Clinical relevance of constant power exercise duration changes in COPD

L. Puente-Maestu*, F. Villar#, J. de Miguel*, W.W. Stringer¶, P. Sanz*, M.L. Sanz+, J. García de Pedro* and Y. Martínez-Abad*

ABSTRACT: The endurance time during constant high work-rate exercise (tLIM) is used to assess exercise capacity in patients with chronic obstructive pulmonary disease and as an outcome measure for pulmonary rehabilitation. Our study was designed to establish the minimum clinically important difference for the tLIM.

tLIM was measured in 105 patients (86 males) before and after an 8-week outpatient pulmonary rehabilitation programme. Subjects were asked to identify, from a five-point Likert scale, the perceived change in their exercise performance immediately upon completion of the exercise tests. The scale ranged from "better" to "worse".

The mean±SD age was 64±5 yrs, forced expiratory volume in 1 s (FEV_1) 47±10% and FEV_1/forced vital capacity 54.7±16.3%. Baseline tLIM at 75% of the peak work rate was 397±184 s, which increased by 62±63% after rehabilitation. In subjects who felt their exercise tolerance was "slightly better", the mean improvement was 34% in the relative improvement over the baseline value (95% CI 29–39)% or 101 (86–116) s compared with 121 (109–134)% in those who reported that their exercise tolerance was "better" and 8 (2–14)% in those who felt their exercise tolerance was "about the same".

Minimum clinically important improvement for tLIM averaged ~33% of baseline. Patients were able to distinguish at least one further additional level of benefit at 120% of baseline.

KEYWORDS: Chronic respiratory questionnaire, clinically relevant changes, constant exercise test reliability

L imitation of physical activity occupies a central role in the symptom complex of patients with chronic obstructive pulmonary disease (COPD) and for that reason, improvement in exercise capacity is a key dimension of response to therapy in COPD and other chronic respiratory diseases [1]. Standardised exercise testing is a reliable and accurate means for assessing exercise performance both in clinical practice and large-scale clinical trials [2]. Several exercise-testing protocols are available, however, endurance time during constant high work-rate exercise (tLIM), *i.e.* above the critical power [3], can be very sensitive in detecting physiological changes induced by interventions [4, 5]. However, we have previously shown that both tLIM and the change in tLIM after leg muscle training vary considerably depending on the intensity set for the constant work rate (CWR) [3]; when above the "critical power", the closer tLIM is to such a physiological point (*i.e.* the less stressful within the high-intensity work rates domain), the larger the expected change [3], making the test more responsive to interventions.

Published reports of clinical investigations generally provide sufficient information to judge the statistical significance of the treatment effect as measured by functional outcomes. However, the clinical interpretation of these changes would be more meaningful if clinically relevant thresholds (*i.e.* the minimum cut-off points that are perceived as beneficial by the patients or subjects) were defined [6–8]. A pilot attempt to estimate the minimal clinically important difference (MCID) of tLIM at 75% of the peak work rate (tLIM,75) arrived at a tentative figure of 105 s [9].

The aim of this study was, therefore, to define MCID for tLIM on a cycle-ergometer. A secondary aim was to test whether different CWR intensities (75% and 85% of the peak ramp work rate) performed differently.

METHODS

Patients diagnosed with COPD [10] according to the reference values used in our laboratory were sent to our centre for rehabilitation at the Hospital Universitario Gregorio Marañón, Madrid, Spain, and were selected for the study

if they were currently nonsmoking, stable and not in need of chronic oxygen therapy [11]. Subjects were excluded if desaturation (<85% by pulse oximetry) was accompanied by arrhythmia or chest pain, or if the patient did not complete the training programme. All subjects signed an informed consent form as approved by our Institutional Committee for Ethics in Human Research (Comité de Ética en la Investigación Clínica del Área 1, Madrid).

The outpatient rehabilitation programme consisted of leg training on the cycle-ergometer for 45 min·day^{-1} divided into as many as three bouts, 3 days·week^{-1}, for 8 weeks. The training began at a work rate equal to 70% of the maximal work rate (WR$_{max}$) achieved on the baseline incremental exercise test (described below). When the subjects were able to tolerate the work rate for the 45 min session, we attempted to incrementally increase the work rate by 5 W every week.

