Comparison of laboratory- and field-based exercise tests for COPD: a systematic review

Iain Fotheringham1
Georgina Meakin1
Yogesh Suresh Punekar2
John H Riley2
Sarah M Cockle2
Sally J Singh3

1Value Demonstration Practice, Oxford PharmaGenesis, Oxford, 2GlaxoSmithKline, Uxbridge, 3Department of Respiratory Medicine, University Hospitals of Leicester NHS Trust, Leicester, UK

Abstract: Exercise tests are often used to evaluate the functional status of patients with COPD. However, to the best of our knowledge, a comprehensive systematic comparison of these tests has not been performed. We systematically reviewed studies reporting the repeatability and/or reproducibility of these tests, and studies comparing their sensitivity to therapeutic intervention. A systematic review identified primary manuscripts in English reporting relevant data on the following exercise tests: 6-minute walk test (6MWT) and 12-minute walk test, incremental and endurance shuttle walk tests (ISWT and ESWT, respectively), incremental and endurance cycle ergometer tests, and incremental and endurance treadmill tests. We identified 71 relevant studies. Good repeatability (for the 6MWT and ESWT) and reproducibility (for the 6MWT, 12-minute walk test, ISWT, ESWT, and incremental cycle ergometer test) were reported by most studies assessing these tests, providing patients were familiarized with them beforehand. The 6MWT, ISWT, and particularly the ESWT were reported to be sensitive to therapeutic intervention. Protocol variations (eg, track layout or supplemental oxygen use) affected performance significantly in several studies. This review shows that while the validity of several tests has been established, for others further study is required. Future work will assess the link between these tests, physiological mechanisms, and patient-reported measures.

Keywords: 6MWT, 12MWT, COPD, walk test, repeatability, reproducibility, shuttle walk test, cycle ergometer test

Introduction

COPD is a leading cause of death worldwide, and the prevalence of the disease is projected to increase as the population ages and as exposure to risk factors, such as smoking, continues.1–3 COPD is characterized by symptoms of breathlessness and reduced exercise capacity.4,5 Decrements in exercise capacity can result in reduced ability to perform activities of daily living, and the resultant inactivity and sedentary lifestyle can further exacerbate exercise impairment (the COPD “vicious circle”).6

In clinical practice, spirometry is recommended by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) for the diagnosis of COPD.5 However, the results of spirometry alone poorly predict disability and quality of life in patients with COPD7 and correlate only weakly with dyspnea, exercise capacity, and health status.8–10 Recent guidelines on the diagnosis and treatment of COPD indicate that assessment of disease severity is improved by using additional functional criteria such as exercise capacity.4,5,11 Quantification of the degree of functional impairment is therefore important for the assessment of response to treatment and as an outcome for clinical trials.

There are a number of laboratory- and field-based tests currently used for the assessment of exercise capacity, including the 6- and 12-minute walk tests (6MWT...
and 12MWT, respectively), the incremental and endurance shuttle walk tests (ISWT and E SWT, respectively), the incremental and endurance cycle ergometer tests (ICET and ECET, respectively), and the incremental and endurance treadmill tests (ITT and ETT, respectively). However, there is no consensus about which test is the most appropriate for use in patients with COPD. These tests have different primary outcomes (eg, endurance time, distance, oxygen consumption) that may reflect different physiological parameters. It is therefore difficult to compare results across studies, limiting interpretation of the published literature in this field. Furthermore, the relative merits of different tests have not been established.

The systematic review presented here therefore evaluated evidence of the “repeatability” (defined as consistency of results when multiple tests are conducted on the same day) and the “reproducibility” (consistency of results when tests are conducted on different days) of the tests. The review also assessed the relative properties of the eight commonly used exercise tests and their sensitivity to therapeutic intervention (such as rehabilitative, pharmacological, or surgical procedures). In addition, the effect of protocol variations within each test was assessed across studies. When possible, results were placed in the context of available minimal clinically important difference (MCIDs) values, which have thus far been ascertained for the 6MWT, ISWT, E SWT, and ICET. Investigation of these factors will be useful in guiding test selection in clinical practice and for outcome measures in clinical trials. As these tests are often also used as interventions, evaluation of exercise testing modalities in patients with COPD will also inform the clinical development of optimal exercise rehabilitation strategies.

**Methods**

**Search strategy**

Literature searches were conducted using Ovid® (Ovid Technologies Inc., New York, NY, USA), incorporating Ovid Medline® (US National Library of Medicine, Bethesda, MD, USA), for the period from 1948 to January 22, 2013, Ovid Embase® (Elsevier Inc., Philadelphia, PA, USA) for 1974 to January 22, 2013, and The Cochrane Library (John Wiley and Sons Ltd, Hoboken, NJ, USA) for 1962 to January 22, 2013 (see Tables S1–S3). Search strings were constructed to identify studies reporting primary data on the outcomes of the following exercise tests in patients with COPD: the 6MWT, 12MWT, ISWT, E SWT, ICET, ECET, ITT, and ETT. The full search strings are presented in the “Supplementary materials” section.

**Study selection**

Study selection followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for performing a systematic literature review. Review articles and studies not published in English were excluded using search-engine filters. Studies confounded by comorbidities (such as cancers, diabetes, and non-COPD respiratory-tract diseases) were excluded on review of title/abstract. The remaining studies were screened based on titles and abstracts, and full articles were reviewed when their relevance was unclear from the abstract. Screening was performed by a single author (GM) and records were initially reviewed by title/abstract; a full paper review was subsequently undertaken for publications that could not be excluded by title/abstract. Included records were verified by a second author (IF). A 30% random sample of excluded records was also reviewed by the second author (IF). Disagreements were settled by consultation with the remaining authors.

When reviewing abstracts or full papers, records were excluded if they were reviews, were not in the English language, studied patients with confounding comorbidities (eg, cancers or diabetes), did not use an exercise test as an outcome measure, or examined an intervention other than our interventions of interest (pulmonary rehabilitation, bronchodilator therapy, and lung-volume reduction surgery). Specific inclusion criteria included any definition of COPD (including emphysema- and bronchitis-specific studies); interventions were included only in our comparison of sensitivity and limited to pulmonary rehabilitation, bronchodilation, and lung-volume reduction surgery. Included test outcomes are outlined in the “Data abstraction” section. Following screening, studies were subsequently included for assessment if they reported data on:

- **repeatability** (studies reporting data from two or more performances of the same test[s] on the same day under the same conditions)
- **reproducibility** (studies reporting data from two or more performances of the same test[s] on different days under the same conditions)
- **comparisons of sensitivity** (studies reporting responses of two or more tests to the following therapeutic interventions: pulmonary rehabilitation, bronchodilator therapy, or lung-volume reduction surgery)
- **protocol variations** (studies reporting two or more performances of a test when protocol parameters have been modified).
Data abstraction
Data were primarily abstracted by a single author (GM) and reviewed by all co-authors. A randomly generated selection of 30% of all articles was reviewed by a second author (IF) for quality-control purposes.

The following outcomes of exercise tests were recorded: distance or stages achieved for the 6MWT, 12MWT, and ISWT; duration of exercise for the ESWT, ECET, and ETT; and the highest recorded volume of oxygen consumption (peak VO$_2$) and/or maximum workload (W$_{max}$) for the ICET and ITT. Articles merited inclusion in this review if they reported: outcomes of the specified tests when performed repeatedly under the same conditions, either on the same day (repeatability) or on different days (reproducibility); changes in response before and after therapeutic intervention (comparison of sensitivity); or effects of within-test variations in protocol (protocol variation).

Studies comparing the sensitivity of tests were also assessed for expression by the authors of preference for any specific test. When distances were reported in feet, values were converted to meters using standard conversion criteria stated by the International Bureau of Weights and Measures (0.3048 meters per foot). Within each publication, tests for which results are available are referred to as “test 1”, “test 2”, etc; occasions on which a test has been described by the authors, but results are not reported (such as for practice tests), are referred to as “familiarizations”.

Results
Overview of identified studies
The search methodology used to identify relevant articles is summarized in Figure 1. Of 1,781 unique articles screened, 71 were ultimately deemed eligible for inclusion in this review.

Studies assessing the repeatability and reproducibility of tests
Clinical practice is influenced by factors such as the repeatability and reproducibility of exercise tests in patients with COPD. These factors have been extensively assessed for the 6MWT and 12MWT; data are more limited for the ISWT, ESWT, ITT, and ETT (23 references for the 6/12MWT; 12 for the IWST, ESWT, ITT, and ETT; seven for the ICET and ECET; this made 37 references in total owing to overlap of these categories). Table 1 summarizes the results of studies assessing the repeatability and reproducibility of the 6MWT and 12MWT; Table 2 focuses on the ISWT, ESWT, ETT, and ITT; and Table 3 presents data on the ICET and ECET.

