

Reliability of Ventilatory Parameters during Cycle Ergometry in Multicentre Trials in COPD

Authors:

Denis E. O'Donnell¹, Justin Travers¹, Katherine A. Webb¹, Zhi He², Yuk-Miu Lam², Alan Hamilton³, Steven Kesten³, François Maltais⁴, Helgo Magnussen⁵

Institutions:

1. Department of Medicine, Queen's University and Kingston General Hospital, Kingston, Ontario, Canada
2. Department of Community Health and Epidemiology, Queen's University, Kingston, Ontario, Canada
3. Boehringer Ingelheim GmbH, Ingelheim, Germany
4. Centre de Recherche de l'Hôpital Laval, Université Laval, Québec, Canada
5. Pulmonary Research Institute at Hospital Grosshansdorf, Center for Pneumology and Thoracic Surgery, Grosshansdorf, Germany

Address for correspondence:

Professor Denis O'Donnell
102 Stuart Street
Kingston, Ontario, Canada K7L 2V6
tel: (613) 548-2339
fax: (613) 549-1459
e-mail: odonnell@queensu.ca

The results of this study were presented, in part, at the European Respiratory Society (ERS) annual congress 2004.

Sources of support: Boehringer Ingelheim GmbH, Ingelheim, Germany;
Dr. O'Donnell held a career scientist award from the Ontario Ministry of Health.

Running head: Reliability of cycle endurance tests in COPD

Body of text word count: 3,599

This manuscript has an online supplement

ABSTRACT

We studied the distribution profiles and repeatability of key exercise performance parameters in the first large multicentre trials to include these measurements in COPD.

After a screening visit, 463 subjects with COPD ($FEV_1 = 43 \pm 13$ %predicted, mean \pm SD) completed two run-in visits before treatment randomization. At the run-in visits, measurements were conducted at rest, a standardized time near end-exercise (isotime) and peak exercise during constant-work rate (CWR) cycle tests at 75% of each individuals' maximum work capacity. The intraclass correlation coefficient was used to evaluate the test-retest repeatability of measurements of endurance time (ET), inspiratory capacity (IC), ventilation and dyspnoea intensity (Borg scale) during exercise.

IC, ventilation and dyspnoea ratings were normally distributed; ET showed rightward skew (median < mean, skewness=10.9 \gg 0) with 16% of the sample exceeding 1 SD of the mean. ET was highly repeatable across run-in visits: 7.9 ± 4.8 and 8.4 ± 5.1 minutes ($R=0.84$). IC values at rest, isotime and peak exercise were all highly repeatable ($R \geq 0.87$); ventilation was repeatable over the same time-points ($R \geq 0.92$); as was dyspnoea intensity at isotime ($R=0.79$) and at peak exercise ($R=0.81$).

In conclusion, key perceptual and ventilatory parameters can be reliably measured during CWR cycle exercise in multicentre clinical trials in moderate to very severe COPD.

Keywords: COPD; exercise testing; cycle endurance; lung hyperinflation; dyspnoea

INTRODUCTION

Measurements of exercise endurance time, dynamic pulmonary hyperinflation and dyspnoea intensity during cardiopulmonary exercise testing are increasingly employed in clinical exercise testing laboratories as important outcome measures in COPD.[1-8] Constant-work rate (CWR) cycle exercise tests conducted at a fixed fraction of each individual's pre-established maximal work rate have been shown to be more responsive than incremental cycle exercise tests or the six-minute walking distance tests for the purpose of bronchodilator evaluation.[7] Indeed, CWR cycle endurance tests appear to be highly responsive to interventions that therapeutically alter dynamic ventilatory mechanics or ventilatory requirements, or both in tandem, in COPD.[2-5,7,9] However, formal evaluation of the test-retest repeatability of this experimental technique has never previously been undertaken in the context of multicentre clinical trials. This information becomes critical for refining future study designs and exercise testing protocols and for interpreting the results of such studies. Moreover, validation of these measurements is a prerequisite for their future incorporation into testing protocols in clinical exercise testing laboratories.

Evidence from small single-centre studies of COPD patients has suggested satisfactory test-retest repeatability of measurements of perceptual and ventilatory parameters during CWR cycle exercise.[10-19] In this study, a pooled analysis of data from two large multicentre, multinational clinical trials allowed us the unique opportunity to study the reliability of these perceptual and ventilatory responses to CWR cycle exercise conducted at 75% of peak work capacity (W_{max}) in a large group of COPD patients. Our first objective, therefore, was to evaluate test-retest repeatability of measurements of endurance time (ET), inspiratory capacity (IC), ventilation (V_E) and dyspnoea intensity during CWR exercise in a large patient population with moderate to very severe COPD.

Our second objective was to evaluate frequency distribution profiles of ET, dynamic pulmonary hyperinflation and dyspnoea intensity in this large population of patients. Endurance time during CWR protocols in COPD is determined by the proximity of the targeted work rate during testing to the individuals' critical power asymptote on the power-duration relationship.[20] This could result in a high degree of variation in ET across individuals so as to make the test unsuitable for the purpose of physiological and subjective measurements in some patients. Frequency distribution analysis of ET in our population cohort allowed us to evaluate the general utility of this exercise testing protocol in COPD. Although dynamic pulmonary hyperinflation is an inevitable consequence of increased ventilatory demands in patients with expiratory flow limitation, very little information is available about the behaviour and distribution profile of operating lung volumes in COPD populations. This is the first large study to chart the variability of operating lung volumes during CWR cycle exercise in COPD and provides new insights into the heterogeneous pathophysiology of this condition. Finally, a number of studies have indicated that during cycle exercise, leg discomfort and not dyspnoea is likely to be the proximate exercise-limiting symptom in patients with moderate to severe COPD.[21,22] We therefore evaluated the suitability of this exercise modality for the purpose of dyspnoea assessment.

