

# Inter-Day Test–Retest Reproducibility of the CAT, CCQ, HADS and EQ-5D-3L in Patients with Severe and Very Severe COPD

Henrik Hansen <sup>1</sup>Nina Beyer<sup>2</sup>Anne Frølich<sup>3,4</sup>Nina Godtfredsen <sup>1,2</sup>Theresa Bieler <sup>5</sup>

<sup>1</sup>Department of Respiratory Medicine, Respiratory Research Unit, Hvidovre University Hospital, Hvidovre, Denmark;

<sup>2</sup>Institute for Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; <sup>3</sup>Innovation and Research Centre for Multimorbidity, Slagelse Hospital, Slagelse, Denmark; <sup>4</sup>Section of General Practice, Department of Public Health, University of Copenhagen, Copenhagen, Denmark; <sup>5</sup>Department of Physical & Occupational Therapy, Bispebjerg and Frederiksberg Hospital, University of Copenhagen, Copenhagen, Denmark

**Introduction:** In patients with COPD, the COPD Assessment Test (CAT), Clinical COPD Questionnaire (CCQ), Hospital Anxiety and Depression Scale (HADS) and EuroQol 5D (EQ-5D-3L) are widely used patient reported outcome measures (PROMs) of respiratory symptoms, anxiety, depression and quality of life. Despite established validity, responsiveness and minimal important change (MIC), the reproducibility and especially important agreement parameters remain unreported in these frequently used PROMs. The aim of this study was to investigate the inter-day test–retest reliability and agreement of the CAT, CCQ, HADS and EQ-5D-3L in patients with severe and very severe COPD (FEV<sub>1</sub> <50%) eligible for hospital-based pulmonary rehabilitation.

**Patients and Methods:** Fifty patients (22 females, mean [SD] age 67 [9] yrs.; FEV<sub>1</sub> 32[9] %; 6-minute walk distance 347 [102] meters; CAT 21 [6] points; BMI: 26 [6] kg/m<sup>2</sup>) completed the questionnaires (CAT, CCQ, HADS, EQ-5D-3L) in combination with functional performance test instructed by one assessor on test-day one (T1) and by another assessor 7–10 days later on test-day two (T2).

**Results:** The inter-day test–retest reliability ICC was 0.88 (LL<sub>95CI</sub>: 0.80) for CAT; 0.69 (LL<sub>95CI</sub>: 0.46) for CCQ; 0.86 (LL<sub>95CI</sub>: 0.75) and 0.90 (LL<sub>95CI</sub>: 0.82) for HADS-anxiety (A) and depression (D) and 0.87 (LL<sub>95CI</sub>: 0.76) for EQ-5D-VAS. The corresponding agreements within a single measurement (standard error of measurement, SEM) and for repeated measurement errors (smallest real difference, SRD) were respectively 2.1 and 2.9 points for CAT; 0.5 and 0.7 points for CCQ total; 1.3 and 1.9 points for HADS-A; 0.9 and 1.3 points for HADS-D and 6.8 and 9.7 VAS-score for EQ-5D-3L, respectively. Ceiling/flooring effect was present in <5% for all questionnaires.

**Conclusion:** In patients with severe and very severe COPD, the CAT, CCQ, HADS and EQ-5D-3L questionnaires presented moderate to excellent inter-day test–retest reliability, and no floor or ceiling effect was documented for any of the questionnaires. Only CAT and HADS had an acceptable SRD below the established MIC for assessing change over time on group level, and none of the PROMs were fit to assess individual changes over time.

**Keywords:** COPD, questionnaires, patient reported outcomes, reproducibility of results

Correspondence: Henrik Hansen  
Department of Respiratory Medicine,  
Respiratory Research Unit, Hvidovre  
University Hospital, Kettegård Alle 30,  
Center 2, Section 255, Hvidovre, 2650,  
Denmark  
Tel +45 28946780  
Email henrik.hansen.09@regionh.dk

## Introduction

In chronic obstructive pulmonary disease (COPD), patient reported outcome measures (PROMs) of respiratory symptoms, other symptoms (eg anxiety), and health-related quality of life are increasingly used both as descriptive instruments and as effect outcome measures.<sup>1–4</sup> In addition, the use of PROMs as critical effect outcomes are being endorsed by health authorities and scientific societies.<sup>5</sup> Especially,

symptom relief is a warranted core outcome in COPD care and pulmonary rehabilitation (PR),<sup>1,5</sup> because COPD is an incurable disease with increasing severity of symptoms as the disease progresses. Both validity and reproducibility are essential requirements of PROMs to be used as outcomes measures of symptoms and health-related quality of life.<sup>6</sup>

Reproducibility concerns the degree to which repeated measurements provide similar results in a specific population.<sup>6</sup> Reproducibility comprises reliability parameters that assess how well patients can be distinguished from each other despite measurement errors, and agreement parameters that assess exactly how close the results of repeated measures are.<sup>7,8</sup> Agreement parameters indicate systematic and random errors attributed to the measure itself.<sup>6</sup> Therefore, agreement parameter of PROMs is paramount in research and clinical settings given the importance of detecting individual and group changes over time, eg after intervention. The smaller the measurement error, the smaller the changes that can be detected beyond measurement error.<sup>7–9</sup> For reproducibility studies of PROMs with continuous scales scores, the COnsensus based Standards for the selection of health Measurement INstruments guideline (COSMIN) recommends that agreement parameters, ie the standard error of measurement (SEM), limits of agreement (LOA) or smallest detectable change (SDC), be calculated and reported.<sup>9,10</sup> Nevertheless, the reproducibility and notably measurement errors have only been sparsely reported in some of the commonly used PROMs to evaluate eg PR.<sup>11–18</sup>

