

## Original Article

# Measuring True Change in Individual Patients: Reliable Change Indices of Cardiac Rehabilitation Outcomes, and Implications for Quality Indicators

Megan E. O'Connell, PhD,<sup>a</sup> Neville Suskin, MbChB, MSc,<sup>b,c,d</sup> and Peter L. Prior, PhD<sup>b,c</sup>

<sup>a</sup> Department of Psychology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

<sup>b</sup> St Joseph's Hospital Cardiac Rehabilitation & Secondary Prevention Program, London, Ontario, Canada

<sup>c</sup> Lawson Health Research Institute, London, Ontario, Canada

<sup>d</sup> Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

### ABSTRACT

**Background:** Mediated by outcomes such as improved exercise capacity, cardiac rehabilitation (CR) reduces morbidity and mortality. For accuracy, an individual CR patient's change must be measured reliably, an issue not typically considered in practice. Drawing from psychometric theory, we calculated reliable change indices (RCIs), to measure *individual* CR patients' *true clinical change*, apart from that from error and test practice/exposure, in exercise capacity, anxiety, and depression.

**Methods:** Indirectly calculated exercise capacity (peak metabolic equivalents [METs]) and psychological symptoms were each measured twice, 1 week apart, by administering treadmill tests or the Hospital Anxiety and Depression Scale (HADS) to separate samples of 35 (mean age: 59.0 years; 6 women) and 96 (mean age: 64.4 years; 32 women) CR patients, respectively. Using test-retest reliability and mean difference scores from these samples to estimate error and practice/exposure effects, we calculated RCIs for a separate cohort (n = 2066; mean age: 62.0 years; 533 women) who completed 6-month CR, and compared change distributions (worsened/unchanged/improved) based on critical RCIs, mean and percent changes, cut-off scores, and standard deviations.

Cardiac rehabilitation (CR) improves morbidity and mortality among cardiovascular patients,<sup>1</sup> mediated by improved exercise capacity.<sup>2</sup> Depression<sup>3</sup> and anxiety<sup>4</sup> are associated with cardiovascular risk, so psychological functioning is an important CR outcome,<sup>5</sup> and psychosocial interventions may contribute to morbidity and mortality improvements.<sup>6,7</sup> Comparisons of means, for example, of a CR cohort at discharge vs entry, might elucidate programmatic outcomes, while revealing little about individual participants. Clinically

### RÉSUMÉ

**Contexte :** Grâce à des résultats tels que l'amélioration de la capacité d'exercice, la réadaptation cardiaque (RC) réduit la morbidité et la mortalité. Pour être précis, le changement de RC d'un patient doit être mesuré de manière fiable, un aspect qui n'est généralement pas pris en compte dans la pratique. En nous inspirant de la théorie psychométrique, nous avons déterminé des indices de changement fiables afin de mesurer le véritable *changement clinique individuel* des patients en RC en lien avec la capacité d'entraînement, l'anxiété et la dépression et indépendamment des erreurs et de la pratique du test/l'exposition au test.

**Méthodologie :** La capacité d'entraînement calculée indirectement (l'équivalent métabolique maximum [MET]) et les symptômes psychologiques ont été mesurés deux fois, à une semaine d'intervalle, par le biais de tests sur tapis roulant ou de l'échelle de l'évaluation de l'anxiété et de la dépression en milieu hospitalier (HADS), auxquels ont respectivement été soumis des échantillons de 35 personnes en RC (âge moyen : 59,0 ans; 6 femmes) et de 96 personnes en RC (âge moyen : 64,4 ans; 32 femmes). En recourant à la fiabilité test-retest et aux écarts moyens dans les scores de ces échantillons pour faire une approximation des effets d'erreur et de pratique/exposition, nous

therefore, consistent with personalizing health interventions,<sup>8</sup> it is important to determine whether *individual* CR patients improve meaningfully in exercise capacity and psychological symptoms.

Suppose an individual patient's exercise capacity increased from 6.5 to 7.5 metabolic equivalents (METs; 1 MET = 3.5 ml O<sub>2</sub> /kg body-weight/min), 1.0 MET or 15%, from CR entry to discharge. This increase matches a mean change measured directly with cardiopulmonary exercise testing,<sup>9</sup> or a mean between-groups difference estimated indirectly from peak speed and grade of treadmill testing,<sup>10</sup> associated with fewer clinical events. This increase is double the 0.5-MET pre-post CR exercise capacity quality indicator of the Canadian Cardiovascular Society (CCS-QI),<sup>11</sup> which we treat here as a minimal clinically important difference (MCID).<sup>12</sup> A critical question is whether such an individual's performance reflects truly improved exercise capacity, over and above test practice and measurement error.<sup>13-16</sup> Pre-/

Received for publication August 24, 2020. Accepted May 10, 2021.

**Ethics Statement:** This research was approved by the Health Sciences Research Ethics Board, Western University, London, Ontario.

Corresponding author: Dr Peter Prior, St Joseph's Hospital Cardiac Rehabilitation and Secondary Prevention Program, 268 Grosvenor St, Rm B3-689, London, Ontario N6A 4V2, Canada. Tel.: +1-519-646-6000 x75725; fax: +1-519-667-6532

E-mail: [peter.prior@sjhc.london.on.ca](mailto:peter.prior@sjhc.london.on.ca)

See page 1147 for disclosure information.

**Results:** Practice/exposure effects were nonsignificant, except the mean HADS anxiety score decreased significantly ( $P \leq 0.013$ ;  $d = 0.17$ , small effect). Test-retest reliabilities were high (METs  $r = 0.934$ ; HADS anxiety score  $r = 0.912$ ; HADS depression score  $r = 0.90$ ;  $P < 0.001$ ). Among 2066 CR patients, RCI distributions differed ( $P < 0.001$ ) from those of most other change criteria.

**Conclusions:** Change ascertainment depends on criterion choice. A Canadian Cardiovascular Society CR quality indicator of increase by 0.5 MET may be too small to assess individuals' functional capacity change. RCIs offer a pragmatic approach to benchmarking reliable change frequency, and pending further validation, could be used for feedback to individual patients.

post-intervention change scores, including differences, percentages, and standard deviations (SDs) may be inadequate to answer this question,<sup>17</sup> because they may reflect measurement error or prior test exposure/practice effects, in addition to clinically relevant change, such as that from treatment, recovery, or deterioration.

We address the question of individual CR participants' change in exercise capacity and psychological symptoms, through an innovative application of reliable change methodology from psychometric theory. Reliable change indices (RCIs) can account for error and practice in difference scores,<sup>18</sup> potentially facilitating measurement of *true clinical change* of *individual CR patients*, offering a direct clinical application.