METHODS

Subjects

Subjects were participants in one of two multicentre, randomized clinical trials examining the effects of tiotropium on exercise performance, dynamic lung volumes and exertional dyspnoea in COPD.[2,3] Subject eligibility criteria were identical in each trial and have been described previously.[2,3] Subjects were clinically stable patients with COPD, aged 40-70 years, had a cigarette smoking history >10 pack-years, a forced expiratory volume in one second (FEV₁) ≤65 % predicted and a plethysmographic functional residual capacity (FRC) ≥120 % predicted. Subjects with asthma or who had a contraindication to participation in pulmonary function or exercise testing were excluded. Subjects who had participated in the first trial were excluded from the second trial.

Study Design

The two trials incorporated a similar study design. Studies were approved by the medical ethics committees of all sites (36 sites across 5 countries) and all subjects gave written informed consent before undertaking any study procedures. Eligibility criteria were assessed during an initial screening visit (*Visit 1*) conducted 15 days prior to randomization. At this visit, patients performed pulmonary function tests followed by a symptom-limited incremental cycle exercise test to determine a maximal work rate (W_{\max}). Eligible patients completed another two visits (*Visit 2*, day -10; *Visit 3*, day -5) during a “run-in” phase to familiarize them with all testing procedures and establish a standardized series of pre-randomization familiarization tests (i.e., training history). On these days, pulmonary function tests were followed by a symptom-limited CWR exercise test at 75% W_{\max} . On day 0 (*Visit 4*), patients were subsequently randomized to the treatment portion of the studies. Visit 3 values were considered the baseline prior to the post-dose testing on Visit 4. All analyses in the present study were performed on data obtained at the pre-randomization visits. Subjects were instructed not to use their inhaled bronchodilators prior to testing based on a predefined schedule.

Procedures

Spirometry, body plethysmography and single-breath diffusing capacity (D_{LCO}) were performed in accordance with recommended techniques.[23-25] The symptom-limited cycle exercise testing protocols have been described previously.[2,3] Across centres, various cardiopulmonary exercise testing systems (SensorMedics, Yorba Linda, CA, USA; MedGraphics, St Paul, MN, USA; Viasys Healthcare GmbH, Hoechberg, Germany) were used; all systems underwent stringent quality control and physiological validation prior to the study and were calibrated prior to each test. Incremental exercise tests included a steady-state resting period of ≥3 min followed by 3 min of loadless pedalling followed by stepwise increases of 10W/min; W_{\max} was defined as the greatest work rate that the subject was able to sustain for ≥30 s. CWR exercise tests included a steady-state resting period followed by 1 min of loadless pedalling and then an immediate increase in work rate to 75% of W_{\max} ; ET was recorded from the start of loaded pedalling to the point of symptom-limitation. Inspiratory capacity (IC) was measured at rest, at 2 min intervals during exercise and at end-exercise (*see online supplement*).[2] The 10-point Borg scale[26] was used to assess dyspnoea intensity at similar time-points. *Rest* was defined as the steady-state period after ≥3 min of breathing on the mouthpiece while seated at rest on the cycle ergometer before exercise started; *peak* was defined as the last 30 s of loaded pedalling and; *isotime* was

defined as the duration of the shortest CWR exercise test on all testing days rounded down to the nearest full minute within each individual.[2,3]

Statistical Analysis

Values are presented as means \pm SD. A p value of 0.05 was used as the threshold of statistical significance. Frequency statistics examining the primary reasons for stopping exercise were analyzed using a Pearson chi-square test. Subsequent analyses were conducted for ET, dyspnoea intensity ratings and physiological measurements which included $V'O_2$, minute ventilation (V'_E), tidal volume (V_T), breathing frequency (F) and IC. Frequency distributions were analyzed for normality using the one-sample Kolmogorov-Smirnov test. Non-normal distributions were further examined for skewness (symmetry) or kurtosis (peakedness) which were considered significant if the skewness or kurtosis coefficients were >2 . Test-retest reproducibility was assessed using the intraclass correlation (ICC).[27] A one-way random-effect analysis of variance model was used unless there was a systematic difference between variables measured in the first and second CWR exercise test in which case a two-way random-effect analysis of variance model was used. Systematic differences were assessed by two-tailed, paired t-tests. A lower limit of the 95% confidence interval for the ICC ≥ 0.75 indicates high reproducibility.[28] Reproducibility was also evaluated by calculating the within-subject coefficient of variation (CV). The subject SD should be independent of the subject mean for the CV to be reliable; this assumption was tested using a rank correlation coefficient.

RESULTS

A total of 463 subjects completed the run-in phase of the two clinical trials and were included in this analysis. Subject characteristics are shown in Table 1. There was no significant heterogeneity between the two trials in these subject characteristics. There were also similar means and distributions for exercise outcomes (i.e., ET, $\dot{V}O_2$, \dot{V}_E , IC, dyspnoea ratings) across the two trials, therefore the data was pooled for further analysis.

Peak incremental cycle work rate and $\dot{V}O_2$ indicated significant exercise limitation (Table 1). The primary reasons for stopping incremental exercise was only available for 248 subjects completing the second validation study [3]: breathing discomfort (51% of subjects), a combination of breathing and leg discomfort (33%), and leg discomfort (16%).

For both run-in CWR exercise tests: ET and dyspnoea intensity ratings were analyzed for all 463 subjects; one subject was missing peak values for \dot{V}_E and $\dot{V}O_2$; exercise IC values were available for 454 subjects. Peak $\dot{V}O_2$ for both CWR tests was similar to that in the incremental test (Tables 1 and 2). Similar to the incremental test, the majority of subjects stopped CWR exercise primarily due to dyspnoea or a combination of dyspnoea and leg discomfort (Figure 1). The distribution of reasons for stopping was also consistent across both CWR tests for the 435 subjects with available data in the intent-to-treat group.