Six studies presented repeatability data for the 6MWT (Table 1). Of these, five reported a significant increase in 6MWT distance from the first to the second test; the remaining study found no differences between results, though patients had been previously familiarized with the tests. The three studies clearly reporting the results of three 6MWTs performed on the same day found that there

---

**Figure 1** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram detailing the identification and inclusion process of the articles. Some studies are included in more than one analysis category; consequently, the aggregate number of studies in the repeatability, reproducibility, comparative, and protocol variation groups adds up to more than 71.
Table 1  Repeatability and reproducibility of the 6- and 12-minute walk tests

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients, n (males, n)</th>
<th>Disease severity (mean ± SD FEV₁ [mean ± SD% predicted] or COPD grading, unless otherwise stated)</th>
<th>Comparison</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-minute walk test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeatability⁴</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andersson et al¹⁶</td>
<td>47 (16)</td>
<td>1.20±0.49 (46.0±17.0)</td>
<td>First versus second test</td>
<td>Significant increase in distance from first to second test (Δ22 m [5.3%]) (ICC =0.94)</td>
</tr>
<tr>
<td>Bansal et al¹⁷</td>
<td>27 (15)</td>
<td>0.8±0.2 (38.1±14.3)</td>
<td>First versus second test</td>
<td>Significant increase in distance from first to second test, for linear (Δ14 m [3.5%]) and circuit (Δ12 m [3.0%]) track layouts</td>
</tr>
<tr>
<td>Eiser et al¹⁸</td>
<td>57 (30)</td>
<td>NR (35.0±12.0)</td>
<td>First versus second versus third test</td>
<td>Significant increase in distance from first to second test (Δ7 m [1.6%]); NSD from second to third test</td>
</tr>
<tr>
<td>Jenkins and Cecins¹⁹</td>
<td>245 (162)</td>
<td>1.06±0.5 (41.0±18.0)</td>
<td>First versus second test</td>
<td>Significant increase in distance from first to second test (Δ37 m [9.5%])</td>
</tr>
<tr>
<td>Stevens et al²⁰</td>
<td>21 (9)</td>
<td>1.07±0.53 (NR)</td>
<td>First versus second versus third test</td>
<td>Hallway: significant increase in distance from first to second test (approximate Δ32.9 m [9.0%]); NSD in distance from second to third test</td>
</tr>
<tr>
<td>Vagaggini et al²¹</td>
<td>18 (15)</td>
<td>NR (48.0±14.0)</td>
<td>Familiarization then first versus second test</td>
<td>NSD in distance from first to second test after familiarization (Δ15 m [3.4%])</td>
</tr>
<tr>
<td>Reproducibility⁸</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bansal et al¹⁷</td>
<td>27 (15)</td>
<td>0.8±0.2 (38.1±14.3)</td>
<td>First versus second test</td>
<td>NSD in distance from first to second day for straight (Δ12 m [2.9%]) and circular (25 m [6.0%]) track layouts</td>
</tr>
<tr>
<td>Behnke et al²⁶</td>
<td>TG: 66 (51)</td>
<td>TG: 1.30±0.49 (41.9±13.9)</td>
<td>Familiarization then first versus mean of second and third tests</td>
<td>TG: significant increase in distance from first to mean of second and third test (Δ209.5 m [72.9%])</td>
</tr>
<tr>
<td>CG: 22 (20)</td>
<td>CG: 1.45±0.51 (46.9±15.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brooks et al²⁷</td>
<td>18 (10)</td>
<td>1.0±0.3 (42.0±8.0)</td>
<td>First versus second test</td>
<td>NSD in distance from first to second test (Δ NR)</td>
</tr>
<tr>
<td>Chatterjee et al²⁸</td>
<td>88 (37)</td>
<td>Median ± IQR FEV₁, L (mean ± SD%) 1.2±0.55 (52.0±19.4)</td>
<td>First versus second versus third test</td>
<td>NSD in distance from first to second test (Δ32 m [9.8%]); NSD in distance from second to third test (Δ4 m [1.1%])</td>
</tr>
<tr>
<td>Eiser et al²⁸</td>
<td>57 (30)</td>
<td>NR (35.0±12.0)</td>
<td>First versus second versus third test</td>
<td>Significant decrease in mean distance from first to second day (Δ-9 m [-2.1%]); significant increase in mean distance from second to third day (Δ8 m [1.9%])</td>
</tr>
<tr>
<td>Kozu et al²⁹</td>
<td>45 (38)</td>
<td>1.10±0.50 (45.0±12.0)</td>
<td>First versus second test</td>
<td>Significant increase in distance from first to second day (Δ13 m [4.3%])</td>
</tr>
<tr>
<td>Poulain et al³⁰</td>
<td>10 (NR)</td>
<td>Moderate COPD</td>
<td>First versus second test</td>
<td>NSD in distance from first to second day after familiarization (Δ4.7 m [0.94%])</td>
</tr>
<tr>
<td>Rejeski et al³¹</td>
<td>209 (117)</td>
<td>1.57; 0.58 (57.1±17.0)</td>
<td>First versus second test</td>
<td>Significant increase in distance from first to second day (Δ122 m [4.6%]); significant correlation between tests (r=0.91)</td>
</tr>
<tr>
<td>Roomi et al³²</td>
<td>15 (6)</td>
<td>NR (49.0±5.0)</td>
<td>First versus second test</td>
<td>NSD in distance from first to second day (Δ-1 m [0.0%])</td>
</tr>
<tr>
<td>Sciurba et al³³</td>
<td>470 (287)</td>
<td>0.75±0.24 (26.3±7.4)</td>
<td>First versus second test</td>
<td>Significant increase in distance from first to second day (Δ120 m; baseline NR) (ICC =0.88)</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Distance (Mean ± SD)</td>
<td>Test Comparison</td>
<td>Notes</td>
</tr>
<tr>
<td>------------------------------</td>
<td>--------------</td>
<td>----------------------</td>
<td>--------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>12-minute walk test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Repeatability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spencer et al&lt;sup&gt;14&lt;/sup&gt;</td>
<td>44 (22)</td>
<td>NR (56.0±19.0)</td>
<td>First versus second test</td>
<td>Significant increase in distance from first to second day (∆27 m [5.9%])</td>
</tr>
<tr>
<td>Troosters et al&lt;sup&gt;35&lt;/sup&gt;</td>
<td>20 (NR)</td>
<td>1.36±0.46 (45.0±14.0)</td>
<td>First versus second test</td>
<td>NSD in distance from first to second day (∆15 m [2.6%])</td>
</tr>
<tr>
<td><strong>Reproducibility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’Reilly et al&lt;sup&gt;32&lt;/sup&gt;</td>
<td>10 (10)</td>
<td>Chronic bronchitis or radiological emphysema</td>
<td>Familiarization then first versus second test</td>
<td>NSD in distance from first to second test after familiarization (mean variation 3.1%; distance NR)</td>
</tr>
<tr>
<td>Arnardóttir et al&lt;sup&gt;36&lt;/sup&gt;</td>
<td>EIH group: 19 (8)</td>
<td>EIH group: 0.9±0.1 (34.6±2.4)</td>
<td>First versus second versus third test</td>
<td>EIH group: NSD in distance from first to second to third day</td>
</tr>
<tr>
<td></td>
<td>Non-EIH group: 38 (19)</td>
<td>Non-EIH group: 1.0±0.1 (38.9±1.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beaumont et al&lt;sup&gt;37&lt;/sup&gt;</td>
<td>12 (10)</td>
<td>1.03±0.27 (NR)</td>
<td>First versus second versus third test</td>
<td>Significant increase in distance from first to second day (∆46 m [6.5%]); NSD in distance from second to third day (∆37 m [4.9%]); using values from Beaumont et al&lt;sup&gt;37&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Berger and Smith&lt;sup&gt;38&lt;/sup&gt;</td>
<td>10 (10)</td>
<td>Moderate/severe COPD</td>
<td>Familiarization then first versus second test</td>
<td>NSD in distance from first to second test after familiarization (∆13 m [2.2%])</td>
</tr>
<tr>
<td>McGavin et al&lt;sup&gt;39&lt;/sup&gt;</td>
<td>35 (35)</td>
<td>Chronic bronchitis</td>
<td>First versus second versus third test</td>
<td>Significant increase in distance in third test compared with first and second; NSD distance on three subsequent tests (distances NR)</td>
</tr>
<tr>
<td>Mungall and Hainsworth&lt;sup&gt;40&lt;/sup&gt;</td>
<td>13 (13)</td>
<td>1.5±0.4 (NR); chronic bronchitis or radiological emphysema</td>
<td>First versus second versus third test (with three further tests NR in detail)</td>
<td></td>
</tr>
<tr>
<td>O’Reilly et al&lt;sup&gt;32&lt;/sup&gt;</td>
<td>10 (10)</td>
<td>Chronic bronchitis or radiological emphysema</td>
<td>Familiarization then first versus second test</td>
<td>NSD in distance from first to second day after familiarization (∆38 m [4.8%])</td>
</tr>
<tr>
<td>Swinburn et al&lt;sup&gt;41&lt;/sup&gt;</td>
<td>17 (6)</td>
<td>0.77±0.30 (NR)</td>
<td>First versus second versus third versus fourth test</td>
<td>Progressive, significant increases in distance from first to fourth day (16%; baseline NR)</td>
</tr>
</tbody>
</table>

**Notes:** Repeatability = similarity of test results when performed on the same day; Reproducibility = similarity of test results when performed on different days.

