Journal of Exercise Science & Fitness 21 (2023) 147–156

Contents lists available at ScienceDirect

Journal of Exercise Science & Fitness

journal homepage: www.elsevier.com/locate/jesf

Unsupervised cluster analysis reveals distinct subgroups in healthy population with different exercise responses of cardiorespiratory fitness

Lin Xie [a, 1], Bo Gou [b, 1], Shuwen Bai [a], Dong Yang [a], Zhe Zhang [a], Xiaohui Di [a], Chunwang Su [a], Xiaoni Wang [a], Kun Wang [b, **], Jianbao Zhang [a, *]

[a] Key Laboratory of Biomedical Information Engineering of Education Ministry, Department of Life Science and Technology, Xi'an Jiaotong University, Xi'an, 710049, China
[b] Department of Physical Activity and Public Health, Xi'an Physical Education University, Xi'an, 710068, China

ARTICLE INFO

Article history:
Received 7 October 2022
Received in revised form
9 December 2022
Accepted 28 December 2022
Available online 31 December 2022

Keywords:
Interindividual variability
Oxygen uptake
Training responsiveness
Unsupervised learning

ABSTRACT

Background: Considerable attention has been paid to interindividual differences in the cardiorespiratory fitness (CRF) response to exercise. However, the complex multifactorial nature of CRF response variability poses a significant challenge to our understanding of this issue. We aimed to explore whether unsupervised clustering can take advantage of large amounts of clinical data and identify latent subgroups with different CRF exercise responses within a healthy population.

Methods: 252 healthy participants (99 men, 153 women; 36.8 ± 13.4 yr) completed moderate endurance training on 3 days/week for 4 months, with exercise intensity prescribed based on anaerobic threshold (AT). Detailed clinical measures, including resting vital signs, ECG, cardiorespiratory parameters, echocardiography, heart rate variability, spirometry and laboratory data, were obtained before and after the exercise intervention. Baseline phenotypic variables that were significantly correlated with CRF exercise response were identified and subjected to selection steps, leaving 10 minimally redundant variables, including age, BMI, maximal oxygen uptake (VO_{2max}), maximal heart rate, VO_2 at AT as a percentage of VO_{2max}, minute ventilation at AT, interventricular septal thickness of end-systole, E velocity, root mean square of heart rate variability, and hematocrit. Agglomerative hierarchical clustering was performed on these variables to detect latent subgroups that may be associated with different CRF exercise responses.

Results: Unsupervised clustering revealed two mutually exclusive groups with distinct baseline phenotypes and CRF exercise responses. The two groups differed markedly in baseline characteristics, initial fitness, echocardiographic measurements, laboratory values, and heart rate variability parameters. A significant improvement in CRF following the 16-week endurance training, expressed by the absolute change in VO_{2max}, was observed only in one of the two groups (3.42 ± 0.4 vs 0.58 ± 0.65 ml·kg^{-1}•min^{-1}, $P = 0.002$). Assuming a minimal clinically important difference of 3.5 ml·kg^{-1}•min^{-1} in VO_{2max}, the proportion of population response was 56.1% and 13.9% for group 1 and group 2, respectively ($P<0.001$). Although group 1 exhibited no significant improvement in CRF at group level, a significant decrease in diastolic blood pressure (70.4 ± 7.8 vs 68.7 ± 7.2 mm Hg, $P = 0.027$) was observed.

Conclusions: Unsupervised learning based on dense phenotypic characteristics identified meaningful subgroups within a healthy population with different CRF responses following standardized aerobic training. Our model could serve as a useful tool for clinicians to develop personalized exercise prescriptions and optimize training effects.

1. Introduction

Cardiorespiratory fitness (CRF) refers to the integrated ability of the circulatory and respiratory systems to deliver and utilize

* Corresponding author. Key Laboratory of Biomedical Information Engineering of Education Ministry, Department of Life Science and Technology, Xi'an Jiaotong University, No.28, Xianning West Road, Xi'an, China.
** Corresponding author.
E-mail addresses: 103001@tea.xaipe.edu.cn (K. Wang), zhangjb@mail.xjtu.edu.cn (J. Zhang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jesf.2022.12.005

oxygen during physical activity.[1] Overwhelming evidence has demonstrated that CRF is a powerful, independent predictor of numerous clinical outcomes, and the improvement in CRF substantially reduces the risk of mortality.[1–3] Physical activity and exercise, especially endurance training, is the most effective intervention for increasing CRF.[4] Regular endurance exercise could approximately increase CRF by 3–35%, determined by maximal oxygen uptake (VO_{2max}), at group level.[5] However, in a growing number of studies, the effects of a given dose of exercise training on CRF are not uniform at the individual level, with some individuals gaining large improvements in CRF while others showing small or even no improvements.[6] Indeed, studies specifically designed to assess the individual variation in exercise response demonstrate substantial variability in CRF response.[7,8] This observation has been made in both healthy and patient populations,[9–11] raising compelling challenges to the application of exercise in preventive and therapeutic medicine, as well as the development of precision exercise medicine.[12]

Unfortunately, the mechanisms underlying individual variability in CRF response to standardized exercise training programs are not fully understood. A central challenge in understanding the heterogeneity of CRF training response is the existence of measurement error, day-to-day variability, and other sources of random error in the observed changes in CRF, which will always lead to an overestimation of true CRF response.[6] Furthermore, methodological factors, including the modality, intensity and volume of exercise, as well as the prescription method of the training program, all influence the individual training response.[13,14] Notably, the methodology used to prescribe exercise intensity appears to have a profound effect on CRF training variability.[8,15] Traditional prescription methods that determine exercise intensities using maximal physiological variables, such as maximal heart rate (HR_{max}), heart rate reserve and VO_{2max}, often fail to elicit a homogeneous exercise stimulus among individuals, as they do not account for individual metabolic differences.[16] It has been suggested that the large interindividual variation in training response following a traditionally prescribed exercise program may be to some extent attributed to varied metabolic responses among individuals.[13] Alternatively, establishing exercise intensity relative to physiological thresholds has been reported to create more comparable metabolic strains among individuals.[13,17] In fact, studies have shown that threshold-based exercise programs significantly reduce training variability and increase the response rate of CRF.[8,17–19] Such results suggest that threshold-based training programs may be superior to traditional prescribed exercise when investigating the individual response to interventions.[13] Other than methodological factors, previous studies have identified multiple biological contributors, including genetics, age, sex, blood oxygen-carrying capacity, baseline fitness and autonomic function.[5,13,20] A better understanding of which variable or combination of variables predisposing some individuals to have a better training response than others may allow more effective exercise prescriptions, and thus would be particularly meaningful in clinical practice.[21] Although numerous exercise training studies are available that support this concept, a number of methodological issues impeded the synthesis of data to identify such mechanisms.[12] For example, existing studies investigating exercise responsiveness varied dramatically in the study design, the statistical model used, and even in the definition of the meaningful training response.[13] Therefore, dedicated studies considering multiple phenotypic characteristics in the prediction of the individual response of CRF following a threshold-based exercise program are warranted.

