Validation of Exercise Capacity as a Surrogate Endpoint in Exercise-Based Rehabilitation for Heart Failure

A Meta-Analysis of Randomized Controlled Trials

Oriana Ciani, PhD,[a,b] Massimo Piepoli, MD,[c] Neil Smart, PhD,[d] Jamal Uddin, MSc,[e,f] Sarah Walker, PhD,[a]
Fiona C. Warren, PhD,[a] Ann D. Zwisler, MD,[f,g] Constantinos H. Davos, PhD,[h] Rod S. Taylor, PhD[a,f,g]

ABSTRACT

OBJECTIVES This study sought to validate exercise capacity (EC) as a surrogate for mortality, hospitalization, and health-related quality of life (HRQOL).

BACKGROUND EC is often used as a primary outcome in exercise-based cardiac rehabilitation (CR) trials of heart failure (HF) via direct cardiorespiratory assessment of maximum oxygen uptake (Vo_2peak) or through submaximal tests, such as the 6-min walk test (6MWT).

METHODS After a systematic review, 31 randomized trials of exercise-based CR compared with no exercise control (4,784 HF patients) were included. Outcomes were pooled using random effects meta-analyses, and inverse variance weighted linear regression equations were fitted to estimate the relationship between the CR on EC and all-cause mortality, hospitalization, and HRQOL. Spearman correlation coefficient (ρ), R^2 at trial level, and surrogate threshold effect (STE) were calculated. STE represents the intercept of the prediction band of the regression line with null effect on the final outcome.

RESULTS Exercise-based CR is associated with positive effects on EC measured through Vo_2peak (+3.10 ml/kg/min; 95% confidence interval [CI]: 2.01 to 4.20) or 6MWT (+41.15 m; 95% CI: 16.68 to 65.63) compared to control. The analyses showed a low level of association between improvements in EC (Vo_2peak or 6MWT) and mortality and hospitalization. Moderate levels of correlation between EC with HRQOL were seen (e.g., $R^2 <52\%$; $|\rho| < 0.72$). Estimated STE was an increase of 5 ml/kg/min for Vo_2peak and 80 m for 6MWT to predict a significant improvement in HRQOL.

CONCLUSIONS The study results indicate that EC is a poor surrogate endpoint for mortality and hospitalization but has moderate validity as a surrogate for HRQOL. Further research is needed to confirm these findings across other HF interventions. (J Am Coll Cardiol HF 2018;6:596–604) © 2018 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

E nhancement of exercise capacity (EC) is a key aspect of the lifestyle and management of patients with heart failure (HF) (1,2). The gold standard approach for measuring EC is a maximal (or symptom-limited) exercise test with direct cardiorespiratory assessment of peak oxygen uptake (Vo_2peak) or via indirect submaximal tests, including the 6-min walk test (6MWT) (3).

EC is often used as a primary outcome in HF trials and is accepted by the United States Food and Drug

---

Administration (4). Although it is a measure of function and clinical benefit, EC is a surrogate endpoint rather than a final patient-relevant outcome, such as mortality, hospital admission, or health-related quality of life (HRQOL).

For a surrogate endpoint to be considered a valid surrogate endpoint (i.e., an adequate substitute for the final outcome), several levels of evidence must be provided (5). First, there needs to be biological plausibility of the relationship between the surrogate and the final outcome. Second, observational or epidemiological studies are required to show a consistent association between the surrogate and the final outcome. Third, the treatment effect on the surrogate must correspond with the treatment effect on the final outcome, preferably in the setting of a meta-analysis of randomized controlled trials (RCTs). Epidemiological studies have shown that a 1.0 metabolic equivalent (MET) (1 MET = 3.5 ml/kg/min) increase in Vo_2peak translates into a 12% risk reduction in mortality in individuals with existing cardiovascular disease, including HF (6). However, to our knowledge no previous study has assessed the validity of EC as a surrogate endpoint for HF in a RCT setting.

Using RCTs of exercise-based cardiac rehabilitation (CR) in patients with HF, we sought to address the following 2 research questions: 1) is there an association between the intervention effect of CR in HF on EC, and each of mortality, hospitalization, and HRQOL?; and 2) can we reliably quantify the expected effect on mortality, hospitalization, and HRQOL that may follow in future HF trials?

METHODS

This study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guideline (7).