### Measuring Exercise Capacity, Anxiety, and Depression

Reliability (which estimates error) of exercise capacity measured with stress-testing is moderate to high, assuming consistent test-termination criteria<sup>19</sup> and standardized protocols.<sup>20</sup> Direct measurement of peak oxygen uptake capacity (peak  $\text{VO}_2$ ) is less sensitive to practice<sup>15</sup> and error<sup>16</sup> than functional capacity (FC) measures dependent on treadmill time, determined from peak speed and grade. Although FC can improve due to merely test practice,<sup>14,16,21-23</sup> giving nonreproducible results,<sup>14</sup> CR programs have often relied upon such "indirect" exercise-capacity measurement.

The Hospital Anxiety and Depression Scale (HADS) measures self-reported depression and anxiety symptoms, with good reliability and validity.<sup>24-26</sup> Anxiety and depression subscales range from 0 to 21 (higher = worse), with case identification  $> 7$ .<sup>24</sup> HADS depression scores predicted 1-year mortality after acute coronary syndrome, with a 20% risk increment/point.<sup>27</sup> In the Ontario Cardiac Rehabilitation Pilot Project (OCRPP), the mean HADS depression score and HADS anxiety score each improved significantly

avons déterminé des indices de changement fiables pour une cohorte distincte ( $n = 2\ 066$ ; âge moyen : 62,0 ans; 533 femmes) qui a terminé une RC de six mois, et comparé la répartition des changements (aggravé/inchangé/amélioré) sur la base des indices de changement fiables critiques, des changements moyens et en pourcentage, des scores-seuils et des écarts types.

**Résultats :** Les effets de la pratique/exposition étaient peu significatifs, sauf une diminution du score moyen de l'anxiété sur l'échelle HADS ( $p \leq 0,013$ ;  $d = 0,17$ , petit effet). La fiabilité test-retest était élevée (MET,  $r = 0,934$ ; score de l'anxiété sur l'échelle HADS,  $r = 0,912$ ; score de la dépression sur l'échelle HADS,  $r = 0,90$ ;  $p < 0,001$ ). Parmi les 2 066 patients en RC, la répartition des indices de changement fiables différait ( $p < 0,001$ ) de celle de la plupart des autres critères de changement.

**Conclusions :** Le constat du changement dépend du choix du critère. Une augmentation de 0,5 MET selon l'indicateur de qualité en RC de la Société canadienne de cardiologie est peut-être trop faible pour évaluer le changement de capacité fonctionnelle des individus. Les indices de changement fiables offrent une approche pragmatique pour comparer la fréquence du changement fiable et, en attendant une vérification plus poussée, ils pourraient être utilisés comme rétroaction fournie individuellement aux patients.

( $n = 1554$ ;  $P < 0.0001$ ), by 1 point over 6 months (Cardiac Care Network [CCN] 2002).<sup>28</sup>

In the current set of 3 studies, we calculated RCIs for CR entry-to-discharge FC changes in peak METs, and HADS anxiety and depression scores. Specific objectives were to (i) estimate stability and test practice/exposure effects for FC, and depression and anxiety measures, for RCI calculations; (ii) illustrate clinical use of RCIs for individuals; (iii) in a large clinical cohort, compare RCI distributions to those of other change criteria based upon mean and percent changes, cut-off scores, and SDs, including the 0.5-MET CCS-QI.<sup>11</sup> We used 1-week retest intervals to estimate practice/exposure effects relatively free of treatment-related, maturational, or other systematic influences, hypothesizing that distributions of RCIs would differ from those of other change criteria, thereby providing unique information in determining change in individual patients.

### Methods

This research was approved by the Health Sciences Research Ethics Board, Western University, London, Ontario. Separate samples for studies 1 and 2 were recruited from participants in our Cardiac Rehabilitation & Secondary Prevention Program, which accepts referrals of cardiovascular patients according to standard recommendations.<sup>29</sup>

#### Inclusion criteria: studies 1 and 2

The inclusion criteria were as follows: (i) referral to CR within 1 year of an acute coronary syndrome event, percutaneous coronary intervention, coronary artery bypass grafting, stable angina, or heart valve repair/replacement; (ii) age  $\geq 19$  years; (iii) ability to participate in stress-testing and exercise; (iv) written, informed consent to use data in research; (v) residence location within 1 hour of our CR program; (vi) no medical contraindications to CR participation. For studies 1 and 2, the 1-week interval was chosen to minimize

spontaneous change of exercise capacity or anxiety/depression, which could impact practice/exposure effects and test–retest reliability estimates.

Study 3, in which the RCI formulae were applied retrospectively, used a sample with inclusion criteria similar to those in studies 1 and 2. Additionally, subjects (i) had consecutively entered our CR program from March 24, 2003 to February 23, 2012; (ii) had both entry and discharge measurements available for stress-testing, HADS anxiety score, or HADS depression score (at least 1 of the 3); (iii) did not have recurrent CR intakes; and (iv) participated in neither study 1 nor study 2.

### Study 1: exercise

For time 1, we used results from usual-care stress-testing at CR entry, patients routinely undergo a peak symptom–limited exercise test using a modified Bruce protocol. Peak METs were derived algorithmically from maximum speed and grade attained.<sup>30</sup> Patients completing an entry stress test were asked to undergo an identical test 1 week later (time 2). This interval was chosen to minimize spontaneous change in exercise capacity. Testing was physician-supervised, occurred before exercise program entry to avoid treatment effects, and was scheduled at similar times of day, at least 2 hours after eating, with room environment and equipment conforming to the American Heart Association Guidelines for Exercise Testing Laboratories.<sup>31</sup> We analyzed peak METs. At our centre, CR stress-testing is supervised by physician CR specialists, using usual American Heart Association/American College of Sports Medicine stopping rules.<sup>32</sup> To limit observer effects, physicians were asked not to review the first test results prior to the second test. We strove to maintain consistency of supervising physicians between tests, although we did not require that for technicians. Detailed, standardized instructions for administration of a modified Bruce protocol were provided for each stress test. This protocol had 2 preliminary 3-minute stages (stage 1: 1.7 mph, 0% grade; stage 2: 1.7 mph, 5% grade) added prior to the usual stage 1 of the Bruce protocol; thus, stage 3 was the usual stage 1 of the Bruce protocol.<sup>33</sup> Participants were administered the Borg Rating of Perceived Exertion (RPE) Scale (range in integers: 6 = no exertion at all, to 20 = maximal exertion).<sup>34</sup> A usual-care goal was to have patients exercise to peak exertion by achieving or exceeding a Borg RPE of 17 (very hard), with light handrail holding permitted for safety. Location, use of treadmill, and medication regime were permitted to vary between tests, as a function of scheduling or clinical contingencies.

### Study 2: psychometrics

All subjects were involved in CR exercise programming, conducted by CR program kinesiologists at a local YMCA. To minimize potential effects of early post-event adjustment, program habituation, and imminent discharge, we planned to test patients within the middle third of 6-month CR. Participants completed the HADS (time 1), repeated at the same time of day 1 week later (time 2), well within the minimum recommended 2-week interval.<sup>24</sup> Participants had typically completed the HADS previously at CR entry.

**Sample sizes.** For both studies 1 and 2, we aimed to recruit at least 30 men and 30 women, assuming the confidence interval around a reliability score of 0.60 would not include 0.