However, as the numbers of input variables and possible associations among them increase, inference from a statistical model becomes less precise.[22] Consequently, undertaking such studies using standard statistical methods becomes problematic. On the other hand, machine learning generally outperforms traditional statistical methods in the ability to account for complex relationships between multiple inputs and has been increasingly used in clinical research in the past decade.[23,24] One major category of machine learning approaches, unsupervised learning, explores the data to learn intrinsic patterns and associations without investigator supervision. This helps derive a robust set of variables for discovering natural subgroups within a population, such as the analysis of clinical data to classify a disease or clinical syndrome into novel subtypes that show different outcomes.[25,26] In this regard, unsupervised learning may be invaluable in phenotyping various exercise responses. To the best of our knowledge, no prior study has applied unsupervised cluster analysis to variables from multiple clinical domains to investigate the heterogeneity of CRF response to endurance training. Therefore, the purpose of the present study was to classify CRF training responsiveness based on dense phenotypic characteristics using unsupervised cluster analysis. We hypothesized that applying unbiased cluster algorithms to variables related to the individual variability of exercise response would allow the detection of novel subgroups among healthy populations with different characteristic profiles and CRF exercise responses.

2. Methods

2.1. Study participants

Between March 2013 and December 2014, 460 study participants were prospectively enrolled in an aerobic exercise training trial as part of the National Science & Technology Program, Expert Exercise Guidance System Development (supported by the Ministry of Science and Technology of China: 2012BAK23B01). The inclusion criteria were age 20-59 years old, healthy enough to participate in exercise training, and not engaging in structured exercise during the last year. Potential participants underwent a structured interview regarding medical history and lifestyle habits (e.g., smoking, sleep, physical activity, and sedentary behaviors). Participants with abnormal resting and exercising electrocardiogram (ECG), blood pressure >130/90 mm Hg, regular use of medications for metabolic syndrome (such as hypertension, insulin resistance, and dyslipidemia), sleep disorders, and chronic pain, history of cardiovascular and neurological disorders, or a diagnosis of diabetes were excluded. All participants provided written informed consent and the study protocol was approved by the Institutional Review Board of Xi'an Physical Education University.

2.2. Clinical measures

All clinical variables were obtained at baseline and after 4 months of the exercise intervention. Participants were instructed to avoid exercise and consumption of caffeine and alcohol for 24 h before testing. Each test session comprised two clinical visits on separate days. On test day 1, participants completed a series of assessments, including anthropometric measurements, resting vital signs, heart rate variability, standard 12-lead ECG, echocardiography, spirometry, and fasting blood chemistry analysis. All assessments were conducted by trained personnel, and standardized laboratory techniques were used. On test day 2, CRF was evaluated by a maximal exercise test using an electronically braked cycle (Ergoselect 100, Ergoline, Bitz, Germany) and an open-circuit gas analyzer (MetaMax 3B, Cortex Biophysik, Leipzig, Germany). The tests were supervised by exercise physiologists according to an incremental work rate protocol with a 2-min warm-up and a 20/25 W increase in workload every 2 min for women/men.

Electrocardiogram was monitored continuously during the test, while the Borg rating of perceived exertion (RPE) and blood pressure were collected every 2 min. Anaerobic threshold (AT) was measured using the ventilatory equivalents method. The attainment of maximal effort was considered when a plateau in oxygen consumption was observed (increase in oxygen consumption with increasing exercise intensity <2 ml·kg^{-1}•min^{-1}) or a heart rate (HR) within 10 beats of the age-predicted HR_{max}, RPE ≥17/20, and peak respiratory exchange ratio ≥1.1. Variables derived from the cardiopulmonary exercise test in the present study included VO_{2max}, VO_2 at AT (VO_{2AT}), VO_{2AT} as a percentage of VO_{2max} ($VO_{2AT/max}$), HR_{max}, HR at AT (HR_{AT}), maximal minute ventilation (VE_{max}), VE at AT (VE_{AT}), O_2 pulse at AT, workload achieved at AT, and exercise time at AT.

2.3. Exercise intervention

Considering the untrained nature of the study sample participants, moderate intensity continuous training was prescribed instead of high intensity interval training in this study. After the assessment of baseline CRF, participants started the prescribed exercise program, which included moderate aerobic training 3 sessions per week for 4 months. The frequency of the training program was determined according to the definition of sport population provided by the Chinese Sports Ministry, which is the proportion of population that participate in physical training more than 30 min on at least 3 days every week. Exercise intensity was prescribed based on HR_{AT}. During the 16-week exercise program, the training intensity and duration progressively increased. In the first month, the training lasted 30 min at an intensity eliciting 85% of the HR_{AT}. From the second month, the duration of the exercise increased to 40 min, and the intensity increased by 10% every month. The intensity and duration of the aerobic training program for each month are listed in Table 1. The aerobic exercise modalities included cycling and running. In the original trial, participants were randomized to either a cycling or running exercise modality (there were no differences in participant characteristics and exercise effects between the two aerobic exercise modalities, see Supplementary Table 1). Participants were instructed to maintain their regular lifestyle during the 4-month intervention period. Exercise specialists supervised each training session to ensure adherence to the prescribed exercise intensity (monitored using Polar RS800cx, Polar Electro, Kempele, Finland) and duration. Compliance with the target exercise intensity within 95%–105% and duration of 100% were accepted as actual attendance. Participants were excluded for downstream analysis if they had less than 90% attendance (N = 181) to the supervised protocol or incomplete data of baseline measurements (N = 7 for heart rate variability analysis, N = 20 for laboratory test), yielding a study sample of 252 participants for the intervention group. Furthermore, participants who deviated from the exercise protocol were strongly encouraged to participate in the follow-up assessment. In those participants, follow-up tests were available for 82 participants, which served as a control group in the current study for the purpose of estimating technical error (TE) and the true individual response variability (SD_{IR}) that elicited by exercise training. TE was estimated by calculating the square root of the sum of squared differences between baseline and follow-up CRF values of the control group divided by the total number of measurements and multiplied by 2.[10] SD_{IR} was obtained by $\sqrt{SD^2_{intervention} - SD^2_{control}}$. The proportion of responders in the population of interest was estimated using an approach developed by Swinton.[27,28] Briefly, the differences in CRF responses to exercise across individuals could be fit into a normal distribution with a mean of the observed change score and an SD of SD_{IR}. With an established minimal clinically important difference (MCID) threshold, the proportion of response could be estimated by calculating the area of the normal distribution that lies beyond the given threshold. An increase of 3.5 ml·kg^{-1}•min^{-1} in VO_{2max} was adopted as the MCID in this study based on its significant correlation with the decrease of all-cause mortality.[29]

2.4. Agglomerative hierarchical clustering

The phenotypic domains in the present study included physical characteristics, heart rate variability parameters, laboratory data, echocardiographic measurements, cardiorespiratory variables obtained by cardiopulmonary exercise test, and pulmonary function parameters. Pearson's correlation was performed between baseline phenotypic characteristics and absolute changes in VO_{2max}. Variables that significantly correlated with the absolute change in VO_{2max} or are known to contribute to the individual variability of exercise response (e.g., age, sex, BMI, baseline fitness) served as candidate inputs for unsupervised clustering. Next, a correlation matrix of variables based on Pearson's coefficient was generated to select the pivotal variables and reduce the dimensions. Variables that were substantially correlated (correlation coefficient >0.6) were filtered (keeping the variable that is more widely used in clinical practice for prognostics or risk predictions), leaving 10 variables for the final clustering. Table 2 summarizes the phenotypic domains and the 10 variables used for the phenomapping analysis. The 10 remaining variables were standardized to a mean of 0 and a standard deviation (SD) of 1. Agglomerative hierarchical clustering, a widely used unsupervised learning method, was performed to group participants with similar baseline characteristics. Briefly, this algorithm considers each sample as an individual cluster initially, and merges the two clusters that are the most similar into a larger cluster at each step, until all samples are in a single large cluster. The current study used Euclidean distance to compute similarity and Ward's method to combine the clusters. A dendrogram diagram was generated to show the progressive grouping of the datasets and gain an idea of a suitable number of clusters (Fig. 1). The number of clusters was chosen based on 30 clustering criteria, using the NbClust package in R (Supplementary Fig. 1). In our implementation, 2 was the optimal number in the range 1–10.