**Abbreviations:** CG, control group; CI, confidence interval; EIH, exercise-induced hypoxia; FEV<sub>1</sub>, forced expiratory volume (L) in 1 second; ICC, intra-class correlation coefficient; IQR, interquartile range; NR, not reported; NSD, no significant difference; SD, standard deviation; TG, training group; m, meters; min, minutes.
Table 2 Repeatability and reproducibility of incremental shuttle walk test, endurance shuttle walk test, incremental treadmill test, and endurance treadmill test

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients, n (males, n)</th>
<th>Disease severity (mean ± SD FEV₁ [mean ± SD% predicted] or COPD grading, unless otherwise stated)</th>
<th>Comparison</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Incremental shuttle walk test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeatability&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eiser et al&lt;sup&gt;18&lt;/sup&gt;</td>
<td>57 (30)</td>
<td>NR (35.00±12.00)</td>
<td>First versus second</td>
<td>Significant increase in distance from first to second test (∆13 m; 4.7%); NSD in distance from second to third test (∆6 m [2.1%])</td>
</tr>
<tr>
<td>McKeough et al&lt;sup&gt;22&lt;/sup&gt;</td>
<td>53 (34)</td>
<td>NR (55.00±19.00)</td>
<td>First versus second</td>
<td>Significant increase in distance from first to second test (∆20 m [6.3%])&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Vagaggini et al&lt;sup&gt;21&lt;/sup&gt;</td>
<td>18 (15)</td>
<td>NR (48.00±14.00)</td>
<td>Familiarization then first versus second test</td>
<td>Significant increase in distance from first to second test, after familiarization (∆40 m [14.6%])&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Reproducibility&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arnardóttir et al&lt;sup&gt;44&lt;/sup&gt;</td>
<td>93 (26)</td>
<td>Moderate/severe COPD</td>
<td>First versus second</td>
<td>NSD in distance from first to second day (∆9 m [2.9%])</td>
</tr>
<tr>
<td>Campo et al&lt;sup&gt;42&lt;/sup&gt;</td>
<td>30 (18)</td>
<td>Mild/moderate/severe/very severe COPD</td>
<td>Familiarization then first versus second test</td>
<td>NSD in distance from first to second day (ICC = 0.88) (distances NR)</td>
</tr>
<tr>
<td>Eiser et al&lt;sup&gt;18&lt;/sup&gt;</td>
<td>57 (30)</td>
<td>NR (35.00±12.00)</td>
<td>First versus second versus third test</td>
<td>Significant increase in distance from first to second day (∆13 m; 4.6%); NSD in distance from first to second day (∆1 m [0.3%])</td>
</tr>
<tr>
<td>Perrault et al&lt;sup&gt;43&lt;/sup&gt;</td>
<td>43 (36)</td>
<td>1.40±0.50</td>
<td>Familiarization then first versus second test</td>
<td>NSD in VO&lt;sub&gt;2&lt;/sub&gt; HR, V&lt;sub&gt;e&lt;/sub&gt;, V&lt;sub&gt;T&lt;/sub&gt; from first to second test at each of the four cadences (ICC &gt;0.93); distance NR</td>
</tr>
<tr>
<td>Singh et al&lt;sup&gt;45&lt;/sup&gt;</td>
<td>35 (25)</td>
<td>Group A: 0.50 (0.36–1.45)</td>
<td>First versus second versus third test</td>
<td>Downgraded protocol: NSD in distance on first, second, and third day</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group B: 1.10 (0.60–2.10)</td>
<td></td>
<td>Modified protocol: significant increase in distance from first to second day (∆31 m [9.0%]); NSD in distance from second to third day</td>
</tr>
<tr>
<td><strong>Endurance shuttle walk test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeatability&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revill et al&lt;sup&gt;44&lt;/sup&gt;</td>
<td>44 (33)</td>
<td>0.94±0.40</td>
<td>First versus second</td>
<td>NSD in duration from first to second test (∆12 s [6.2%])</td>
</tr>
<tr>
<td>McKeough et al&lt;sup&gt;33&lt;/sup&gt;</td>
<td>53 (34)</td>
<td>NR (55.00±19.00)</td>
<td>First versus second</td>
<td>NSD in duration from first to second test (∆2 s [5.8%])</td>
</tr>
<tr>
<td></td>
<td>31 (NR)</td>
<td>NR (51.00±16.00)</td>
<td>Third versus fourth test</td>
<td>NSD in duration from third to fourth test (∆44 s [8.7%])</td>
</tr>
<tr>
<td><strong>Reproducibility&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revill et al&lt;sup&gt;47&lt;/sup&gt;</td>
<td>44 (22)</td>
<td>Group A: 1.01±0.36 (35.00±4.00)</td>
<td>First versus second versus third test</td>
<td>Significant increase in duration from first to second (∆59 s; 23.5%); NSD in duration from second to third day (∆15 s [4.8%])</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group B: 0.79±0.21 (34.00±4.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group C: 0.80±0.18 (35.00±8.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revill et al&lt;sup&gt;46&lt;/sup&gt;</td>
<td>23 (13)</td>
<td>0.81±0.27</td>
<td>First versus second</td>
<td>NSD in duration from first to second test after familiarization (tests performed with supplemental oxygen)</td>
</tr>
<tr>
<td><strong>Endurance treadmill test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reproducibility&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooper et al&lt;sup&gt;42&lt;/sup&gt;</td>
<td>470 (NR)</td>
<td>Moderate/severe/very severe COPD</td>
<td>First versus second</td>
<td>NSD in duration from first to second test (∆24 s [7.6%]) (ICC = 0.85)</td>
</tr>
<tr>
<td><strong>Incremental treadmill test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reproducibility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathur et al&lt;sup&gt;10&lt;/sup&gt;</td>
<td>8 (6)</td>
<td>0.69±0.16</td>
<td>First versus second</td>
<td>Numerical increase in peak VO&lt;sub&gt;2&lt;/sub&gt; from first to second test (∆1.0 mL/min/kg [8.3%]; statistical test NR)</td>
</tr>
</tbody>
</table>

**Notes:**<sup>a</sup>Repeatability = similarity of test results when performed on the same day;<sup>b</sup>reproducibility = similarity of test results when performed on different days; P<0.001; *P<0.05; **P<0.01.

**Abbreviations:** FEV₁, forced expiratory volume (L) in 1 second; HR, heart rate; ICC, intra-class correlation coefficient; NR, not reported; NSD, no significant difference; SD, standard deviation; V<sub>e</sub>, expired volume; VO<sub>2</sub>, oxygen consumption; V<sub>T</sub>, tidal volume; m, meters; min, minutes; s, time in seconds.
was no significant difference between the second and third tests.\textsuperscript{18,20,21} The only study that presented intra-class correlation coefficients (ICCs) for repeated 6MWTs on the same day reported excellent repeatability (ICC = 0.94), but also observed that the second test was significantly higher.\textsuperscript{16} For the 12MWT, the distance achieved was reported to be repeatable in the only study in which patients were retested on the same day.\textsuperscript{22} Equivocal results were reported for the repeatability of the ISWT by three studies;\textsuperscript{18,21,23} one found that the distance was repeatable after familiarization,\textsuperscript{18} but the other two reported poor repeatability even after familiarization (Table 2).\textsuperscript{21,23} For the ESWT, exercise duration was reported to be repeatable in the two studies in which patients were retested on the same day.\textsuperscript{23,24} One study reported that peak VO\textsubscript{2} and W\textsubscript{max} were repeatable for the ICET (Table 3).\textsuperscript{25} No repeatability data were found for the ECET, ITT, or ETT.