### Study 3: CR clinical cohort

We categorized the CR clinical cohort of 2066 patients according to whether each individual had worsened, not changed, or improved from CR entry to discharge. For FC, the change criteria in either direction were:  $\geq 0.5$  MET, the absolute value of which corresponded with the CCS-QI<sup>11</sup>; 1.0 MET, associated with morbidity and mortality<sup>9,10</sup>; 17.5%, the mean percentage change from the OCRPP (CCN, 2002)<sup>28</sup>; meeting or crossing the 7-MET mark, a functional cut-off point<sup>35</sup>; distribution-based changes of 0.5 SD or 1.0 SD; and the RCI.<sup>18</sup> For HADS scores, criteria were as follows: change in either direction of  $\geq 1$  point, corresponding with the mean improvement in the OCRPP (CCN, 2002),<sup>28</sup> associated with mortality<sup>27</sup>; meeting or crossing 7 points, a clinical screening criterion<sup>24</sup>; distribution-based changes of 0.5 SD or 1.0 SD; and the RCI.<sup>18</sup>

### Statistical analyses

For studies 1 and 2, given the relatively small sample sizes, we employed robust methods (bootstrapped, 2000 samples) to obviate concerns about potential deviations from normality. In study 1, to assess consistency and termination of the modified Bruce protocol stress-test administration, we compared time 1 and time 2 with respect to mean peak heart rate and RPE using dependent-measures *t*-tests, and calculated Pearson test-retest reliability. For METs and HADS scores, we expressed central tendency as means (SD), and mean changes at time 2 ( $M_2$ ) vs time 1 ( $M_1$ ) as effect size, Cohen's  $d = (M_2 - M_1)/SD_1$ ; we used dependent-measures *t*-tests to compare means at time 2 vs time 1; and we estimated test-retest stability with Pearson product-moment correlation coefficients.

### RCIs

From cardiology, based on data from 14 male patients with exercise-induced angina and ST-segment depression, who were exercise stress-tested once on each of 2 consecutive days, Sullivan et al.<sup>16</sup> recommended multiplying the standard error of the difference  $\times 1.96$ , to determine the minimum statistically reliable change ( $P < 0.05$ , 2-tailed) in an individual patient. Jacobson and Truax<sup>36</sup> calculated a *reliable change index* (RCI) to determine whether psychometric difference scores of individuals undergoing psychotherapy were statistically reliable, noting that the standard error of the difference estimates the probability of a change score occurring by chance. To correct for practice over repeated neuropsychological test administrations, Chelune et al.<sup>18</sup> subtracted a constant from the expression of Jacobson and Truax,<sup>36</sup> derived from the mean change observed in a “control” sample unexposed to treatment.

To the extent that a “control” sample has retest intervals and characteristics matching those of a patient or cohort receiving an intervention, the approach of Chelune et al.<sup>18</sup> permits estimation of *treatment-specific* effects, apart from error and test practice, but also distinct from spontaneous recovery or deterioration; which are sources of true change.

However, a treatment-free, no-CR “control” condition of cardiovascular patients with a retest interval comparable to the duration of usual-care CR might be affected by non-CR factors, including spontaneous recovery or deterioration, or by CR-independent medication changes or exercise, which would be challenging to exclude, and potentially of interest, as they contribute to “true” clinical change. In any case, withholding CR from eligible patients to form a control condition would be unethical.

We reasoned that clinicians would be interested primarily in gauging individual CR patients’ *overall* true improvements in exercise capacity and psychological well-being, whether caused specifically by CR, already known to be efficacious,<sup>1</sup> or by other true-change factors such as spontaneous recovery or deterioration, CR-independent exercise, or medications. Therefore, we modified the approach of Chelune et al.,<sup>18</sup> by using a 1-week retest interval (in contrast to intervals corresponding with actual CR programming duration) with CR-referred patients, to minimize treatment-related, maturational, or other systematic influences, while correcting for measurement error and test practice/exposure. The RCI of Chelune et al.<sup>18</sup> is calculated as

$$RCI = [(X_2 - X_1) - (M_2 - M_1)] / SD_1 \sqrt{2(1 - r_{12})} \quad (1)$$

As applied here,  $X_2$  and  $X_1$  represent individuals’ test scores in the large clinical cohort at CR program discharge and entry, respectively (retest intervals: approximately 7.3-9.5 months). From the study 1 (exercise) or study 2 (psychometrics) research samples,  $r_{12}$  is the test-retest stability estimate (to estimate error; retest interval = 1 week);  $M_1$  and  $M_2$  are the mean test and retest scores (to estimate mean test exposure/practice), respectively;  $SD_1$  is the time-1 standard deviation. The Pearson product-moment correlation coefficient ( $r$ ) was considered the most appropriate correlation coefficient to use to estimate stability, for 2 reasons: (i)  $r$  was used in the development of RCIs,<sup>18,36</sup>; and (ii) other possibilities, such as using the intraclass correlation coefficient in the RCI of Chelune et al.,<sup>18</sup> would result in 2 corrections for systematic differences such as practice effects, potentially resulting in a mathematical overcorrection.

For CR clinical cohort individuals (study 3), we calculated RCIs using  $(X_2 - X_1)$  for (i) indirectly measured peak MET change, corrected using variables (i.e.,  $M_2$ ,  $M_1$ ,  $SD_1$ ,  $r_{12}$ ) from study 1; and (ii) HADS anxiety score and HADS depression score changes, corrected with study 2 variables. We calculated RCI variables based on whole samples for both study 1 and study 2; where sample size permitted ( $n \geq 30$ , study 2), we developed sex-specific RCI variables. Whole-sample, male-specific, and female-specific variables were subsequently applied to the CR clinical cohort and each of its male and female subsamples, respectively, for whom entry and discharge data were available. As an RCI is a  $z$ -score, reliable change for an individual was defined as an RCI  $\leq -1.96$  or  $\geq 1.96$  ( $P < 0.05$ ; 2-tailed).

For exercise and psychometric outcomes of the CR clinical cohort, we used analysis of variance with time (program discharge vs entry) and sex as within- and between-subjects factors, respectively, and  $t$ -tests (2-tailed, bootstrapped, 2000 samples) for dependent measures to compare time points within sexes. For each HADS subscale, we set 2-tailed significance at  $\alpha = 0.025$  to control type 1 error. Distributions of

CR cohort patients categorized as worsened, unchanged, or improved were compared with nonparametric statistics<sup>18</sup> using a Bonferroni correction to the alpha level for number of comparisons. Analyses were performed using SPSS v.24 or v.26 (IBM, Armonk, NY).

## Results

### Study 1: exercise

A total of 43 subjects consented to participate in study 1. Two stress tests were completed by each of 35 participants (mean age 59.0 years, SD = 8.3, range = 47.9-81.4 years; 6 women = 17.1%; ethnicity Caucasian,  $n = 28$  [80.0%], 5 missing; modal education level = university/college completed,  $n = 12$  [35.3%], 1 missing; see Table 1, clinical characteristics).

