and MVIP$_{FATIGUE}$ – the capacity of the neuromuscular system to voluntarily generate the required task power has been reached at $\dot{V}O_{2peak}$, and fatigue is preventing continuation of the ramp-incremental task, i.e. exercise is muscle-limited.
- *Power reserve*: MVIP$_{FATIGUE}$ exceeds P-$\dot{V}O_{2peak}$ – there is capacity within the neuromuscular system to voluntarily generate more power than is required by the task and therefore the exercise is symptom-limited.

In young (~30 yr; n = 20) and older (~65 yr; n = 12) healthy participants we have demonstrated that standard ramp-incremental exercise is commonly terminated at $\dot{V}O_{2peak}$ with no meaningful power reserve,[6,7,8] (MVIP$_{FATIGUE}$ ~30 % greater than P-$\dot{V}O_{2peak}$ at task failure; 274±73 vs. 212±84 Watts in healthy older participants[8]), consistent with fatigue determining task failure at $\dot{V}O_{2peak}$. Thus, in health during standard ramp-incremental exercise, physiologic and sensory responses are closely

associated and increasing effort, even for a very brief time (~5 s) is unable to increase $\dot{V}O_{2peak}$ or instantaneous power. Overcoming this muscle limitation thus requires interventions such as exercise training that increase physiologic capacity and fatigue resistance by increasing muscle O_2 delivery, capillarity and oxidative capacity.

Influence of chronic disease on limitations to $\dot{V}O_{2peak}$: While the triggering cardiac event leading to CHF impairs cardiac function, the absence of a relationship between $\dot{V}O_{2peak}$ and heart function,[11] highlights the predominance of systemic consequences of CHF in contributing to the reduction in $\dot{V}O_{2peak}$. Whilst limits to $\dot{V}O_{2peak}$ were considered to be primarily the consequence of reduced skeletal muscle O_2 delivery due to cardiac dysfunction it is now clear that there are considerable peripheral influences. Peripheral haemodynamics are adversely affected by persistent sympathetic activation and humorally-mediated vasoconstriction impair skeletal muscle O_2 delivery during exercise. However, skeletal muscle maladaptations also contribute to the reduction in $\dot{V}O_{2peak}$ by reducing O_2 utilisation. For example, in CHF there is loss of type I fibres, reductions in mitochondrial volume, density, function and muscle capillarity that reduce muscle oxidative capacity. These pathophysiological changes throughout the O_2 cascade result in a more fatigable skeletal muscle phenotype (greater muscle limitation).[12] In CHF there is also decreased ventilatory efficiency (high $\dot{V}_E/\dot{V}CO_2$), manifest as a consequence of altered ventilatory control (low $PaCO_2$) and high ventilation-perfusion mismatch (high V_D/V_T), that increases the ventilatory requirements of exercise, work of breathing and dyspnoea (greater symptom limitation).[13]

Thus, exercise limitation in CHF is a complex interplay between increased fatigue and increased dyspnoea. Which operant mechanism predominates is currently unknown, cannot be identified by traditional CPX-tests, and might be different between patients in whom the severity of exercise limitation is the same. Using our innovative CPX approach can now make the distinction between intolerable fatigue (i.e. muscle limited) or dyspnoea (i.e. symptom limited by breathlessness).

CHF pilot data: In contrast to the response seen in health, in CHF patients (n=16), we have shown that ramp-incremental exercise is terminated at $\dot{V}O_{2peak}$ with a large power reserve. $MVIP_{FATIGUE}$ was, on average, 101±122% greater than P-$\dot{V}O_{2peak}$ (213±132 vs. 109±44 Watts, RM-ANOVA with Dunnett *post-hoc,*[7] p = 0.002, 95% $CI_{Difference}$ 38, 171 Watts; Figure 2). This power reserve suggests the physiologic and sensory responses have become dissociated and $\dot{V}O_{2peak}$ is symptom-limited in CHF. This is similar to our observation in COPD patients where $MVIP_{FATIGUE}$ was ~160 % greater than P-$\dot{V}O_{2peak}$ at task failure.[8]