Reproducibility of the 6MWT was assessed in 12 studies.\textsuperscript{17,18,26–35} Six of which reported that distances achieved in the 6MWT demonstrated good reproducibility between the first and second tests (Table 1).\textsuperscript{17,27,28,32,35} Of two studies presenting reproducibility results from three 6MWTs,\textsuperscript{18,28} only one reported reproducibility between the second and third tests.\textsuperscript{28} Two further studies reported results of tests after familiarization;\textsuperscript{26,30} only one found 6MWT results to be reproducible.\textsuperscript{30} The only study presenting ICC data between the first and second 6MWT performance showed high reproducibility (ICC = 0.88), but also that there was a significant increase in distance in the second 6MWT.\textsuperscript{33} Reproducibility of the 12MWT was assessed in seven studies.\textsuperscript{22,36–41} Five of these presented the results of three or more tests, and reported that the 12MWT distance increased significantly from the first to the second test;\textsuperscript{16,37,39,41} one of these studies reported that in a subset of patients who readily experienced

### Table 3: Repeatability and reproducibility of cycling tests

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients, n (males, n)</th>
<th>Disease severity (mean ± SD FEV\textsubscript{1}, [mean ± SD% predicted] or COPD grading, unless otherwise stated)</th>
<th>Comparison</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Incremental cycle ergometer test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repetitability\textsuperscript{a}</td>
<td>Brown et al\textsuperscript{25}</td>
<td>11 (NR)</td>
<td>1.51±0.59 (NR)</td>
<td>First versus second test</td>
</tr>
<tr>
<td>Reproducibility\textsuperscript{a}</td>
<td>Brown et al\textsuperscript{25}</td>
<td>11 (NR)</td>
<td>1.51±0.59 (NR)</td>
<td>First versus third test</td>
</tr>
<tr>
<td></td>
<td>Covey et al\textsuperscript{30}</td>
<td>56 (40)</td>
<td>NR (49.00±16.00)</td>
<td>First versus second test</td>
</tr>
<tr>
<td></td>
<td>Cox et al\textsuperscript{49}</td>
<td>11 (8)</td>
<td>Individually listed in paper</td>
<td>First versus second test</td>
</tr>
<tr>
<td></td>
<td>Mathur et al\textsuperscript{50}</td>
<td>8 (6)</td>
<td>0.69±0.16 (NR)</td>
<td>First versus second test</td>
</tr>
<tr>
<td></td>
<td>Poulain et al\textsuperscript{50}</td>
<td>10 (NR)</td>
<td>Moderate COPD</td>
<td>Familiarization then first versus second test</td>
</tr>
<tr>
<td></td>
<td>Swinburn et al\textsuperscript{51}</td>
<td>17 (6)</td>
<td>0.77±0.30 (NR)</td>
<td>First versus second versus third versus fourth test</td>
</tr>
<tr>
<td><strong>Endurance cycle ergometer test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reproducibility\textsuperscript{a}</td>
<td>van’t Hul et al\textsuperscript{31}</td>
<td>60 (46)</td>
<td>Moderate/severe COPD</td>
<td>First versus second test</td>
</tr>
</tbody>
</table>

Notes: \textsuperscript{a}Repeatability = similarity of test results when performed on the same day; \textsuperscript{b}reproducibility = similarity of test results when performed on different days.

Abbreviations: FEV\textsubscript{1}, forced expiratory volume (L) in 1 second; ICC, intra-class correlation coefficient; NR, not reported; NSD, no significant difference; SD, standard deviation; VO\textsubscript{2}, oxygen consumption; W\textsubscript{max}, maximum workload; W, watts; min, minutes; s, time in seconds.
exercise-induced hypoxia, the 12MWT distance did not significantly change from the first to the second to the third test.36 Two additional studies, in which there had been prior familiarization with the test, reported no significant change in 12MWT distance between subsequent first and second tests.22,38 For the ISWT, five studies presented data assessing reproducibility: from the first to the second test with prior familiarization,42,43 from the first to the second test without prior familiarization,44,45 and from the second to the third test without prior familiarization (Table 2).46 Two further studies evaluating the ESWT reported reproducibility (either from the first to the second test with prior familiarization,47 or from the second to the third test without familiarization).48 Good reproducibility of the ICET from the first to the second test was reported by four studies both with30 and without prior familiarization (Table 3).25,48,49 Two further studies reported an increase in ICET duration from the first to the second test with no familiarization,41,50 with one of these studies reporting progressive increases in ICET duration from the first through to the fourth test.41 In the only study reporting data for the ECET, duration was found to have excellent reproducibility (ICC = 0.85).51 Reproducibility was also found to be excellent in the only study reporting such data for the ETT (no significant increases from the first to the second test, ICC = 0.85),52 but less so for the ITT (increased peak VO\textsubscript{2} from the first to the second test, statistical test not reported).50

Several studies were identified that compared the repeatability and/or reproducibility of two or more exercise tests. One study observed that repeatability for the 6MWT and the ISWT was comparable, but that the ISWT was more reproducible.18 However, another study showed that the ISWT was more repeatable than the 6MWT.23 One study reported that both the 6MWT and the ICET were reproducible.30 The ESWT was reported to be more repeatable than the ISWT in one study, when measured in two sessions before and two sessions after pulmonary rehabilitation.23 In another study, both 12MWT distance and ICET performance were found to increase significantly and progressively over four tests; the ICET was found to have no obvious advantages over the 12MWT when assessing exercise performance.41 The final study to report reproducibility of more than one test reported that peak VO\textsubscript{2} increased from test 1 to test 2 in both the ICET and the ITT; however, the authors did not report the statistical tests used.50 Three studies18,28,30 were found that compared the reproducibility of two or more exercise tests. Of these, two reported that the 6MWT was found to have similar reproducibility to the ISWT18 and the ICET.30

### Studies comparing responses to interventions among exercise tests

In total, 23 studies were identified that compared responses of two or more exercise tests after one of the following interventions: pulmonary rehabilitation (16 studies),23,47,53–66 administration of bronchodilators (six studies),14,18,67–70 and lung-volume reduction surgery (one study)71 (Table 4). Of the 16 studies that assessed pulmonary rehabilitation in patients with COPD, the most commonly assessed test was the 6MWT, which was reported by eleven studies.53–55,57–59,61,63–66

Two studies compared the response to the 6MWT and ITT after pulmonary rehabilitation; both reported significant increases in 6MWT distance and ITT performance (peak VO\textsubscript{2} and work-level completed).57 However, the latter study did not observe a significant response in peak VO\textsubscript{2} during the ITT after pulmonary rehabilitation.57 One further study assessed the 6MWT and ITT during nutritional supplementation and placebo, and reported that the 6MWT distance was sensitive to pulmonary rehabilitation (>MCID); but these authors did not present peak VO\textsubscript{2} or W\textsubscript{max} data for the ITT.54 Another study assessed the 6MWT, ETT, and ITT, and found that both the ETT and the ITT were sensitive to pulmonary rehabilitation, whereas the 6MWT was not (again, the authors did not report peak VO\textsubscript{2} or W\textsubscript{max} data for the ITT).58 Several further studies reported equivocal findings when comparing the 6MWT with the ICET55,63–65 after pulmonary rehabilitation. All three studies comparing the ECET with the 6MWT found the ECET to be more responsive to pulmonary rehabilitation.63,65,66 One study assessing responses to pulmonary rehabilitation reported similar sensitivities for the 6MWT and the ISWT, with both giving responses that exceeded the MCID.61 All four studies assessing the sensitivity of the ISWT and the ESWT to pulmonary rehabilitation reported a significant improvement in performance for both tests;23,47,60,62 however, in all four studies the response of the ESWT was greater and in two the ISWT response did not reach the MCID.23,47 An additional study suggested that although both tests showed a significant response that was above the MCID, the ESWT was more responsive to pulmonary rehabilitation than the 6MWT.59 Equivocal sensitivity was observed in response to pulmonary rehabilitation when using the 12MWT and the ICET.56

Of the six studies comparing the responses of two or more exercise tests to bronchodilator therapy,14,18,67–70 one reported the 6MWT to be more responsive to pharmacological intervention than the 12MWT70 and one reported the 6MWT to be more responsive to pharmacological intervention...
Table 4 Between-test comparisons of sensitivity to interventions