STUDY IDENTIFICATION. We updated the Cochrane systematic review of RCTs of exercise-based CR in HF up to February 2017 (8). This search included the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, MEDLINE In-Process, Embase, CINAHL, PsycINFO, conference proceedings via Web of Science Core Collection, and trial registries (World Health Organization International Clinical Trials Registry Platform and ClinicalTrials.gov). We included RCTs of adults, age 18 years or older, comparing exercise-based CR and control in HF patients with follow-up of 6 months or longer for at least 1 of the outcomes. Exercise-

based CR was defined as an intervention that includes exercise training, either alone or in addition to psychosocial and/or educational interventions. Controls could receive standard medical care without any form of structured exercise training or advice. We sought to include all RCTs that reported EC at baseline and follow-up, whether measured using Vo_2peak or 6MWT, and at least 1 of the final patient-relevant outcomes of interest (i.e., mortality, hospitalization, or HRQOL).

Screening of full study reports was undertaken by 1 of the authors (O.C., J.U.) and checked by a second author (R.S.T.).

DATA EXTRACTION. For each study, we extracted the following information: first author, publication year, geographical location, sample size and ratio of intervention to control, study follow-up duration, setting of exercise training intervention (center- or home-based exercise program), age (mean), sex (percentage male), left ventricular ejection fraction (mean), and New York Heart Association functional class of the patient population. The continuous outcomes of EC and HRQOL were extracted at baseline and at the latest reported follow-up as mean ± SD for both exercise and control groups. The binary outcomes of mortality and hospitalization were extracted as the number of patient events at the latest follow-up relative to the number of patients randomized to each group. Whenever necessary, an online digitizer (Web-PlotDigitizer) was used. Missing SDs at baseline or follow-up were imputed from confidence intervals, interquartile ranges, or SEs, and missing SDs for follow-up minus baseline change were estimated using Cochrane Handbook recommended methods (9). For trials with more than 1 exercise intervention arm, we followed the Cochrane Handbook's approach for combining groups (9). All data was first extracted by 1 of the authors (O.C.) and then checked by another (R.S.T.). Trial quality was assessed using the Cochrane risk of bias tool (10).

STATISTICAL ANALYSIS. Binary outcomes at the latest follow-up were expressed as an odds ratio (OR), where OR <1.0 indicated a beneficial effect of exercise-based CR compared with control. Given the variation in reporting of outcomes across studies, we first expressed the between-group difference in EC and HRQOL for each study as a standardized mean difference and 95% confidence interval (CI). For the subset of studies that reported HRQOL using the Minnesota Living With Heart Failure (MLwHF)

FIGURE 1   Study Selection Flow Diagram

6MWT = 6-min walk test; HF = heart failure; HRQoL = health-related quality of life; MLwHF = Minnesota Living With Heart Failure questionnaire; Vo_2peak = peak oxygen uptake.
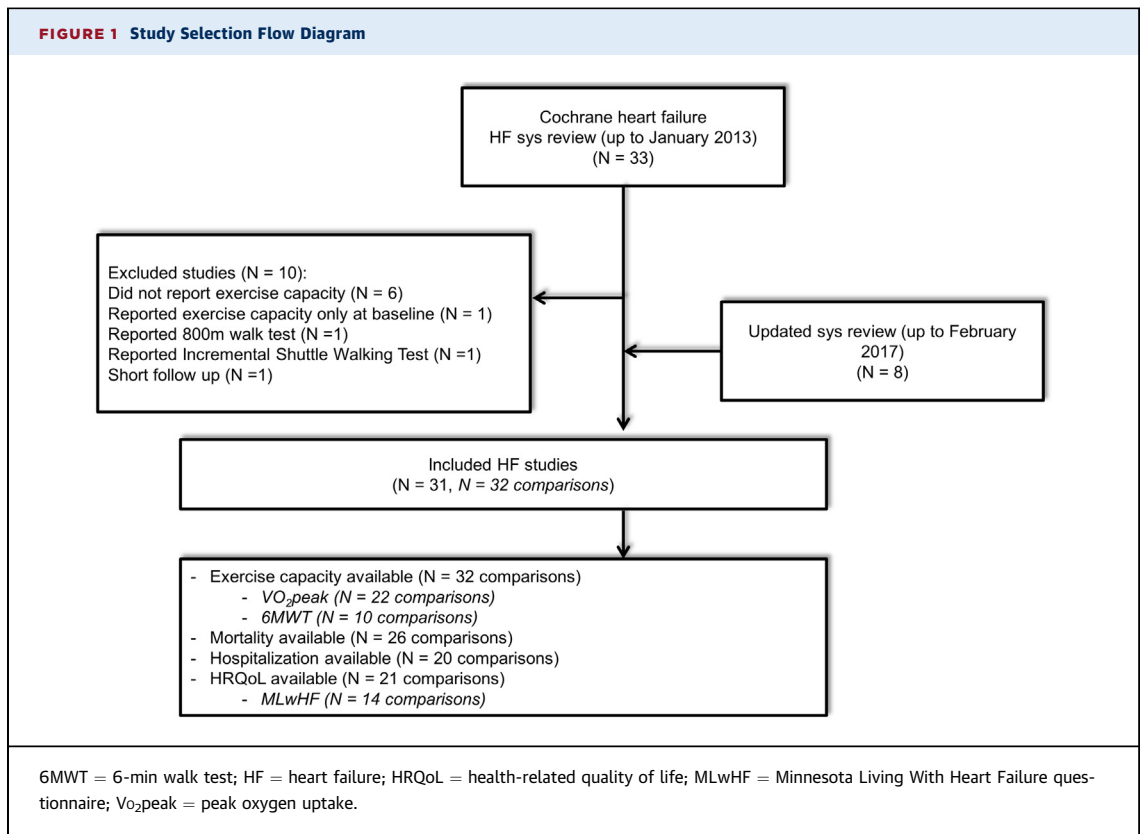
questionnaire, weighted mean differences were calculated, where a negative between group mean difference (exercise minus control) in MLwHF indicates a greater HRQOL score in the training group compared to the control group. In the subset of studies that reported Vo_2peak (in ml/kg/min) or 6MWT (in meters), we calculated the mean difference between groups and 95% CI in original units. A positive mean difference (exercise minus control) indicated greater EC in the training group compared to the control group.