A variety of PROMs are being used in all types of study designs related to COPD as well as in clinical practice.<sup>2–4,19–22</sup> St. George Respiratory Questionnaire (SGRQ) is considered the gold-standard questionnaire covering patients self-reported respiratory symptoms.<sup>23</sup> However, both the COPD Assessment Test (CAT) and the COPD Clinical Questionnaire (CCQ) are frequently preferred as they are considered less time consuming, easier to complete for patients and easier to interpret for healthcare professionals.<sup>23</sup> Both the CAT and the CCQ have proved excellent concurrent validity with the SGRQ.<sup>11,16,23–25</sup> Reliability parameters, which are highly dependent of the heterogeneity of the study sample,<sup>6</sup> have been investigated for the CAT and the CCQ questionnaires in several studies among patients with varying severity of COPD. The reported intraclass correlation coefficient (ICC) ranged from 0.80 to 0.96 for CAT and 0.70 to 0.99 for CCQ indicating moderate to excellent reliability.<sup>11–18,24,26–28</sup> The agreement parameters have been investigated for the CCQ in

four studies. Three studies reported SEM ranging from 0.10 to 0.21 points for the total score,<sup>24,27,28</sup> and one study<sup>28</sup> reported a 95% LOA from –1.87 to 1.35 points. Regarding CAT, only one study has reported agreement parameters, ie a SEM of 1.92 points, mainly in patients with mild to moderate airflow obstruction, low symptom score, and high walking capacity.<sup>24</sup> Reliability and agreement parameters are disease specific,<sup>8</sup> and because COPD is a heterogeneous disease, parameters determined in patients with mild to moderate airflow obstruction, low symptom score, and high walking capacity may not necessarily apply for patients with severe and very severe COPD referred to hospital-based PR.

Among other symptoms frequently reported in COPD are anxiety and depression. The Hospital Anxiety and Depression Scale (HADS) questionnaire is a generic PROM, which is widely used across medical conditions. In patients with COPD, HADS is used for both symptom screening and evaluation of changes in symptoms following an intervention.<sup>3,22,29,30</sup> To our knowledge no study has reported reliability and agreement parameters for the HADS in patients with COPD. Likewise, we were unable to find any study in patients with COPD concerning reproducibility for the widely used generic questionnaire EuroQol 5D (EQ-5D-3L), which assesses health related quality of life.

The reproducibility of a questionnaire is usually assessed using a test–retest design with repeated administration (at least two) of the questionnaire over a period of time when the underlying construct (eg respiratory symptoms) is stable.<sup>9,31</sup> Consequently, it is important to select patients whose symptoms are not expected to change, and to carefully choose a between-administration time gap that is neither too short nor too long. A too short period might allow patients to recall their earlier responses and a too long period might allow for a true change in the status of the patient.<sup>9,10,32</sup>

The primary aim of this study was to investigate the inter-day test–retest reliability and agreement of commonly used PROMs, ie CAT, CCQ, HADS and EQ-5D-3L, in patients with severe and very severe COPD ( $FEV_1 < 50\%$ ) eligible for hospital-based PR.

## Patients and Methods

### Study Design

This inter-day test–retest reproducibility study was planned as one of two separate reproducibility studies, which both were part of a randomized controlled multicenter trial (RCT) (ClinicalTrials.gov-identifier:

NCT02667171) investigating the effect of pulmonary tele-rehabilitation and conventional PR in patients with severe and very severe ( $FEV_1 < 50\%$ ) COPD.<sup>33,34</sup> The purpose of conducting this nested reproducibility study was to obtain knowledge about how much difference is needed to detect a real change in the PROM outcomes used in the RCT, considering the measurement errors. We followed the Guideline for Reporting Reliability and Agreement Studies (GRRAS)<sup>8</sup> and the COSMIN standards for studies on reliability.<sup>10</sup>

## Participants

Eligible patients for the RCT were identified and recruited by respiratory nurses during outpatient COPD control visits from the University Hospitals Amager, Hvidovre, Bispebjerg, Frederiksberg, Herlev, Gentofte, Frederikssund and Hillerød. All patients provided written and informed consent. The RCT was approved by the Ethics Committee of the Capital Region of Denmark (H-15019380) and the Danish Data Protection Agency (jr.no.: 2012-58-0004) and conducted in accordance with the ethical principles of the Helsinki Declaration.

All patients who agreed to participate in the RCT were consecutively asked to participate in the reproducibility study, which required an extra assessment visit prior to randomization and intervention start. Recruitment for the reproducibility study commenced on March 18, 2016 and continued until 50 patients were recruited in March 20, 2017. A consecutive convenience sample size of 50 patients was chosen according to the recommendation from COSMIN.<sup>32</sup>

Inclusion and exclusion criteria<sup>33</sup> corresponded to the criteria for outpatient hospital-based routine PR in the Capital Region of Copenhagen, Denmark and pertained to adults with a clinical diagnosis of COPD defined as  $FEV_1/FVC$  ratio  $< 0.70$ ;  $FEV_1 < 50\%$ ;  $MRC \geq 2$ ; able to communicate in Danish; no cognitive impairments; no contraindication to exercise intervention; and no participation in PR within the prior six months.<sup>33</sup>

## Study Setting

Administration of the questionnaires was conducted at the Respiratory and Physical Therapy Departments of five different University Hospitals (Hvidovre, Bispebjerg, Herlev, Gentofte and Frederikssund) in Greater Copenhagen by ten raters who were familiar with the questionnaires from clinical practice and had obtained accreditation to be raters.

## Assessment/Test Procedures

The raters followed the exact same procedures (Figure 1) at test-day one (T1) and test-day two (T2), and administration of the questionnaires were conducted in the same location and at the same time during the outpatient clinics' opening hours from 10am to 2pm, Monday to Friday. The administration on the first test-day (T1) was conducted by one rater, and another rater completed the administration on the second test-day (T2). To ensure that the first administration of the questionnaires (T1) had no influence on the second administration (T2), patients and raters were blinded to the previous responses, and the interval between the two administrations was 7–10-days. This interval was chosen and appraised as long enough to prevent recall bias and short enough to ensure that the patients had not changed on the constructs that were to be measured. The patients completed the questionnaires in a pause between two sets of performance tests, ie, the six-minute walk test and the 30-second sit-to-stand test (Figure 1). The CAT, CCQ, HADS and EQ-5D-3L were administered to all patients in the same order, and the patients filled out the questionnaires in an undisturbed room without interference from the rater. All patients got a brief, standardized instruction in how to complete the questionnaires from the rater;

Answer the questionnaires and questions consecutively in the prepared order. If you have difficulty understanding a question, I will help you with the clarification of the specific question when all other questions are answered. Take the time you need; you do not need to hurry.

Patients were instructed not to do any vigorous activities three hours prior to the appointment and to take their prescribed medication as usual. The administration procedure reflects the conditions in everyday clinical practice, where several performance tests and questionnaires are conducted within a narrow time frame (Figure 1).