Incremental exercise tests were performed on an electromagnetically braked cycle-ergometer (ER-900; Jaeger, Hochberg, Germany) using a 1-min step protocol [12] at 10 W·min^{-1} to a symptom-limited maximum. Ventilation and pulmonary gas exchange were measured breath-by-breath by an cardiopulmonary exercise system (Oxycon α; Jaeger). WR$_{max}$ was defined as the maximum work rate that could be sustained for at least 30 s. At least 1 day after the incremental test, the subjects randomly performed the two CWR tests on different days at 75% and 85% of the WR$_{max}$ from the pre-training incremental test. The order of the CWR tests was randomised in sequence. All the exercise tests (both incremental and constant) were standardised with respect to the proper seat adjustment relative to leg length and pedalling cadence (60 rpm). The CWR endurance test protocol was 1 min of rest and 1 min of unloaded cycling before the pre-set work rate was instituted. The tests were terminated when, after standardised encouragement, the subjects were unable to continue because of symptoms. If the initial tLIM was <120 s or >1,200 s in either of the two tests, the work rate was increased or decreased in both the 75% and 85% tests by 10% of WR$_{max}$ or by 5 W, whichever was smaller. In total, 13 (12%) of the study subjects needed work-rate adjustment, but none needed more than one adjustment. Test–retest reliability of the 75% and 85% CWR endurance and ventilatory responses were assessed in 25 consecutive subjects by repeating the tests within 2–3 days after the first one. All these tests were performed prior to the start of the leg training programme. Post-rehabilitation tests were performed within a week of the last session.

Subjective ratings of exercise capacity change after the intervention were measured by a Likert scale containing seven categories: ''no change'', ''a little bit better (or worse)'', ''somewhat better (or worse)'', and ''much better (or worse)'' [7, 13], and administered by the therapist after the training programme. As done previously by SINGH et al. [8], we grouped the categories ''somewhat better'' and ''much better'' into ''better'' and ''somewhat worse'' and ''much worse'' into ''worse'' to increase the power of the sample.

Quality of life was measured in all subjects by the same investigator using a version of the Chronic Respiratory Questionnaire (CRQ) [14] translated into, and validated, in Spanish [15]. The CRQ was administered just before the

exercise tests. Clinically relevant changes in the dyspnoea score (CRQ-D) have been found to be 0.5 per item [13], so that a sum total of 2.5 points for the five items was considered relevant. Accordingly, four categories of CRQ-D response were defined: $\leqslant$ -2.5, -2.4–2.4, 2.5–4.9 and >5 of the baseline CRQ-D score [16].

We integrated certain statistical indexes with the two clinical benchmarks described above (subjective ratings of improvement and CRQ-D score) to ''triangulate'' the minimum relevant changes, as suggested by LEIDY and WYRWICH [16]. The statistical indices utilised were the percentage of change from baseline and a distribution-based approach, the effect size (defined as change from baseline to the end of the treatment divided by the SD of the whole group at baseline). Therefore, an effect of 1 corresponds to an increase equal to one SD of the parameter at baseline. In addition, we calculated the ''optimal'' tLIM cut-off points (i.e. those that maximised the area under the receiving operating characteristic curve (AUC)) to detect a subject's changes in their self-rating from an ''about the same'' or lower score to an ''a little bit better'' or higher score, for the subjective ratings of improvement, and changes of <2.5 points versus improvements of $\geqslant$2.5 for the CRQ-D. Sensitivities and specificities of such cut-off points were also calculated.

tLIM exercise variables and their changes after intervention were normally distributed and are described well by their mean$\pm$SD values. The relationships between changes in tLIM, subjective global scale of exercise capacity and CRQ were assessed using the nonparametric Spearman's rank correlation coefficient ''ρ''. Reliability was evaluated by means of Bland–Altman graphical analysis [17], intra-class correlation coefficients and paired t-tests. Variation coefficients were also calculated. The sample size was estimated for an α error of 0.05 and a power of 0.8 from the mean differences observed after recruiting the first 70 subjects (mean values for ''worse'', ''slightly worse'', ''about the same'', ''slightly better'' and ''better'' were (in s): -100 (-134- -65), -23.8 (-33- -14), 39 (26–53), 102 (89–116) and 544 (474–614), respectively, with an overall standard deviation of 262 s). From these results we considered it reasonable to power the study to detect differences with one-way ANOVAs of at least 60 s between all groups except for the last two (slightly better and better), for which detection of a difference of 120 s was considered enough due to the large differences seen in the pilot study and a distribution of the sample over the categories of 10, 10, 20, 30 and 30%, respectively. The estimated sample size needed was 12, 12, 23, 34, 34 or 105 subjects, respectively. The required sample size was calculated using Ene 2.0 software (GSK Servei d'Estadística de la UAB; www.e-biometria.com/ene-ctm/index.htm).