Measurement distribution

By visual inspection and by the Kolmogorov-Smirnov test ($p > 0.05$), physiological variables ($\dot{V}O_2$, \dot{V}_E , F , V_T , IC) measured at isotime during exercise and at peak exercise all followed a normal distribution. Frequency histograms for resting IC and the magnitude of change in IC during exercise are provided in Figure 2.

At both run-in visits, median ET values (6.4 and 6.8 min) were less than mean values (7.9 and 8.4 min), indicating a non-normal distribution. ET had an asymmetric distribution such that there was a significant skewness with a long right tail (skewness coefficient = 10.9) (Figure 3). Log transformation of the data normalized the ET distribution (Kolmogorov-Smirnov test, $p = 0.46$) (Figure 3). The distribution of the difference between ET at each run-in visit was also normal.

At both run-in visits, mean and median values were similar for dyspnoea intensity ratings measured at both isotime and peak exercise. The distribution of dyspnoea ratings measured at isotime and at peak exercise was symmetrical but not normal: measurements at both time points showed significant kurtosis with flattening of the distribution curve (kurtosis coefficient > 2).

Reliability of measurements

Repeatability of resting spirometric and plethysmographic lung volume measurements across visits 1 through 3 was excellent: IC (ICC=0.89), SVC (ICC=0.92), FRC (ICC=0.90). Repeatability of exercise variables is shown in Table 2. The assumption that the subject SD was independent of the subject mean was not violated for any variable except the ET where there was a clear relationship, i.e., variability increased as ET increased ($r = 0.122$, $p = 0.011$). There was a small but significant increase in ET in the second constant-load test suggesting a learning effect (mean difference = 0.56 min; $p < 0.001$); therefore, the two-way analysis of variance model was used to calculate the ICC for this variable. The CV is not valid as a measure of the reproducibility of ET measurement because of this effect. There was also a very small decrease in isotime measurements of dyspnoea intensity (mean difference = -0.19 Borg units; $p = 0.007$).

The increase in ET from the first to second run-in visit correlated significantly with the corresponding decrease in dyspnoea intensity at isotime during exercise ($r=0.46$, $p<0.0005$).

Cycle endurance time

Of the 463 subjects, 43 (9.3%) had an ET of more than one SD below the mean (i.e., <202 seconds) and 72 (15.6%) had an ET greater than one SD above the mean (i.e., >810 seconds). Patients with a low ET had a significantly lower FEV₁, IC and peak incremental exercise capacity (work rate and VO₂); those with the highest ET had a better preserved resting IC and higher peak exercise capacity (Table 3). The distribution of reasons for stopping exercise was also significantly different across these subgroups at the second run-in visit ($p=0.001$): the low-ET group was more likely to stop primarily due to dyspnoea while the high-ET group stopped primarily due to leg discomfort (Table 3).

There was no significant relationship between ET variability across the two run-in visits (i.e., test-retest difference) and any of the following parameters: baseline FEV₁ (%predicted), FEV₁/FVC and IC (%predicted); peak incremental VO₂ (mL/kg/min or %predicted) and peak work rate (%predicted); height, weight and body mass index; or smoking history and COPD duration. Correlates of ET at the second run-in visit included (all $p<0.0005$): the peak incremental work rate ($r=0.370$), peak incremental VO₂ ($r=0.336$), pre-exercise resting IC ($r=0.313$), and FEV₁ ($r=0.274$). The combination of the peak work rate and the resting IC accounted for 15% of the variability in ET ($r=0.38$).

Dynamic lung hyperinflation

Of the 457 subjects with exercise IC measurements at the second run-in visit, 15% of subjects did not change or increased IC during exercise at the second run-in visit, i.e., the rest-to-peak exercise change in IC was outside 1 SD from the mean (greater than a -0.04 L change). The remaining 85% of the sample who hyperinflated by at least 0.04 L had a significantly ($p<0.05$) lower FEV₁ (by 8%), a larger resting IC (by 17%), lower exercise capacity (by 7%), and greater dyspnoea/V'O₂ slopes (by 16%). Interestingly, there was no significant difference across these subgroups for baseline ET or for their primary reasons for stopping exercise.

The best correlate of the exercise-induced change in IC during the second run-in visit was the resting pre-exercise IC ($r=-0.400$, $p<0.0005$): the larger the IC, the more it decreased during exercise; those with the smallest resting IC had the least change in IC (little room to reduce IRV further). The exercise-induced change in IC also correlated with FEV₁/FVC ratio ($r=0.281$, $p<0.0005$): those with the worst ratios experienced greater reductions in IC. The combination of resting IC and FEV₁ explained 40% of the variance in the exercise-induced change in IC.

DISCUSSION

This study is the first to demonstrate high test-retest repeatability of endurance time, ventilation, IC and dyspnoea intensity ratings during CWR exercise in a large population of patients with COPD. Our results also provide novel information on the distribution characteristics of physiological and subjective responses to CWR cycle exercise testing in this COPD cohort.

Repeatability of endurance time measurements

The average peak VO_2 during incremental and CWR cycle tests were similar, thus confirming that the cycle endurance test (at $75\%W_{\max}$) is a valid test of maximal exercise performance in COPD. Peak symptom-limited VO_2 was also similar across the two run-in CWR tests ($R=0.93$). The high level of reproducibility of ET reported here is in general accordance with previous small single-centre studies.[9-14] Reproducibility of ET was high despite a variety of cardiopulmonary exercise testing systems being employed in multiple centres in diverse locations.