Between test 1 and test 2, there was 100% consistency with respect to supervising physicians ( $n = 34$ ; 1 missing). As shown in Table 2, mean peak heart rate and RPE did not differ significantly. For each, there was a significant correlation between test 1 and test 2 ( $P < 0.001$ ; for RPE,  $n = 28$ , 7 missing).

Mean peak METs did not change significantly (Table 2), showed a small effect size ( $d_{\text{METs}} = 0.12$ ), and significant test-retest reliability ( $P < 0.001$ ).

### Study 2: psychometrics

A total of 114 subjects consented to study 2 participation; 100 completed 2 administrations of at least one psychometric subscale; none participated in study 1. Of these, retest intervals in days (number of subjects) were 2 (2), 7 (88), 8 (2), 9 (4), 14 (3), and 37 (1). Post hoc, we excluded the 4 subjects who met/exceeded the minimum 14-day retest interval specified by the test developer.<sup>24</sup> Characteristics of the remaining  $n = 96$  (Table 1) were as follows: mean age = 64.4 (9.7) years, range = 35.9-84.9 years; 32 women (33.3%); 93 Caucasians (97%); modal education level = university/college completed,  $n = 23$  (24.0%). The mean (SD) inter-test interval was 7.0 (0.85) days. The mean (SD) interval from CR referral to test 1 was 25.9 (6.8) weeks, and from test 2 to CR discharge was 12.3 (4.6) weeks. This sample included one subject who had been formally discharged from CR 9 weeks before his test 1.

The mean HADS anxiety (HADS-A) score (Table 3) decreased significantly overall and within sexes ( $P \leq 0.013$ ); with a whole-sample small effect size ( $d_{\text{HADS-A}} = 0.17$ ). The mean HADS depression (HADS-D) score showed no significant changes (Table 3), with a whole-sample trivial effect size ( $d_{\text{HADS-D}} = 0.02$ ) Test-retest reliability was significant for all measures ( $P < 0.001$ ).

### Study 3: CR cohort

The clinical cohort (Table 1) was comprised of 2066 patients (mean age = 62.0 years, SD = 10.7, range = 23.0-89.7; 533 women = 25.8%; 1941 Caucasian = 94.0%, 2 missing; modal education level = university/college completed,  $n = 564$  or 27.5%, 18 missing). Mean (SD) intervals (weeks), based on available data, were: referral to CR entry ( $n = 2058$ ), 12.8 (6.8); CR entry to discharge, ( $n = 1969$ ), 32.7 (5.5); stress-testing retest ( $n = 2024$ ), 31.4 (6.3); psychometric retest ( $n = 1956$ ), 40.6 (7.4).

**Table 1. Clinical characteristics of samples at cardiac rehabilitation entry**

Characteristic	Study 1: exercise stress-testing (n = 35)	Study 2: psychometrics (n = 96)	Study 3: cardiac rehabilitation clinical cohort (n = 2066)
<b>Risk factors</b>			
Body mass index $\geq 25$ , kg/m <sup>2*</sup>	33/34 (97.1)	74/96 (77.1)	1593/2066 (77.1)
Sedentary <sup>*,†</sup>	24/35 (68.6)	64/96 (66.7)	1264/2058 (61.4)
Dyslipidemia	20/35 (57.1)	70/96 (72.9)	1317/2059 (64.0)
Hypertension	23/35 (65.7)	64/96 (66.7)	1153/2062 (55.9)
Personal cardiovascular history	7/21 (33.3)	Not available	424/1035(41.0)
Family cardiovascular history	18/35 (51.4)	54/96 (56.3)	1083/2028 (53.4)
Diabetes	4/35 (11.4)	19/96 (19.8)	395/2065 (19.1)
Smoking			
Current	1/34 (2.9)	13/96 (13.5)	294/2062 (14.3)
Past	20/34 (58.8)	45/96 (46.9)	910/2062 (44.1)
<b>Index/referral events<sup>‡</sup></b>			
Stable coronary artery disease	1/35 (2.9)	8/96 (8.3)	158/2066 (7.6)
Myocardial infarction	16/35 (45.7)	40/96 (41.7)	1027/2066 (49.7)
Unstable angina	5/35 (14.3)	6/96 (6.3)	175/2066 (8.5)
Acute coronary syndrome—unclassified	3/35 (8.6)	6/96 (6.3)	50/2066 (2.4)
Percutaneous coronary intervention	22/33 (66.7)	38/96 (39.6)	867/2066 (42.0)
Coronary artery bypass grafting	6/35 (17.1)	26/96 (27.1)	616/2066 (29.8)
Aortic valve surgery	2/35 (5.7)	6/96 (6.3)	84/2066 (4.1)
Mitral valve surgery	0	3/96 (3.1)	53/2066 (2.6)
Congestive heart failure	0	5/96 (5.2)	33/2066 (1.6)
Cardiomyopathy	0	2/96 (2.1)	10/2066 (0.5)
Transient ischemic attack or mild stroke	0	0	6/2066 (0.3)
Other	1/35 (2.8)	4/96 (4.2)	41/2066 (2.0)

Values are n/total n (%; in repeated measures sample for risk factors).

\*Ascertained from measurement (body mass index) or enquiry at cardiac rehabilitation entry (other variables ascertained from chart/referral documentation).

†Sedentary was defined as < 30 min/d moderate physical activity 3 d/wk.<sup>45</sup>

‡Patients may have >1.

**Table 2. Study 1: exercise stress-testing variables**

n = 35	Time 1, mean (SD)	Time 2, mean (SD)	Mean difference	<i>P</i> <sub>difference</sub> (2-tailed)	Percent difference	<i>r</i> <sub>12</sub>	<i>P</i> <sub>r</sub>
Days from referral	77.2 (91.2)	79.2 (91.2)	7.0	n/a	n/a	n/a	n/a
Peak HR (bpm)	131.66 (25.63)	132.43 (26.51)	0.77	0.62	0.59	0.940	< 0.001
Peak RPE (n = 28)	17.36 (1.99)	17.68 (1.76)	0.32	0.10	n/a	0.858	< 0.001
Peak METs	9.78 (3.01)	10.13 (3.15)	0.35	0.09	3.60	0.934	< 0.001

bpm, beats per minute; HR, heart rate; METs, metabolic equivalents; n/a, not applicable; RPE, rate of perceived exertion; SD, standard deviation.

For the whole cohort and each sex (Table 4), all variables changed significantly from CR entry to discharge ( $P < 0.001$ ). The sex main effect was significant for all variables ( $P \leq 0.002$ ), with men showing higher mean exercise capacity, and lower mean anxiety and depression scores. The interaction of time with sex was significant for METs ( $P < 0.001$ ) but not for the HADS anxiety score ( $P = 0.122$ ) or the HADS depression score ( $P = 0.041$ ;  $\alpha_{crit} = 0.025$ ).