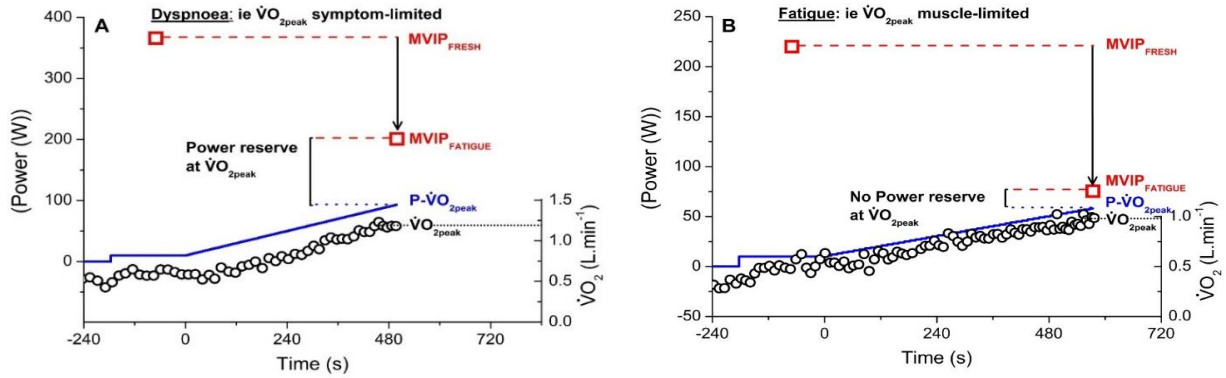


Figure 2: Representative examples of CHF patient responses to our innovative ramp-incremental CPX test. Fresh maximal voluntary isokinetic power (□ **$MVIP_{FRESH}$**) decreases as the test progresses. **A.** Fatigued maximal voluntary isokinetic power (□ **$MVIP_{FATIGUE}$**) at $\dot{V}O_{2peak}$ (o) is greater than the ramp-incremental power (**P-$\dot{V}O_{2peak}$**): the power reserve suggests exercise in this patient is symptom-limited. **B.** **$MVIP_{FATIGUE}$** was not meaningfully different from **P-$\dot{V}O_{2peak}$**: the absence of a power reserve suggests exercise in this patient is muscle-limited.

In our CHF patients, we also observed two key phenotypes:

1. CHF patients (n = 6) without a power reserve ($MVIP_{FATIGUE}$ 22±12 % greater than P-$\dot{V}O_{2peak}$ at task failure; 141±66 vs. 115±47 Watts p = 0.547, $CI_{Difference}$ -44, 97 Watts), suggesting they reached their physiological limits and are muscle-limited (Figure 2B).
2. CHF patients (n = 10) in whom the power reserve was substantial ($MVIP_{FATIGUE}$ 149±134% greater than P-$\dot{V}O_{2peak}$ at task failure; 256±145 vs. 105±44 Watts p = 0.002, $CI_{Difference}$ 55, 246 Watts), suggesting $\dot{V}O_{2peak}$ was symptom-limited (Figure 2A).

Importantly, this objective CPX-test measure of the power reserve that identifies whether fatigue or dyspnoea is the primary mechanism of exercise limitation is unique. The presence of a relevant power reserve cannot be identified from traditional subjective ratings of perceived exertion (RPE). The mechanisms that determine effort, e.g. the effort required to turn the legs during cycling, interact with dyspnoea, meaning that the RPE of leg effort is heightened with dyspnoea, independent of objective measures of leg fatigue, and *vice versa*. Therefore, objectively identifying the primary limitation to exercise is critical and can only be identified using our technique.

Effect of opioids on exercise tolerance: In COPD patients during maximal constant-load exercise, immediate-release oral morphine decreased breathing frequency, ventilation and Borg ratings of dyspnoea. This was associated with a clinically meaningful (+ 2.5±0.9 min) increase in time to task failure,[14] consistent with our suggestion above that dyspnoea can limit exercise tolerance, and when dyspnoea is attenuated in patients who are symptom-limited (and not fatigue-limited), the skeletal muscle is capable of performing more work. While these data are consistent with other studies showing reductions in dyspnoea and increases in exercise tolerance with

systemic opioids in COPD,[15,16] this study also identified ‘responders’ and ‘non-responders’.[14] Whilst our innovative CPX test revealed that mean $MVIP_{FATIGUE}$ was on average ~160% greater than P-$\dot{V}O_{2peak}$ in COPD and that some patients were therefore symptom-limited with a large power reserve, there appeared to be a sub-group of muscle-limited patients with no meaningful power reserve,[8] in whom we hypothesis opioid treatment would have no effect, due to the absence of a power reserve. This heterogeneous response mirrors that seen in our pilot CHF data.