Table 1
A summary of the exercise prescription.

Week	Warm-up, min	Target HR, %HR_{AT}	Training time, min	Cool-down, min
1	5–10	75–85	20–30	5–10
2–4	5–10	85	30	5–10
5–8	5–10	85	40	5–10
9–12	5–10	95	40	5–10
13–16	5–10	100	40	5–10

HR, heart rate; HR_{AT}, HR at the anaerobic threshold.

Table 2
Phenotype domains and variables.

Domains	Variables
Demographics	Age[a]
Physical characteristics	Height, weight, body mass index[a], heart rate, systolic blood pressure, diastolic blood pressure, vital capacity
Cardiorespiratory fitness	Maximal VO_2[a], VO_2 at AT, VO_2 at AT as a percentage of maximal VO_2[a], maximal heart rate[a], heart rate at AT, maximal minute ventilation, minute ventilation at AT[a], O_2 pulse at AT, the work load achieved at AT, exercise time at AT
Laboratory data	White blood cell count, red blood cell count, hemoglobin, hematocrit[a], platelet count, thrombocytocrit, mean corpuscular volume, blood glucose, triglycerides, LDL-cholesterol, HDL-cholesterol, total cholesterol
Echocardiography data	IV septal thickness of end-diastole, IV septal thickness of end-systole[a], LV end diastolic dimension, LV end-systolic dimension, LV posterior wall thickness of end-diastolic, LV posterior wall thickness of end-systole, LV end-diastolic volume, LV end-systolic volume, ejection fraction, stroke volume, fractional shortening, E velocity[a], A velocity
Heart rate variability	RMSSD[a], pNN50, maximal RR interval, minimal RR interval, low frequency, high frequency and total power of the RR interval data
Pulmonary function	Forced vital capacity, forced expiratory volume in 1 s, the ratio of forced expiratory volume in 1 s to forced vital capacity, peak expiratory flow, forced expiratory flow between 25 and 75% of vital capacity, forced expiratory flow at 25%, 50%, and 75%

AT, anaerobic threshold; LDL, low density lipoprotein; HDL, high density lipoprotein; IV, interventricular; LV, left ventricle; RMSSD, root mean square of successive differences of RR intervals; pNN50, percentage of RR intervals differing >50 ms from the preceding one.
[a] The 10 variables used in the final clustering after the feature selection steps.

2.5. Statistical analyses

Data are summarized as mean ± SD for continuous variables and as proportions with percentages for categorical variables. For the sample size calculation, we hypothesized that a 3.5 $ml \cdot kg^{-1} \bullet min^{-1}$ difference in CRF response following the exercise intervention between the two phenogroups was likely and would be clinically meaningful. Based on a previous study, a threshold-based exercise program (30 min/day, 5 days/week, 12-week duration) elicited a 3.93 $ml \cdot kg^{-1} \bullet min^{-1}$ increase in VO_{2max}.[18] Assuming an SD of 5 $ml \cdot min^{-1} \bullet kg^{-1}$, 90% test power, and a 2-sided α value of 0.05, a minimal sample size of 44 participants per group would be required.[30] The normality of the data distribution was confirmed by the Shapiro–Wilk test and, where appropriate, natural log transformation was carried out prior to the analysis to reduce the skewness of the distribution. Group differences in baseline characteristics were analyzed using the independent *t*-test for parametric data and the chi-squared test for non-parametric data. The effects of exercise training on measured variables within each group were examined using the paired *t*-test with Bonferroni correction. The CRF training responsiveness was quantified by the absolute and percentage changes in VO_{2max} before and after the 4-month exercise intervention. ANCOVA was performed with the absolute and percentage changes in VO_{2max} as dependent variables and age, sex, BMI, and baseline VO_{2max} (for absolute change in VO_{2max}) as covariates. The between-group difference in the incidence of population response was examined by the chi-squared test. Levene's test was used to evaluate the between-group difference in the interindividual variability of exercise response of CRF.[31] All statistical analyses were performed using SPSS Statistics (Version 26, IBM Corp., NY, USA), and the significance level was set at 0.05.

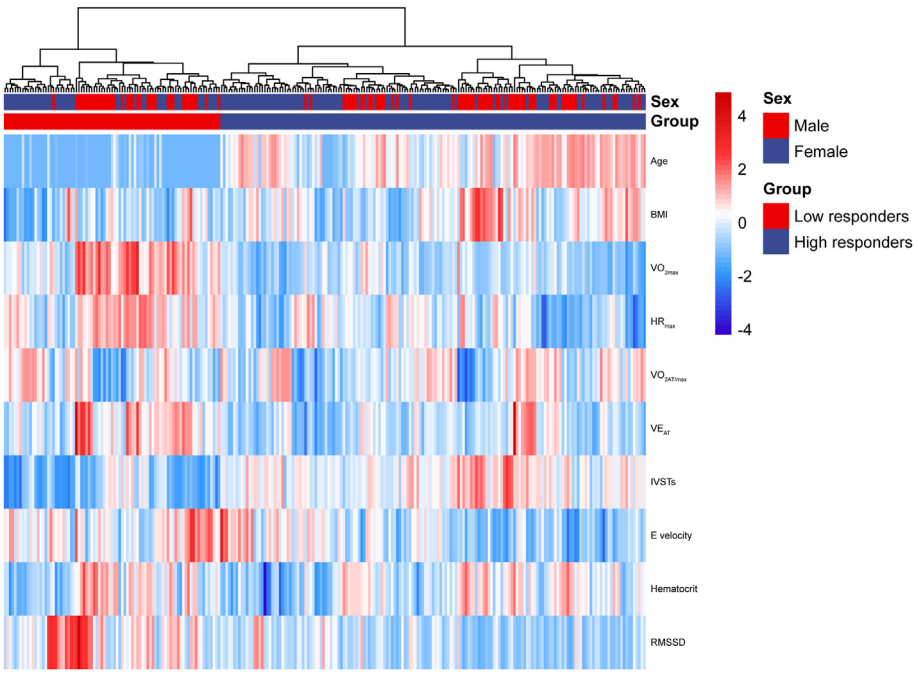

Fig. 1. Heat map of study participants by hierarchical clustering. Columns represent individual participants, and rows represent the selected 10 variables used in the unsupervised clustering. BMI, body mass index; VO_{2max}, maximal oxygen uptake; HR_{max}, maximal heart rate; $VO_{2AT/max}$, VO_2 at anaerobic threshold as a percentage of maximal VO_2; VE_{AT}, minute ventilation at anaerobic threshold; IVSTs, interventricular septal thickness of end-systole; RMSSD, root mean square of successive differences of RR intervals.

3. Results

The characteristics of the study participants are summarized in Table 3. The exercise group (N = 252) had a mean age of 36.8 ± 13.4 yr and a mean BMI of 23.3 ± 2.8 kg·m^{-2}. Most of the study participants were non-smokers, and 60% were women. Similar clinical characteristics were observed for participants in the control group.