Statistical heterogeneity was assessed using the I^2 statistic (9), and small-study effects and publication bias were assessed using the Egger test or the Peters test. All outcomes were pooled using a DerSimonian and Laird random effects model with continuity correction when needed.

We calculated commonly reported indicators of surrogate validation (11). The correlation coefficient (ρ) and the R^2 for the relationship between treatment effect differences in EC and each of the final outcomes individually were estimated using weighting by the inverse of the variance (for the treatment effect on final outcomes). Where possible, the surrogate threshold effect (STE) was calculated. The STE represents the intercept of the prediction band of the regression line with zero effect on the final outcome (12).

We performed sensitivity analyses to assess whether our findings changed when: 1) excluding the largest included trial (HF-ACTION [Heart Failure: A Controlled Trial Investigating Outcomes of Exercise Training]) (13); 2) limiting our analysis to studies of HF with reduced ejection fraction only; or 3) limiting our analysis to studies at low risk of bias as assessed by random sequence generation (14) and allocation concealment (15).

All data analyses were conducted using Stata version 14.2 (Stata Corp., College Station, Texas) software.

RESULTS

STUDY SELECTION AND CHARACTERISTICS. A total of 31 studies and 32 comparison groups (1 study had 2 exercise intervention arms [16]) were included for analysis (Figure 1).

A summary of included studies is given in Table 1 (details are listed in Online Table 1 with supplemental references listed). The nature of exercise

TABLE 1 Summary of Characteristics of Included Trials

Exercise type	
Aerobic	23 (72)
Mixed	9 (28)
Exercise length (weeks)	24 (4–60)
Exercise session duration (min)	30 (15–60)
Exercise dose* (min)	2,160 (480–6,000)
Follow-up (months)	6.5 (6–120)
Setting	
Center-based	13 (41)
Home-based	7 (22)
Both	12 (37)
Sample size	54 (19–2,331)
Publication date	
1990–1999	5 (16)
2000–2009	16 (50)
2010 or later	11 (34)
Single center	27 (84)
Male (%)	81 (42–100)
Age (yrs)	60.5 (51.0–80.5)
Diagnosis	
HFrEF only	26 (81)

Values are n (%) or median (range). *Dose = number of weeks of intervention ×
number of exercise training sessions per week × mean duration of exercise session
(in min).

HFrEF = heart failure with reduced ejection fraction.

TABLE 2 Pooled Exercise Capacity*

	Exercise Capacity (Follow-Up Minus Baseline Change)	
	Vo_2peak	6MWT
No. of studies	22	10
WMD (95% CI)	3.10 (2.01–4.20)	41.15 (16.68–65.63)
p value	<0.001	0.001
I^2	96.6%	85.1%
Egger test p value	0.052	0.806

*DerSimonian and Laird random effects method. Treatment or lower than 1.0
favors exercise intervention; negative WMD favors exercise intervention. Positive
weighted (by inverse variance) changes in exercise capacity values favor the
exercise rehabilitation group. Number of studies, WMD (95% CI), p value for
significance of effect size, I^2 statistic, and Egger test p value for small-study effect
reported.

6MWT = 6-min walk test; CI = confidence interval; Vo_2peak = maximum oxygen
uptake; WMD = weighted mean difference.

training varied across studies with regard to the frequency, duration, and intensity of exercise (Online Table 1).

RISK OF BIAS. The overall risk of bias assessment was moderate, although several studies failed to give sufficient details to assess risk of bias criteria (Online Table 2). Reporting was found to be considerably better in more recently published studies.