## Questionnaires

COPD Assessment Test (CAT) assesses the impact of COPD on self-reported health status and symptoms.<sup>16</sup> It consists of 8 items, each scored between 0 and 5 points (0 = no impact or symptoms, 5 = worst possible impact or symptoms) summing up to a total CAT score ranging from 0 to 40 points.<sup>16</sup> A minimal important change (MIC) of 2–3 points has been reported.<sup>35,36</sup>

Clinical COPD Questionnaire (CCQ) assesses self-reported quality of life.<sup>11</sup> It consists of 10 items, each

### Assessment and progression:

1. Subject history/introduction, while seated: resting blood pressure, resting heart rate, resting SpO<sub>2</sub>, resting dyspnea. Standing: anthropometric measures, i.e. weight and height (30 minutes)
2. Instruction and performing 6MWT, end-heart rate, end-SpO<sub>2</sub>, end-dyspnea (10 minutes)
3. Seated rest (5 minutes)
4. Instruction and performing 30sec-STs (5 minutes)
5. Instruction of four questionnaires. Completion of the questionnaires in the order CAT, CCQ, HADS, EQ5D-3L, in an undisturbed room without interference (30 minutes)
6. Seated: resting blood pressure, resting heart rate, resting SpO<sub>2</sub>, resting dyspnea (5 minutes)
7. Instruction and performing 6MWT, end-heart rate, end-SpO<sub>2</sub>, end-dyspnea (10 minutes)
8. Seated rest for (5 minutes)
9. Instruction and performing 30sec-STs (5 minutes)
10. Assessment session completed. Total time 145 minutes.

**Figure 1** Assessment procedures at test-day one (T1) and test-day two (T2).

**Abbreviations:** SpO<sub>2</sub>, arterial oxygen saturation as measured by pulse oximetry (%); dyspnea, perceived dyspnea (Borg cr-10); 6MWT, six-minute walk test; 30sec-STs, 30 seconds sit-to-stand test (repetitions); end-, immediately measure after test completion; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; HADS-A and D, Hospital Anxiety and Depression Scale (HADS); EQ-5D-3L, EuroQol 5-Dimension 3-likert utility score and VAS score.

scored between 0 and 6 points (0 = no impairment), divided into three domains: Symptoms (4-Items), Functional state (4-Items) and Mental state (2-Items).<sup>11</sup> The total score is calculated by summing the individual items and dividing by 10. A MIC of 0.4 points has been reported.<sup>27,36,37</sup>

Hospital Anxiety and Depression Scale (HADS) assesses the level of anxiety and level of depression in medically ill persons.<sup>38</sup> The scale comprises two subscales, HADS anxiety (HADS-A, 7-items) and HADS depression (HADS-D, 7-items), with each item scored between 0 and 3. A total subscale score of 0–7 is considered normal, 8–10 indicates a risk of anxiety or depression, and 11–21 indicates considerable symptoms of anxiety or depression disorder.<sup>38</sup> A MIC of 1.5 points in each scale has been reported.<sup>36,39</sup>

EuroQol 5-Dimension Questionnaire (EQ-5D), is a generic global questionnaire measuring health-related quality of life.<sup>40</sup> We used the 3-Likert version of the EQ-5D-3L, which has a descriptive and a visual analogue scale. The descriptive system (EQ-5D) comprises five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. In the 3-likert version each dimension has three scoring levels (no problem,

some problem, severe problem). This version comprises a total of 243 utility scores ranging from –0.624 (worst possible health utility) to 1.0 (best possible health utility) based on Danish EQ-5D-3L norm data set. The EQ-5D-VAS records the overall self-rated health on a 20 cm vertical visual analog scale ranging from zero (worst imaginable health) to 100 (best imaginable health).<sup>40</sup> A MIC of 6.5 to 8.0 points in EQ-5D-VAS is suggested in persons with COPD, while MIC has not been reported for the 5-dimension 3-likert questionnaire in patients with COPD.<sup>41,42</sup>

### Demographic and Descriptive Variables

Demographic and descriptive variables including age, gender, body mass index, smoking status, FEV<sub>1</sub>/FVC, FEV<sub>1</sub>, GOLD, A/B/C/D stratification,<sup>43</sup> Charlson Comorbidity Index, BODE-index and oxygen supplement were registered at T1.<sup>44</sup>

### Statistical Analysis

Descriptive data are presented as means with standard deviations (SD) for continuous data and as medians with range for ordinal data and data not normally distributed. Data distribution was inspected by histogram, Q-Q Plots

and verified by Shapiro–Wilk test to determine approximately normal distribution. Independent *t*-test or Mann–Whitney *U*-test was used to compare demographic and descriptive variables between patients included and not included in the study. Paired *t*-test or Wilcoxon signed rank test was used to compare inter-day systematic bias between the patients' completed questionnaires at T1 and T2.

Intra-class correlation coefficient (ICC) was calculated to describe the reliability. The ICC<sub>1,1</sub> model was used because the assessments were conducted at five centers, and all raters did not instruct each patient.<sup>7,45</sup> The ICC<sub>1,1</sub> is a fixed model. ICCs values between 0–0.49 were considered weak, ≥0.50–0.75 moderate, >0.75–0.90 good, and >0.90 excellent reliability.<sup>46</sup>

Agreement between results at T1 and T2 was calculated as standard error of measurement (SEM) and the SEM<sub>95</sub> using the equation  $SD \times \sqrt{1-ICC}$  respectively  $1.96 \times SEM$  (SEM<sub>95</sub>).<sup>7,45</sup> The SEM expresses the measurement error that occurs within a single measurement where no real change has occurred and indicates that there is a 68% likelihood that the “true” score for a group of patients' (or a single patient's SEM<sub>95</sub>) is within this measurement error.<sup>31,46</sup>

The corresponding smallest real difference (SRD) and SRD<sub>95</sub> was calculated by the equation  $\sqrt{2} \times SEM$  (SRD) and  $1.96 \times \sqrt{2} \times SEM$  (SRD<sub>95</sub>) respectively. The SRD represents the smallest real difference to be detected beyond the measurement error of repeated measurement without a real change in a group of patients (or a single patient's SRD<sub>95</sub>).<sup>45,47,48</sup> The SEM, SEM<sub>95</sub>, SRD and SRD<sub>95</sub> are expressed in the same unit as the original measurement. To make comparisons between our agreement parameters and results from other studies easier, these parameters were also expressed as a percentage of the mean from the two subsequent visits (grand mean).