RESULTS

We studied 105 COPD subjects (86 males); 22 met the Global Inititaive for Chronic Obstructive Lung Disease (GOLD) II criteria and the rest, who were moderately hyperinflated and had moderate reductions in exercise capacity, met GOLD III (table 1) [10]. Patients were recruited until the total estimated number was reached, even if the expected number in the ''worse'' and ''a little bit worse'' groups could not be reached. All the participants who tolerated the 75% test were able to tolerate the 85% tests (i.e. were able to endure the 85% test for >2 min).

TABLE 1	Characteristics of 105 study participants
Variables	**Mean $\pm$ SD**
Age yrs	64 ± 5
Height cm	168 ± 4
BMI kg·m^{-2}	23 ± 3
FEV$_1$ % pred	46.8 ± 10
FEV$_1$/FVC %	54.7 ± 16.3
P_{a,O_2} mmHg	65 ± 4
P_{a,CO_2} mmHg	39 ± 4
RV/TLC %	60.1 ± 15.6
TLC %	112 ± 4.6
$V'_{O_2,max}$ L·min^{-1}	1.32 ± 0.27
V'_{O_2} % pred	67.5 ± 14
WR$_{max}$ W	89 ± 23
Dyspnoea CRQ	16.7 ± 2.8

BMI: body mass index; FEV$_1$: forced expiratory volume in 1 s; % pred: % predicted; FVC: forced vital capacity; P_{a,O_2}: arterial oxygen tension; P_{a,CO_2}: arterial carbon dioxide tension; RV: residual volume; TLC: total lung capacity; $V'_{O_2,max}$: maximal oxygen uptake; V'_{O_2}: peak oxygen uptake; WR$_{max}$: maximal work rate; CRQ: Chronic Respiratory Questionnaire.

Both CWR tests showed excellent test–retest reliability (fig. 1), with a tendency to increase tLIM in the second repetition (table 2). Highly significant intra-class correlation coefficients ($\geqslant 0.85$; p<0.001) were found between test and retest assessments of endurance time, as well as of end-exercise ventilatory responses.

For the 75% and 85% CWR endurance tests, pre-treatment tLIM was 397 ± 184 s and 315 ± 194 s. The average increases in tLIM after the intervention were 289 ± 311 s and 138 ± 147 s, or $62 \pm 63\%$ and $48 \pm 57\%$, respectively, for the 75% and 85% CWR tests.

The distribution of responses of the questions about perceived improvement was as follows: ''better'', 33% (n=35); ''slightly better'', 32% (n=33); ''about the same'', 24% (n=25); ''slightly worse'', 6% (n=7); and ''worse'', 5% (n=5). There were significant differences (p<0.05) in baseline CRQ-D scores and tLIM between the ''better'' group and the other groups.

The average improvement in the CRQ-D score was 5.5 ± 5.1. Of the subjects, ~75% improved >2.5 points in their scores (table 3). The changes observed with the CRQ scores correlated well ($\rho=0.65$, p<0.001 for the 75% CWR test and $\rho=0.61$, p<0.001 for the 85% CWR test), with improvements in tLIM and the CRQ-D score. A very good correlation between changes in CRQ-D and perceived improvement was found ($\rho=0.89$).