IMPACT OF EXERCISE TRAINING. Exercise capacity. EC was reported at follow-up times ranging from 1 to 14 months. In the subset of studies reporting Vo_2peak, there was a larger increase in mean pooled EC with exercise-based CR compared with control (3.10 ml/kg/min; 95% CI: 2.01 to 4.20; 22 studies; p < 0.001; I^2 = 96.6%) (**Table 2,** Online Figure 1). A similar positive finding for the exercise-based CR arm was seen in the subset of studies reporting 6MWT (41.15 m; 95% CI: 16.68 to 65.63; 10 studies; p = 0.001; I^2 = 85.1%) (Online Figure 2).

Final patient-relevant outcomes. There was no difference in pooled mortality between exercise-based CR and control (OR: 0.85; 95% CI: 0.71 to 1.01; 26 trials; p = 0.066; I^2 = 0%) (Online Figure 3). The risk of all-cause hospitalization was reduced in the exercise group compared with control (OR: 0.64; 95% CI: 0.44 to 0.93; 20 studies; p = 0.02; I^2 = 60.3%) (**Table 3,** Online Figure 4).

HRQOL at baseline and follow-up was reported in 21 comparisons, of which 14 used the disease-specific measure, the MLwHF questionnaire (17). Other HRQOL questionnaires reported were the Kansas City Cardiomyopathy Questionnaire (13,18), Icelandic Quality of Life (19), the Chronic Heart Failure Questionnaire (20,21), the Likert scale for symptoms (22), and the patients' global assessment of change in quality of life (23). Across all HRQOL outcome measures, the level of HRQOL at follow-up was higher with exercise-based CR compared with control (standardized mean difference: −0.48 SDU; 95% CI: −0.73 to −0.24; 21 comparisons; p < 0.0001; I^2 = 89.5%) (Online Figure 5). When pooling the subgroup of trials that reported MLwHF follow-up scores, the level of HRQOL was higher in the exercise-based CR group (−7.24; 95% CI: −11.84 to −2.63; 14 comparisons; p = 0.002; I^2 = 67.9%) (Online Figure 6). All patient-relevant outcomes were measured over follow-up times from a minimum of 3 months up to 120 months.

Evaluation of EC as surrogate endpoint. Regression coefficients of determination (R^2) and correlation coefficients (ρ) between the change in EC and mortality or hospitalization were relatively low (R^2 ≤ 28%; |ρ| < 0.53) (**Table 4**). The coefficients for the slope of the regression line were not significantly different from zero (p > 0.05), confirming no clear associations between CR intervention effect on EC and clinical outcomes.

Higher correlations were seen between change in EC and HRQOL, for example, R^2 = 32% and ρ = −0.57 for the change in 6MWT and MLwHF; R^2 = 52%; ρ = −0.72 for the change in Vo_2peak and all HRQOL measures. Negative correlation coefficients indicate that larger CR effects on EC are associated with larger CR effects on HRQOL (**Table 4**).

TABLE 3   Pooled All-Cause Mortality, Hospitalization, and HRQOL (Either MLwHF or All Scales)*

	Mortality (All-Cause)	Hospitalization (All-Cause)	MLwHF (Follow-Up)	HRQOL (All Scales) (Follow-Up)
No. of studies	26	20	14	21
Effect estimate (95% CI)	OR: 0.85 (0.71 to 1.01)	OR: 0.64 (0.44 to 0.93)	WMD: −7.24 (−11.84 to −2.63)	SMD: −0.48 (−0.73 to −0.24)
p value	0.066	0.020	0.002	<0.001
I^2	0%	60.3%	67.9%	89.5%
Peters or Egger test: p value	Peters: 0.162	Peters: 0.373	Egger: 0.798	Egger: 0.078

*DerSimonian and Laird random effects method. Treatment or lower than 1.0 favors exercise intervention; negative SMD/WMD favors exercise intervention. Positive weighted (by inverse variance) changes in exercise capacity values favor the exercise rehabilitation group. Number of studies, effect estimates (95% CI), p value for significance of effect size, I^2 statistic, and Peters or Egger test p value for small-study effect reported.

HRQOL = Health-related quality of life; MLwHF = Minnesota Living with Heart Failure questionnaire; OR = odds ratio; SMD = standardized mean difference; other abbreviations as in **Table 2**.

Surrogate threshold effect. Based on the correlation analyses, we estimated STEs for EC and HRQOL. For the subset of studies reporting Vo_2peak (**Figure 2**), we estimated that an average improvement of 5 ml/kg/min in Vo_2peak exercise-based CR versus control is needed to predict a favorable improvement in HRQOL with exercise-based CR compared to control. For the subset of studies reporting 6MWT (**Figure 3**), we estimated an STE of 80 m for 6MWT to predict a significant improvement in MLwHF with exercise-based CR versus control. This pattern of link between EC and final outcomes was consistent across all sensitivity analyses (Online Tables 3 to 6).