We determined a questionnaire suitable for evaluative use, when the SRD was smaller than the established minimal important change (MIC). The MIC, which is derived from longitudinal validity studies and preferably determined by using an anchor-based methods, is the smallest change in an outcome that an individual patient or clinician would identify as important.<sup>31,49</sup> MIC is often referred to as the minimal important difference (MID) or the minimal clinical important difference (MCID), and because that they constitute the same, they are used interchangeably in the literature.<sup>50</sup>

Bland Altman plots were used to visualize potential systematic bias around the zero line as well as heteroscedasticity. The mean difference with 95% CI and limits of agreement (95% LOA) were calculated as  $\text{mean} \pm 1.96 \times SD$  and included in the plots.<sup>7,51</sup> For all analyses P values of less than 0.05 were considered statistically significant.

Finally, we report the proportion of patients with minimum and maximum score for each questionnaire, because this shows the population-specific risk of floor and/or ceiling effects. There is no consensus regarding cut off values for floor or ceiling effects, but it has been suggested that it is present if >15% of the participants achieve the lowest (floor) or highest (ceiling) score.<sup>52</sup> Floor and ceiling effects are of special interest in intervention studies, because patients with the lowest possible scores may not be able to further decline, and patients with the best possible scores may not be able to further improve, following an intervention. Data was analyzed using SPSS version 22.0 (SPSS Inc., Chicago, IL, USA).

## Results

### Participants vs Non-Participants

Of the 108 eligible patients, 50 (22 females, mean [SD] age 67 [9] yrs.; FEV<sub>1</sub> 32 [9] %; 6-minute walk distance (6MWD) 347 [102] meters; CAT 21 [6] points; BMI: 26 [6] kg/m<sup>2</sup>) agreed to participate in the reproducibility study ([Supplementary Figure 1](#)) shows how the final sample was obtained). Twenty-three declined to participate due to the extra testing date, while 35 patients could not be included because they undertook the baseline assessments for the RCT less than one week before the scheduled randomization and intervention. Demographic and descriptive characteristics of the 58 patients, who did not participate in the reproducibility study, did not differ significantly from those who participated ([Table 1](#)).

### Inter-Day Test–Retest Reproducibility

All questionnaires and items were completed at both T1 and T2, and therefore, no values are missing. Test–retest reliability (ICC<sub>1,1</sub>) for the CAT, CCQ-total, HADS-A, HADS-D and EQ-5D-VAS were 0.88, 0.69, 0.86, 0.90 and 0.87, respectively. The test–retest agreement parameters of the questionnaires are presented in [Table 2](#). Agreement between test results on group level within a single measurement (SEM) and for repeated measurement errors (SRD) were respectively 2.1 and 2.9 points for CAT; 0.5 and 0.7 points for CCQ total; 1.3 and 1.9 points for

HADS-A; 0.9 and 1.3 points for HADS-D and 6.8 and 9.7 VAS-score for EQ-5D-3L, respectively. The Bland Altman plots with 95% limits of agreement for the questionnaires are shown in Figure 2 A to F. There was no significant difference between results at T1 and T2 for any of the PROMs (Table 2 and Figure 2A–F). For all questionnaires, less than 5% of the patients achieved the lowest (floor), respectively highest (ceiling) score (Table 2).

## Discussion

To the best of our knowledge this is the first study to report inter-day test–retest reproducibility parameters of the HADS and EQ-5D-3L in patient with COPD, and one of the few studies that have reported agreement parameters of

CAT and CCQ. We found excellent reliability and acceptable agreement for the CAT and HADS suggesting that they can be used for group evaluative purpose in patients with severe and very severe COPD.<sup>53</sup>

## CAT

In line with previous results (ICC ranging from 0.80 to 0.94)<sup>16–18,24,26</sup> we found good reliability for the CAT in patients with severe and very severe COPD. To our knowledge the study by Tsiligianni et al<sup>24</sup> is the only study that has reported agreement parameters for the CAT in patients with COPD. Although their patients had less symptoms (median CAT score 13 points), less disease severity (65% GOLD group I or II) and milder risk profile (BODE index

**Table 1** Characteristics of Eligible Participants

	Included	Not Included
Sex, men/women (n)	28/22	21/37
Age, years, mean (SD)	66.6±9.0	69.4±9.1
Body mass index, mean kg m <sup>-2</sup> , (SD)	25.4±5.6	25.8±5.6
FEV <sub>1</sub> % predicted, mean (SD)	32.0033±9.0	35.1±9.4
FEV <sub>1</sub> /FVC ratio, mean (SD)	41.4±10.6	45.1±11.8
GOLD I/II/III/IV, %	0/0/54/46	0/0/67/33
A/B/C/D, %	0/36/0/64	3/33/7/57
MRC dyspnea scale, median (range)	3.5 (3–5)	3.0 (2–5)
BODE index points, median (range)	5 (3–9)	5 (3–8)
Charlson index 1/2/≥3, (%)	52/30/18	28/47/26
LTOT, n (%)	4 (8)	9 (16)
6MWD in meters, mean (SD)	347 (102)	330 (103)
CAT score, mean (SD)	20.84±6.13	18.69±7.64
CCQ total, mean (SD)	2.90±0.92	2.68±0.98
CCQ-symptoms, mean, (SD)	2.88±1.00	2.75±1.02
CCQ-functional, mean (SD)	2.88±1.16	2.71±1.19
CCQ-Mental, mean (SD)	2.96±1.48	2.48±1.56
HADS-A, mean (SD)	5.72±3.63	5.67±3.88
HADS-D, mean (SD)	4.62±3.00	3.57±3.00
EQ-5D-VAS score, mean (SD)	49.22±19.50	54.22±18.59
EQ-5D-3L Utility score, mean (SD)	0.66±0.17	0.70±0.18

**Notes:** Data are presented as mean ±standard deviation and median (range) or percent in non-normally distributed variables; Independent *t*-test or Mann–Whitney *U*-test was used to compare demographic and descriptive variables between patients included and not included in the study and denoted \*if *p*<0.05.