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Result</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>6-minute walk test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borghi-Silva et al</td>
<td>8</td>
<td>(Car) 8 (Car)</td>
<td>Significant increase in distance: ∆87 m (19.8%) (&gt;MCID) Training with concomitant carnitine versus placebo; ITT conducted but peak VO$<em>2$ and $W</em>{max}$ NR</td>
</tr>
<tr>
<td>Borghi-Silva et al</td>
<td>8</td>
<td>(Pla) 8 (Pla)</td>
<td>Significant increase in distance: ∆34 m (7.3%) (&gt;MCID) Training significantly improved peak VO$_2$ compared with control group</td>
</tr>
<tr>
<td>Carriere-Kohlman et al</td>
<td>20</td>
<td></td>
<td>Significant increase in distance: ∆105 m (27.8%) (&gt;MCID) Improvements were similar in both tests whether patients were given a monitored or coached pulmonary rehabilitation program</td>
</tr>
<tr>
<td>Cooper</td>
<td>7</td>
<td>(GPT) 7 (GPT)</td>
<td>Significant increase in distance: ∆32 m (6.9%) (&gt;MCID) Patients in IMT group showed significant improvement in pulmonary function test; increase in GPT group was not significant</td>
</tr>
<tr>
<td>Cortopassi et al</td>
<td>71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eaton et al</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>12-minute walk test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arnardottir et al</td>
<td>28</td>
<td>(IT) 28 (IT)</td>
<td>Significant increase in distance: ∆75 m (9.0%) Similar improvements in 12MWT distance and $W_{max}$ whether training was IT or CT; peak VO$_2$ was higher in CT, but VO$_2$ values at isotime were lower in IT</td>
</tr>
<tr>
<td><strong>Incremental shuttle walk test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greening et al</td>
<td>601</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McKeough et al</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ngage et al</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O'Farrell et al</td>
<td>85</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4 (Continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Result</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revill et al**</td>
<td>44</td>
<td>Significant increase in distance: ∆36 m (20.2%) (non-MCID)</td>
<td>Clinically relevant response after PR in ESWT but not ISWT; greater sensitivity in ESWT than in ISWT (effect size 2.90 versus 0.41)</td>
</tr>
<tr>
<td>Endurance shuttle walk test</td>
<td>20</td>
<td>Significant increase in distance: ∆302 m (92%) (&gt;MCID); significant increase in duration: ∆270 s (88%) (&gt;MCID)</td>
<td>Standardized mean differences after PR: 6MWT, 0.32 versus ESWT, 0.54 Values taken from text (do not match Eaton et al)**</td>
</tr>
<tr>
<td>Eaton et al**</td>
<td>60</td>
<td>Significant increase in duration: ∆408 s (205.4%) (&gt;MCID)</td>
<td>The authors note: “The ESWT has potential advantages in that it may be more responsive than the 6MWT” Values taken from text (do not match Eaton et al)**</td>
</tr>
<tr>
<td>Greening et al**</td>
<td>634</td>
<td>Significant increase in distance: ∆271.0 m (106.7%) (non-MCID)</td>
<td>Clinically relevant response after PR for both tests; greater sensitivity in the ESWT than in the ISWT (not assessed statistically)</td>
</tr>
<tr>
<td>McKeough et al**</td>
<td>31</td>
<td>Significant increase in duration: ∆182 s (58%)</td>
<td>Results are calculated from the better of two pre-PR test results versus the better of two post-PR test results Values taken from McKeough et al**</td>
</tr>
<tr>
<td>O’Farrell et al**</td>
<td>85</td>
<td>Significant increase in distance: ∆271.0 m (106.7%) (non-MCID)</td>
<td>ESWT more responsive than ISWT to determine improvements in exercise capacity following PR</td>
</tr>
<tr>
<td>Revill et al**</td>
<td>44</td>
<td>Significant increase in distance: ∆334 m (140%) (&gt;MCID); significant increase in duration: ∆404 s (150%) (&gt;MCID)</td>
<td>Clinically relevant response after PR in ESWT but not ISWT; greater sensitivity in ESWT than in ISWT (effect size 2.90 versus 0.41)</td>
</tr>
<tr>
<td>Incremental cycle ergometry test</td>
<td></td>
<td>Significant increase in peak VO$<em>2$: ∆53 mL/min (5.4%); significant increase in W$</em>{max}$: ∆11 W (18.0%) (&gt;MCID); Significant increase in peak VO$<em>2$: ∆146 mL/min (15.0%); significant increase in W$</em>{max}$: ∆11 W (17.2%) (&gt;MCID)</td>
<td>Similar improvements in 12MWT distance and W$_{max}$ whether training was IT or CT; peak VO$_2$ was higher in CT, but VO$_2$ values at isotime were lower in IT</td>
</tr>
<tr>
<td>Arnardóttir et al**</td>
<td>28 (IT)</td>
<td>Significant increase in peak VO$<em>2$: ∆53 mL/min (5.4%); significant increase in W$</em>{max}$: ∆11 W (18.0%) (&gt;MCID)</td>
<td>Patients in IMT group showed significant improvement in pulmonary function test, increase in GPT group was not significant</td>
</tr>
<tr>
<td></td>
<td>32 (CT)</td>
<td>Significant increase in peak VO$<em>2$: ∆146 mL/min (15.0%); significant increase in W$</em>{max}$: ∆11 W (17.2%) (&gt;MCID)</td>
<td></td>
</tr>
<tr>
<td>Cooper**</td>
<td>7 (GPT)</td>
<td>NSD in peak VO$<em>2$ or W$</em>{max}$</td>
<td>6MWT did not correlate with ICET or ECET responses to PR; moderate correlation between ICET W$<em>{max}$ and ECET duration (r=0.406; P=0.013) Patients who had never previously received PR responded more favorably than those who had ∆6MWT distance: 31.0 versus 18.5 m; ICET W$</em>{max}$: 4.3 versus 2.4 W) 6MWT, ICET, and ECET had greater validity in assessing responses to PR than cardiorespiratory measures</td>
</tr>
<tr>
<td>Ong et al**</td>
<td>9 (IMT)</td>
<td>NSD in peak VO$<em>2$ or W$</em>{max}$</td>
<td>6MWT did not correlate with ICET or ECET responses to PR; moderate correlation between ICET W$_{max}$ and ECET duration (r=0.406; P=0.013) 6MWT, ICET, and ECET had greater validity in assessing responses to PR than cardiorespiratory measures Clinically relevant response after PR for both tests; greater sensitivity in the ECET than in the 6MWT (not assessed statistically)</td>
</tr>
<tr>
<td>Ries et al**</td>
<td>37</td>
<td>Significant increase in peak VO$<em>2$: ∆172 mL/min (20.0%); significant increase in W$</em>{max}$: ∆6 W (12.0%) (&gt;MCID)</td>
<td>Training with concomitant carnitine versus placebo; ITT conducted but peak VO$<em>2$ and W$</em>{max}$ NR</td>
</tr>
<tr>
<td>Van Helvoort et al**</td>
<td>1,218</td>
<td>Significant increase in W$_{max}$: ∆3.1 W (8.6%) (non-MCID)</td>
<td></td>
</tr>
<tr>
<td>Van Helvoort et al*</td>
<td>18</td>
<td>NSD in W$_{max}$ and peak VO$_2$</td>
<td></td>
</tr>
<tr>
<td>Endurance cycle ergometry test</td>
<td></td>
<td>Significant increase in duration: ∆322 s (73.5%)</td>
<td></td>
</tr>
<tr>
<td>Ong et al**</td>
<td>37</td>
<td>Significant increase in duration: ∆6.5 min (166%)</td>
<td></td>
</tr>
<tr>
<td>Van Helvoort et al**</td>
<td>18</td>
<td>Significant increase in duration: ∆241 s (84.6%)</td>
<td></td>
</tr>
<tr>
<td>Van Ranst et al**</td>
<td>389</td>
<td>Significant increase in duration: ∆322 s (73.5%)</td>
<td></td>
</tr>
<tr>
<td>Incremental treadmill test</td>
<td></td>
<td>Peak VO$<em>2$ and W$</em>{max}$ NR</td>
<td></td>
</tr>
<tr>
<td>Borghi-Silva et al**</td>
<td>8 (Car)</td>
<td>Peak VO$<em>2$ and W$</em>{max}$ NR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 (Pla)</td>
<td>Peak VO$<em>2$ and W$</em>{max}$ NR</td>
<td></td>
</tr>
</tbody>
</table>
Significant increase in peak VO\textsubscript{2}: ∆2 mL/min/kg (14.3%)