In CHF opioids reduced symptoms of dyspnoea,[17] increase exercise time (~12 % increase in tolerable duration) and $\dot{V}O_{2peak}$ [18] (+1.6 $ml.min^{-1}.kg^{-1}$; MCID = 1 $ml.min^{-1}.kg^{-1}$[19]). However, similar to the variable response seen in COPD, the effect is inconsistent across CHF patients, possibly due to the heterogeneity in the mechanism of exercise limitation that has previously not been discernible. We believe that we can now divide CHF patients into two groups using our novel CPX approach (Figure 2).

Compared with alternative interventions (e.g. device, medical, rehabilitation) that affect *both* fatigue and dyspnoea in combination, opioid treatment specifically targets dyspnoea. Thus, objective 2 aims to demonstrate that our CPX-test can for the first time identify likely ‘responders’ (symptom-limited) and ‘non-responders’ (muscle-limited) to this mechanism-specific, symptom-alleviating opioid intervention (*Objective 2*) whilst also confirming our hypothesis of two phenotypes within the exercise intolerance CHF population. This will also provide the first evidence that the mechanisms limiting exercise tolerance in CHF are directly modifiable and can be selectively targeted by personalised interventions to maximise therapeutic efficacy.

3 Study design and randomisation

Design: This study is a randomised, double-blinded, crossover, multi-group trial conducted at one centre in the UK (Leeds Teaching Hospitals NHS Trust).

Participants: The study will recruit a total of 100 subjects of whom 75 will have HF with reduced ejection fraction (HFrEF) and 25 will be control subjects.

Inclusion criteria: Patients with CHF due to left ventricular systolic dysfunction (NYHA II or III, and left ventricular ejection fraction < 50%), optimally tolerated stable medical therapy, able and willing to provide written informed consent and perform CPX-tests.

Exclusion criteria: CHF without symptoms or class IV symptoms, any contraindications to exercise or severe neuromuscular disability limiting exercise, significant COPD (FEV1 < 50%) or other significant respiratory disease contributing to symptoms, or renal disease (eGFR < 20).

Study activities and exercise protocols: After baseline assessments (resting echocardiography, demographics, resting ECG, medical history, medication review) are complete, each participant will perform one maximum effort ramp-incremental cardiopulmonary exercise test on an electromagnetically braked cycle ergometer which allows for the measurement of maximal voluntary isokinetic power (at baseline and immediately end ramp-incremental exercise $MVIP_{FRESH}$, $MVIP_{Fatigue}$) along with the traditional measure of ramp-incremental power at peak exercise.

At three subsequent visits each approximately one week apart, participants will perform an exercise test at a constant power of around 70-80% ramp-incremental

power at peak exercise until they cannot continue. The exercise power of these tests has been chosen aiming for a duration of the test to be between approximately 4 and 8 minutes.

Of these three tests, tests one and two will be maximum effort constant power exercise and will be continued until the point of exercise intolerance, and will be performed following a single dose of either dihydrocodeine or placebo (randomised).

The third test will be performed at the same constant power, following the same condition during which the participant achieved the longest exercise time, but will be terminated at the time of the shorter test (the iso-time test). The investigators will be blinded to whether this was with dihydrocodeine or placebo ensuring that the rigours of double-blinding are maintained. This third test is important because, as is increasingly appreciated, the important variables at the end of an exercise test ($\dot{V}O_{2peak,}$ $MVIP_{FATIGUE}$, dyspnoea) may be the same at the point of exercise intolerance with dihydrocodeine and placebo, but the rate at which these end-points are achieved may be different.

Intervention: Prior to the first two constant power tests, subjects will be asked to take a sweetened oral solution containing 0.5mg/kg dihydrocodeine up to a maximum of 60mg, or placebo solution with matching colour/taste, a minimum of 30 minutes prior to the test. Prior to the third constant power test, subjects will be asked to take the intervention that was associated with the longest exercise time during the first two tests.