**Abbreviations:** FEV<sub>1</sub>, forced expiratory volume in the first second; FVC, forced vital capacity; GOLD, Global initiative for Chronic Obstructive Lung; A/B/C/D, risk stratification; MRC, Medical Research Council; BODE index, body mass index, airflow obstruction, dyspnea and exercise capacity; LTOT, long-term oxygen therapy; 6MWD, six-minute walk distance; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; HADS-A and D, Hospital Anxiety and Depressions Scale (HADS); EQ-5D VAS score, EuroQol 5-Dimension Questionnaire Visual Analogue Scale; EQ-5D-3L Utility, EuroQol 5-Dimension 3-likert utility score.

**Table 2** Inter-Day Test–Retest Reproducibility of Results from the Patient Reported Questionnaires

Variables	Test-Day 1 (T1)	Test-Day 2 (T2)	Difference	Floor/Ceiling (n)	ICC <sub>1,1</sub> (LL <sub>95</sub> )	SEM (SEM%)	SEM <sub>95</sub> (SEM <sub>95</sub> %)	SRD (SRD%)	SRD <sub>95</sub> (SRD <sub>95</sub> %)
CAT total, points	20.84±6.13	20.00±5.89	-0.84 [-1.93; 0.25]	0/0	0.88 (0.80)	2.08 (10)	4.08 (20)	2.94 (14)	5.77 (28)
CCQ-Total, points	2.90±0.92	2.72±0.82	-0.18 [-0.41; 0.06]	0/0	0.69 (0.46)	0.48 (17)	0.94 (34)	0.68 (24)	1.41 (51)
CCQ-symptoms	2.88±1.00	2.70±0.92	-0.18 [-0.45; 0.08]	0/0	0.71 (0.49)	0.51 (19)	1.00 (36)	0.72 (26)	1.44 (51)
CCQ-functional	2.88±1.16	2.78±1.08	-0.10 [-0.37; 0.17]	2/3	0.78 (0.62)	0.52 (19)	1.02 (36)	0.74 (26)	2.47 (88)
CCQ-Mental	2.96±1.48	2.65±1.22	-0.31 [-0.73; 0.11]	0/0	0.57 (0.25)	0.90 (32)	1.76 (63)	1.27 (45)	1.33 (47)
HADS-A, points	5.72±3.63	5.80±3.52	-0.10 [-0.64; 0.80]	1/0	0.86 (0.75)	1.33 (23)	2.61 (45)	1.88 (32)	3.69 (64)
HADS-D, points	4.62±3.00	4.36±2.69	-0.20 [-0.76; 0.24]	0/0	0.90 (0.82)	0.90 (20)	1.76 (39)	1.27 (28)	2.49 (55)
EQ-5D-VAS, score	49.22±19.5	50.28±18.64	1.10 [-2.69; 4.81]	0/1	0.87 (0.76)	6.84 (14)	13.4 (27)	9.67 (19)	18.41 (37)
EQ-5D-3L Utility	0.66±0.17	0.68±0.13	-0.02 [-0.02; 0.05]	0/3	0.77 (0.59)	0.07 (11)	0.14 (21)	0.10 (15)	0.19 (28)

**Notes:** Results from test-day 1 and test-day 2 are presented as mean±SD and difference between days as mean±[SE CI95%]; Paired *t*-test or Wilcoxon signed rank test was used to compare inter-day systematic bias between the patients' completed questionnaires at T1 and T2 and denoted \**p*<0.05.

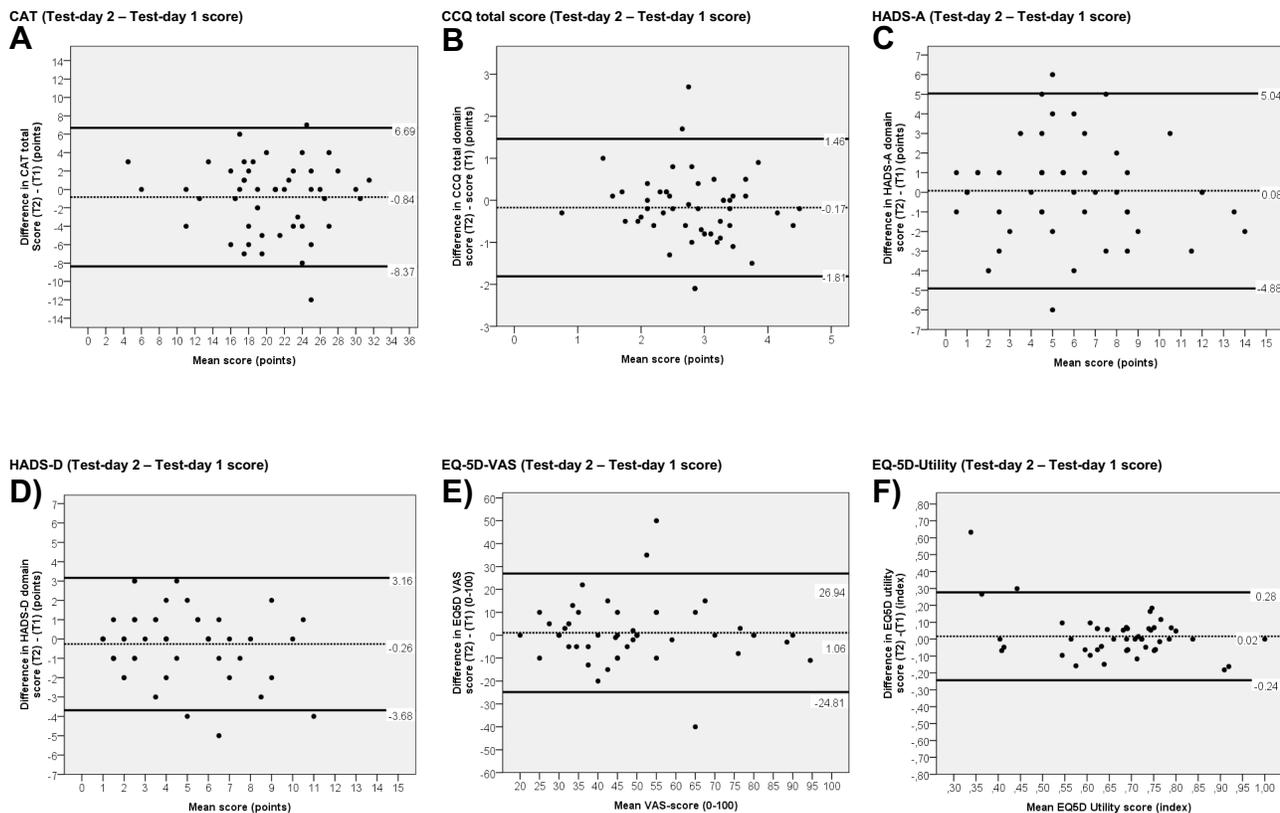
**Abbreviations:** Floor; lowest score; Ceiling, highest score; n, number; ICC<sub>1,1</sub>, intraclass correlation coefficient model 1,1; LL 95, lower limit 95% confidence; SEM, standard error of measurement; SEM%, standard error of measurement expressed as a percentage of the mean; SEM<sub>95</sub>, standard error of the measurement at the 95% confidence level; SEM<sub>95</sub>%, standard error of measurement expressed as a percentage of the mean; SRD, smallest real difference; SRD%, smallest real difference as a percentage of the mean; SRD<sub>95</sub>, smallest real difference at the 95% confidence level; SRD<sub>95</sub>%, smallest real difference as a percentage of the mean; CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; HADS-A and -D, Hospital Anxiety and Depression Scale (HADS); EQ-5D VAS, EuroQol 5-Dimension Questionnaire Visual Analogue Scale; EQ-5D-3L Utility, EuroQol 5-Dimension 3-Item utility score.