Figures 1 (exercise) and 2 (psychometrics) display frequency distributions of clinical cohort patients who

worsened, did not change, or improved from CR program entry to discharge, according to the aforementioned change criteria. Most change criteria produced distributions differing significantly ( $P < 0.001$ ) from each other, and in particular, from those of the RCIs. Among all distributions, only the HADS anxiety “meet/cross 7 points” vs 1.0 SD ( $P = 0.18$ ) and; vs RCI ( $P = 0.019$ ) did not differ significantly. HADS anxiety score ( $P = 0.58$ ) and HADS depression score ( $P = 0.30$ ) RCI distributions (not shown) did not differ significantly between sexes.

**Table 3. Study 2: psychometric variables**

Variable	Time 1, mean (SD)	Time 2, mean (SD)	Mean change	<i>P</i> <sub>change</sub> (2-tailed)	<i>r</i> <sub>12</sub>	<i>P</i> <sub>r</sub>
<b>Whole sample</b>						
HADS anxiety score (n = 93)	5.91 (3.80)	5.26 (3.81)	-0.66	< 0.001	0.912	< 0.001
HADS depression score (n = 93)	3.59 (3.34)	3.53 (3.43)	-0.065	0.679	0.900	< 0.001
<b>Women</b>						
HADS anxiety score (n = 32)	6.28 (3.84)	5.41 (4.03)	-0.88	0.003	0.922	< 0.001
HADS depression score (n = 31)	2.90 (2.66)	3.10 (3.22)	-0.19	0.508	0.876	< 0.001
<b>Men</b>						
HADS anxiety score (n = 61)	5.72 (3.80)	5.18 (3.73)	-0.54	0.013	0.908	< 0.001
HADS depression score (n = 62)	3.94 (3.61)	3.74 (3.54)	-0.194	0.311	0.913	< 0.001

HADS, Hospital Anxiety and Depression Scale; SD, standard deviation.

**Table 4. Study 3: exercise stress-testing and psychometric outcomes among cardiac rehabilitation cohort of n = 2066**

Measure	n	Entry, mean (SD)	Discharge, mean (SD)
<b>Exercise: METs</b>			
Overall	2028	7.56 (3.30)	9.29 (3.62)
Men	1508	8.15 (3.28)	9.97 (3.56)
Women	520	5.83 (2.59)	7.31 (3.01)
<b>HADS anxiety score</b>			
Overall	1722	5.94 (3.88)	4.99 (3.48)
Men	1272	5.58 (3.76)	4.67 (3.38)
Women	450	6.98 (4.04)	5.88 (3.59)
<b>HADS depression score</b>			
Overall	1722	3.91 (3.30)	2.82 (2.84)
Men	1272	3.75 (3.21)	2.72 (2.83)
Women	450	4.36 (3.49)	3.10 (2.86)

HADS, Hospital Anxiety and Depression Scale; METs, metabolic equivalents; SD, standard deviation.

**Discussion**

We propose a method for clinical application, correcting for measurement error and mean test practice/exposure effects, to determine whether an individual patient has manifested true change from CR entry to discharge. Our exploration of different approaches demonstrates that ascertainment of change depends strongly upon the change criteria applied to individual patients. Present findings suggest the 0.5-MET CCS-QI<sup>11</sup> may be too small or liberal to ascertain true change over and above that from error and test practice in individuals, when using “indirect” FC measurement.

**Study 1: exercise**

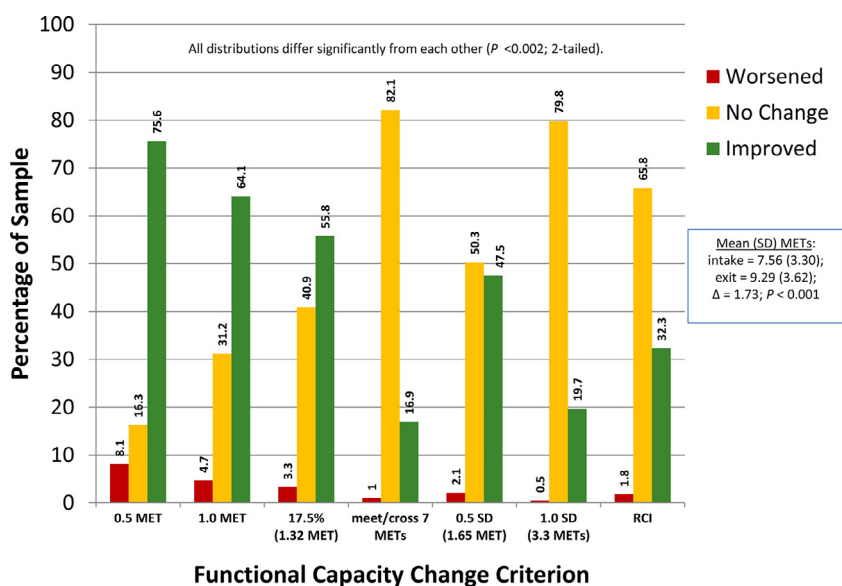
Stress-test protocols were administered very consistently between time 1 and time 2, including consistency of supervising physicians, very similar levels of mean peak heart rates and Borg scale ratings, the latter corresponding with “very hard” RPEs, and high FC test-retest reliability. We achieved uniformity of 7-day retest intervals, and both measurements occurred before exercise program entry. Consequently, any

test-retest FC differences were not due to CR exercise training, and were unlikely to have been caused by spontaneous change in fitness, but instead probably reflected test practice. The FC improvement was not statistically significant, although this was likely due to type II error, given combined small effect and sample sizes. Nonetheless, it is noteworthy that the estimated practice effect of 0.35 METs equates to 70% of the 0.5-MET CCS-QI,<sup>11</sup> assuming that they were measured indirectly as FC.