The ICC reflects the degree of variance in values between tests within a given subject, referenced to the degree of variance in values between subjects.[33] Perfect agreement between tests within subjects results in an ICC of 1 and values ≥ 0.75 are accepted as being very highly reproducible. There are many sources of within-subject variance aside from random measurement errors. These include: day-to-day variations in airway function, differences in medication use, subject motivation and extraneous factors such as subject encouragement, subject distraction and prior pre-test activity. In addition, learning or fatigue effects may contribute to within-subject variance between the two tests. Regardless of all these theoretical considerations, however, the ICC for ET in this study was 0.84, confirming high reproducibility. Confounding factors were minimised by performing tests using the same technician at the same time of day (on two separate days) with avoidance of inhaled medication and large meals prior to testing. Although the lung function inclusion criteria were generally broad, some disease heterogeneity was removed by requiring subjects to meet eligibility criteria for baseline levels of static hyperinflation. The within-subject CV is also presented. In the case of the ET measurement, the CV is likely to be less accurate than the ICC as a measure of reproducibility since the variability was large, the variability was significantly dependent on the mean and the mean values changed significantly (albeit modestly) across run-in visits.

The fixed work rate of 75% of each individual's pre-determined W_{\max} was selected because previous experience has taught us that this was likely to provide a desirable duration of ~6-8 minutes of physiological data collection in patients with moderate to severe COPD.[6] Indeed, the central tendency (mean, median) for ET fell within this range. The somewhat arbitrary use of 75% W_{\max} for all subjects would perhaps result in greater variability in ET than the choice of a work rate based on a fixed proportion of a narrower intensity range such as between the lactate threshold (or critical power asymptote) and peak. However, selection of a CWR intensity using this latter method would not be practical in many contexts.

Unlike the normal distribution of the physiological parameters tested, the frequency distribution for ET was significantly skewed to the right; however, a log transformation of the ET data normalized this distribution.[34] Of the 463 enrolled patients, 9.3% had a low ET of more than one SD below the mean (i.e., $\text{ET} < 202$ seconds) and 15.6% had a high ET greater than one SD above the mean (i.e., $\text{ET} > 810$ seconds). Patients with a low ET had a lower FEV_1 , IC and peak exercise capacity (work rate and VO_2) with evidence of intolerable dyspnoea at the

peak of exercise. Those with the highest ET had a better preserved resting FEV₁, IC and higher peak exercise capacity in the incremental test. Patients who occupied the extremes of the frequency distribution for ET could, for the most part, be readily identified during the screening visit based on peak incremental work rate (or VO₂) and baseline resting IC. The rightward skew of the ET distribution suggests that the 16% of patients with the best exercise endurance were likely exercising at a work rate below their critical power.[20] For these patients, a higher work rate (relative to maximum) may be more appropriate for repeat testing. The corollary of this is that selection of a lower constant work rate might be preferable for severely breathless patients with an excessively low peak incremental work rate (or VO₂) and resting IC during the screening visit.

Measurement of dynamic pulmonary hyperinflation

Although many centres had little or no experience with IC measurements during exercise, test-retest reproducibility for measurement of operating lung volumes (IC) was excellent, as were measurements of ventilation and breathing pattern. IC values at rest, isotime and peak exercise were all highly repeatable ($R > 0.87$) between run-in tests 1 and 2. Changes in IC during exercise (from the pre-exercise resting value) reliably reflect changes in end-expiratory lung volume (EELV), provided TLC does not change during exercise.[35] A number of mechanical studies have confirmed stability of TLC as reflected by consistency of peak inspiratory esophageal pressures during serial IC manoeuvres during exercise, thereby validating this method of measuring dynamic lung hyperinflation.[12,36] Recent studies employing optoelectronic plethysmography to measure chest wall and abdominal motion during exercise have suggested that some patients with COPD do not exhibit dynamic hyperinflation during exercise at least as inferred from lower chest wall displacement.[37,38] The current study is the first to analyze the behaviour of operating lung volumes during exercise, measured by change in IC, in a large COPD cohort. The change in IC from rest to peak exercise, inversely reflecting the change in dynamic EELV, was normally distributed. The majority (85%) of the sample showed an increase in EELV, with an average increase of 0.42 L above resting values, in accordance with previous reports.[2,3,9,10,12,19,39] The change in dynamic IC during exercise was directly related to the resting baseline IC: those with lowest resting IC (the greatest resting hyperinflation) showed the least amount of dynamic hyperinflation.

Measurement of exertional dyspnoea

Treadmill walk tests appear to be more popular in clinical exercise testing laboratories than cycle endurance tests and may better mimic daily activity. However, precise measurements of power output and operating lung volumes are more difficult during treadmill exercise, and arterial oxygen desaturation is more pronounced during such weight-bearing activity in COPD patients.[40] Previous studies have provided evidence that leg discomfort is a more prominent exercise limiting symptom during cycle exercise than dyspnoea.[21,22] In the current study, however, the majority (over 80%) of patients stated on direct questioning that dyspnoea was the primary or co-primary symptom limiting exercise, and Borg ratings of dyspnoea intensity at peak exercise were “severe” or “very severe.”

Since dyspnoea is a subjective experience, it is reasonable to anticipate that intensity ratings would be less reproducible than measurement of objective physiological parameters. Given the stability of physiological measurements across run-in tests, a very small but statistically significant improvement in standardized measurements of dyspnoea intensity (by a

mean of ~0.2 Borg units) and a resultant small improvement in ET (by a mean of ~30 s) over the course of the two run-in visits suggests that subjective measurements may be more susceptible to familiarization effects on repeated testing. However, an ICC of 0.79 for Borg ratings at isotime exercise and of 0.81 at peak exercise was reassuring. The dominance of severe dyspnoea and the proven repeatability of its measurement during cycle exercise in COPD, suggests that this testing protocol which includes two run-in tests before randomization is suitable for evaluating the impact of dyspnoea-relieving interventions.