Primary outcome measures: During each test we will measure power, pulmonary gas exchange ($\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}_E$, RER, tidal volume, breathing frequency), haemodynamaic variables such as blood pressure and heart rate ECG, oxygen saturation (Ultima Cardio2, Medical Graphics) throughout. As described above, the use of an electromagnetically braked cycle ergometer, will provide the opportunity to measure $MVIP_{FRESH}$, $MVIP_{Fatigue}$ from power measurements made directly at the crank of the cycle ergometer(Figure 3).

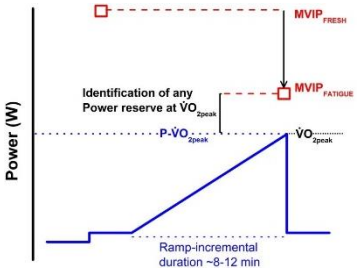


Figure 3: Schematic of the innovative ramp-incremental CPX protocol (objective 1). Ramp-incremental exercise will be replaced by constant-work rate exercise to address objective 2. All other measurements and procedures will be identical.

Borg scale ratings of perceived exertion (leg fatigue and breathlessness) will be recorded every 2 minutes during exercise, and at $\dot{V}O_{2peak}$.

Blinding: Subjects and observers will be blinded to allocation for all three tests. Leeds Pharmacy will be provided with the exercise duration for the first two constant power tests and allocate either dihydrocodeine or placebo at the third test based on which gave the longest exercise duration in the first two constant power tests.

4 Aims and objectives

Overarching objective: To determine whether a novel CPX-test can identify the primary mechanism of exercise limitation in CHF, and be used to target mechanism-specific interventions personalised to the driving pathophysiology in individuals.

Objective 1: Characterise fatigue vs. dyspnoea limitations to $\dot{V}O_{2peak}$ in CHF using novel power measurements made during the CPX.

Objective 2: Investigate the efficacy of symptom-alleviating opioid treatment to increase exercise tolerance in CHF.

5. Sample size calculation

5.1 Objective 1: Phenotyping the mechanism of exercise limitation

To identify if the power reserve is different between CHF patients vs. controls (two-tailed independent t-test), assuming a standard deviation of 40% (based on our CHF pilot data) we require 60 CHF and 20 control patients to have 90% power to detect an absolute power reserve difference of 30% at $p < 0.05$. The 3:1 ratio allows adequate power for our secondary analysis comparing the 3 different severities of CHF, where 20 in each group gives us 80% power to detect a linear trend over the 3 categories of increasing severity, with a mean power reserve difference of 20 percentage points in absolute terms between each group, e.g. 20 % vs 40 % vs 60 % power reserve with increasing severity. In anticipation that we may need to log-transform our primary outcomes due to non-normally distributed data, our power calculation is conservative. To allow for any drop-out, and any unsuccessful test, we propose recruiting an additional 25% (*Objective 1*). Hence, we will invite 75 people with CHF and 25 controls to participate.

5.2: Objective 2: Efficacy of intervention

Based on exercise on a conservative effect size (f) of 0.300,[18] and a conservative correlation between repeated measures of 0.5,[14] 17 patients in each group gives us 90% power at $p<0.05$. Thus, we will be powered to detect a smaller increase in

exercise duration than that found in COPD.[14] We will recruit 20 patients to allow for any drop-out *(Objective 2)*.

6. Planned analyses

6.1 Objective 1

6.1.1 Endpoints for objective 1

In *Objective 1* we will use the power reserve to describe the frequency of muscle-limited vs. dyspnoea (symptom)-limited CHF patients and controls and describe these phenotypes by clinical severity and cardiopulmonary variables.

6.1.2 Derivation of outcome measures

The key outcome measure will utilise three measures from the cycle ergometer. MVIP prior to exercise at the baseline test baseline ($MVIP_{FRESH}$), the standard cycle ergometer measure of power at $\dot{V}O_{2peak}$ (P-$\dot{V}O_{2peak}$) and MVIP in the fatigued state, instantaneously at the point of $\dot{V}O_{2peak}$ at task failure ($MVIP_{FATIGUE}$) will be recorded as described in Figure 1.