3.1. Hierarchical clustering

Among the 252 healthy participants, hierarchical clustering based on 10 phenotypic variables (As stated in the methods, variables that were correlated at r > 0.6 were filtered, leaving 10 minimally redundant variables) yielded 2 phenogroups. Fig. 1 shows a heat map created using agglomerative hierarchical clustering. Within the heat map, the values of the 10 features were presented in varying patterns, and individuals with shared characteristic patterns were clustered together. For example, most of the individuals in group 1 were younger and had a lower interventricular septal thickness of end-systole (IVSTs) and higher CRF, but most individuals in group 2 were older, with higher IVSTs and lower CRF. Additionally, low CRF seemed to occur in some individuals with higher resting vagal activity in group 1, but lower vagal function in group 2.

3.2. Comparisons of clinical characteristics and baseline laboratory, echocardiographic, and cardiorespiratory variables between phenogroups

The baseline characteristics were significantly different between the two phenogroups. As shown in Table 4, group 1 was significantly younger and had a lower BMI (22.1 ± 2.7 vs 23.9 ± 2.7 kg•m^{-2}, $P < 0.001$). Sedentary time was slightly higher in group 2, but group 1 slept longer than group 2 (7.9 ± 1.0 vs 7.4 ± 0.9 h, $P = 0.001$). Leisure time physical activity level was similar between the two groups. With respect to vital signs, the two groups presented with similar resting HR and systolic blood pressure (SBP), but diastolic blood pressure (DBP) was higher in group 2 (70.4 ± 7.8 vs 73.2 ± 8.4 mm Hg, $P = 0.014$). Furthermore, group 1 had a significantly greater vital capacity than group 2. In the heart rate variability analysis, group 1 had a higher root mean square of successive differences of RR intervals (RMSSD) than group 2 (53.4 ± 27.6 vs 26.9 ± 14.4 ms, $P < 0.001$). Furthermore, laboratory data also exhibited significance between the two groups. For example, the level of low density lipoprotein cholesterol (LDL-cholesterol) and total cholesterol were significantly higher in group 2.

Table 4 also shows the differences in cardiac structure and function between the two groups. As expected, group 2 had higher interventricular septal thickness and posterior wall thickness compared to group 1. Furthermore, the left ventricular (LV) relaxation was worse in group 2, as indicated by lower E velocity and A velocity. On the other hand, group 1 had lower systolic function than group 2 (i.e., lower ejection fraction, stroke volume, and fractional shortening, but higher LV dimension and volume at systolic, $P < 0.001$). Despite these differences, the LV end-diastolic dimensions and volumes were similar between the groups.

At baseline, there was a significant difference in CRF between the two phenogroups (Table 4). Both maximal and submaximal aerobic capacities were higher in group 1 than in group 2. For example, the mean VO_{2max} and VO_{2AT} for group 1 were significantly higher than group 2 (36.0 ± 7.4 vs 25.9 ± 4.3 ml•kg^{-1}•min^{-1}, 25.4 ± 5.6 vs 18.5 ± 3.7 ml•kg^{-1}•min^{-1}, $P < 0.001$). Nevertheless, the two groups achieved similar percentages of maximal VO_2 at AT (71.6 ± 11.3 vs 71.8 ± 10.9 %, $P = 0.865$).

Table 3
Baseline participant characteristics.

Clinical characteristic	All participants ($N = 460$)	Control group ($N = 82$)	Exercise group ($N = 252$)
Age, y	36.1 ± 13.2	37.2 ± 11.1	36.8 ± 13.4
Female, n (%)	266 (57)	46 (56)	153 (60)
Height, cm	165.2 ± 8.1	165.3 ± 8.2	164.7 ± 7.9
Weight, kg	63.5 ± 10.8	65.0 ± 10.3	63.4 ± 10.1
BMI, kg•m^{-2}	23.2 ± 3.1	23.7 ± 3.3	23.3 ± 2.8
Sleep, h	7.6 ± 1.0	7.5 ± 0.9	7.6 ± 1.0
Leisure time physical activity, h/week	24.7 ± 61.5	25.0 ± 66.6	22.0 ± 55.0
Sedentary time, h	6.2 ± 2.7	6.6 ± 2.4	6.2 ± 2.8
Current smoking, n (%)	79 (17)	17 (20)	44 (17)
Former smoking, n (%)	27 (5)	6 (7)	13 (5)
Treadmill running exercise, n (%)	245 (53)	35 (42)	120 (47)
Vital signs			
HR, bpm	67.4 ± 9.3	68.5 ± 9.7	66.6 ± 9.2
SBP, mm Hg	110.6 ± 10.9	110.6 ± 11.8	110.6 ± 10.9
DBP, mm Hg	72.7 ± 8.4	73.4 ± 9.2	72.3 ± 8.3
VC, ml	3377.1 ± 933.5	3314.1 ± 847.6	3334.4 ± 918.9
Laboratory data			
WBC count, × 10^9 L^{-1}	5.8 ± 1.3	6.3 ± 1.5	6.0 ± 1.3
RBC count, × 10^{12} L^{-1}	4.6 ± 0.4	4.6 ± 0.5	4.5 ± 0.4
Hemoglobin, g•L^{-1}	138.9 ± 13.5	141.6 ± 15.2	139.8 ± 14.4
Hematocrit, %	42.2 ± 4.0	42.8 ± 4.5	42.2 ± 4.2
Platelet count, × 10^{11} L^{-1}	191.5 ± 43.2	191.5 ± 42.5	199.0 ± 46.0
Glucose, mmol•L^{-1}	4.2 ± 0.5	4.2 ± 0.4	4.2 ± 0.6
Triglycerides, mmol•L^{-1}	1.3 ± 0.8	1.3 ± 0.9	1.2 ± 0.6
LDL-cholesterol, mmol•L^{-1}	2.8 ± 0.7	2.6 ± 0.6	2.7 ± 0.8
HDL-cholesterol, mmol•L^{-1}	1.4 ± 0.3	1.4 ± 0.3	1.4 ± 0.3
Total cholesterol, mmol•L^{-1}	4.4 ± 0.8	4.3 ± 0.7	4.4 ± 0.9
Cardiorespiratory fitness			
VO_{2max}, ml•kg^{-1}•min^{-1}	30.4 ± 7.8	29.6 ± 6.5	29.3 ± 7.3
HR_{max}, bpm	168.1 ± 15.2	169.1 ± 14.3	167.2 ± 15.1

Data are expressed as mean ± SD for continuous variables and n (%) for categorical variables. BMI, body mass index; HR, heart rate; SBP, systolic blood pressure; DBP, diastolic blood pressure; VC, vital capacity; WBC, white blood cell; RBC, red blood cell; LDL, low density lipoprotein; HDL, high density lipoprotein.

Table 4
Baseline characteristics and changes in cardiorespiratory fitness and cardiometabolic parameters stratified by phenogroup.