In conclusion, this is the first large study to demonstrate that the measurement of endurance time, operating lung volumes, ventilation and ratings of dyspnoea intensity is highly reproducible in COPD patients during symptom-limited, high intensity, constant work rate exercise. Measurement of dynamic lung hyperinflation and dyspnoea intensity is potentially clinically important and can now be reliably incorporated into cardiopulmonary exercise test protocols. This approach allows a more comprehensive characterization of the individual patient presenting with exercise limitation. Moreover, our results attest to the feasibility of conducting accurate measurement of these key exercise performance parameters using the CWR cycle test in the setting of multicentre clinical trials in patients with COPD.

REFERENCES

1. ATS/ACCP Statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med* 2003; 167: 211-277.
2. O'Donnell DE, Flüge T, Gerken F, Hamilton A, Webb KA, Aguilaniu B, Magnussen H. Effects of tiotropium on lung hyperinflation, dyspnoea and exercise tolerance in COPD. *Eur Respir J* 2004; 23: 832-840.
3. Maltais F, Hamilton A, Marciniuk D, Hernandez P, Sciruba FC, Richter K, Kesten S, O'Donnell D. Improvements in symptom-limited exercise performance over 8 h with once-daily tiotropium in patients with COPD. *Chest* 2005; 128: 1168-1178.
4. O'Donnell DE, Lam M, Webb KA. Spirometric correlates of improvement in exercise performance after anticholinergic therapy in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1999; 160: 542-549.
5. O'Donnell DE, Voduc N, Fitzpatrick M, Webb KA. Effect of salmeterol on the ventilatory response to exercise in chronic obstructive pulmonary disease. *Eur Respir J* 2004; 24: 86-94.
6. Palange P, Ward SA, Carlsen K-H, Casaburi R, Gallagher CG, Gosselink R, O'Donnell DE, Puente-Maestu L, Schols AM, Singh S, Whipp BJ; ERS Task Force. Recommendations on the use of exercise testing in clinical practice. *Eur Respir J* 2007; 29: 185-209.
7. Oga T, Nishimura K, Tsukino M, Hajiro T, Ikeda A, Izumi T. 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.
8. Cazzola M, MacNee W, Martinez FJ, et al; on behalf of the American Thoracic Society/European Respiratory Society Task Force on outcomes in COPD. Outcomes for COPD pharmacological trials: from lung function to biomarkers. *Eur Respir J* 2008; 31: 416-468.
9. Peters MM, Webb KA, O'Donnell DE. Combined physiological effects of bronchodilators and hyperoxia on dyspnea and exercise performance in normoxic COPD. *Thorax* 2006; 61: 559-567.
10. O'Donnell DE, Lam M, Webb KA. Measurement of symptoms, lung hyperinflation, and endurance during exercise in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998; 158: 1557-1565.
11. 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.
12. Yan S, Kaminski D, Sliwinski P. Reliability of inspiratory capacity for estimating end-expiratory lung volume changes during exercise in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1997; 156: 55-59.
13. Dolmage TE, Goldstein RS. Repeatability of inspiratory capacity during incremental exercise in patients with severe COPD. *Chest* 2002; 121: 708-714.

14. Nosedá A, Carpiáux JP, Prigogine T, Schmerber J. Lung function, maximum and submaximum exercise testing in COPD patients: reproducibility over a long interval. *Lung* 1989; 167: 247-257.
15. Silverman M, Barry J, Hellerstein H, Janos J, Kelsen S. Variability of the perceived sense of effort in breathing during exercise in patients with chronic obstructive pulmonary disease. *Am Rev Respir Dis* 1988; 137: 206-209.
16. Owens MW, Kinasewitz GT, Strain DS. Evaluating the effects of chronic therapy in patients with irreversible air-flow obstruction. *Am Rev Respir Dis* 1986; 134: 935-937.
17. Brown SE, Fischer CE, Stansbury DW, Light RW. Reproducibility of VO₂max in patients with chronic air-flow obstruction. *Am Rev Respir Dis* 1985; 131: 435-438.
18. Wergel-Kolmert U, Wisen A, Wohlfart B. Repeatability of measurements of oxygen consumption, heart rate and Borg's scale in men during ergometer cycling. *Clin Physiol Funct Imaging* 2002; 22: 261-265.
19. Belman MJ, Brooks LR, Ross DJ, Mohsenifar Z. Variability of breathlessness measurement in patients with chronic obstructive pulmonary disease. *Chest* 1991; 99: 566-571.
20. Neder JA, Jones PW, Nery LE, Whipp BJ. Determinants of the exercise endurance capacity in patients with chronic obstructive pulmonary disease: the power-duration relationship. *Am J Respir Crit Care Med* 2000; 162: 497-504.
21. Killian KJ, Leblanc P, Martin DH, Summers E, Jones NL, Campbell EJM. Exercise capacity and ventilatory, circulatory and symptom limitation in patients with chronic airflow limitation. *Am Rev Respir Dis* 1992; 146: 935-940.
22. Hamilton AL, Killian KJ, Summers E, Jones NL. Muscle strength, symptom intensity, and exercise capacity in patients with cardiorespiratory disorders. *Am J Respir Crit Care Med* 1995; 152: 2021-2031.
23. American Thoracic Society. Standardization of spirometry - 1994 update. *Am J Respir Crit Care Med* 1995; 152: 1107-1136.
24. Coates AL, Peslin R, Rodenstein D, Stocks J. Measurement of lung volumes by plethysmography. *Eur Respir J* 1997; 10: 1415-1427.
25. Cotes JE, et al. Standardization of the measurement of transfer factor (diffusing capacity). *Eur Respir J* 1993; 6: 41-52.
26. Borg GAV. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc* 1982; 14: 377-381.
27. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1996; 1: 30-46.
28. Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring a quantitative variable. *Comput Biol Med* 1989; 19: 61-70.
29. Morris JF, Koski A, Temple WP, Claremont A, Thomas DR. Fifteen-year interval spirometric evaluation of the Oregon predictive equations. *Chest* 1988; 92: 123-127.
30. Crapo RO, Morris AH, Clayton PD, Nixon CR. Lung volumes in healthy nonsmoking adults. *Bull Europ Physiopath Resp* 1982; 18: 419-425.