≤2 points), the agreement parameters (SEM: 1.9 points; LOA 95%: -8.0; 12.0 points) were very similar to ours. We could not find any other study that has reported the SRD. Our results suggest that a change of 2.9 points on group level, respectively 5.7 points on individual level, is required before we can be confident that a real change has occurred. In patients with moderate to severe COPD, the MIC for the CAT has been reported to be from 1 to 3.8 points depending on study design and method.<sup>35,36</sup> The MIC can be calculated in different ways and there is often uncertainty surrounding the calculation and interpretation of MIC.<sup>50</sup> In addition the MIC estimate may differ on the patients initial health status or symptom burden and the specific intervention delivered.<sup>50</sup> We found that the SRD for the CAT is lower than the previously reported MIC based on rehabilitations studies,<sup>35,36</sup> and this suggests that the MIC can be distinguished from repeated measurement error on a group level. Thus, it appears that CAT is acceptable for evaluative purposes in a group of patients with severe and very severe COPD. In contrast, our results at the individual level, SRD<sub>95</sub> of 5.8 points, suggest that the MIC cannot be distinguished from repeated measurement error in single patients. Substantial fluctuation in daily symptoms in patients with severe and very severe COPD might be a contributing factor.

To our knowledge floor and ceiling effects have not been investigated before. We did not find any floor or ceiling effects for the CAT, and thus this cannot have influenced the results.

### CCQ

The reliability of the CCQ total score was moderate (ICC: 0.69), which is in the lower end of what has previously been reported (ICC 0.70 to 0.99) in patients with mild to severe COPD.<sup>11–18,24,26–28</sup> Similarly, we found SEM (0.5 points) in the higher end than previously reported (SEM respectively reported as 0.2 points,<sup>27</sup> 0.4 points<sup>24</sup> and 0.6 points<sup>28</sup>). However, it must be noted that the previously reported SEM of 0.2 points appears to be estimated by using an ICC from an unrelated study sample. None of these previous studies reported the SRD, but Berkhoff et al<sup>28</sup> reported LOA (mean difference of -0.3 points with a 95% LOA from -1.9 to 1.4 points),<sup>28</sup> which is very similar to our results for LOA. The study by Berkhoff et al<sup>28</sup> collected data based on routine inclusion criteria similar to ours, the sample size was similar, baseline CCQ scores were comparable and the included patients had multimorbidity as most had ≥2 comorbidities. The study only differed from ours regarding FEV<sub>1</sub>% predicted



**Figure 2** Bland and Altman plots of the CAT, CCQ-total, HADS-A, HADS-D, EQ-5D-VAS and EQ-5D-Utility.

**Notes:** Mean difference between result from test-day 1 and test-day 2 (dotted line) with limits of agreement 95% CI (black lines). **(A)** CAT score difference obtained on two separate days (T2 vs T1). **(B)** CCQ total score difference obtained on two separate days (T2 vs T1). **(C)** HADS-A score difference obtained on two separate days (T2 vs T1). **(D)** HADS-D score difference obtained on two separate days (T2 vs T1). **(E)** EQ-5D-VAS score difference obtained on two separate days (T2 vs T1). **(F)** EQ-5D Utility score difference obtained on two separate days (T2 vs T1).

**Abbreviations:** CAT, COPD Assessment Test; CCQ, Clinical COPD Questionnaire; HADS-A and D, Hospital Anxiety and Depressions Scale (HADS); EQ-5D VAS score, EuroQol 5-Dimension Questionnaire Visual Analogue Scale; EQ-5D-3L Utility, EuroQol 5-Dimension 3-likert utility score.

mean, which was 51.0 (SD 15.0) in the Berkhoff study<sup>28</sup> and 32.3 (SD 9.0) in our study. We did not find any floor or ceiling effects for the CCQ.

The previously reported MIC for the CCQ total score is 0.4 points in patients with moderate to severe COPD.<sup>36,37</sup> Our result for the SRD (0.7 points) and SRD<sub>95</sub> (1.4 points) in patients with severe and very severe COPD (Table 2) suggests that the previously reported MIC cannot be distinguished from repeated measurement error. In that perspective, the CCQ may be less suitable compared to the CAT questionnaire for evaluating changes in respiratory symptoms over time, both on group and individual level, in patients with severe and very severe COPD. These findings need to be confirmed in future studies, before any appraisal can be made.

## HADS and EQ-5D-3L

Both HADS and EQ-5D-3L are commonly used outcomes in clinical research,<sup>3,4,19,22,54,55</sup> clinical practice<sup>56</sup> and for public health evaluative purposes.<sup>57</sup>

Although HADS has been used in patients with COPD, to our knowledge this is the first study to investigate the reproducibility of the HADS in this patient group. We found that the HADS questionnaire showed good reliability (ICC: 0.86 to 0.90) in patients with severe and very severe COPD. The agreement parameter SRD for the HADS-D (1.3 points) is below the established MIC of 1.5 points, while it was exceeded for the HADS-A subscale (SRD: 1.8 points).<sup>39</sup> The results indicate that the HADS-D subscale is acceptable for evaluative purposes in a group of patients with severe COPD. In contrast, the SRD<sub>95</sub> of 3.7 points (HADS-A) and 2.5 points (HADS-D) is greater than the established MIC suggesting that the HADS questionnaire is less suitable for evaluation of changes over time in single patients. We found no floor or ceiling effect for the HADS-A and HADS-D in patients with severe and very COPD.