	Group 1 Low responders ($N = 85$)	Group 2 High responders ($N = 167$)	*P* Value
Age, y	22.6 ± 4.7	44.0 ± 10.2	<0.001
Female, n (%)	47 (55)	106(63)	0.209
Height, cm	166.8 ± 8.5	163.7 ± 7.3	0.003
Weight, kg	61.8 ± 10.7	64.2 ± 9.8	0.076
BMI, kg•m^{-2}	22.1 ± 2.7	23.9 ± 2.7	<0.001
Sleep, h	7.9 ± 1.0	7.4 ± 0.9	0.001
Leisure time physical activity, h/week	21.5 ± 44.9	22.3 ± 59.6	0.911
Sedentary time, h	5.7 ± 2.9	6.5 ± 2.7	0.047
Current smoking, n (%)	14(16)	30(17)	0.322
Former smoking, n (%)	2 (2)	11(6)	0.322
Treadmill running exercise, n (%)	42(49)	79(47)	0.752
Vital signs			
HR, bpm	65.8 ± 9.4	67.0 ± 9.1	0.337
SBP, mm Hg	110.0 ± 11.3	110.8 ± 10.7	0.557
DBP, mm Hg	70.4 ± 7.8	73.2 ± 8.4	0.014
VC, ml	3722.3 ± 1072.9	3130.6 ± 753.9	<0.001
Heart rate variability			
RMSSD, ms	53.4 ± 27.6	26.9 ± 14.4	<0.001
nLF, nu	0.45 ± 0.17	0.56 ± 0.18	<0.001
nHF, nu	0.54 ± 0.17	0.43 ± 0.18	<0.001
Laboratory data			
WBC count, × 10^{9} L^{-1}	6.5 ± 1.4	5.8 ± 1.3	<0.001
RBC count, × 10^{12} L^{-1}	4.6 ± 0.4	4.5 ± 0.4	0.116
Hemoglobin, g•L^{-1}	142.4 ± 14.6	138.5 ± 14.2	0.046
Hematocrit, %	42.9 ± 4.3	41.9 ± 4.1	0.065
Platelet count, × 10^{11} L^{-1}	213.6 ± 42.6	191.6 ± 46.1	<0.001
Glucose, mmol•L^{-1}	4.0 ± 0.4	4.3 ± 0.7	<0.001
Triglycerides, mmol•L^{-1}	0.9 ± 0.4	1.3 ± 0.6	<0.001
LDL-cholesterol, mmol•L^{-1}	2.3 ± 0.7	2.9 ± 0.8	<0.001
HDL-cholesterol, mmol•L^{-1}	1.3 ± 0.2	1.4 ± 0.3	0.129
Total cholesterol, mmol•L^{-1}	3.9 ± 0.7	4.6 ± 0.9	<0.001
Echocardiography			
IVSTd, mm	7.1 ± 1.1	8.3 ± 1.5	<0.001
IVSTs, mm	8.8 ± 2.9	13.3 ± 2.1	<0.001
LVDd, mm	45.9 ± 4.6	45.9 ± 4.8	0.981
LVDs, mm	32.4 ± 3.6	30.3 ± 4.2	<0.001
LVPWTd, mm	7.0 ± 1.5	8.0 ± 1.6	<0.001
LVPWTs, mm	9.5 ± 3.1	12.5 ± 2.2	<0.001
LVEDV, ml	98.1 ± 22.2	98.7 ± 24.5	0.851
LVESV, ml	43.3 ± 11.2	37.2 ± 12.8	<0.001
Ejection fraction, %	55.5 ± 7.8	62.2 ± 7.7	<0.001
Stroke volume, ml•min^{-1}	54.7 ± 15.9	61.4 ± 15.8	0.002
Fractional shortening, %	28.9 ± 5.4	33.9 ± 5.4	<0.001
E velocity, cm•s^{-1}	84.0 ± 10.3	75.5 ± 11.0	<0.001
A velocity, cm•s^{-1}	43.9 ± 14.4	40.1 ± 12.4	0.033
Cardiorespiratory fitness			
VO_{2max}, ml•kg^{-1}•min^{-1}	36.0 ± 7.4	25.9 ± 4.3	<0.001
HR_{max}, bpm	176.9 ± 13.2	162.3 ± 13.6	<0.001
VE_{max}, L•min^{-1}	74.0 ± 24.2	62.5 ± 16.7	<0.001
VO_{2AT}, ml•kg^{-1}•min^{-1}	25.4 ± 5.6	18.5 ± 3.7	<0.001
$VO_{2AT/max}$, %	71.6 ± 11.3	71.8 ± 10.9	0.865
HR_{AT}, bpm	142.5 ± 15.8	126.9 ± 13.0	<0.001
VE_{AT}, L•min^{-1}	44.7 ± 12.6	35.0 ± 10.2	<0.001
Changes in cardiorespiratory fitness			
ΔVO_{2max}, ml•kg^{-1}•min^{-1}	0.58 ± 0.65	3.42 ± 0.40	0.002
% Δ in VO_{2max}, %	1.45 ± 2.38	14.9 ± 1.46	<0.001
Proportion of population response, %	56.1%	13.9%	<0.001
Changes in metabolic parameters			
ΔSBP, mm Hg	–	–	
ΔDBP, mm Hg	−1.7 ± 7.2	–	
ΔLDL-cholesterol, mmol•L^{-1}	–	–	
ΔHDL-cholesterol, mmol•L^{-1}	0.07 ± 0.21	0.12 ± 0.26	0.099
ΔTotal cholesterol, mmol•L^{-1}	0.10 ± 0.50	0.22 ± 0.64	0.141

Data are expressed as mean ± SD for continuous variables and n (%) for categorical variables. ΔVO_{2max} values were adjusted for age, sex, BMI, and baseline $VO_{2max.}$ %Δ in VO_{2max} values were adjusted for age, sex, and BMI. BMI, body mass index; HR, heart rate; SBP, systolic blood pressure; DBP, diastolic blood pressure; VC, vital capacity; RMSSD, root mean square of successive differences of RR intervals; nLF, normalized low frequency power of the RR interval data; nHF, normalized high frequency power of the RR interval data; WBC, white blood cell; RBC, red blood cell; LDL, low density lipoprotein; HDL, high density lipoprotein; IVSTd, interventricular septal thickness of end-diastole; IVSTs, interventricular septal thickness of end-systole; LVDd, Left ventricular end-diastolic dimension; LVDs, left ventricular end-systolic dimension; LVPWTd, left ventricular posterior wall thickness of end-diastolic; LVPWTs, left ventricular posterior wall thickness of end-systole; LVEDV, left ventricle end-diastolic volume; LVESV, left ventricle end-systolic volume; VO_{2max}, maximal oxygen uptake; HR_{max}, maximal heart rate; VE_{max}, maximal minute ventilation; VO_{2AT}, oxygen uptake at anaerobic threshold; $VO_{2AT/max}$, VO_2 at anaerobic threshold as a percentage of maximal VO_2; HR_{AT}, heart rate at anaerobic threshold; VE_{AT}, minute ventilation at anaerobic threshold.

3.3. Changes in CRF

After the 16-week intervention, group 2 showed significant improvement in CRF at group level. VO_{2max} significantly increased from 25.9 ± 4.3 to 29.6 ± 5.7 ml•kg^{-1}•min^{-1} (N = 167, $P < 0.001$) for group 2. However, the change in VO_{2max} (36.0 ± 7.4 vs 35.9 ± 7.0 ml•kg^{-1}•min^{-1}, N = 85, $P = 0.875$) was not statistically significant in group 1. There were significant between-group differences in absolute and percent changes in VO_{2max} after adjusting for age, sex, BMI and baseline VO_{2max} (Table 4). The absolute (0.58 ± 0.65 vs 3.42 ± 0.40 ml•kg^{-1}•min^{-1}, $P = 0.002$) and percent changes in VO_{2max} (1.45 ± 2.38 vs 14.9 ± 1.46 %, $P < 0.001$) after the exercise intervention was significantly larger in group 2 compared with group 1. The individual changes in CRF following the 4-month exercise training for group 1 and group 2 are illustrated in Fig. 2. Levene's test revealed a significantly reduced response variability in VO_{2max} in group 2 compared to group 1 ($P < 0.001$). Furthermore, the proportion of population response for group 2 was significantly higher than for group 1 (56.1% vs 13.9%, $P < 0.001$).