31. Burrows B, Kasik JE, Niden AH, Barclay WR. Clinical usefulness of the single-breath pulmonary diffusing capacity test. *Am Rev Respir Dis* 1961; 84: 789-806.
32. Jones NL. *Clinical Exercise Testing*, 3rd edition. Philadelphia, PA: Saunders, 1988; p.306.
33. Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. *Stat Med* 1994; 13: 2465-76.
34. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; 8: 135-160.
35. Stubbing DG, Pengelly LD, Morse JLC, Jones NL. Pulmonary mechanics during exercise in subjects with chronic airflow obstruction. *J Appl Physiol* 1980; 49: 511-515.
36. O'Donnell DE, Webb KA. Exertional breathlessness in patients with chronic airflow limitation: the role of hyperinflation. *Am Rev Respir Dis* 1993; 148: 1351-1357.
37. Aliverti A, Stevenson N, Dellacà RL, Lo Mauro A, Pedotti A, Calverley PMA. Regional chest wall volumes during exercise in chronic obstructive pulmonary disease. *Thorax* 2004; 59: 210-216.
38. Vogiatzis I, Georgiadou O, Golemati S, Aliverti A, Kosmas E, Kastanakis E, Geladas N, Koutsoukou A, Nanas S, Zakynthinos S, Roussos C. Patterns of dynamic hyperinflation during exercise and recovery in patients with severe chronic obstructive pulmonary disease. *Thorax* 2005; 60: 723-729.
39. O'Donnell DE, Revill SM, Webb KA. Dynamic hyperinflation and exercise intolerance in COPD. *Am J Respir Crit Care Med* 2001; 164: 770-777.
40. Man WD, Soliman MGG, Gearing J, Radford SG, Rafferty GF, Gray BJ, Polkey MI, Moxham J. Symptoms and quadriceps fatigability after walking and cycling in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2003; 168: 562-567.

TABLE 1. Subject characteristics (n=463)

Nationality, n (% of group):	
Germany	154 (33)
Canada	145 (31)
USA	102 (22)
Australia	38 (8)
France	24 (5)
Male, n (%)	339 (73)
Female, n (%)	124 (27)
Ex-smoker, n (%)	278 (60)
Current smoker, n (%)	185 (40)
Cigarette exposure, pack-yr	53 ± 28
Age, yr	62 ± 7
Body mass index, kg/m ²	26.4 ± 5.0
FEV ₁ , L	1.23 ± 0.44
FEV ₁ , % predicted	42 ± 13
FEV ₁ /FVC, %	45 ± 11
FRC, % predicted	164 ± 33
DL _{CO} , % predicted	56 ± 19
Peak incremental work rate:	
Watts	78 ± 31
% predicted	50 ± 18
Peak incremental VO ₂ :	
L/min	1.20 ± 0.43
% predicted	60 ± 19

Data are presented as means ± SD unless otherwise specified.

FEV₁, forced expiratory volume in one second; FVC, forced vital capacity; FRC, functional residual capacity; DL_{CO}, diffusing capacity of the lung for carbon monoxide; VO₂, oxygen consumption.

Predicted normal values for FEV₁, FRC, DL_{CO} and peak exercise parameters (work rate and VO₂) are from references 29, 30, 31 and 32, respectively.

TABLE 2. Reproducibility of exercise test variables

	Run-in visit 1, 10 days before randomization mean ± SD	Run-in visit 2, 5 days before randomization mean ± SD	Within- subject CV %	Intraclass correlation (95% CI)
Exercise endurance time, min	7.88 ± 4.78	8.44 ± 5.07 *	23.7	0.84 (0.81,0.87)
Dyspnoea/time slope, Borg/min	0.98 ± 0.64	0.94 ± 0.60	32.4	0.75 (0.71, 0.79)
Dyspnoea/V [̇] O ₂ slope, Borg/(L/min)	6.85 ± 4.17	6.96 ± 4.16	32.3	0.71 (0.67, 0.75)
Dyspnoea/V [̇] _E slope, Borg/(L/min)	0.22 ± 0.13	0.23 ± 0.15	32.7	0.73 (0.68, 0.77)
Pre-exercise rest: IC, L	2.23 ± 0.64	2.23 ± 0.63	9.5	0.89 (0.87, 0.91)
Isotime exercise:				
V [̇] O ₂ , L/min	1.15 ± 0.41	1.14 ± 0.42	8.2	0.95 (0.94, 0.96)
V [̇] _E , L/min	39.7 ± 13.0	39.5 ± 13.4	7.4	0.95 (0.94, 0.96)
F, breaths/min	29.1 ± 6.5	29.1 ± 6.4	7.5	0.88 (0.86, 0.90)
V _T , L	1.39 ± 0.45	1.39 ± 0.46	8.1	0.94 (0.93, 0.95)
IC, L	1.84 ± 0.59	1.85 ± 0.58	10.8	0.88 (0.86, 0.90)
IRV, L	0.45 ± 0.34	0.45 ± 0.31	45.5	0.60 (0.54, 0.66)
Dyspnoea, Borg	5.3 ± 2.3	5.1 ± 2.3 *	19.9	0.79 (0.76, 0.82)
Peak exercise:				
V [̇] O ₂ , L/min	1.19 ± 0.41	1.19 ± 0.42	9.0	0.93 (0.92, 0.94)
V [̇] _E , L/min	41.9 ± 13.3	42.1 ± 13.4	9.1	0.92 (0.90, 0.89)
F, breaths/min	31.1 ± 6.5	31.3 ± 6.5	8.7	0.83 (0.79, 0.85)
V _T , L	1.37 ± 0.43	1.37 ± 0.43	9.4	0.91 (0.89, 0.93)
IC, L	1.81 ± 0.58	1.81 ± 0.59	11.6	0.87 (0.85, 0.89)
IRV, L	0.44 ± 0.32	0.44 ± 0.34	47.3	0.61 (0.55, 0.67)
Dyspnoea, Borg	6.8 ± 2.3	6.8 ± 2.3	14.7	0.81 (0.77, 0.84)

* p<0.05 Significant difference between run-in visits.