The mean changes in daily living dyspnoea and perceived exercise tolerance, as measured by the CRQ-D, are described in tables 3 and 4, respectively. Patients needed to improve by 34 (29–39)% or 101 (86–116) s in the 75% CWR test and by 31 (25–34)% or 67 (61–85) s in the 85% CWR test to perceive their tolerance to exercise as ''slightly better''. A further difference was found at 121 (109–134)% or 521 (472–571) s for the 75% CWR and at 92 (81–103)% or 299 (263–335) s for the 85% CWR.
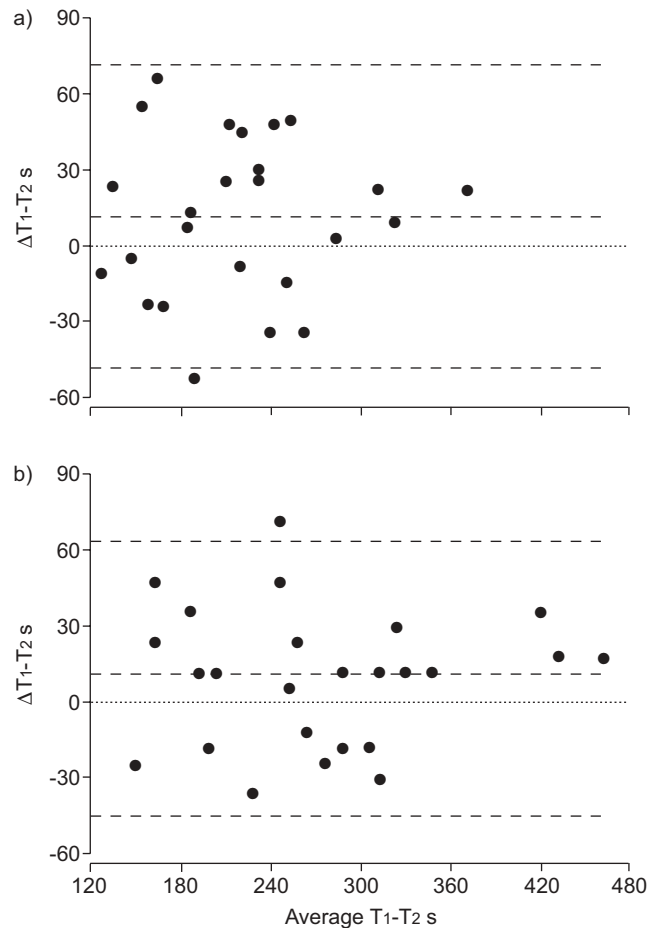
FIGURE 1. Bland and Altman diagram showing the reproducibility of the consant work-rate tests at a) 85% and b) 75% of maximal work rate. T1: first trial; T2: second repetition of trial 3 days after T1. - - - -: the middle line represents the average difference between the two determinations and the two outer lines are the upper and lower limits of the 95% confidence interval of the average difference.

When the effects were reported as absolute time, a relationship ($r\approx0.4$) between baseline tLIM and the magnitude of the improvements was observed (fig. 2); the apparent influence of the initial performance on the magnitude of the effects disappeared when they were standardised as a percentage over the baseline.

The ''optimum'' tLIM cut-off points, by receiver operating characteristic analysis (sensitivity 0.97, specificity 0.93), to detect a subject's changes in their self-rating from ''about the same'' to ''a little bit better'' were 90 s for tLIM,75 (sensitivity 0.9, specificity 1.0) and 76 s for tLIM,85. On analysing the percentages, the best cut-off points were 27% (sensitivity 0.87, specificity 0.9) for tLIM,75 and 23% (sensitivity 0.83, specificity, 0.9) for tLIM,85. From the point of view of their classifying power, the AUC was slightly better (AUC 0.94 (95% CI 0.92-0.98)) for the 75% CWR test compared with the 85% CWR test (AUC 0.85 (95% CI 0.82–0.88)).

The cut-off points were similar: 30 (26–35)% or 90 (74–106) s for the 75% CWR test and 26 (21–30)% or 55 (39–73) s for the 85% CWR test when using changes in CRQ-D $\geqslant 2.5$ points as the clinical benchmark of improvement.

TABLE 2	Reliability of endurance time during constant high-work-rate exercise (tLIM)	
Variables	**Mean $\pm$ SD**	**p-value**
Pre-training tLIM,75 s	420 $\pm$ 126	
Pre-training tLIM,85 s	294 $\pm$ 90	
tLIM,75 (T2–T1) s	12.5 $\pm$ 27.7	0.054
tLIM,85 (T2–T1) s	11.4 $\pm$ 30.9	0.051
VC tLIM,75 %	5.2 $\pm$ 12.5	
VC tLIM,85 %	7.3 $\pm$ 14.8	
Pre-training V'E,75 L·min^{-1}	40.1 $\pm$ 9.6	
Pre-training V'E,85 L·min^{-1}	41.3 $\pm$ 10.6	
V'E,75 (T2–T1) L·min^{-1}	0.84 $\pm$ 1.09	NS
V'E,85 (T2–T1) L·min^{-1}	0.76 $\pm$ 1.22	NS
VC V'E,75 %	2.1 $\pm$ 2.7	
VC V'E,85 %	1.3 $\pm$ 2.8	

tLIM,75 and tLIM,85: endurance to constant work-rate test at 75% and 85%, respectively, of the peak work rate; T1 and T2: the first and second trials, respectively, separated by 3 days; VC: vital capacity; V'E,75 and V'E,85: minute ventilation on exercise at 75% and 85%, respectively, of the peak work rate; NS: nonsignificant. Note: the tLIM and V'E values shown here correspond to only the 25 subjects involved in the test–retest reliability substudy.