Training significantly improved peak VO\textsubscript{2} compared with control group

Improvements were similar in both tests whether patients were given a monitored or coached pulmonary rehabilitation program

ITT and ETT both responded significantly to PR, but 6MWT did not

ITT and ETT both responded significantly to PR, but 6MWT did not

"6-MWT seems to be a more appropriate instrument than 12-MWT for assessing the exercise response to a bronchodilator"

Sensitivity index: 6MWT, 0.84; ISWT, 0.76; distance calculated by number of shuttles performed (1 shuttle = 10 m)

"There is no important difference in either the reproducibility or sensitivity of self-paced or externally paced walking tests"

Percentage improvement: 6MWT, 1%; ICET, 3%; ECET, 19%

"Among the frequently used post-PR exercise tests, the most responsive index, as measured by the percentage change from baseline, is the endurance time. The correlation between the post-PR changes in these exercise indices is poor"

The endurance shuttle walk is more responsive to the acute effects of bronchodilation than the 6MWT

"6-MWT seems to be a more appropriate instrument than 12-MWT for assessing the exercise response to a bronchodilator"

Sensitivity index: 6MWT, 0.84; ISWT, 0.76; distance calculated by number of shuttles performed (1 shuttle = 10 m)

"There is no important difference in either the reproducibility or sensitivity of self-paced or externally paced walking tests"

Percentage improvement: 6MWT, 1%; ICET, 3%; ECET, 19%

"The endurance shuttle walk is more responsive to the acute effects of bronchodilation than the 6MWT"

"Standardized response mean larger for walking than cycling (0.93 and 0.20, respectively); % change and baseline NR

Percentage improvement: 6MWT, 1%; ICET, 3%; ECET, 19%

"Among the frequently used post-PR exercise tests, the most responsive index, as measured by the percentage change from baseline, is the endurance time. The correlation between the post-PR changes in these exercise indices is poor"
Table 4 (Continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Result</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Endurance cycle ergometry test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oga et al[6]</td>
<td>38</td>
<td>Significant increase in duration: 34 s (18.0%)</td>
<td>Percentage improvement: 6MWT, 1%; ICET, 3%; ECET, 19%</td>
</tr>
<tr>
<td>Pepin et al[69]</td>
<td>17</td>
<td>NSD in duration</td>
<td>“Among the frequently used post-PR exercise tests, the most responsive index, as measured by the percentage change from baseline, is the endurance time. The correlation between the post-PR changes in these exercise indices is poor”</td>
</tr>
<tr>
<td>Zhang et al[70]</td>
<td>20</td>
<td>Significant increase in duration: ∆157 s (47.1%)</td>
<td>Values taken from Zhang et al[70] (percentage estimates in text are different); greater percentage change in ECET than in TT</td>
</tr>
<tr>
<td><strong>Endurance treadmill test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang et al[70]</td>
<td>20</td>
<td>Significant increase in duration: ∆110 s (30.4%); W&lt;sub&gt;max&lt;/sub&gt; NR</td>
<td>Values taken from Table 4 (percentage estimates in text are different); greater percentage change in ECET than in TT</td>
</tr>
<tr>
<td><strong>Lung-volume reduction surgery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-minute walk test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lederer et al[71]</td>
<td>23</td>
<td>Significant increase in distance: ∆38 m (9.1%) (MCID)</td>
<td>Measurements taken 1 year after surgery; clinically relevant change in 6MWT; greater percentage change in ICET than in 6MWT</td>
</tr>
<tr>
<td>Incremental cycle ergometry test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lederer et al[71]</td>
<td>23</td>
<td>NSD in peak VO&lt;sub&gt;2&lt;/sub&gt;; significant increase in W&lt;sub&gt;max&lt;/sub&gt;; 6 W (15.4%)</td>
<td>Measurements taken 1 year after surgery; clinically relevant change in 6MWT; greater percentage change in ICET than in 6MWT</td>
</tr>
</tbody>
</table>

**Abbreviations:** 6MWT, 6-minute walk test; 12MWT, 12-minute walk test; Car, patients receiving carnitine supplements; CT, continuous training; ECET, endurance cycle ergometer test; ESWT, endurance shuttle walk test; ETT, endurance treadmill test; GPT, general physical training; ICET, incremental cycle ergometer test; IMT, inspiratory muscle training; ISWt, incremental shuttle walk test; IT, interval training; ITT, incremental treadmill test; MCID, minimal clinically important difference; NR, not reported; NSD, no significant difference; pla, patients receiving placebo; VO<sub>2</sub>, oxygen consumption; PR, pulmonary rehabilitation; TT, treadmill test; V<sub>a</sub>, ventilation; W<sub>max</sub>, maximum workload; W, watt.
than the ISWT. The standardized (percentage) increase in response to the ECET was higher than that of either the 6MWT or the ICET in one study, and higher than that of the ETT in another. One study reported the response of the ESWT to bronchodilation to be greater than that of the 6MWT and one study reported the response of the ESWT to bronchodilation to be greater than that of the ECET.

Finally, one study assessed exercise test performance 1 year after lung-volume reduction surgery, and reported a 9.1% increase in 6MWT distance and a 15.4% increase in ICET \(W_{\text{max}}\), but noted that ICET peak \(V_O_2\) did not increase significantly.

**Minimal clinically important differences in responses to interventions**

MCIDs have been thus far ascertained for the 6MWT (26 meters), the ISWT (48 meters), the ESWT (45 seconds–85 seconds or 60 meters–115 meters [MCID calculated after bronchodilatory intervention]), and the ICET (4 watts). Of the eleven studies assessing the 6MWT in response to pulmonary rehabilitation, nine reported an increase in excess of the recognized MCID, with another reporting a significant increase in duration of less than the MCID (the remaining study reported no significant change in 6MWT distance after pulmonary rehabilitation); these increases ranged in magnitude from 4.8% to 36.3%. Five studies reported a significant response of the ISWT to pulmonary rehabilitation ranging from 15.0% to 52.9% with three finding that the distance observed reached the MCID. All five studies assessing the sensitivity of the ESWT to pulmonary rehabilitation reported that distance and/or duration increased in excess of the MCID (increases ranged from 88.0% to 205.4% when expressed as time [seconds]); and from 92.0% to 140.0% when expressed as distance [meters]). Of five studies assessing the ICET before and after pulmonary rehabilitation, two reported \(W_{\text{max}}\) responses in excess of the MCID, and another reported significant changes in \(W_{\text{max}}\) that did not reach the MCID (the increased ICET performance observed across these studies ranged from 5.4% to 20.0% for peak \(V_O_2\) and 8.6% to 18.0% for \(W_{\text{max}}\)). The two remaining studies did not observe a significant change in ICET peak \(V_O_2\) or \(W_{\text{max}}\) after pulmonary rehabilitation.

All three studies reporting the response of the ECET to pulmonary rehabilitation observed significant increases in duration (ranging from 73.5% to 166.0%). Four studies assessed the ITT before and after pulmonary rehabilitation; two did not present data for either peak \(V_O_2\) or \(W_{\text{max}}\).

Of the remaining two studies, one reported a significant increase in peak \(V_O_2\) of 14.3%, while the other found no significant increase in peak \(V_O_2\) after pulmonary rehabilitation. The only study assessing the ETT before and after pulmonary rehabilitation reported significant increases in duration (27.4%) and distance (22.1%).

Of the four studies assessing the 6MWT before and after bronchodilator therapy, two reported improvements in distance that exceeded the MCID (one found an increase in 6MWT distance of 8.7%; the other reported a higher absolute distance increase [53.6 meters], but did not present baseline values; therefore the percentage increase cannot be calculated). The third study reported a small (1.2%) but significant increase in 6MWT distance, while the remaining study found no significant difference in 6MWT distance.

One study found a significant increase in ISWT performance after bronchodilators (30 meters) that did not reach the MCID. Two studies reported significant increases in ESWT performance above the MCID, of 144 meters and 164 seconds, respectively. One study reported a small but significant improvement in ICET \(W_{\text{max}}\) of 3.4% after bronchodilator therapy, but noted that peak \(V_O_2\) did not increase significantly. Three studies assessed ECET performance after bronchodilators; two found significant increases in duration of 18.0% and 47.1%, with the remaining study reporting no significant improvement in ECET duration.

In the only study assessing lung-volume reduction surgery, a 6MWT improvement in excess of the MCID was seen; however, ICET improvements (whether peak \(V_O_2\) or \(W_{\text{max}}\)) did not reach MCID.