CV, coefficient of variation; CI, confidence interval; V[̇]O₂, oxygen uptake; V[̇]_E, minute ventilation; F, breathing frequency; V_T, tidal volume; IC, inspiratory capacity; IRV, inspiratory reserve volume.

TABLE 3. Cycle exercise endurance time (ET) subgroups

	Low-ET n=43	Mid-ET n=348	High-ET n=72	ANOVA p value
Gender, %male	67 *	71	85 †	0.032 chi-square
Age, yr	64.2 ± 6.3 *	61.5 ± 7.5	60.6 ± 6.7	0.032
Height, cm	169 ± 9 *	171 ± 9	174 ± 8 †	0.019
Weight, kg	70.4 ± 13.1 * †	78.1 ± 17.0	79.4 ± 15.5	0.010
Body mass index, kg/m ²	24.6 ± 4.5 †	26.6 ± 5.0	26.3 ± 4.8	0.044
COPD duration, yr	8.1 ± 6.2	8.8 ± 7.0	9.1 ± 7.8	0.756
Smoking pack-yr	58 ± 39	52 ± 27	51 ± 26	0.407
FEV ₁ , L	0.95 ± 0.39 * †	1.23 ± 0.44	1.35 ± 0.41	<0.0005
FEV ₁ , %pr	36 ± 12 * †	43 ± 13	44 ± 11	0.001
FEV ₁ /FVC, %	38 ± 9 * †	45 ± 11	46 ± 11	<0.0005
Peak work rate, W	50 ± 22 * †	78 ± 30	93 ± 28 †	<0.0005
Peak work rate, %pr	37 ± 16 * †	53 ± 18	58 ± 17 †	<0.0005
Peak VO ₂ , L/min	0.83 ± 0.23 * †	1.21 ± 0.43	1.39 ± 0.41 †	<0.0005
Peak VO ₂ , %pr	54 ± 21 * †	71 ± 22	75 ± 22 †	<0.0005
Baseline ET, sec	150 ± 36 * †	432 ± 159	1088 ± 203 †	<0.0005
IC rest, L	1.82 ± 0.50 * †	2.22 ± 0.50	2.55 ± 0.57 †	<0.0005
IC peak, L	1.38 ± 0.41 * †	1.80 ± 0.58	2.13 ± 0.53 †	<0.0005
ΔIC peak-rest, L	-0.45 ± 0.30	-0.41 ± 0.38	-0.41 ± 0.43	0.848
Reasons for stopping exercise, % of group:				0.001
Dyspnoea	59.5	47.7	31.4	
Legs	13.5	14.2	20.0	
Dyspnoea + legs	27.0	37.5	41.4	
Other	0.0	0.6	7.1	

Data are presented as means ± SD unless otherwise specified.

Distribution of reasons for stopping analyzed by Pearson chi-square test.

* p<0.05 low-ET vs high-ET, † p<0.05 vs mid-ET (after Bonferroni adjustment).

FIGURE LEGENDS

Figure 1. The patient-reported main reasons for stopping exercise are shown for the initial incremental cycle exercise test (n=248 in the second validation study only) and for both run-in constant work rate cycle exercise tests (n=435 in both studies).

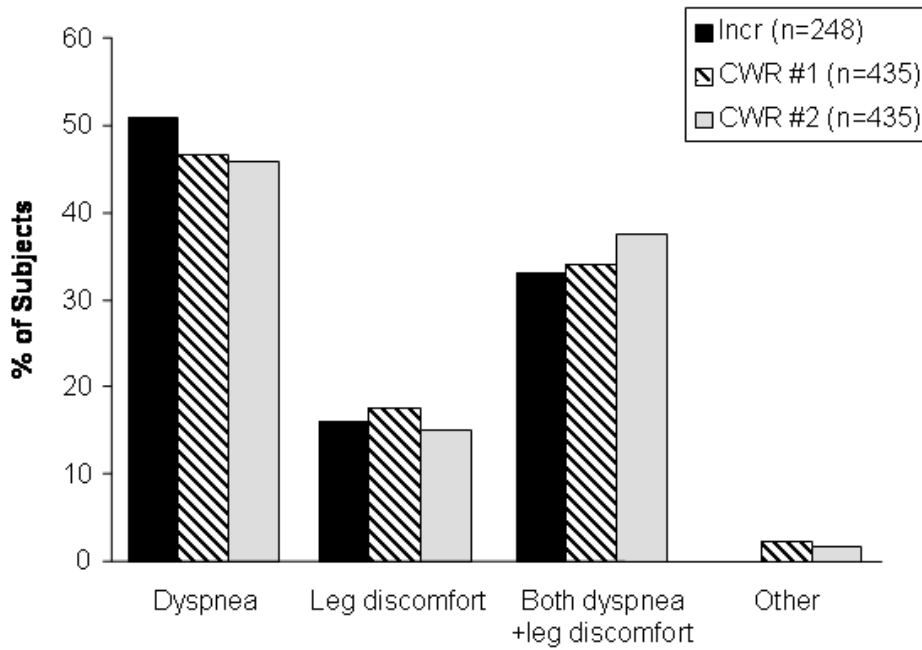


Figure 2. Frequency distributions for resting inspiratory capacity (IC) (*panels a and c*) and the rest-to-peak exercise change in IC (*panels b and d*) are shown for the two pre-randomization run-in visits. Results show a normal distribution and high reproducibility across visits. DH, dynamic pulmonary hyperinflation as indicated by a reduction in IC during exercise.