3.4. Changes in other parameters

After the 16 weeks of exercise training, changes in SBP, glucose, triglycerides, and LDL-cholesterol were not significantly different within or between the groups. However, in both groups, high density lipoprotein cholesterol (HDL-cholesterol) significantly increased after exercise intervention. The increments in HDL-cholesterol were similar between the two groups (0.07 ± 0.21 vs 0.12 ± 0.26 mmol•L^{-1}, $P = 0.099$). Interestingly, for group 1, there was a significant decrease in DBP from 70.4 ± 7.8 to 68.7 ± 7.2 mm Hg ($P = 0.027$) while no similar findings were noted for group 2.

Exercise intervention also elicited significant improvements in cardiac function in both groups. For example, parameters reflecting systolic (ejection fraction, stroke volume, and fractional shortening) and diastolic functions (LV end-diastolic dimension and LV end-diastolic volume) were significantly increased in both groups (Supplementary Fig. 2). However, only group 2 exhibited improvements in LV relaxation parameters, such as E velocity and A velocity, following exercise training.

4. Discussion

The present study investigated the variability of CRF responses to a moderate 16-week training protocol in 252 healthy individuals using unsupervised cluster analysis. We report three novel findings. First, within a healthy population, unsupervised clustering successfully identified 2 groups with distinct baseline characteristics and exercise responses of CRF. Second, one of the two clusters (group 2) exhibited a marked improvement in CRF following the given dose of aerobic training, which was associated with not only a larger mean change but also reduced interindividual variation. Third, although group 1 exhibited no improvement in CRF after 4 months of endurance training, other exercise benefits, including a decrease in DBP and increase in cardiac function, were observed. To our knowledge, this is the first study to apply unsupervised learning techniques to resolve the heterogeneity in CRF response to exercise using multidimensional clinical measures. Our innovative findings provide novel insights into the growing body of literature on exercise response variability and are of particular interest for personalized exercise medicine.

Variability in CRF responsiveness following a standardized training intervention has been linked to many phenotypic variables.[13] In line with this, our study identified multiple pre-training phenotypes that were significantly correlated with the exercise response of CRF, and 10 variables were used in the unsupervised learning analysis, including markers of fitness, cardiac function, autonomic function and blood oxygen transport capacity. Investigating which phenotypes or combinations of a number of phenotypes could be used to predict an individual's training responsiveness will be practically meaningful for clinicians to identify and design personalized training regimens, given the considerable interindividual variation in exercise responsiveness.[20] However, the presence of a surge of input factors and their potential interactions not only makes it difficult to develop an accurate statistical model but also brings the multiple testing problem.[6] Consequently, inferences from traditional statistical methods become less precise. On the other hand, machine learning techniques, including supervised and unsupervised methods, make predictions by finding patterns within large volumes of data and can be effective even when complicated nonlinear associations exist. It is important to note that unsupervised learning may be more appropriate than supervised learning when considering the heterogeneity of CRF response following exercise training owing to the issue of random error. The observed variance of CRF between baseline and after exercise training will always be contaminated by random variation, which contains the measurement error from the experimental tool/protocol and the within-subject variation from biological and environmental sources.[32,33] Especially, the random

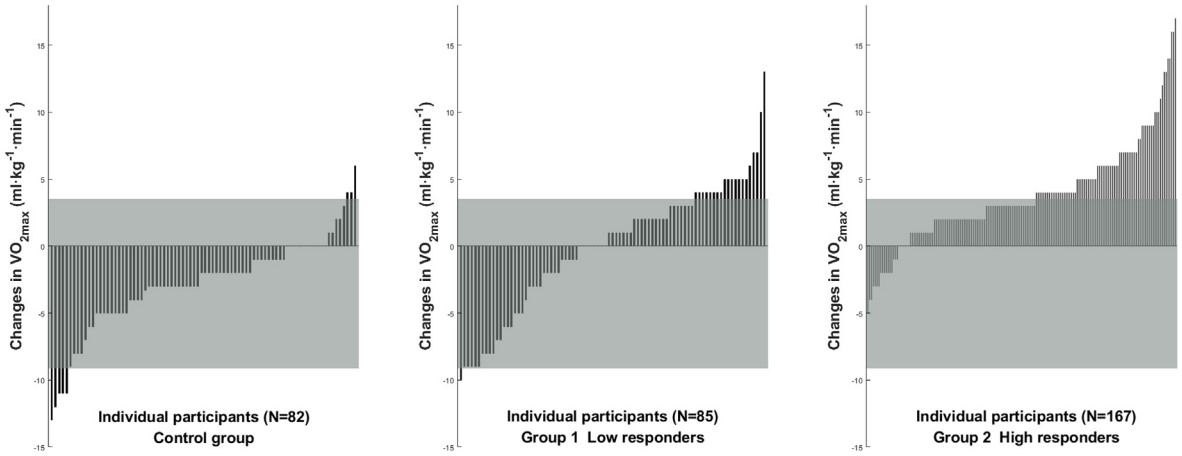

Fig. 2. Change in VO_{2max} following the 16-week exercise training for each participant in control and exercise groups. The technical error for CRF measurement is illustrated by the shaded area.

within-subject biological variation is particularly large when the time period between baseline and follow-up measurements is relatively long, which is often the case in exercise training studies.[32] It is plausible that supervised approaches using the observed difference of CRF measured before and after exercise training as the learning target are grouping individuals entirely based on random error.[32] By contrast, unsupervised learning does not require a target output, and learns to find intrinsic patterns within the data on its own. Such methods could discover homogeneous subgroups within a given population according to the baseline attributes they have.[34] Using unsupervised learning, we were able to take advantage of the dense phenotypic data to discover natural subgroups of study participants that were associated with different CRF training responses.

The two groups in our study exhibited a significant difference in VO_{2max} response following 4 months of standardized aerobic training. Specifically, the proportion of population response for group 2 (high responders) was 56.1% with MCID set at 3.5 ml•kg^{-1}•min^{-1}, whereas group 1 (low responders) only had a 13.9% incidence of population response. Additionally, group 1 demonstrated no improvements in VO_{2max} at the group level, while a significant increase in group mean was found in group 2. The ratio of the number of participants in groups 1 and 2 is approximately 1:2, which is consistent with previous estimations that 20-30% of individuals may fail to benefit from an exercise training intervention in terms of CRF.[5,11,35] Several biological factors have been studied that may affect the response rate of CRF, but no conclusive results have been achieved.[13] For example, while the effects of sex and initial fitness on CRF trainability have been suggested in some studies,[29,36] others reported no associations between sex, initial fitness, and CRF response.[13,37] Likewise, an association between age and the response rate of CRF following endurance training was reported in postmenopausal women (45-75 yr),[35] whereas no significant effect of age on CRF exercise response was found in an elderly cohort (60-85 yr).[38] The HERITAGE study,[37] which incorporated a relatively large age range (17-65 yr), suggests a lower CRF response for subjects aged 50–65 yr (4.5 ± 2.3 ml•kg^{-1}•min^{-1}) compared to the other two younger groups (17-29 yr: 5.3 ± 2.6 ml•kg^{-1}•min^{-1}, 30-49 yr: 5.7 ± 3.1 ml•kg^{-1}•min^{-1}).[37] By pooling data of the current study into the same age ranges of the HERITAGE study, similar results were observed after adjusting for baseline fitness with the older group showing slightly lower CRF response (50-59 yr: 1.9 ± 0.6 ml•kg^{-1}•min^{-1}) than the two younger groups (20-29 yr: 2.1 ± 0.5 ml•kg^{-1}•min^{-1}, 30-49 yr: 3.3 ± 0.4 ml•kg^{-1}•min^{-1}), but the difference did not the reach statistical significance ($P = 0.069$). In this study, the two groups identified by unsupervised cluster analysis exhibited a significant difference in training response of VO_{2max} after adjusting for age, sex, BMI, and baseline fitness. However, age seemed to play a prominent role in phenotyping participants into the two groups in the current study. One possible explanation for this observation may be that age was significantly correlated with several other input variables. Although highly correlated variables were excluded using a correlation coefficient threshold of 0.6 in accordance with previous studies,[25,26] the redundancy across age and other variables may still exist. Selection of the optimal set of variables for unsupervised learning may be less straightforward than for supervised learning when the importance of input variables could be evaluated by their contributions in predicting the output variable.[24] Future studies assessing whether incorporating data-driven dimension reduction techniques, such as principal component analysis and singular value decomposition, could mitigate collinearity and promote CRF response phenotyping are of great interest.