DISCUSSION

Using trial-level data from RCTs of exercise-based CR for HF, we formally evaluated the evidence for EC as a surrogate endpoint for the final outcomes of mortality, hospitalization, and HRQOL. Our results show an increase in Vo_2peak or 6MWT with exercise-based CR to be associated with improvements in clinical outcomes (24). However, the observed levels of association indicate EC is a poor surrogate endpoint for mortality and hospitalization and has moderate validity for HRQOL. We found an STE for Vo_2peak of 5 ml/kg/min and 6MWT of 80 m. Thus, exercise-based CR would need to increase Vo_2peak and 6MWT by this level (or more) to have 95% confidence interval to be able to demonstrate significant improvement in HRQOL in a future trial. Sensitivity analyses confirmed Vo_2peak to have moderate surrogate validity for HRQOL.

IMPLICATIONS FOR PRACTICE. Our study findings have significant implications for future HF trials. Contrary to epidemiological (observational) evidence, our results show that intervention effects on EC are not predictive of treatment effects on the clinical events of mortality or hospitalization. However, we also showed that improvements in EC, if large enough, can be predictive of important gains in patient HRQOL.

Surrogate endpoints generally accrue more quickly than final endpoints, thus allowing for RCTs with

TABLE 4   Surrogacy Validity of Vo_2peak and 6MWT Vs. Mortality, Hospitalization, and HRQOL Measures*

	Exercise Capacity (Follow-Up Minus Baseline Change)	
	Vo_2peak	6MWT
Mortality (all-cause, log OR)		
Coeff (SE)	−0.19 (0.08)	−0.009 (0.005)
p value	0.035	0.118
R^2	28%	28%
P	−0.53	−0.53
Hospitalization (all-cause, log OR)		
Coeff (SE)	−0.17 (0.12)	−0.01 (0.02)
p value	0.181	0.509
R^2	12%	24%
P	−0.35	−0.49
MLwHF (follow-up value)		
Coeff (SE)	−1.59 (1.64)	−0.13 (0.08)
p value	0.388	0.141
R^2	19%	32%
ρ	−0.43	−0.57
STE		80 m
HRQOL (all scales, follow-up value)		
Coeff (SE)	−0.22 (0.07)	−0.008 (0.004)
p value	0.013	0.120
R^2	52%	27%
ρ	−0.72	−0.52
STE (ml/kg/min)	5	

*Coeff (SE) and related p value for the slope of the regression line and R^2 derived from least-squares linear regressions weighted by the inverse of variance of the treatment effect on final outcomes. Weighted correlation coefficient (ρ) is shown. STE calculated as the value of the treatment effect on the surrogate outcome at which the linear regression prediction line crosses the null effect line for the treatment effect on patient-relevant outcomes. Only trials contributing to the within-study paired analysis were considered.

Coeff = coefficient; STE = surrogate threshold effect; other abbreviations as in **Tables 2 and 3**.