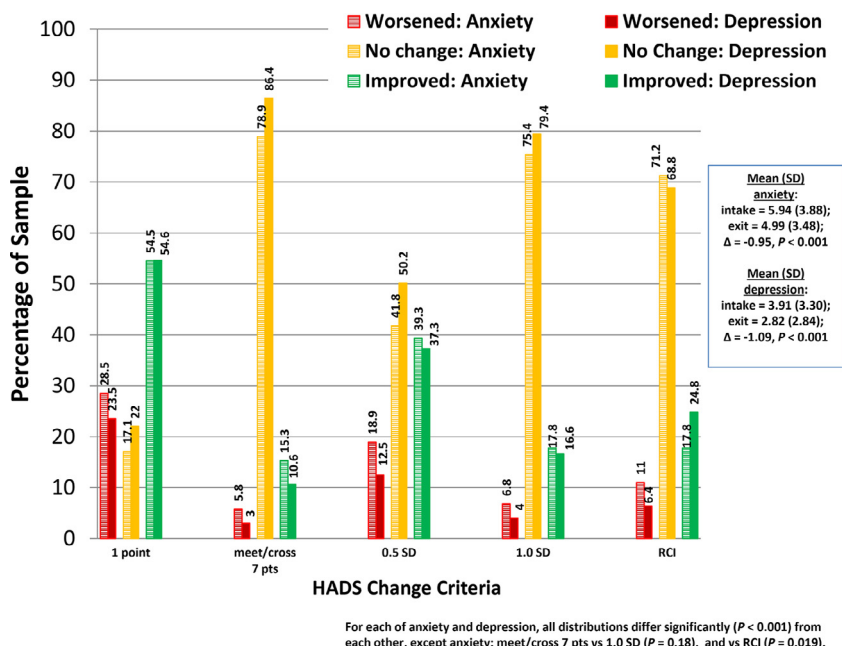
Peak METs demonstrated very high stability, providing a strong basis to calculate RCIs. For practical illustration, applying study 1 variables to equation 1 and requiring  $RCI_{METs} \leq -1.96$  or  $\geq 1.96$  (2-tailed) or  $RCI_{METs} \geq 1.64$  (1-tailed) for significance, an individual patient would have to change by at least  $\pm 2.50$  METs, or gain 2.15 METs, respectively, to show true FC change, over and above that from error and test practice. Thus, hypothetically, *individual Y*, who gained 3.0 METs from CR entry to discharge, improved over and above the change from error and practice. However, for *individual Z*, whose performance increased by 1.0 MET, twice the CCS-QI of 0.5 METs,<sup>11</sup> and matching the significant overall mean increase from the OCRPP,<sup>28</sup> we are unable to conclude that FC changed reliably.

**Study 2: psychometrics**

The mean retest interval was well within the recommended minimum retest interval of 2 weeks<sup>37</sup>; tests 1 and 2 occurred well after referral and before CR-discharge, respectively. Therefore, the significant improvements obtained in mean anxiety scores were unlikely to reflect treatment, true spontaneous change, acute adjustment, or anticipation of discharge. This finding suggests the importance of considering test exposure even in cases in which responses do not depend on psychomotor performance. Scores from the HADS demonstrated high to very high stabilities, providing a strong basis to calculate RCIs.



**Figure 1.** Percentages of patients in cardiac rehabilitation cohort (n = 2066) with worsened, unchanged, or improved functional capacity, by change criterion. METs, metabolic equivalents; RCIs, reliable change indices; SD, standard deviation.



**Figure 2.** Percentages of patients in cardiac rehabilitation cohort (n = 1722) with worsened, unchanged, or improved Hospital Anxiety and Depression Scale (HADS) anxiety or depression scores, by change criterion. RCIs, reliable change indices; SD, standard deviation.

Applying study 1 variables to equation 2 and requiring  $RCI_{HADS} \leq -1.96$  or  $\geq 1.96$  (2-tailed), or  $RCI_{HADS} \geq 1.64$  (1-tailed) for significance, an individual patient would have to change by at least  $\pm 2.42$ , or 1.91 HADS anxiety points, respectively, to evidence true change in psychometrically measured anxiety symptoms, over and above that from error and test exposure. Corresponding values for the HADS depression score are  $\pm 2.80$  or 2.32 points, respectively. Recently, Lemay et al. reported an MCID of 1.7 points for each HADS subscale.<sup>38</sup> As  $1.7 < 1.91$  points, this MCID is within the potential measurement error and test exposure effects we estimated for the HADS, and it therefore may be too liberal for application with individual patients.

### Study 3: CR cohort

The samples for studies 1 and 2 were recruited from the same CR population as the large clinical cohort. Consequently, variables determined from the study samples were generalizable to the large clinical cohort. All samples were referred according to standard criteria<sup>29</sup> and thus were broadly similar to CR populations elsewhere, being composed substantially of patients with coronary heart disease with or without acute coronary syndrome, large proportions of whom had undergone revascularization procedures.<sup>28,39</sup> Mean initial FC in study 1 and the CR cohort suggested relatively high exercise capacity for CR patients. However, indirect measurement with light hand-rail-holding, as conducted in our centre, probably overestimates direct peak  $VO_2$  by approximately 2 METs.<sup>40</sup> Subtracting 2 METs from the CR cohort FC levels at intake would give  $VO_2$  estimates comparable to those in Alter et al.'s<sup>41</sup> real-world CR outcomes study, as determined by direct peak  $VO_2$  measurement. The higher baseline level in study 1 could represent a selection bias in

favour of healthier volunteers. In general, we believe that the external validity of our study is reasonable.

This cohort, overall and by sex, showed robustly significant mean improvements in exercise capacity, anxiety, and depression, which we used to illustrate a clinical problem: given statistically significant group changes and no control condition, how can a clinician infer what constitutes true change for individuals? Comparisons of the change distributions show that answering this question depends crucially on the criterion applied.

With respect to exercise, the criterion least sensitive to change (highest percentage of unchanged patients) was meeting or crossing the 7-MET mark. Although this level may be functionally important,<sup>35</sup> we recommend against using it as a simple benchmark of individual *change*.

The criterion giving the most favourable ascertainment of change in exercise capacity (highest percentage improved) was  $\pm 0.5$  MET, corresponding with the CCS-QI.<sup>11</sup> Notably, this QI does not consider measurement error and test practice effects. Furthermore, it (intentionally) does not specify the method of measurement, whether through gas-exchange techniques, or indirect calculation from peak treadmill speed and grade, which is more subject to practice<sup>15</sup> and error effects.<sup>16</sup> These issues are crucial, given that the mean (nonsignificant) practice effect in study 1 alone could account for 70% of the 0.5-MET QI in a larger sample with more statistical power. Although the CCS-QI may meet criteria for an MCID, an FC increase of 0.5 METs by an individual would fall within the range of measurement error and practice effect, as illustrated above; in other words, that change would not be reliable. Therefore, we recommend against interpreting a gain of 0.5 METs measured *indirectly*, as meeting the CCS-QI, *for individuals*. We also recommend measuring exercise capacity (peak  $VO_2$ ) directly, with cardiopulmonary exercise testing.

When direct cardiopulmonary measurement of peak  $\text{VO}_2$  is not feasible, we suggest application of the critical RCI scores ( $\pm 2.50$  METs, 2-tailed; or 2.15 METs, 1-tailed) derived from RCI analyses reported here to ascertain reliable change by individuals.

We also suggest a review to determine whether the CCS-QI<sup>11</sup> suffices as a reliable criterion of change for *individual* patients. As direct peak  $\text{VO}_2$  uptake measures are less sensitive to practice<sup>15</sup> and error<sup>16</sup> effects than indirect FC measures based on peak treadmill speed and grade,<sup>14,16,21-23</sup> such a review should consider the “direct” and “indirect” exercise-capacity measurement situations separately.

With respect to HADS anxiety and depression scores, the change criterion of 1 point<sup>27</sup> produced the most favourable ascertainment of improvement. Comparison with RCI distributions and critical RCI values suggests that using a 1-point change criterion may overestimate the frequency of reliable improvement for individuals.