**Studies assessing within-test variations in protocol**

Eighteen studies were identified that assessed minor variations in protocol within a specific exercise test (Table 5). Variations (such as track environment or layout and the type of encouragement provided by the investigators to the patient) affected test outcomes and, consequently, their repeatability and reproducibility.

**Discussion**

A number of laboratory- and field-based exercise tests are used to assess the degree of functional impairment in patients with COPD. However, the choice of which test to use in clinical trials historically seems to have been made.
on a practical basis (tests such as the 6MWT and 12MWT require little time, organization, or equipment), or without necessarily taking into account how representative the exercise modality used is to activities of daily living for these patients (eg, cycling tests) or the likely impact of the intervention on the outcome of the test. Equally there are few data describing the relative merits of these tests employed simultaneously to evaluate interventions such as rehabilitation and bronchodilator therapy. Results from this systematic review indicate that there is an extensive body of published literature regarding the performance of the eight exercise tests that are widely used.

Repeatability data were found for only the 6MWT, 12MWT, ISWT, ESWT, and ICET. As could be anticipated, we did not identify studies assessing the repeatability of the ECET, ITT, or ETT; these studies test to exhaustion and would be impractical for patients with COPD to perform repeatedly on the same day. Of those reported, the 6MWT was by far the most thoroughly assessed. This may reflect its simplicity and relevance to daily life. However, a substantial proportion of the reported data does not explicitly support 6MWT repeatability. Some studies suggested that the ISWT was more repeatable and reproducible than the 6MWT, while another found that the ESWT was, in turn, more repeatable than the ISWT. The comparative results of exercise tests are inconsistent. Furthermore, there are only very limited data to support the repeatability and/or reproducibility of all cycle tests and treadmill tests. Repeatability and reproducibility were generally improved with familiarization in all types of test. Although this review assesses the influence of protocol variations, it is possible that some studies in which variations were not the primary focus may not have been identified. However, it is clear that even in ostensibly identical tests (eg, two studies reporting the 6MWT), responses can be significantly affected by subtle variations in track layout or environment, or by encouragement from the researchers conducting the test.

When studies reporting the sensitivity of two or more exercise tests to therapeutic intervention were reviewed, there was no consistent evidence supporting the use of one test over any other. Nine of the eleven studies assessing the 6MWT after pulmonary rehabilitation, two of the four studies assessing 6MWT after bronchodilators, and the only study assessing the 6MWT after lung-volume reduction surgery all reported distance improvements greater than the MCID. Additionally, three of the five studies assessing ISWT and all five of the studies assessing ESWT after pulmonary rehabilitation reported distance improvements greater than the MCID. Performance improvements were also observed to be in excess of the MCID in the only two studies assessing the ESWT following bronchodilator therapy; this was therefore the only exercise test reported by multiple papers that consistently responded to therapeutic interventions to a clinically relevant degree. It must be considered that these exercise tests have differing physiological demands, and it may be that the benefits of each intervention are measured differently by each test. However, limited data for bronchodilator therapy and lung-volume reduction surgery make it difficult to identify any obvious differences in the responses of tests to these interventions.

The review has several limitations that must be acknowledged. As well as identifying whether or not test responses have exceeded the MCID, we have also reported percentage changes in exercise test performance whenever possible to enable a very crude comparison of test outcomes recorded in different units. However, we are aware that the validity of this comparison relies on a direct relationship between these test outcomes, which is unlikely to be true: a large percentage change in one test result may not be equivalent to a large percentage change in another.

In an attempt to assess the validity of exercise tests in patients with COPD as comprehensively as possible, we have collated data from studies that have used various definitions of COPD, which include distinct subcategories
such as emphysema and chronic bronchitis, and that have employed diverse methods of assessing diagnosis and severity. Moreover, the study designs included are very diverse. These issues make meaningful meta-analysis difficult. We have, however, tried to provide the ranges of responses (both absolute and percentage changes) when possible, to provide an indication of the magnitude of exercise test responses. A further consequence of the comprehensive nature of this review is that the sample sizes of the identified studies vary widely. For this reason, we have included study sample sizes in our tables.

Decisions regarding which test to use are also influenced by the practicalities of routine clinical practice. It is reasonable to assume that walking is more representative of daily life than cycling for patients with COPD. Furthermore, given the equivocal evidence for the use of the ESWT over the ISWT, clinicians may wish to consider that the ESWT requires a prior “workload setting” ISWT to be performed by the patient, requiring additional time and resource considerations. Any test of exercise capacity should be highly repeatable and reproducible and also should be able to detect changes in performance after interventions aiming to improve exercise capacity.

**Conclusion**

This review of the published literature has found good evidence to support the repeatability and reproducibility of all tests, particularly the 6MWT, as long as a prior familiarization is conducted. There is consistent evidence to suggest that the ESWT is highly sensitive to therapeutic intervention. Sensitivity data that are available for other tests are largely inconsistent, and the 6MWT and ISWT appear to be less sensitive to intervention than the ESWT and ICET. These factors, allied to practical aspects, must be considered when planning interventional trials.

**Author contributions**

All authors contributed to the conception and design of the study, the analysis and interpretation of data, and revision of the manuscript. All authors approved the final version of the manuscript for publication.

**Acknowledgments**

The study was funded by GlaxoSmithKline UK. Writing support was provided by Martin Bell of Oxford PharmaGenesis™ Ltd, funded by GlaxoSmithKline UK. Georgina Meakin and Iain Fotheringham are employees of Oxford PharmaGenesis Ltd, which has received funding from GlaxoSmithKline UK. Yogesh Suresh Punekar, John Riley, and Sarah Cockle are current employees of GlaxoSmithKline, Uxbridge, UK. Sally J Singh was involved with the development of the incremental shuttle walk test, and has served on advisory boards for GlaxoSmithKline. Sally J Singh was part funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care East Midlands. Support was also provided by the NIHR Leicester Respiratory Biomedical Research Unit. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health.

**Disclosure**

Other than the funding outlined in the “Acknowledgments” section, the authors declare no conflicts of interest in this work.

**References**


**Supplementary materials**

Complete Ovid search strings

**Table S1** Embase™ search strings, search conducted January 22, 2013

<table>
<thead>
<tr>
<th>Search</th>
<th>Search strings</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>chronic obstructive pulmonary disease.mp. or exp chronic obstructive lung disease/</td>
<td>68,878</td>
</tr>
<tr>
<td>2</td>
<td>chronic obstructive lung disease.mp.</td>
<td>65,527</td>
</tr>
<tr>
<td>3</td>
<td>chronic obstructive respiratory disease.mp. or copd.mp.</td>
<td>119</td>
</tr>
<tr>
<td>4</td>
<td>chronic obstructive airway disease.mp.</td>
<td>33,438</td>
</tr>
<tr>
<td>5</td>
<td>lung emphysema.mp. or exp lung emphysema/</td>
<td>16,146</td>
</tr>
<tr>
<td>6</td>
<td>chronic obstructive airway disease.mp.</td>
<td>19,606</td>
</tr>
<tr>
<td>7</td>
<td>chronic airflow obstruction.mp.</td>
<td>378</td>
</tr>
<tr>
<td>8</td>
<td>1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9</td>
<td>103,379</td>
</tr>
<tr>
<td>9</td>
<td>exp cardiopulmonary exercise test/or exp exercise tolerance/or exp aerobic exercise/or exp exercise/or exp exercise test/</td>
<td>228,181</td>
</tr>
<tr>
<td>10</td>
<td>(exercise or exercise capacity or exercise tolerance).mp.</td>
<td>302,038</td>
</tr>
<tr>
<td>11</td>
<td>(clinically important difference or mcid or clinically meaningful).mp.</td>
<td>6,903</td>
</tr>
<tr>
<td>12</td>
<td>11 or 12 or 13</td>
<td>311,378</td>
</tr>
<tr>
<td>13</td>
<td>(‘6 minute walk’ or ‘6m walk’ or ‘6 m walk’ or ‘six minute walk’ or ‘6mwd’).mp.</td>
<td>4,644</td>
</tr>
<tr>
<td>14</td>
<td>(‘12 minute walk’ or ‘12m walk’ or ‘12 m walk’ or ‘twelve minute walk’ or ‘12mwd’).mp.</td>
<td>90</td>
</tr>
<tr>
<td>15</td>
<td>(‘shuttle walk’ or ‘shuttle walking’ or ‘iswt’ or ‘10 metre walking’ or ‘10 meter walking’ or ‘eswt’).mp.</td>
<td>1,010</td>
</tr>
<tr>
<td>16</td>
<td>cycle ergometer.mp. or exp bicycle ergometer/</td>
<td>5,540</td>
</tr>
<tr>
<td>17</td>
<td>bicycle ergometry.mp. or exp bicycle ergometry/</td>
<td>6,840</td>
</tr>
<tr>
<td>18</td>
<td>(cycle ergometry or cycle ergometric or bicycle ergometric).mp.</td>
<td>1,721</td>
</tr>
<tr>
<td>19</td>
<td>treadmill.mp. or exp treadmill/</td>
<td>29,379</td>
</tr>
<tr>
<td>20</td>
<td>(short physical performance battery or sppb).mp.</td>
<td>367</td>
</tr>
<tr>
<td>21</td>
<td>15 or 16 or 17 or 18 or 19 or 20 or 21 or 22</td>
<td>45,996</td>
</tr>
<tr>
<td>22</td>
<td>10 and 14 and 23</td>
<td>1,626</td>
</tr>
<tr>
<td>23</td>
<td>(test or assessment or capacity).mp.</td>
<td>3,225,394</td>
</tr>
<tr>
<td>24</td>
<td>10 and 14 and 23</td>
<td>1,408</td>
</tr>
<tr>
<td>25</td>
<td>(test or assessment or capacity).mp.</td>
<td>1,099</td>
</tr>
</tbody>
</table>