Similar to the HADS questionnaire, we could not find any study that has investigated the reproducibility of the

EQ-5D-3L questionnaire in patients with COPD. We found that EQ-5D-VAS showed good reliability (ICC: 0.87) in patients with severe and very severe COPD. The agreement parameter SRD (9.7 points) exceeded the established MIC of 6.5 to 8.0 points.<sup>41,42</sup> None of the studies performed analysis for reproducibility when MIC was established.<sup>41,42</sup> The study by Nolan et al which was a prospective responder-blinded 8-week outpatient pulmonary rehabilitation intervention, reported an average increase in EQ-5D-VAS of 8.6 points (CI95%: 6.5 to 10.7) and thus did not exceed the SRD of 9.7 for groups. The study by Zanini et al reported an average EQ-5D-VAS improvement of 14 points (CI95%: 12.8 to 15.1) exceeding the SRD for groups. The effect size from this study is however limited by its retrospective and unblinded study design and the study was additionally based on data from a 3-week inpatient rehabilitation program in patients admitted with COPD exacerbation and changes due to medical treatment and recovery cannot be separated from the rehabilitation intervention. This indicates some cautiousness for the use of the EQ-5D-3L questionnaire for evaluative purposes unless it is used in large population-based studies.<sup>57</sup> We found no floor or ceiling effect for the EQ-5D-3L utility and VAS score.

The key messages from our study are that in general the PROMs can be used for evaluative purposes in groups of patients with severe and very severe COPD, but they are less suitable for assessing individual changes over time. Patients with severe and very severe COPD may experience significant fluctuations in daily symptoms without a clinical exacerbation. As a consequence, it has been suggested that agreement parameters of less stable measurements can be improved if the average of several measurements is used.<sup>46</sup> Thus, for individual patients, completion of consecutive questionnaires could be considered in the days or weeks before consultations or measurement time-points. This could feasibly be solved by using electronic surveys, although this potentially impacts the psychometric properties of questionnaires.<sup>58</sup> The agreement parameters of such a measurement procedure must therefore be investigated in future studies.

## Strength and Limitations

This study followed the guideline for reporting reliability and agreement studies (GRRAS), including reports on all relevant reproducibility domains, and in accordance to COSMIN recommendations, a moderate to good sample

size of 50 patients. We used a rigorous standardized methodological assessment approach, which included using the same conditions to reduce the effect of diurnal fluctuations in symptoms, the same rest intervals and order of questionnaires and functional tests, and a standardized instruction from trained raters. Furthermore, we reassured that patients were stable and did not have an exacerbation, defined by the Global Initiative for Chronic Obstructive Lung Disease as: “an acute worsening of respiratory symptoms that results in additional therapy”<sup>43</sup> during the reproducibility study. Retrospectively, it would have been valuable if we additionally had used the global rating scale between test and retest to ensure that the patients perceived themselves as stable. We cannot rule out that the functional tests performed before completion of the questionnaire may have influenced the reported symptoms at both visits. To limit any influence of dyspnea and fatigue we ensured that every patient felt rested and that oxygen saturation, heart rate and perceived dyspnea was fully normalized before the patients filled out the questionnaires. The disclosed limitations to restrict a possible recall bias are similar to those known from existing publications.<sup>11–18,24,26–28</sup> Finally, due to our inclusion criteria, our results cannot be generalized to all patients with COPD per se.

## Conclusion

In conclusion, the inter-day test–retest reliability of the CAT, CCQ, HADS and EQ-5D-3L were moderate to excellent. The agreement parameter SRD were smaller than the previously reported MICs for CAT and HADS, indicating that these PROMs on group level are suitable for evaluating changes over time in patients with severe and very severe COPD. In contrast to previous studies, we found that the CCQ was less suitable for assessing self-reported respiratory symptoms, because the SEM and SRD exceeded the previously reported MIC for CCQ total score. None of the PROMs were suitable for measuring individual changes over time.

## Data Sharing Statement

All relevant data are within the paper. Anonymous raw data will be available if application and requirements are approved by The Danish Data Protection Agency and the ethics committee of the capital region. Proposal for data use should be addressed to henrik.hansen.09@regionh.dk.

## Ethics Approval and Consent to Participate

The trial protocol was approved by the ethics committee of the capital region of Denmark (h-15019380) and the Danish Data Protection agency (jr. no.: 2012–58–0004).

## Acknowledgments

The authors would like to thank the patients for taking part in this study and all the raters who assisted with the blinded data collection. We thank statistician Thomas Kallemsø, Clinical Research Center, Copenhagen University Hospital Hvidovre for analytical support.

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

This work was supported by the Danish lung Foundation (charitable funding), Telemedical center regional capital Copenhagen (governmental funding), TrygFonden foundation (charitable funding).

## Disclosure

HH received personal grants from the Danish Lung Foundation (charitable funding), Telemedical Center Regional Capital Copenhagen (governmental funding), TrygFonden foundation (charitable funding). The grants covered expenses conducting the trial, salary and university fee for the PhD education. The authors report no other conflicts of interest in this work.