Due to the "black-box" nature of machine learning techniques, we cannot make a comprehensive interpretation of how modulators used in the current study interact with each other and contribute to CRF response variability. Nevertheless, our model could be used to predict whether a participant is a relatively high or low responder in terms of CRF, following moderate endurance training. In this case, our analysis could serve as a powerful tool in clinical settings for health professionals to develop effective and personalized training strategies. For example, if phenomapping classifies a certain participant that may have a low training response for CRF following moderate aerobic training, strategies demonstrated to decrease or even eliminate the CRF "non-response", such as increasing exercise intensity/volume or changing training modality, could be considered to create an individually optimal training prescription.[14,39,40] This is of great interest for promoting population health, considering the large dropout caused by modest returns on training investment, as well as the concerns arising from increasing exercise intensities for all individuals.[41–43] Notably, although our study and a number of previous studies considered CRF as the key training adaptation to determine a higher or lower training response, exercise clearly exerts other health benefits.[44–46] The HART-D study demonstrated that exercise elicited a significant improvement in glycemic control in diabetic individuals that exhibited no improvement in CRF following 9-month exercise training.[11] In our study, although group 1 exhibited no significant improvement in VO_{2max} following the 16-week endurance training program, a significant reduction in DBP (1.7 ± 7.2 mm Hg) was observed. Even small reductions in blood pressure could significantly decrease the risk of cardiovascular events and mortality.[47] For instance, a reduction of 2 mm Hg in SBP was associated with approximately 10% and 7% lower mortality from stroke and ischemic heart disease, respectively.[47] Although the predictive power of DBP for mortality is slightly less than SBP (e.g. informativeness of SBP and DBP for the prediction of stroke was 89% and 83%, respectively), it is plausible that a comparable reduction in DBP as observed in the current study as well as other lifestyle interventions (e.g. reducing dietary sodium) may have important clinical significance in the prevention of cardiovascular mortality.[47,48] Collectively, evaluating exercise response in CRF is clearly important, as improvements in CRF appear to have more health benefits than other parameters.[14,49] Nevertheless, it may be important to pay attention to other favorable effects of exercise when evaluating the success of a training intervention, especially for the patient population, in which parameters most relevant to certain diseases should be considered, such as fasting glucose in type 2 diabetes or blood pressure in hypertension.[14] As such, future studies targeting effective exercise interventions for other parameters with different patient populations are warranted.

4.1. Limitations

The present study had several limitations. First, all participants were enrolled from a single center, and our study lacked validation in an external cohort. Future studies that replicate our phenomapping techniques using datasets from other institutions are warranted. Second, some information was gathered based on participant's self-report (e.g., smoking status, leisure time physical activity, sedentary time, etc.). Although these variables were not included in the clustering analysis, the comparisons of these factors between the two phenotypic groups may be confounded. Third, other biological factors (such as genetics and circulating hemoglobin mass) and lifestyle factors (such as changes in physical activity level, food intake, and sleep time during the exercise intervention), which could also influence CRF trainability, were not considered in this study. Further studies are needed to quantify the contributions of these factors to CRF training response. Finally, the current model requires a certain amount of clinical measures for

efficient analysis, this may limit its widespread clinical application as many facilities may not have accesses to all the measures performed in this study.

5. Conclusions

In conclusion, this is the first study to conduct unsupervised learning on dense clinical data to investigate the heterogeneity of CRF response to endurance training. We have shown that unsupervised clustering of multiple phenotypic variables related to the CRF exercise response can result in meaningful, clinically relevant subgroups in a healthy population with significant differences in CRF response to a 16-week aerobic training program. Given the considerable individual differences in CRF training response, incorporation of phenomapping into clinical practice could help physicians design more efficacious and individualized exercise prescriptions and may lead to the development of precision exercise medicine. Furthermore, individuals who experienced no significant improvements in CRF following aerobic training could also benefit from exercise in other clinical measures, such as blood pressure and cardiac function.

Author statement

Lin Xie: Conceptualization, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, **Bo Gou:** Conceptualization, Investigation, Resources, Writing - Review & Editing, **Shuwen Bai:** Resources, Investigation, **Dong Yang:** Resources, Investigation, **Zhe Zhang:** Investigation, **Xiaohui Di:** Formal analysis, Visualization, **Chuanwang Su:** Software, Writing - Review & Editing, **Xiaoni Wang:** Formal analysis, Writing - Review & Editing, **Kun Wang:** Investigation, Resources, Writing - Review & Editing, **Jianbao Zhang:** Conceptualization, Writing - Original Draft, Writing - Review & Editing, Supervision.

Declaration of competing interest

The authors have no conflicts of interest relevant to this article.

Acknowledgments

This work was supported by the National Natural Science Foundation for Young Scholars of China (grant number: 32000978) and the National Natural Science Foundation of China (grant number: 11872297). The authors would like to thank all the technicians and participants for their time and effort devoted to this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesf.2022.12.005.