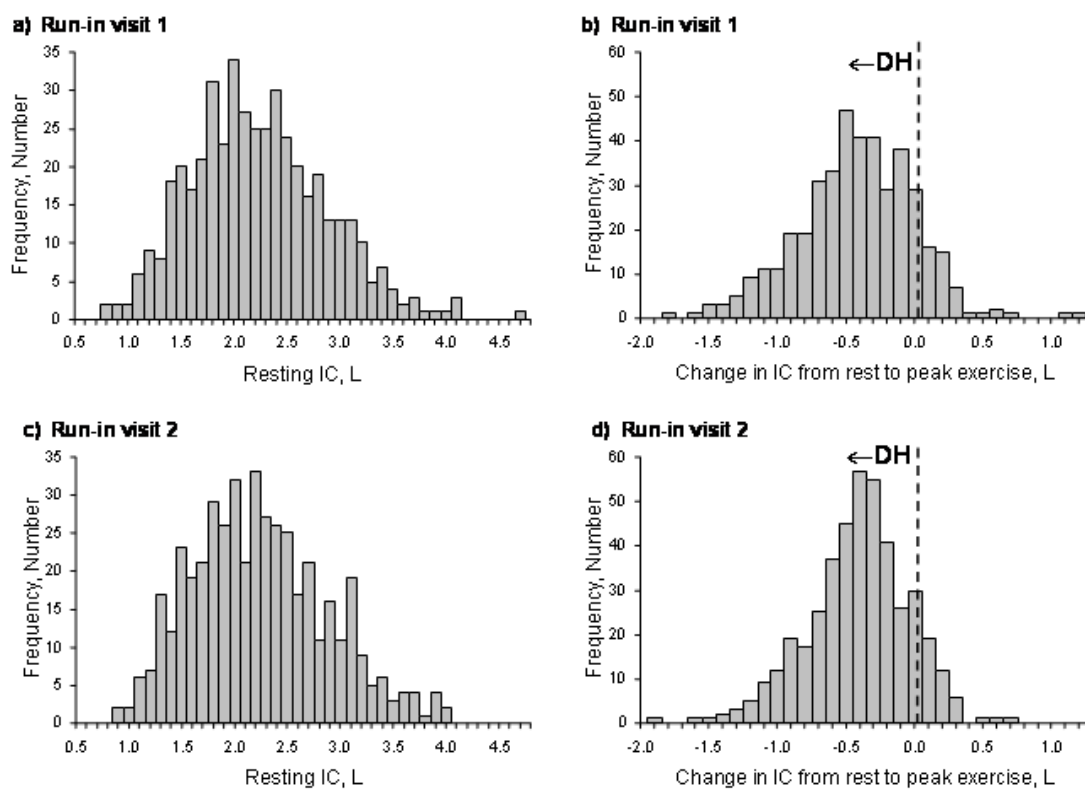
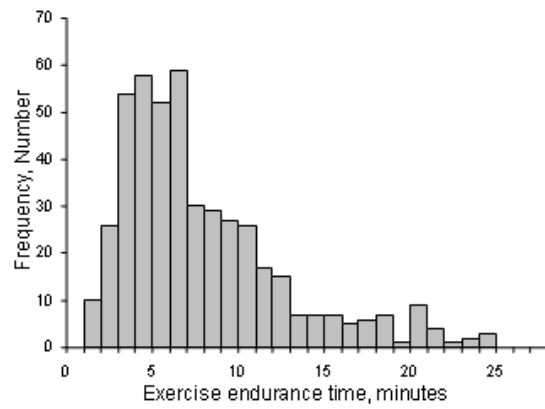
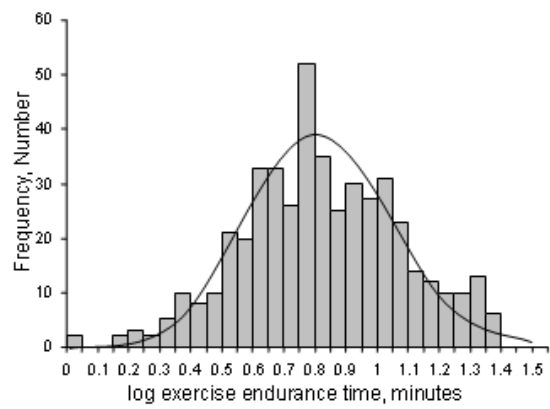


Figure 3. Distribution of constant-work rate (CWR) cycle exercise endurance time (ET) showing significant rightward skew but good reproducibility across run-in visits (*panels a and c*) in 463 subjects with moderate to very severe COPD. A log transformation normalized the distribution of ET (*panels b and d*).

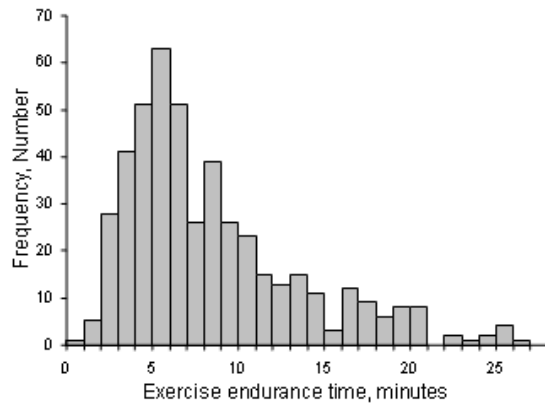
a) Run-in visit 1



b) Run-in visit 1



c) Run-in visit 2



d) Run-in visit 2