The presence of a power reserve or not will be determined by a difference of >40% difference between P-$\dot{V}O_{2peak}$ and $MVIP_{FATIGUE}$.

6.1.3 Primary outcome analyses

This will describe the clinical and exercise phenotype of participants in whom fatigue (no power reserve; muscle-limited) or dyspnoea (large power reserve; symptom-limited) predominate as the operant mechanism, and how the groups (patients v controls and participants with dyspnoea or fatigue as limiting symptoms) differ.

6.1.4 Secondary outcome analyses

This will explore relationships between the power variables (e.g. $MVIP_{FRESH}$, P-$\dot{V}O_{2peak}$, $MVIP_{FATIGUE}$), ventilatory and pulmonary gas exchange variables (e.g. $\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}_E$, RER, tidal volume, breathing frequency), and clinical variables in the CHF participants and controls.

6.2 Objective 2

6.2.1 Endpoints for objective 2

Exercise time sustained at the required power during the maximum effort constant power tests.

6.2.2 Derivation of outcome measures

This will be measured from when the required power is applied to the cycle ergometer power until the point of intolerance, when the required power will be removed.

6.2.3 Primary outcome analysis

This will be the change in exercise tolerance time between the two constant power tests (dihydrocodeine vs. placebo).

6.2.4 Secondary outcome analyses

The secondary outcomes will include a comparison of the power variables (e.g. $MVIP_{FRESH}$, P-$\dot{V}O_{2peak}$, $MVIP_{FATIGUE}$), ventilatory and pulmonary gas exchange variables (e.g. $\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}_E$, RER, tidal volume, breathing frequency) and perceived symptom scores at baseline and intolerance between the tests following

dihydrocodeine and placebo. We will also include comparison of the power variables (e.g. $MVIP_{FRESH}$, P-$\dot{V}O_{2peak}$, $MVIP_{FATIGUE}$), ventilatory and pulmonary gas exchange variables (e.g. $\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}_E$, RER, tidal volume, breathing frequency) and perceived symptom scores at baseline and isotime to investigate whether the rate at which variables are achieved is different between dihydrocodeine and placebo.

6.3 Missing data

The percent of missing data for each variable will be recorded. If there is a large number of missing data on variables, an evaluation on the missing data will be conducted before analysis. Patients who did not attend all tests will be reported along with their reasons for not attending.

7 Populations

7.1 Intention-to-treat population

The primary method for analysing the trial data (objective 2) will be the Intention-to-treat analysis (ITT). All subjects who completed the first two constant power tests will be included in the intention-to-treat population (ITT) according to intention-to-treat principles, in which subjects will be analysed according to the test assigned by randomization (dihydrocodeine or placebo). Participants who withdraw consent for their data to be used in the trial will not be included in the ITT population.

7.2 As-treated population

The as treated population is all patients being reported with the treatment in which they received.

8 Data handling

8.1 Data validation

Baseline and follow up data will be captured using an excel spreadsheet. The data will be exported to specific statistical software for analysis (e.g. STATA, SPSS, Origin, and MATLAB). Data cleaning will involve checking for incompleteness, inconsistencies and inaccuracies. Inaccuracies and inconsistencies will be reported back to the primary investigators. Validation checks will include:

- Checking eligibility of all randomised patients
- Checking on 1:1 cross-over randomisation to see if it was achieved
- Checking for outliers
- Checking on missing data

8.2 Data analysis

8.2.1 General calculations

Categorical data will be summarised in terms of the number of patients and percentages. All percentages will be calculated using as the denominator the total number of patients within the specified analysis population; that is including patients with missing data. Baseline grouping will be patients with CHF and control subjects.

Normality for all continuous variables will be tested using the Shapiro-Wilk test, histograms, Q–Q plots and boxplots. Normally distributed continuous variables will be reported as mean and mean ± 1 SD, and if there are non-normally distributed continuous variables these will be reported as median (interquartile range). Subsequently, associations between groups or interventions and baseline characteristics will be assessed using either analysis of variance and the 2-sample Student t test for normally distributed values. In cases of non-normally distributed data

log-transformations or equivalent non-parametric test will be considered. Similar associations with categorical variables will be analysed using the chi-squared test for contingency tables.