TABLE 3	Change after treatment in the different groups according to the changes in chronic respiratory questionnaire initial dyspnoea score (CRQ-D)			
	CRQ-D response categories			
	$\leqslant$ -2.5	**-2.5–2.4**	**2.5–4.9**	**$\geqslant$5**
Subjects n	6	22	37	40
M/F n	5:1	17:5	30:7	33:7
Baseline dyspnoea[#]	16.8 $\pm$ 5.2	15.5 $\pm$ 3.9	15.5 $\pm$ 2.5	17.5 $\pm$ 1.7
Baseline tLIM,75 s	316 $\pm$ 113	313 $\pm$ 187	315 $\pm$ 156	452 $\pm$ 172
Baseline tLIM,85 s	214 $\pm$ 117	260 $\pm$ 177	238 $\pm$ 157	377 $\pm$ 203
ΔtLIM,75 s	-90 $\pm$ 54	11 $\pm$ 34	90 $\pm$ 28	546 $\pm$ 249
ΔtLIM,85 s	-60 $\pm$ 30	1 $\pm$ 23	55 $\pm$ 21	245 $\pm$ 124
ΔtLIM,75 %	-30 $\pm$ 17	3 $\pm$ 14	30 $\pm$ 13	126 $\pm$ 54
ΔtLIM,85 %	-34 $\pm$ 16	1 $\pm$ 16	26 $\pm$ 12	75 $\pm$ 41
Effect size at 75%[¶]	-0.5 $\pm$ 0.3	0.1 $\pm$ 0.2	0.5 $\pm$ 0.2	3.0 $\pm$ 1.4
Effect size at 85%[+]	-0.3 $\pm$ 0.2	0.0 $\pm$ 0.1	0.6 $\pm$ 0.3	1.3 $\pm$ 0.6

Data are presented as mean $\pm$ SD, unless otherwise indicated. M: male; F: female; tLIM,75 and tLIM,85: endurance to constant work-rate test at 75% and 85%, respectively, of the peak work rate; ΔtLIM,75 and ΔtLIM,85: improvement after leg training in the 75% and 85% tests, respectively. [#]: CRQ-D score; [¶]: ΔtLIM,75 divided by the SD of the baseline 75% test; [+]: ΔtLIM,85 divided by the SD of the baseline 85% test.

DISCUSSION

Our study showed that both tLIM,75 and endurance to constant work-rate test at 85% of the peak work rate (tLIM,85) were quite reproducible (fig. 1), which could be explained, in part, by the subjects becoming acquainted with the testing environment and procedures in the previously performed baseline incremental test. In addition, we integrated several clinical and statistical approaches to ''triangulate'' the MCID [16], and the different methods rendered narrow ranges of values of important differences or cut-off points for both tLIM,75 and tLIM,85 (tables 3 and 4). For tLIM,75, for instance, subjects needed to improve by 101 (86–116) s or 34 (29–39)% to rate themselves as clinically slightly improved (table 4). Results were quite similar: 30 (26–35)% or 90 (74–106) s for a significant increase in the CRQ-D score (table 3) and the ''optimum'' cut-off point, calculated as described in the methods, was 90 s. Additional levels of benefit could be perceived by those with larger improvement (tables 3 and 4). Finally, we found that decreases in tLIM,75 of 12 (-28–3.5)% or -29 (-36- -22) s were perceived by the subjects as a slight deterioration, but the sample size of the group who deteriorated was very small.

We found that, when measured in seconds, the MCID was dependent on the baseline state, but this was not the case when it was expressed as a percentage of baseline (fig. 2). Furthermore, when expressed as a percentage of baseline, the MCID were similar in both intensities, suggesting that the percentage of change over baseline is a better way of standardising the response to interventions.

tLIM during high CWR exercise (*i.e.* at 75–85% of the WRmax) has previously been shown to be reliable [18] and highly responsive to therapeutic interventions, such as bronchodilator therapy [5, 19–22], oxygen [23], heliox administration during exercise [24] and rehabilitation [3, 4]. Work rates utilised in

recent clinical studies were 75–85% of maximal oxygen uptake or WRmax measured during symptom-limited incremental cardiopulmonary exercise testing [5, 18–22, 24]. Changes much larger than the proposed cut-off points are typically seen in patients after rehabilitation programmes that include intense leg training [2–4].