FIGURE 2 Results of Regression Analyses Showing the Relationship Between Changes in Vo_2peak Between Baseline and FU on Log Odds of Clinical Event (Mortality or Hospitalization) or HRQoL Reported as MLwHF or All Scales
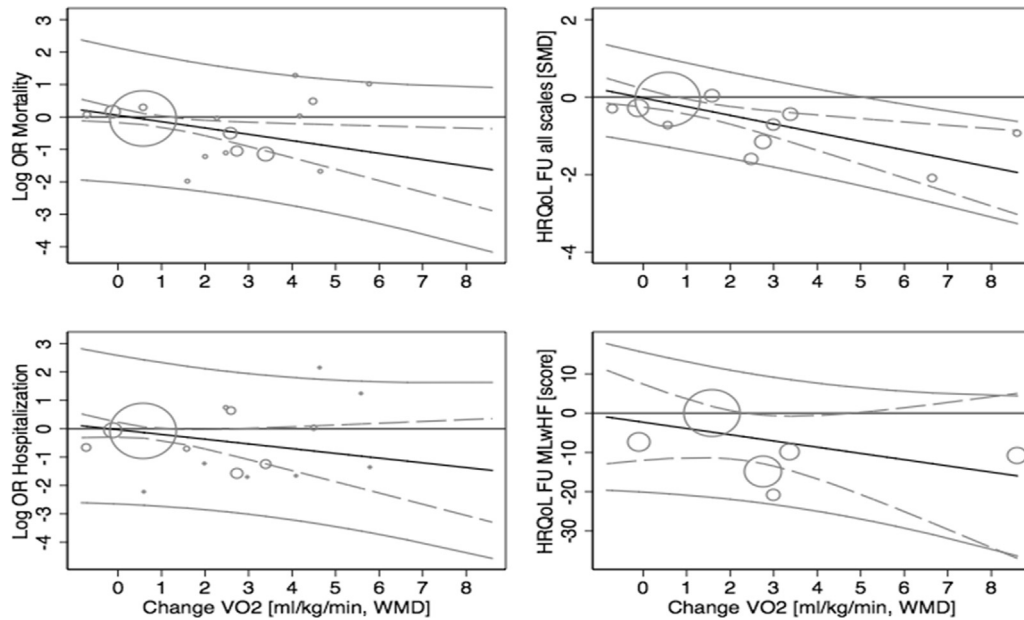
Circles represent a study-level comparison, with sizes proportionate to study weights (based on inverse variance weighting). **Dashed gray lines** correspond to the bounds of the 95% confidence interval for the regression line. **Solid gray lines** correspond to the bounds of the 95% prediction interval for the regression line. FU = follow-up; OR = odds ratio; SMD = standardized mean difference; Vo_2 = peak oxygen uptake; WMD = weighted mean difference; other abbreviations as in **Figure 1**.
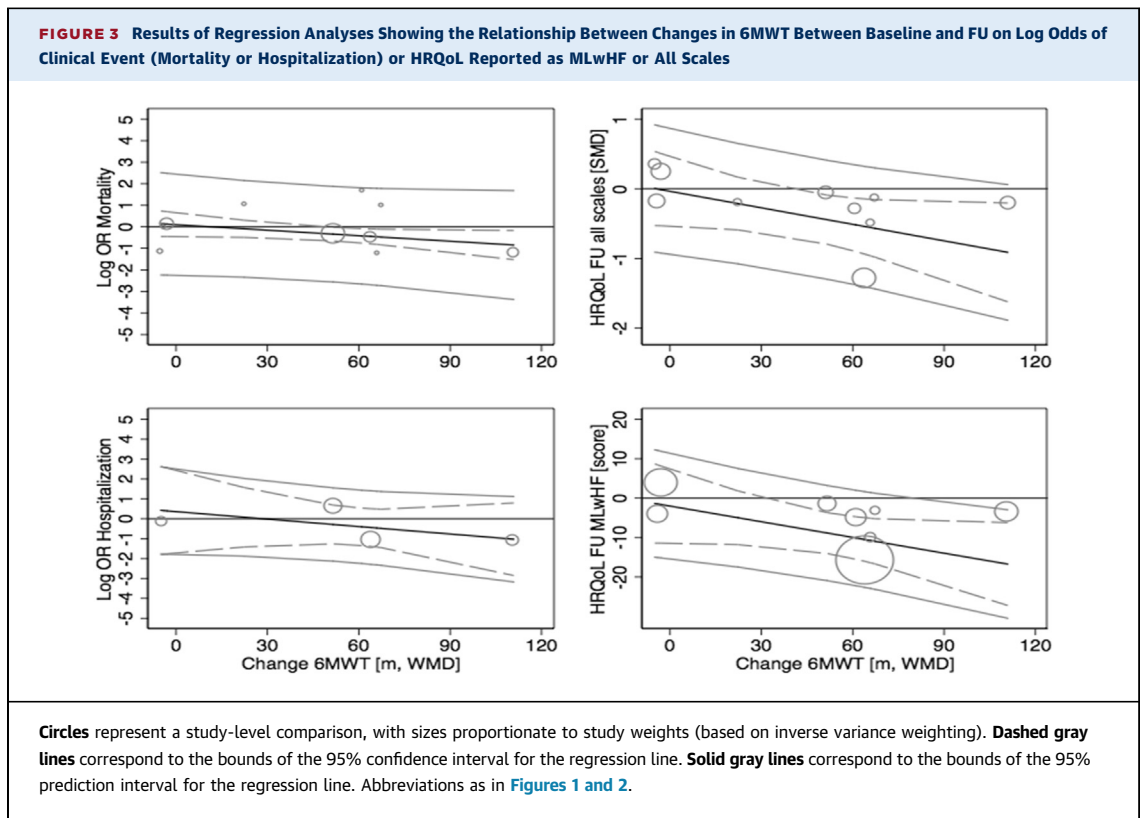
shorter follow-up periods and smaller sample sizes. Reducing trial sample size and duration ensures faster patient access to new therapies and means that trials are less expensive, which makes surrogate endpoints attractive to manufacturers and research sponsors alike. However, it is important that surrogate endpoints be carefully validated, as evidenced by the dramatic failures of surrogates used for regulatory purposes.