All statistical tests performed will be two-sided tests at a 5% significance level and giving the 2-sided 95% confidence intervals. These tests and confidence intervals will be displayed to 2 decimal places, along with standard errors, standard deviations and parameter estimates. Any presented p-value will be given to 3 decimal places, any p-value lower than 0.001 will be shown as <0.001. Analysis will be conducted using specific statistical software for analysis (e.g. STATA, SPSS, Origin, and MATLAB)

8.2.2 Baseline characteristics

Demographic baseline characteristics will be presented in a table.

8.2.3 Recruitment and compliance

The flow of patients throughout the study from randomisation to analysis will be displayed in a CONSORT diagram.

Summaries of the compliance of the patients and the treatment in which they all received will be shown in tables for each randomised arm. If any reasons why patients were not compliant or why they received a different treatment are collected then they will also be summarised.

8.2.4 Objective 1 analyses

These will use the power reserve to describe the frequency of fatigue vs. dyspnoea limitations to $\dot{V}O_{2peak}$ in CHF compared with controls and also compare the magnitude of the power reserve along with relevant clinical characteristics of participants with either limiting factor using appropriate tests depending upon the nature of the data.

We will then proceed to explore the relationships between the relevant clinical characteristics, the power reserve, other power variables (e.g. $MVIP_{FRESH}$, P-$\dot{V}O_{2peak}$, $MVIP_{FATIGUE}$) and ventilatory and pulmonary gas exchange variables (e.g. $\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}_E$, RER, tidal volume, breathing frequency).

All model assumptions will be checked to ensure the residuals are normal, and that the residuals are independent with no evidence of homoscedasticity.

8.2.5 Objective 2 analyses

8.2.5.1 Principles of analysis of cross-over data will be followed.

Once a familiarization test has been performed, a peak exercise test is not a training stimulus. We previously performed up to 5 exercise tests in consecutive weeks in patients with CHF and controls, with no longitudinal effects (15). However, to account for any carryover effects, the interventional crossover study will be analysed using a linear mixed model with a random effect for subject. For each endpoint Y_{ak} (e.g., exercise time) under consideration in the study:

$$Y_{ak} = \mu + d_i + \pi_j + i_n + \alpha_k + i_{nk}$$

where $i_{nk} \sim N(0, \sigma^2_{\varepsilon})$, $\alpha_k \sim N(0, \sigma^2_{\alpha})$ and μ is the overall mean, s is the treatment effect, p is the period effect, and l is the carryover effect (which is mathematically identical to an interaction term between treatment and period). This model will be estimated using PROC MIXED in SAS, and least squares means will be estimated for

each of these terms and their differences. All statistical tests will be 2-sided, and any p value <0.05 will be called statistically significant.

8.2.5.2 Secondary analyses within objective 2

The secondary outcomes will include a description and comparison of the power variables (e.g. $MVIP_{FRESH}$, P-$\dot{V}O_{2peak}$, $MVIP_{FATIGUE}$), ventilatory and pulmonary gas exchange variables (e.g. $\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}_E$, RER, tidal volume, breathing frequency) and perceived symptom scores at baseline and intolerance between the tests following dihydrocodeine and placebo, and between CHF fatigue and dyspnoea limited groups. We will also include a description and comparison of the power variables (e.g. $MVIP_{FRESH}$, P-$\dot{V}O_{2peak}$, $MVIP_{FATIGUE}$), ventilatory and pulmonary gas exchange variables (e.g. $\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}_E$, RER, tidal volume, breathing frequency) and perceived symptom scores at baseline and isotime to investigate whether the rate at which variables are achieved is different between dihydrocodeine and placebo in CHF fatigue and dyspnoea limited groups. For these analyses we will use a 2x2 repeated-measures ANOVA. All statistical tests will be 2-sided, and any p value <0.05 will be called statistically significant

8.3 Multiplicity

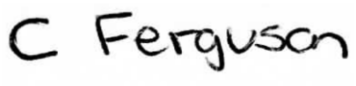
This study does not consider multiplicity issues and therefore does not adjust significance levels based on multiplicity tests, unless specified otherwise.

9 Reporting and dissemination of the results

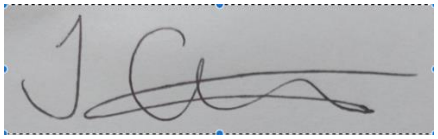
The results of this study will be published in medical peer-reviewed journals and presented at cardiovascular conferences.