**Table S2** Medline® search strings, search conducted January 22, 2013

<table>
<thead>
<tr>
<th>Search</th>
<th>Search string</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>chronic obstructive pulmonary disease.mp. or exp pulmonary disease, chronic obstructive/</td>
<td>31,239</td>
</tr>
<tr>
<td>2</td>
<td>chronic obstructive lung disease.mp.</td>
<td>2,686</td>
</tr>
<tr>
<td>3</td>
<td>lung diseases, obstructive/or chronic obstructive respiratory disease.mp. or copd.mp.</td>
<td>150,257</td>
</tr>
<tr>
<td>4</td>
<td>chronic obstructive airway disease.mp.</td>
<td>21,567</td>
</tr>
<tr>
<td>5</td>
<td>chronic bronchitis.mp. or exp bronchitis, chronic/</td>
<td>8,867</td>
</tr>
<tr>
<td>6</td>
<td>lung emphysema.mp.</td>
<td>298</td>
</tr>
<tr>
<td>7</td>
<td>chronic obstructive airway disease.mp.</td>
<td>239</td>
</tr>
<tr>
<td>8</td>
<td>chronic airflow obstruction.mp.</td>
<td>502</td>
</tr>
<tr>
<td>9</td>
<td>1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9</td>
<td>161,227</td>
</tr>
<tr>
<td>10</td>
<td>exp exercise tolerance/or exp exercise test/or exercise.mp. or exp exercise/or exercise test/</td>
<td>239,508</td>
</tr>
<tr>
<td>11</td>
<td>(exercise capacity or exercise tolerance).mp.</td>
<td>16,495</td>
</tr>
<tr>
<td>12</td>
<td>(clinically important difference or mcid or clinically meaningful).mp.</td>
<td>4,675</td>
</tr>
<tr>
<td>13</td>
<td>11 or 12 or 13</td>
<td>243,911</td>
</tr>
<tr>
<td>14</td>
<td>(‘6 minute walk’ or ‘6m walk’ or ‘6 m walk’ or ‘six minute walk’ or ‘6mwd’).mp.</td>
<td>2,563</td>
</tr>
<tr>
<td>15</td>
<td>(‘12 minute walk’ or ‘12m walk’ or ‘12 m walk’ or ‘twelve minute walk’ or ‘12mwd’).mp.</td>
<td>62</td>
</tr>
<tr>
<td>16</td>
<td>(‘shuttle walk’ or ‘shuttle walking’ or ‘iswt’ or ‘10 metre walking’ or ‘10 meter walking’ or ‘eswt’).mp.</td>
<td>644</td>
</tr>
<tr>
<td>17</td>
<td>cycle ergometer.mp.</td>
<td>4,838</td>
</tr>
<tr>
<td>18</td>
<td>treadmill.mp.</td>
<td>20,912</td>
</tr>
<tr>
<td>19</td>
<td>(short physical performance battery or sppb).mp.</td>
<td>247</td>
</tr>
<tr>
<td>20</td>
<td>15 or 16 or 17 or 18 or 19 or 20</td>
<td>28,642</td>
</tr>
<tr>
<td>21</td>
<td>10 and 14 and 21</td>
<td>1,278</td>
</tr>
<tr>
<td>22</td>
<td>(test or assessment or capacity).mp.</td>
<td>1,939,771</td>
</tr>
<tr>
<td>23</td>
<td>10 and 14 and 21 and 23</td>
<td>1,065</td>
</tr>
<tr>
<td>24</td>
<td>limit 24 to (English language and humans)</td>
<td>928</td>
</tr>
</tbody>
</table>
Table S3 The Cochrane Library search strings, search conducted January 22, 2013

<table>
<thead>
<tr>
<th>Search</th>
<th>Search string</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MeSH descriptor: [Pulmonary Disease, Chronic Obstructive] explode all trees</td>
<td>1,845</td>
</tr>
<tr>
<td>2</td>
<td>chronic obstructive pulmonary disease</td>
<td>4,929</td>
</tr>
<tr>
<td>3</td>
<td>MeSH descriptor: [Lung Diseases, Obstructive] explode all trees</td>
<td>12,759</td>
</tr>
<tr>
<td>4</td>
<td>chronic obstructive respiratory disease</td>
<td>3,340</td>
</tr>
<tr>
<td>5</td>
<td>copd</td>
<td>6,244</td>
</tr>
<tr>
<td>6</td>
<td>chronic obstructive airway disease</td>
<td>4,348</td>
</tr>
<tr>
<td>7</td>
<td>MeSH descriptor: [Bronchitis, Chronic] explode all trees</td>
<td>77</td>
</tr>
<tr>
<td>8</td>
<td>chronic bronchitis</td>
<td>1,969</td>
</tr>
<tr>
<td>9</td>
<td>lung emphysema</td>
<td>517</td>
</tr>
<tr>
<td>10</td>
<td>chronic obstructive airway disease</td>
<td>4,348</td>
</tr>
<tr>
<td>11</td>
<td>chronic airflow obstruction</td>
<td>416</td>
</tr>
<tr>
<td>12</td>
<td>#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11</td>
<td>18,973</td>
</tr>
<tr>
<td>13</td>
<td>MeSH descriptor: [Exercise Tolerance] explode all trees</td>
<td>1,431</td>
</tr>
<tr>
<td>14</td>
<td>MeSH descriptor: [Exercise Test] explode all trees</td>
<td>5,816</td>
</tr>
<tr>
<td>15</td>
<td>MeSH descriptor: [Exercise] explode all trees</td>
<td>11,528</td>
</tr>
<tr>
<td>16</td>
<td>exercise or exercise capacity or exercise tolerance clinically important difference or mcid or clinically meaningful</td>
<td>38,220</td>
</tr>
<tr>
<td>17</td>
<td>#13 or #14 or #15 or #16 or #17</td>
<td>7,450</td>
</tr>
<tr>
<td>18</td>
<td>‘6 minute walk’ or ‘6m walk’ or ‘6 m walk’ or ‘six minute walk’ or ‘6mwd’</td>
<td>45,480</td>
</tr>
<tr>
<td>19</td>
<td>‘12 minute walk’ or ‘12m walk’ or ‘12 m walk’ or ‘twelve minute walk’ or ‘12mwd’</td>
<td>3,377</td>
</tr>
<tr>
<td>20</td>
<td>‘shuttle walk’ or ‘shuttle walking’ or ‘iswt’ or ‘10 metre walking’ or ‘10 meter walking’ or ‘eswt’</td>
<td>716</td>
</tr>
<tr>
<td>21</td>
<td>‘cycle ergomet’</td>
<td>2,392</td>
</tr>
<tr>
<td>22</td>
<td>treadmill</td>
<td>3,759</td>
</tr>
<tr>
<td>23</td>
<td>short physical performance batter or sppb</td>
<td>41</td>
</tr>
<tr>
<td>24</td>
<td>#19 or #20 or #21 or #22 or #23 or #24</td>
<td>9,217</td>
</tr>
<tr>
<td>25</td>
<td>test or assessment or capacity</td>
<td>211,756</td>
</tr>
<tr>
<td>26</td>
<td>#12 and #18 and #25 and #26</td>
<td>779</td>
</tr>
</tbody>
</table>