Our results are consistent with the review by Ferreira et al. (4), which recommended the use of EC as outcome in HF trials based on qualitative analysis of results from HF trials showing an improvement in 6MWT was associated with a favorable treatment effect of morbidity and mortality. This analysis noted that an increase of 30 to 50 m in 6MWT has been used in cardiac resynchronization therapy trials in order to gain pre-market approval. This magnitude of improvement in EC is somewhat smaller than our estimated STE of 80 m in the 6MWT. Similarly, our finding of an STE of 5 ml/kg/min in Vo_2peak is considerably larger than the increase of 6% in Vo_2peak (i.e., ~1.0 ml/kg/min) needed to predict an improvement in the primary outcome (time to

all-cause mortality or all-cause hospitalization) reported by Swank et al. (25) based on HF-ACTION. Although based on patient-level data analysis from a large RCT of exercise-based CR (HF-ACTION), this later analysis does not take into account the role of treatment; therefore, it can be used to establish the prognostic validity of Vo_2peak but not the association between treatment effects on Vo_2peak and treatment effects on patient-relevant outcomes as measured across a number of RCTs. In contrast, and in accordance with contemporary recommendations for surrogate endpoint validation (26), the present study derived STE values from the prediction interval around the regression line rather than from the 95% CI based on a meta-analysis of RCTs of exercise-based CR in HF.

STUDY LIMITATIONS. First, the methods of EC assessment varied considerably across studies. Whereas a number of studies reported Vo_2peak, a small proportion directly measured Vo_2peak using cardiorespiratory testing, and others predicted Vo_2peak using a submaximal exercise. However, this limitation does not apply to the subgroup of studies that assessed EC using 6MWT. In addition, although

FIGURE 3   Results of Regression Analyses Showing the Relationship Between Changes in 6MWT Between Baseline and FU on Log Odds of Clinical Event (Mortality or Hospitalization) or HRQoL Reported as MLwHF or All Scales

Circles represent a study-level comparison, with sizes proportionate to study weights (based on inverse variance weighting). **Dashed gray lines** correspond to the bounds of the 95% confidence interval for the regression line. **Solid gray lines** correspond to the bounds of the 95% prediction interval for the regression line. Abbreviations as in **Figures 1 and 2**.

most studies reported EC at baseline and follow-up, they did not report the SD of the baseline-follow up change in EC. For these studies, we imputed the SD of change (using a correlation coefficient 0.74 from included trials) (9). Both of these factors might have introduced measurement error and masked an underlying association between EC with final outcomes. Second, there was considerable heterogeneity in exercise-based intervention applied in the included RCTs: a wide range in exercise training dose and variation in whether trials included educational and psychological cointerventions. Here, we would argue that such heterogeneity is implicit in a systematic review and meta-analysis of a complex intervention such as exercise-based CR (27). However, it could be argued that restricting our analyses to exercise-based CR trials limits the generalizability of our results. However, it is recommended that the surrogate validation be undertaken in trials across the same intervention (5). Third, a number of included RCTs studies had methodological issues or poor reporting that may have resulted in their high risk of bias. However, reassuringly, our findings were consistent when limited to the subgroup of trials of low risk of bias. Fourth, despite excluding trials with follow-up <6 months, the timing of longest assessment across

studies varied from 1 to 14 months for EC and from 3 months to 10 years for final outcomes. Unfortunately, individual studies did not consistently report EC and final outcomes at repeated time points, so we could not explore this issue.

CONCLUSIONS

This systematic review and meta-analysis of RCTs of exercise-based CR shows that treatment effect on EC is a poor surrogate endpoint for treatment effect on mortality and hospitalization. However, we found the effect of exercise-based CR on EC to have moderate validity as a surrogate endpoint for treatment effect on HRQOL. Further research is needed to determine whether our findings are generalizable across other HF interventions. Given that the participant-level fitness response to exercise-based CR can be highly heterogeneous (28,29), our findings also need confirmation from individual participant data meta-analyses of exercise-based CR (5).

ADDRESS FOR CORRESPONDENCE: Dr. Oriana Ciani, Institute of Health Research, University of Exeter Medical School, South Cloisters, St. Luke's Campus, Heavitree Road, Exeter EX1 2LU, United Kingdom. E-mail: o.ciani@exeter.ac.uk.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: Using data from randomized trials of exercise-based CR for HF, we formally evaluated the evidence for EC as a surrogate endpoint for mortality, hospitalization, and HRQOL. EC is a poor surrogate endpoint for mortality and hospitalization but has moderate validity as a surrogate for HRQOL. In severe and advanced stages of HF, improvements in independence and the ability to perform daily tasks become even more important than improvements in morbidity and mortality. In this respect, linking increases in EC to HRQOL improvements can be seen as a clinically relevant finding.