One limitation of our study is that we did not perform a prospective validation of the criteria obtained in this work for an independent population in different settings. Secondly, in the present study, we did not test the minimum clinically important difference for bronchodilator therapy. Prior work has demonstrated that small improvements in tLIM can be achieved with this therapy [5, 19–22], but effects >100 s are unusual; for example, the mean increase in endurance time in a large-scale tiotropium trial was observed to be 110 s compared with placebo [19]. The physiological determinants of improvement after leg training are likely to be different from those after bronchodilator therapy [3, 19]. Thirdly, we had to readjust the work rate in 13 (12%) subjects. Based on previous experience [3, 4] we considered that tests of <2 min might be unreliable and tests of >20 min might not be stressful enough for the individual and therefore the subject might terminate them for different physiological or psychological reasons (boredom, seat pain, exhaustion of the muscles, *etc.*). Furthermore, while tests <2 min do not allow time for certain additional measurements, tests >20 min are impractical. Therefore, we faced a dilemma: either to exclude the subjects or to retest them. It is doubtful that these patients represent ''outliers'' to the COPD population, rather they represent either an under- or

TABLE 4	Change after treatment in the different groups according to their perceived improvement				
	"Worse"	"A little bit worse"	"About the same"	"A little bit better"	"Better"
Subjects n	5	7	25	33	35
M/F n	5:0	6:1	21:4	25:8	28:7
Baseline dyspnoea[#]	16.2 ± 5.5	15.5 ± 4.1	15.4 ± 3.3	15.4 ± 3.3	$17.6 \pm 1.7*$
Baseline $t_{LIM,75}$ s	281 ± 81	293 ± 176	308 ± 180	384 ± 191	456 ± 176
Baseline $t_{LIM,85}$ s	169 ± 45	265 ± 143	248 ± 177	291 ± 177	379 ± 207
$\Delta t_{LIM,75}$ s	-100 ± 54	-29 ± 16	31 ± 32	101 ± 26	521 ± 166
$\Delta t_{LIM,85}$ s	-68 ± 22	-25 ± 16	13 ± 17	67 ± 28	299 ± 120
$\Delta t_{LIM,75}$ %	-35 ± 15	-12 ± 8	12 ± 12	34 ± 16	121 ± 41
$\Delta t_{LIM,85}$ %	-40 ± 17	-15 ± 14	8 ± 14	31 ± 14	92 ± 38
Effect size at 75%[¶]	-0.8 ± 0.2	-0.2 ± 0.1	0.2 ± 0.1	0.6 ± 0.2	3.2 ± 1.2
Effect size at 85%[+]	-0.5 ± 0.1	-0.2 ± 0.1	0.2 ± 0.1	0.7 ± 0.3	1.4 ± 0.6
ΔCRQ-D	-5.2 ± 2.7	-1.8 ± 0.7	1.5 ± 1.5	4.2 ± 1.2	10 ± 7