10 Approved

Dr Klaus Witte Date: 16th October 2024



Dr Carrie Ferguson Date: 21st October 2024



Dr John Gierula Date: 22nd October 2024

10. References

1. National Institute for Health and Clinical Excellence. New NICE guidance will improve diagnosis and treatment of chronic heart failure. London: NICE, 2010.
2. Weber KT, Kinasewitz GT, Janicki JS & Fishman AP. (1982). Oxygen utilization and ventilation during exercise in patients with chronic cardiac failure. *Circulation* **65,**1213-1223.
3. The Criteria Committee of the New York Heart Association. Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels. Vol. 9th ed. Little, Brown & Co; Boston, Mass: 1994. p.253-256.
4. Keteyian SJ, Patel M, Kraus WE, Brawner CA, McConnell TR, Pina IL, Leifer ES, Fleg JL, Blackburn G, Fonarow GC, Chase PJ, Piner L, Vest M, O'Connor CM, Ehrman JK, Walsh MN, Ewald G, Bensimhon D, Russell SD & Investigators H-A. (2016). Variables Measured During Cardiopulmonary Exercise Testing as Predictors of Mortality in Chronic Systolic Heart Failure. *J Am Coll Cardiol* **67,** 780-789.
5. O'Connor CM, Whellan DJ, Lee KL, Keteyian SJ, Cooper LS, Ellis SJ, Leifer ES, Kraus WE, Kitzman DW, Blumenthal JA, Rendall DS, Miller NH, Fleg JL, Schulman KA, McKelvie RS, Zannad F, Pina IL & Investigators H-A. (2009). Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial. *JAMA* **301,**1439-1450.
6. Coelho AC, Cannon DT, Cao R, Porszasz J, Casaburi R, Knorst MM & Rossiter HB. (2015). Instantaneous quantification of skeletal muscle activation, power production, and fatigue during cycle ergometry. *J Appl Physiol* **118** 646-654.
7. Ferguson C, Wylde LA, Benson AP, Cannon DT & Rossiter HB. (2016). No reserve in isokinetic cycling power at intolerance during ramp incremental exercise in endurance-trained men. *J Appl Physiol* **120,**70-77.
8. Cannon DT, Coelho AC, Cao R, Cheng A, Porszasz J, Casaburi R & Rossiter HB. (2016). Skeletal muscle power and fatigue at the tolerable limit of ramp-incremental exercise in COPD. *J Appl Physiol* **121,**1365-1373.
9. Palange P, Ward SA, Carlsen KH, Casaburi R, Gallagher CG, Gosselink R, O'Donnell DE, Puente-Maestu L, Schols AM, Singh S & Whipp BJ. (2007). Recommendations on the use of exercise testing in clinical practice. *Eur Respir J* **29,**185-209.
10. Beelen A & Sargeant AJ. (1991). Effect of fatigue on maximal power output at different contraction velocities in humans. *J Appl Physiol* **71,**2332-2337.
11. Witte KK, Nikitin NP, De Silva R, Cleland JG & Clark AL. (2004). Exercise capacity and cardiac function assessed by tissue Doppler imaging in chronic heart failure. *Heart* **90,**1144-1150.
12. Poole DC, Hirai DM, Copp SW & Musch TI. (2012). Muscle oxygen transport and utilization in heart failure: implications for exercise (in)tolerance. *Am J Physiol Heart Circ Physiol* **302,**H1050-1063.
13. Woods PR, Olson TP, Frantz RP & Johnson BD. (2010). Causes of breathing inefficiency during exercise in heart failure. *J Card Fail* **16,**835-842.
14. Abdallah SJ, Wilkinson-Maitland C, Saad N, Li PZ, Smith BM, Bourbeau J & Jensen D. (2017). Effect of morphine on breathlessness and exercise endurance in advanced COPD: a randomised crossover trial. *Eur Respir J* **50**.
15. Witte KK, Thackray SD, Nikitin NP, Cleland JG, Clark AL. The effects of alpha and beta blockade on ventilatory responses to exercise in chronic heart failure. *Heart* 2003;**89:**1169–7