References

1. Ross R, Blair S, Arena R, et al. Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement from the American Heart Association. *Circulation*. 2016;134:e653–e699.
2. Laukkanen JA, Kurl S, Salonen R, et al. The predictive value of cardiorespiratory fitness for cardiovascular events in men with various risk profiles: a prospective population-based cohort study. *Eur Heart J*. 2004;25:1428–1437.
3. Kodama S, Saito K, Tanaka S, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. *JAMA*. 2009;301:2024–2035.
4. McGee SL, Hargreaves M. Exercise adaptations: molecular mechanisms and potential targets for therapeutic benefit. *Nat Rev Endocrinol*. 2020;16:495–505.
5. Montero D, Lundby C. Refuting the myth of non-response to exercise training: 'non-responders' do respond to higher dose of training. *J Physiol*. 2017;595:3377–3387.
6. Ross R, Goodpaster BH, Koch LG, et al. Precision exercise medicine: understanding exercise response variability. *Br J Sports Med*. 2019;53:1141–1153.
7. Bouchard C, Rankinen T. Individual differences in response to regular physical activity. *Med Sci Sports Exerc*. 2001;33:S446–S453.
8. Weatherwax RM, Harris NK, Kilding AE, et al. Incidence of VO2max responders to personalized versus standardized exercise prescription. *Med Sci Sports Exerc*. 2019;51:681–691.
9. Yu F, Salisbury D, Mathiason MA. Inter-individual differences in the responses to aerobic exercise in Alzheimer's disease: findings from the FIT-AD trial. *J Sport Health Sci*. 2021;10:65–72.
10. Ross R, Lannoy L, Stotz PJ. Separate effects of intensity and amount of exercise on interindividual cardiorespiratory fitness response. *Mayo Clin Proc*. 2015;90:1506–1514.
11. Pandey A, Swift DL, McGuire DK, et al. Metabolic effects of exercise training among fitness-nonresponsive patients with Type 2 Diabetes: the HART-D Study. *Diabetes Care*. 2015;38:1494–1501.
12. Buford TW, Roberts MD, Church TS. Toward exercise as personalized medicine. *Sports Med*. 2013;43:157–165.
13. Meyler S, Bottoms L, Muniz-Pumares D. Biological and methodological factors affecting VO2max response variability to endurance training and the influence of exercise intensity prescription. *Exp Physiol*. 2021;106:1410–1424.
14. Pickering C, Kiely J. Do non-responders to exercise exist-and if so, what should we do about them? *Sports Med*. 2019;49:1–7.
15. Kirton MJ, Burnley MT, Ramos JS, et al. The effects of standardised versus individualised aerobic exercise prescription on fitness-fatness index in sedentary adults: a randomised controlled trial. *J Sports Sci Med*. 2022;21:347–355.
16. Iannetta D, Keir D, Fontana F, et al. Evaluating the accuracy of using fixed ranges of METs to categorize exertional intensity in a heterogeneous group of healthy individuals: implications for cardiorespiratory fitness and health outcomes. *Sports Med*. 2021;51:2411–2421.
17. Weatherwax RM, Harris NK, Kilding AE, et al. The incidence of training responsiveness to cardiorespiratory fitness and cardiometabolic measurements following individualized and standardized exercise prescription: study protocol for a randomized controlled trial. *Trials*. 2016;17:1–12.
18. Wolpern AE, Burgos DJ, Janot JM, et al. Is a threshold-based model a superior method to the relative percent concept for establishing individual exercise intensity? a randomized controlled trial. *BMC Sports Science, Medicine and Rehabilitation*. 2015;7:1–9.
19. Kirton MJ, Burnley MT, Ramos JS, et al. The effects of standardised versus individualised aerobic exercise prescription on fitness-fatness index in sedentary adults: a randomised controlled trial. *J Sports Sci Med*. 2022;21:347–355.
20. Mann TN, Lamberts RP, Lambert MI. High responders and low responders: factors associated with individual variation in response to standardized training. *Sports Med*. 2014;44:1113–1124.
21. Voisin S, Jacques M, Lucia A, et al. Statistical considerations for exercise protocols aimed at measuring trainability. *Exerc Sport Sci Rev*. 2019;47:37–45.
22. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15:233–234.
23. Deo RC. Machine learning in medicine. *Circulation*. 2015;132:1920–1930.
24. Stevens LM, Mortazavi BJ, Deo RC, et al. Recommendations for reporting machine learning analyses in clinical research. *Circ Cardiovasc Qual Outcomes*. 2020;13, e006556.
25. Shah SJ, Katz DH, Selvaraj S, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*. 2015;131:269–279.
26. Kwak S, Lee Y, Ko T, et al. Unsupervised cluster analysis of patients with aortic stenosis reveals distinct population with different phenotypes and outcomes. *Circ Cardiovasc Imaging*. 2020;13, e009707.
27. Swinton P, Hemingway B, Saunders B, et al. A statistical framework to interpret individual response to intervention: paving the way for personalized nutrition and exercise prescription. *Front Nutr*. 2018;5:41.
28. Atkinson G, Williamson P, Batterham A. Issues in the determination of 'responders' and 'non-responders' in physiological research. *Exp Physiol*. 2019;104:1215–1225.
29. Williams CJ, Gurd BJ, Bonafiglia JT, et al. A multi-center comparison of O2peak trainability between interval training and moderate intensity continuous training. *Front Physiol*. 2019;10:19.
30. Faul F, Erdfelder E, Lang A, et al. *Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39:175–191.
31. Bonafiglia JT, Preobrazenski N, Islam H, et al. Exploring differences in cardiorespiratory fitness response rates across varying doses of exercise training: a retrospective analysis of eight randomized controlled trials. *Sports Med*. 2021;51:1785–1797.
32. Atkinson G, Batterham AM. True and false interindividual differences in the physiological response to an intervention. *Exp Physiol*. 2015;100:577–588.
33. Dankel SJ, Loenneke JP. A method to stop analyzing random error and start analyzing differential responders to exercise. *Sports Med*. 2020;50:231–238.
34. Kagiyama N, Shrestha S, Farjo PD, et al. Artificial intelligence: practical primer for clinical research in cardiovascular disease. *J Am Hear Assoc Cardiovasc Cerebrovasc Dis*. 2019;8, e012788.
35. Sisson SB, Katzmarzyk PT, Earnest CP, et al. Volume of exercise and fitness nonresponse in sedentary, postmenopausal women. *Med Sci Sports Exerc*.

2009;41:539–545.
36. Diaz-Canestro C, Montero D. Sex dimorphism of VO2max trainability: a systematic review and meta-analysis. *Sports Med*. 2019;49:1949–1956.
37. Skinner JS, Jaskólski A, Jaskólska A, et al. Age, sex, race, initial fitness, and response to training: the HERITAGE Family Study. *J Appl Physiol*. 1985;90:1770–1776, 2001.
38. Huang G, Wang R, Chen P, et al. Dose-response relationship of cardiorespiratory fitness adaptation to controlled endurance training in sedentary older adults. *Eur J Prev Cardiol*. 2016;23:518–529.
39. Hautala AJ, Kiviniemi AM, Mäkikallio TH, et al. Individual differences in the responses to endurance and resistance training. *Eur J Appl Physiol*. 2006;96:535–542.
40. Marsh CE, Thomas HJ, Naylor LH, et al. Fitness and strength responses to distinct exercise modes in twins: studies of Twin Responses to Understand Exercise as a THerapy (STRUETH) study. *J Physiol*. 2020;598:3845–3858.
41. Zubin Maslov P, Schulman A, Lavie CJ, et al. Personalized exercise dose prescription. *Eur Heart J*. 2018;39:2346–2355.
42. Biddle SJ, Batterham AM. High-intensity interval exercise training for public health: a big HIT or shall we HIT it on the head? *Int J Behav Nutr Phys Activ*. 2015;12:1–8.
43. Williamson W, Foster C, Reid H, et al. Will exercise advice be sufficient for treatment of young adults with prehypertension and hypertension? A systematic review and meta-analysis. *Hypertension*. 2016;68:78–87.
44. Egan BM. Physical activity and hypertension: knowing is not enough; we must apply. Willing is not enough; we must do-von Goethe. *Hypertension*. 2017;69:404–406.
45. Pedersen BK. The physiology of optimizing health with a focus on exercise as medicine. *Annu Rev Physiol*. 2018;81:1–21.
46. Tanaka H. Antiaging effects of aerobic exercise on systemic arteries. *Hypertension*. 2019;74:237–243.
47. Lewington S, Clarke R, Qizilbash N, et al. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet (London, England)*. 2002;360:1903–1913.
48. Sacks F, Svetkey L, Vollmer W, et al. Effects on blood pressure of reduced dietary sodium and the dietary approaches to stop hypertension (DASH) diet. DASH-sodium collaborative research group. *N Engl J Med*. 2001;344:3–10.
49. Raghuveer G, Hartz J, Lubans DR, et al. Cardiorespiratory fitness in youth: an important marker of health: a scientific statement from the American Heart Association. *Circulation*. 2020;142:e101–e118.