Data are presented as mean ± SD, unless otherwise indicated. M: male; F: female; $t_{LIM,75}$ and $t_{LIM,85}$: endurance to constant work-rate test at 75% and 85%, respectively, of the peak work rate; $\Delta t_{LIM,75}$ and $\Delta t_{LIM,85}$: improvement after leg training in the 75% and 85% tests, respectively; ΔCRQ-D: change in the breathlessness score of the chronic respiratory questionnaire initial dyspnoea score. [#]: CRQ-D score; [¶]: $\Delta t_{LIM,75}$ divided by the SD of the baseline 75% test; [+]: $\Delta t_{LIM,85}$ divided by the SD of the baseline 85% test. *: p<0.05.
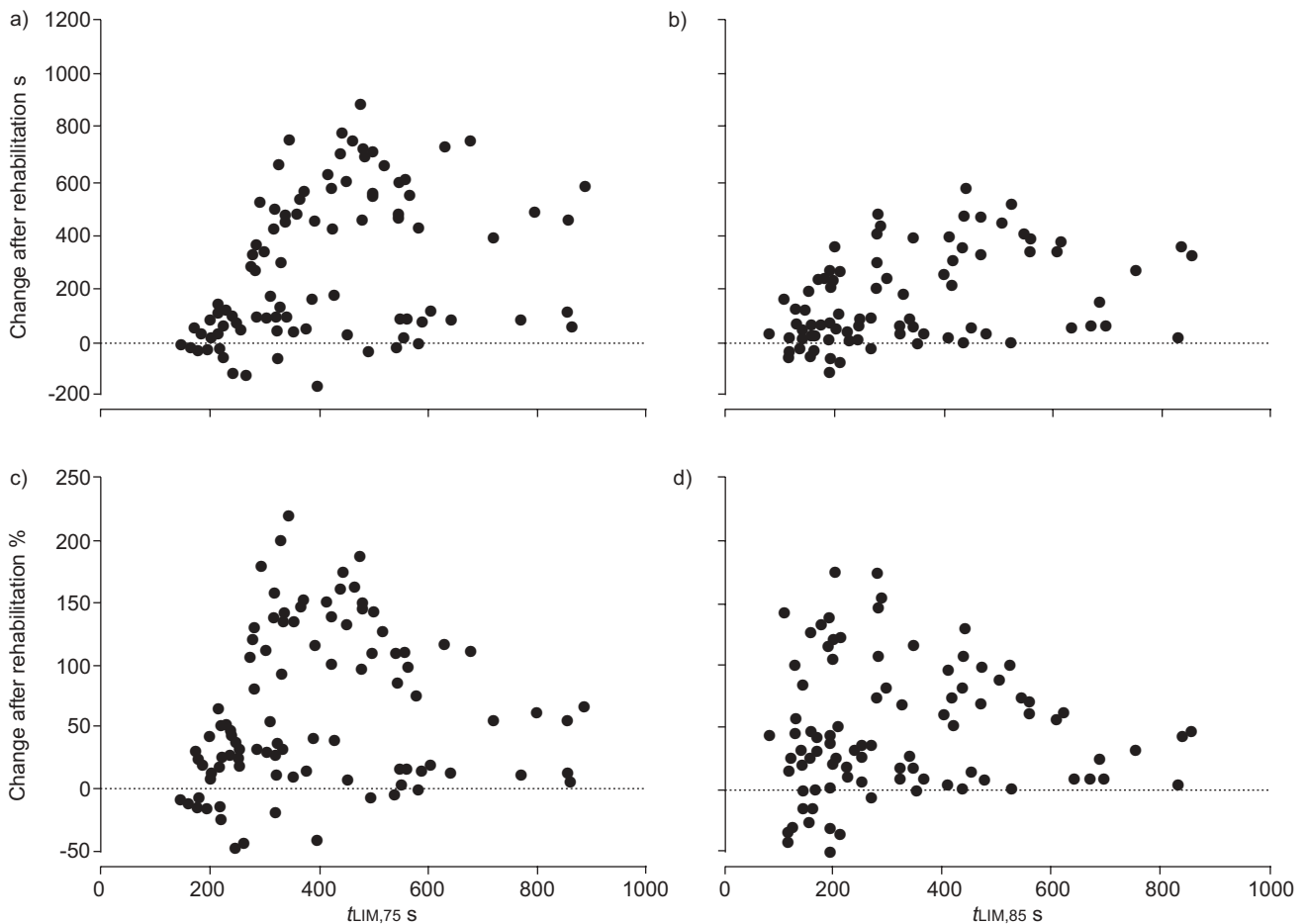
FIGURE 2. Relationship between baseline tolerance and treatment effect for duration of the initial test at 75% of the maximal work rate ($t_{LIM,75}$) with respect to a) time and c) percentage and for endurance to constant work-rate test at 85% of the peak work rate ($t_{LIM,85}$) with respect to b) time and d) percentage. a) r=0.41, p<0.001; b) r=0.40, p<0.001; c) r=0.16; and d) r=0.02.

overestimation of the target work rate; moreover, we believe that in both clinical practice and trials, similar adjustments are and will be common and thus our data could be more readily extrapolated if we included such subjects. Fourthly, only ~20% of subjects enrolled were female, although this proportion is consistent with the published gender distribution of COPD in Spain [25, 26]. Finally, in the present paper, we studied symptomatically mostly severe (GOLD III) patients and, therefore, inferences upon other populations of COPD subjects have to be made with caution. However, the types of patients included in the present study are those most frequently included in rehabilitation programmes.

In summary, our study has found that CWR tLIM is highly reproducible. We also have identified that an increase of 33% in baseline (~100 s in the 75% CWR test, and ~70 s in the 85% CWR test) are important thresholds for clinically relevant changes of tLIM.

STATEMENT OF INTEREST
None declared.

REFERENCES

1 Troosters T, Casaburi R, Gosselink R, *et al.* Pulmonary rehabilitation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2005; 172: 19–38.