TRANSLATIONAL OUTLOOK: The use of surrogate endpoints can improve the efficiency of clinical trials by reducing sample size and duration, thereby promoting quicker patient access to new therapies. However, to ensure reliable prediction of final patient-related outcomes, it is important that surrogate endpoints be validated. Our results show an increase of 5 ml/kg/min in VO$_2$peak and 80 m for 6MWT can significantly predict an improvement in HRQOL. Further research is needed to determine whether our findings can be replicated for other types of interventions in chronic HF patients and in other cardiovascular disease populations.

REFERENCES

1. McMurray JJ, Adamopoulos S, Anker SD, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. Eur Heart J 2012;33:1787–847.

2. Yancy CW, Jessup M, Bozkurt B, et al. 2017 ACC/AHA/HFSA Focused Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. J Am Coll Cardiol 2017;70:776–803.

3. Guyatt GH, Sullivan MJ, Thompson PJ, et al. The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure. Can Med Assoc J 1985;132:919–23.

4. Ferreira JP, Duarte K, Graves TL, et al. Natriuretic peptides 6-min walk test, and quality-of-life questionnaires as clinically meaningful endpoints in HF trials. J Am Coll Cardiol 2016;68:2690–707.

5. Ciani O, Buyse M, Drummond M, Rasi G, Saad ED, Taylor RS. Time to review the role of surrogate end points in health policy: state of the art and the way forward. Value Health 2017;20:487–95.

6. Kokkinos P, Myers J. Exercise and physical activity: clinical outcomes and applications. Circulation 2010;122:1637–48.

7. Shamseer L, Moher D, Clarke M, et al. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015;349:g7647.

8. Sagar VA, Davies EJ, Briscoe S, et al. Exercise-based rehabilitation for heart failure: systematic review and meta-analysis. Open Heart 2015;2:e000163.

9. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011].

Higgins J, Green S, editors: The Cochrane Collaboration; 2011. Available at: http://handbook-5-1.cochrane.org/. Accessed February 1, 2017.

10. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.

11. Ciani O, Davis S, Tappenden P, et al. Validation of surrogate endpoints in advanced solid tumours: systematic review of statistical methods, results, and implications for policy makers. Int J Technol Assess Health Care 2014;30:13.

12. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006;5:173–86.

13. O'Connor CM, Whellan DJ, Lee KL, et al. Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial. JAMA 2009;301:1439–50.

14. Page MJ, Higgins JP, Clayton G, Sterne JA, Hrobjartsson A, Savovic J. Empirical evidence of study design biases in randomized trials: systematic review of meta-epidemiological studies. PLoS One 2016;11:e0159267.

15. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008;336:601–5.

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

17. Rector TS, Cohn JN. Assessment of patient outcome with the Minnesota Living with Heart Failure questionnaire: reliability and validity during a randomized, double-blind, placebo-controlled trial of pimobendan. Pimobendan Multicenter Research Group. Am Heart J 1992;124:1017–25.

18. Norman JF, Pozehl BJ, Duncan KA, Hertzog MA, Krueger SK. Effects of exercise training versus attention on plasma B-type natriuretic peptide, 6-minute walk test and quality of life in individuals with heart failure. Cardiopulm Phys Ther J 2012;23:19–25.

19. Jonsdottir S, Andersen KK, Sigurosson AF, Sigurosson SB. The effect of physical training in chronic heart failure. Eur J Heart Fail 2006;8:97–101.

20. Wall HK, Ballard J, Troped P, Njike VY, Katz DL. Impact of home-based, supervised exercise on congestive heart failure. Int J Cardiol 2010;145:267–70.

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

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

23. Willenheimer R, Rydberg E, Cline C, et al. Effects on quality of life, symptoms and daily activity 6 months after termination of an exercise training programme in heart failure patients. Int J Cardiol 2001;77:25–31.

24. Ciani O, Buyse M, Garside R, et al. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. BMJ 2013;346:f457.

25. Swank AM, Horton J, Fleg JL, et al. Modest increase in peak VO$_2$ is related to better clinical outcomes in chronic heart failure patients: results from heart failure and a controlled trial to investigate outcomes of exercise training. Circ Heart Fail 2012;5:579–85.

26. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points—the challenge of statistical validation. Nat Rev Clin Oncol 2010;7:309-17.

27. Smart NA, Waldron M, Ismail H, et al. Validation of a new tool for the assessment of study quality and reporting in exercise training studies: TESTEX. Int J Evid Based Healthc 2015; 13:9-18.

28. Pandey A, Kitzman DW, Brubaker P, et al. Response to endurance exercise training in older adults with heart failure with preserved or reduced ejection fraction. J Am Geriatr Soc 2017;65: 1698-704.

29. Sisson S, Katzmarzyk P, Earnest C, Bouchard C, Blair S, Church T. Volume of exercise and fitness non-response in sedentary, post-menopausal women. Med Sci Sports Exerc 2009;41:539-45.