There are depression-related CCS-QIs for programmatic processes<sup>11</sup> but not outcome. Challenges to standardized QI implementation include the variety of psychometric instruments used to measure depression.<sup>11</sup> These factors, and scientific challenges to defining clinically meaningful change, could complicate any depression-outcome QI. RCIs, which are standard (“z”) scores, could enable meaningful comparisons with respect to depression outcomes, across psychometric instruments and CR programs. We propose that any future outcome QI should at minimum be capable of measuring individual-level reliable change.

Practical implications of our findings include benchmarking for program evaluation or to inform quality improvement, as, for example, with respect to frequencies of participants who either show reliable decrements, do not change, or reliably improve over CR, with respect to key outcomes including exercise capacity and emotional functioning. In light of the potential impact on individual CR participants, direct feedback to them might be considered, after further validation of RCIs against clinical criteria. These RCIs can be used clinically, similar to how they are applied in other disciplines (for example, psychology) to help determine if a patient seen for at least 2 testing sessions has improved, declined, or stayed the same on a standardized measure. This determination is impossible to make by eyeballing raw score changes. In contrast, RCIs allow a clinician to determine if any changes in raw scores are beyond expected fluctuations in test scores based on measurement error and practice effects.

### Strengths and limitations

Strengths of our research include the following: short retest intervals, minimization of potential contributions of “true” treatment or maturational change to our practice/exposure estimates; high stability estimates; and use of RCIs, which permits correction of individual change scores for measurement error and test practice/exposure.

The principal limitation is that we have not validated RCIs against clinical criteria such as future cardiovascular event risk. Using 1-week retest intervals may not accurately estimate error or practice/exposure over CR. If treadmill practice effects decayed, 1-week retest intervals might overcorrect for practice over longer intervals. Conversely, treadmill use in CR might

preserve or enhance practice, independently of true exercise-capacity change. Short retest intervals might then underestimate practice in CR. Further studies should test the validity of generalizing practice estimates derived from short retest intervals to 6-month CR.

Further challenges to our estimation of error and practice effects are that we did not control for medication changes between time 1 and time 2, which could have occurred for clinical reasons after the first stress test; or for retest changes in technician, location, or treadmill. However, changes to medications would have been difficult to control, as they reflect normal clinical practice. Also, all testing was done at London Health Sciences Centre locations, reflecting commonality of equipment and maintenance. The high levels of inter-test consistency with respect to peak heart rate and RPE, and test-retest reliability of peak METs, suggest that these issues were not influential.

This method of assessing practice effects assumes uniform benefit from practice/exposure, but baseline fitness could interact with test repetition. This possibility could be explored in future work, by comparison of RCIs as calculated here<sup>18</sup> to regression-based approaches for measuring individuals’ change.<sup>42,43</sup> Our work was also limited by small sample sizes, which preclude sex-specific exercise estimates, and by low ethnic diversity, which limits its generalizability.

### Conclusions

Although requiring further validation, reliable change methodology offers a promising and pragmatic approach to measuring true change in individuals in CR. RCIs can provide complementary information to external criteria of clinically important change.<sup>44</sup> Ascertaining change from CR entry to discharge strongly depends upon the specific criteria applied, as RCI distributions differed significantly from those of other change criteria we tested. Our data suggest that improvement by 0.5 MET in exercise capacity, a CCS-QI,<sup>11</sup> is too liberal a criterion, when indirect measurement is used, to ascertain true change in individual patients over and above error and practice effects. We recommend use of direct, cardiopulmonary measurement of peak  $\text{VO}_2$  to ascertain exercise capacity, when feasible. Although the 0.5 CCS-QI may meet criteria for an MCID, we recommend a review of its reliability and validity when it is applied to individuals, particularly in situations of indirect measurement. In any such review, we recommend separate consideration of direct  $\text{VO}_2$  measurement vs indirect algorithmic estimation of exercise capacity. We propose that RCI methodology may have utility in development of future outcome-related psychometric QIs. The RCI critical values reported here may be applied to frequencies of individual CR participants for program evaluation or quality assurance purposes.

### Acknowledgements

We thank Karen Unsworth, MSc, and the CR Program clinical team for facilitating data collection.

### Funding Sources

This research was funded by grant IRF-018-06 to PLP, MEO, NS, and Karen Unsworth from the Lawson Health Research Institute’s (LHRI) Internal Research Fund. PLP



receives salary support from LHRI and the Division of Cardiology, Department of Medicine, Western University. NS is supported by the Program of Experimental Medicine of the Department of Medicine, Western University. The funding sources had no role in the study design, data collection, analysis or interpretation, writing of the article, or the decision to submit this article for publication.

## Disclosures

The authors have no conflicts of interest to disclose.