2 Nici L, Donner C, Wouters E, *et al.* American Thoracic Society/ European Respiratory Society statement on pulmonary rehabilitation. *Am J Respir Crit Care Med* 2006; 173: 1390–1413.

3 Puente-Maestu L, SantaCruz A, Vargas T, *et al.* Effects of training on the tolerance to high-intensity exercise in patients with severe COPD. *Respiration* 2003; 70: 367–370.

4 Puente-Maestu L, Sanz ML, Sanz P, *et al.* Comparison of effects of supervised *versus* self-monitored training programmes in patients with chronic obstructive pulmonary disease. *Eur Respir J* 2000; 15: 517–525.

5 Oga T, Nishimura K, Tsukino M, *et al.* The effects of oxitropium bromide on exercise performance in patients with stable chronic obstructive pulmonary disease. A comparison of three different exercise tests. *Am J Respir Crit Care Med* 2000; 161: 1897–1901.

6 Redelmeier DA, Goldstein RS, Min ST, *et al.* Spirometry and dyspnea in patients with COPD. When small differences mean little. *Chest* 1996; 109: 1163–1168.

7 Redelmeier DA, Bayoumi AM, Goldstein RS, *et al.* Interpreting small differences in functional status: the six minute walk test in chronic lung disease patients. *Am J Respir Crit Care Med* 1997; 155: 1278–1282.

8 Singh SJ, Jones PW, Evans R, *et al.* Minimum clinically important improvement for the incremental shuttle walking test. *Thorax* 2008; 63: 775–777.

9 Casaburi R. Factors determining constant work rate exercise tolerance in COPD and their role in dictating the minimal clinically important difference in response to interventions. *COPD* 2005; 2: 131–136.

10 Rabe KF, Hurd S, Anzueto A, *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007; 176: 532–555.

11 Roca J, Burgos F, Sunyer J, *et al.* References values for forced spirometry. Group of the European Community Respiratory Health Survey. *Eur Respir J* 1998; 11: 1354–1362.

12 Whipp BJ, Davis JA, Torres F, *et al.* A test to determine parameters of aerobic function during exercise. *J Appl Physiol* 1981; 50: 217–221.

13 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10: 407–415.

14 Guyatt GH, Berman LB, Townsend M, *et al.* A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987; 42: 773–778.

15 Guell R, Casan P, Sangenis M, *et al.* The Spanish translation and evaluation of a quality-of-life questionnaire in patients with chronic obstructive pulmonary disease. *Arch Bronconeumol* 1995; 31: 202–210.

16 Leidy NK, Wyrwich KW. Bridging the gap: using triangulation methodology to estimate minimal clinically important differences (MCIDs). *COPD* 2005; 2: 157–165.

17 Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; 8: 135–160.

18 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: 143–150.

19 O'Donnell DE, Fluge T, Gerken F, *et al.* Effects of tiotropium on lung hyperinflation, dyspnoea and exercise tolerance in COPD. *Eur Respir J* 2004; 23: 832–840.

20 O'Donnell DE, Voduc N, Fitzpatrick M, *et al.* Effect of salmeterol on the ventilatory response to exercise in chronic obstructive pulmonary disease. *Eur Respir J* 2004; 24: 86–94.

21 Maltais F, Hamilton A, Marciniuk D, *et al.* Improvements in symptom-limited exercise performance over 8 h with once-daily tiotropium in patients with COPD. *Chest* 2005; 128: 1168–1178.

22 Oga T, Nishimura K, Tsukino M, *et al.* A comparison of the effects of salbutamol and ipratropium bromide on exercise endurance in patients with COPD. *Chest* 2003; 123: 1810–1816.

23 Emtner M, Porszasz J, Burns M, *et al.* Benefits of supplemental oxygen in exercise training in nonhypoxemic chronic obstructive pulmonary disease patients. *Am J Respir Crit Care Med* 2003; 168: 1034–1042.

24 Palange P, Valli G, Onorati P, *et al.* Effect of heliox on lung dynamic hyperinflation, dyspnea, and exercise endurance capacity in COPD patients. *J Appl Physiol* 2004; 97: 1637–1642.

25 Pena VS, Miravitlles M, Gabriel R, *et al.* Geographic variations in prevalence and underdiagnosis of COPD: results of the IBERPOC multicentre epidemiological study. *Chest* 2000; 118: 981–989.

26 Nowak D, Berger K, Lippert B, *et al.* Epidemiology and health economics of COPD across Europe: a critical analysis. *Treat Respir Med* 2005; 4: 381–395.