## References

1. Anderson L, Thompson R, Oldridge N, et al. Exercise-based cardiac rehabilitation for coronary heart disease. *Cochrane Database Syst Rev* 2016;2016:CD001800.
2. De Schutter A, Kachur S, Lavie CJ, et al. Cardiac rehabilitation fitness changes and subsequent survival. *Eur Heart J Qual Care Clin Outcomes* 2018;4:173–9.
3. Lichtman JH, Froelicher ES, Blumenthal JA, et al. Depression as a risk factor for poor prognosis among patients with acute coronary syndrome: systematic review and recommendations: a scientific statement from the American Heart Association. *Circulation* 2014;129:1350–69.
4. Grace SL, Abbey SE, Irvine J, Shnek ZM, Stewart DE. Prospective examination of anxiety persistence and its relationship to cardiac symptoms and recurrent cardiac events. *Psychother Psychosom* 2004;73:344–52.
5. Prior PL, Francis JA, Reitav J, Stone JA. Behavioural and psychosocial issues in cardiovascular disease. In: Stone JA, Arthur H, Suskin N, eds. *Canadian Guidelines for Cardiac Rehabilitation and Cardiovascular Disease Prevention: Translating Knowledge into Action*, 3rd ed., Winnipeg, Manitoba: Canadian Association of Cardiac Rehabilitation; 2009.
6. Kabboul NN, Tomlinson G, Francis TA, et al. Comparative effectiveness of the core components of cardiac rehabilitation on mortality and morbidity: a systematic review and network meta-analysis. *J Clin Med* 2018;7:514.
7. Albus C, Herrmann-Lingen C, Jensen K, et al. Additional effects of psychological interventions on subjective and objective outcomes compared with exercise-based cardiac rehabilitation alone in patients with cardiovascular disease: a systematic review and meta-analysis. *Eur J Prev Cardiol* 2019;26:1035–49.
8. Shaffer JA, Kronish IM, Falzon L, Cheung YK, Davidson KW. N-of-1 randomized intervention trials in health psychology: a systematic review and methodology critique. *Ann Behav Med* 2018;52:731–42.
9. Hambrecht R, Walther C, Mobius-Winkler S, et al. Percutaneous coronary angioplasty compared with exercise training in patients with stable coronary artery disease: a randomized trial. *Circulation* 2004;109:1371–8.
10. Myers J, Prakash M, Froelicher V, et al. Exercise capacity and mortality among men referred for exercise testing. *N Engl J Med* 2002;346:793–801.
11. Grace SL, Poirier P, Norris CM, et al. Pan-Canadian development of cardiac rehabilitation and secondary prevention quality indicators. *Can J Cardiol* 2014;30:945–8.
12. Turner D, Schunemann HJ, Griffith LE, et al. The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol* 2010;63:28–36.
13. Strauss E, Sherman EMS, Spreen O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. 3rd ed. Oxford: Oxford University Press; 2006.
14. Mossberg KA, Greene BP. Reliability of graded exercise testing after traumatic brain injury: submaximal and peak responses. *Am J Phys Med Rehabil* 2005;84:492–500.
15. Froelicher Jr. VF, Brammell H, Davis G, et al. A comparison of the reproducibility and physiologic response to three maximal treadmill exercise protocols. *Chest* 1974;65:512–7.
16. Sullivan M, Genter F, Savvides M, et al. The reproducibility of hemodynamic, electrocardiographic, and gas exchange data during treadmill exercise in patients with stable angina pectoris. *Chest* 1984;86:375–82.
17. Collie A, Darby DG, Falletti MG, Silbert BS, Maruff P. Determining the extent of cognitive change after coronary surgery: a review of statistical procedures. *Ann Thorac Surg* 2002;73:2005–11.
18. Chelune G, Naugle RI, Luders H, Sedlak J, Awad IA. Individual change after epilepsy surgery: practice effects and base-rate information. *Neuropsychology* 1993;7:41–52.
19. Sidney KH, Shephard RJ. Maximum and submaximum exercise tests in men and women in the seventh, eighth, and ninth decades of life. *J Appl Physiol* 1977;43:280–7.
20. Myers J, Froelicher VF. Optimizing the exercise test for pharmacological investigations. *Circulation* 1990;82:1839–46.
21. Foster VL, Hume GJ, Dickinson AL, Chatfield SJ, Byrnes WC. The reproducibility of VO<sub>2</sub> max, ventilatory, and lactate thresholds in elderly women. *Med Sci Sports Exerc* 1986;18:425–30.
22. van 't Hul A, Gosselink R, Kwakkel G. Constant-load cycle endurance performance: test-retest reliability and validity in patients with COPD. *J Cardiopulm Rehabil* 2003;23:143–50.
23. De Cock CC, Visser FC, Roos JP, Van Eenige MJ. Reproducibility of symptom-limited exercise-stress testing performed 2 and 6 weeks after myocardial infarction. *J Intern Med* 1992;232:147–54.
24. Snaith RP, Zigmond AS. *The Hospital Anxiety and Depression Scale Manual*. London: nferNelson; 1994.
25. Herrmann C. International experiences with the Hospital Anxiety and Depression Scale—a review of validation data and clinical results. *J Psychosom Res* 1997;42:17–41.
26. Elliott D. Comparison of three instruments for measuring patient anxiety in a coronary care unit. *Intensive Crit Care Nurs* 1993;9:195–200.
27. Doyle F, McGee HM, De La, Harpe D, Shelley E, Conroy R. The Hospital Anxiety and Depression Scale depression subscale, but not the Beck Depression Inventory-Fast Scale, identifies patients with acute coronary syndrome at elevated risk of 1-year mortality. *J Psychosom Res* 2006;60:461–7.
28. Swabey T, Suskin N, Arthur HM, Ross J. The Ontario Cardiac Rehabilitation Pilot Project. *Can J Cardiol* 2004;20:957–61.
29. Thomas RJ, King M, Lui K, et al. AACVPR/ACC/AHA 2007 performance measures on cardiac rehabilitation for referral to and delivery of cardiac rehabilitation/secondary prevention services endorsed by the American College of Chest Physicians, American College of Sports Medicine, American Physical Therapy Association, Canadian Association of Cardiac Rehabilitation, European Association for Cardiovascular Prevention and Rehabilitation, Inter-American Heart Foundation, National Association of Clinical Nurse Specialists, Preventive Cardiovascular

- Nurses Association, and the Society of Thoracic Surgeons. *J Am Coll Cardiol* 2007;50:1400–33.
30. American College of Sports Medicine. *Guidelines for Exercise Testing and Prescription*. 6th ed. Baltimore: Williams and Wilkins; 2000.
  31. Pina IL, Balady GJ, Hanson P, et al. Guidelines for clinical exercise testing laboratories. A statement for healthcare professionals from the Committee on Exercise and Cardiac Rehabilitation, American Heart Association. *Circulation* 1995;91:912–21.
  32. American College of Sports Medicine Thompson WR, Gordon NF, Pescatello LS. *ACSM's Guidelines for Exercise Testing and Prescription*. 8th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2010.
  33. Fletcher GF, Ades PA, Kligfield P, et al. Exercise standards for testing and training: a scientific statement from the American Heart Association. *Circulation* 2013;128:873–934.
  34. Borg GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc* 1982;14:377–81.
  35. Fletcher GF, Balady GJ, Amsterdam EA, et al. Exercise standards for testing and training: a statement for healthcare professionals from the American Heart Association. *Circulation* 2001;104:1694–740.
  36. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991;59:12–9.
  37. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983;67:361–70.
  38. Lemay KR, Tulloch HE, Pipe AL, Reed JL. Establishing the minimal clinically important difference for the Hospital Anxiety and Depression Scale in patients with cardiovascular disease. *J Cardiopulm Rehabil Prev* 2019;39:E6–11.
  39. Poffley A, Thomas E, Grace SL, et al. A systematic review of cardiac rehabilitation registries. *Eur J Prev Cardiol* 2017;24:1596–609.
  40. McConnell TR, Clark BA. Prediction of maximal oxygen consumption during handrail-supported treadmill exercise. *J Cardiopulm Rehabil Prev* 1987;7:324–31.
  41. Alter DA, Oh PI, Chong A. Relationship between cardiac rehabilitation and survival after acute cardiac hospitalization within a universal health care system. *Eur J Cardiovasc Prev Rehabil* 2009;16:102–13.
  42. Attix DK, Story TJ, Chelune GJ, et al. The prediction of change: normative neuropsychological trajectories. *Clin Neuropsychol* 2009;23:21–38.
  43. Maassen GH, Bossema E, Brand N. Reliable change and practice effects: outcomes of various indices compared. *J Clin Exp Neuropsychol* 2009;31:339–52.
  44. O'Connell ME, Gould B, Ursenbach J, Enright J, Morgan DG. Reliable change and minimum clinically important difference (MCID) of the Repeatable Battery for the Assessment of Neuropsychology Status (RBANS) in a heterogeneous dementia sample: support for reliable change methods but not the MCID. *Appl Neuropsychol Adult* 2019;26:268–74.
  45. Stone JA, ed. *Canadian Guidelines for Cardiac Rehabilitation and Cardiovascular Disease Prevention*, Winnipeg: Canadian Association of Cardiac Rehabilitation; 1999.