## References

- Spruit MA, Singh SJ, Garvey C, et al. An official American thoracic society/European respiratory society statement: key concepts and advances in pulmonary rehabilitation. *Am J Respir Crit Care Med.* 2013;188(8):e13–e64. doi:10.1164/rccm.201309-1634ST
- McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic obstructive pulmonary disease. *Cochrane Database Syst Rev.* 2015;2.
- Horton EJ, Mitchell KE, Johnson-Warrington V, et al. Comparison of a structured home-based rehabilitation programme with conventional supervised pulmonary rehabilitation: a randomised non-inferiority trial. *Thorax.* 2018;73(1):29–36. doi:10.1136/thoraxjnl-2016-208506
- Demeyer H, Louvaris Z, Frei A, et al. Physical activity is increased by a 12-week semiautomated telecoaching programme in patients with COPD: a multicentre randomised controlled trial. *Thorax.* 2017;72(5):415–423. doi:10.1136/thoraxjnl-2016-209026
- Bausewein C, Daveson BA, Currow DC, et al. EAPC white paper on outcome measurement in palliative care: improving practice, attaining outcomes and delivering quality services - recommendations from the European association for palliative care (EAPC) task force on outcome measurement. *Palliat Med.* 2016;30(1):6–22. doi:10.1177/0269216315589898
- De Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Health and quality of life outcomes minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes.* 2006;4:54. doi:10.1186/1477-7525-4-54
- De Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006;59(10):1033–1039. doi:10.1016/j.jclinepi.2005.10.015
- Kottner J, Audige L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Int J Nurs Stud.* 2011;48(6):661–671. doi:10.1016/j.ijnurstu.2011.01.016
- Mokkink LB, Boers M, van der Vleuten CPM et al. COSMIN risk of bias tool to assess the quality of studies on reliability and measurement error of outcome measurement instrument.; 2020.
- Mokkink LB, Boers M, van der Vleuten CPM, et al. COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med Res Methodol.* 2020;20(1):1–13. doi:10.1186/s12874-020-01179-5
- van der Molen T, Willemse BWM, Schokker S, Ten Hacken NHT, Postma DS, Juniper EF. Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health Qual Life Outcomes.* 2003;1:13. doi:10.1186/1477-7525-1-13
- Damato S, Bonatti C, Frigo V, et al. Validation of the clinical COPD questionnaire in Italian language. *Health Qual Life Outcomes.* 2005;3(1):1–7. doi:10.1186/1477-7525-3-9
- Ställberg B, Nokela M, Ehlers PO, Hjemdal P, Jonsson EW. Validation of the clinical COPD questionnaire (CCQ) in primary care. *Health Qual Life Outcomes.* 2009;7:1–9. doi:10.1186/1477-7525-7-26
- Papadopoulos G, Vardavas CI, Limperi M, Linardis A, Georgoudis G, Behrakis P. Smoking cessation can improve quality of life among COPD patients: validation of the clinical COPD questionnaire into Greek. *BMC Pulm Med.* 2011;11. doi:10.1186/1471-2466-11-13
- Antoniou SA, Puiu A, Zaharia B, Azoicai D. Health status during hospitalisations for chronic obstructive pulmonary disease exacerbations: the validity of the Clinical COPD Questionnaire. *Expert Rev Pharmacoecon Outcomes Res.* 2014;14:283–287. doi:10.1586/14737167.2014.887446
- Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD assessment test. *Eur Respir J.* 2009;34(3):648–654. doi:10.1183/09031936.00102509
- Al-Moamary MS, Al-Hajjaj MS, Tamim HM, Al-Ghobain MO, Al-Qahtani HA, Al-Kassimi FA. The reliability of an Arabic translation of the chronic obstructive pulmonary disease assessment test. *Saudi Med J.* 2011;32(10):1028–1033.
- Agustí A, Soler JJ, Molina J, et al. Is the CAT questionnaire sensitive to changes in health status in patients with severe COPD exacerbations. *COPD J Chronic Obstr Pulm Dis.* 2012;9(5):492–498. doi:10.3109/15412555.2012.692409
- Arbillaga-Etxarri A, Gimeno-Santos E, Barberan-Garcia A, et al. Long-term efficacy and effectiveness of a behavioural and community-based exercise intervention (Urban Training) to increase physical activity in patients with COPD: a randomised controlled trial. *Eur Respir J.* 2018;52(4):3. doi:10.1183/13993003.00063-2018
- Lipson DA, Barnhart F, Brealey N, et al. Once-daily single-inhaler triple versus dual therapy in patients with COPD. *N Engl J Med.* 2018;378(18):1671–1680. doi:10.1056/NEJMoa1713901

21. Maddocks M, Lovell N, Booth S, Man WDC, Higginson IJ. Series chronic obstructive pulmonary disease 2 palliative care and management of troublesome symptoms for people with chronic obstructive pulmonary disease. *Lancet*. 2017;390:988–1002. doi:10.1016/S0140-6736(17)32127-X
22. Hansen H, Bieler T, Beyer N, et al. Supervised pulmonary tele-rehabilitation versus pulmonary rehabilitation in severe COPD: a randomised multicentre trial. *Thorax*. 2020;75(5):413–421. doi:10.1136/thoraxjnl-2019-214246
23. Ringbaek T, Martinez G, Lange PA. Comparison of the assessment of quality of life with CAT, CCQ, and SGRQ in COPD patients participating in pulmonary rehabilitation. *COPD J Chronic Obstr Pulm Dis*. 2012;9(1):12–15. doi:10.3109/15412555.2011.630248
24. Tsiligianni IG, Van Der Molen T, Moraitaki D, et al. Assessing health status in COPD. A head-to-head comparison between the COPD assessment test (CAT) and the clinical COPD questionnaire (CCQ). *BMC Pulm Med*. 2012;12:1. doi:10.1186/1471-2466-12-20
25. Gupta N, Pinto LM, Morogan A, Bourbeau J. The COPD assessment test: a systematic review. *Eur Respir J*. 2014;44(4):873–884. doi:10.1183/09031936.00025214
26. Pinheiro Ferreira da Silva G, Tereza Aguiar Pessoa Morano M, Maria Sampaio Viana C, Bentes de Araujo Magalhães C, Delgado Barros Pereira E. Portuguese-language version of the COPD assessment test: validation for use in Brazil. *J Bras Pneumol*. 2013;39(4):402–408. doi:10.1590/S1806-37132013000400002
27. Kocks J, Tuinenga M, Uil S, van den Berg J, Ståhl E, van der Molen T. Health status measurement in COPD: the minimal clinically important difference of the clinical COPD questionnaire. *Respir Res*. 2006;7(1):62. doi:10.1186/1465-9921-7-62
28. Berkhof FF, Metzemaekers L, Uil S, Kerstjens H, van den Berg JW. Health status in patients with coexistent COPD and heart failure: a validation and comparison between the clinical COPD questionnaire and the minnesota living with heart failure questionnaire. *Int J Chron Obstruct Pulmon Dis*. 2014;9:999–1008. doi:10.2147/COPD.S66028
29. Sibilitz KL, Berg SK, Rasmussen TB, et al. Cardiac rehabilitation increases physical capacity but not mental health after heart valve surgery: a randomised clinical trial. *Heart*. 2016;102(24):1995–2003. doi:10.1136/heartjnl-2016-309414
30. Quist M, Langer SW, Lillilund C, et al. Effects of an exercise intervention for patients with advanced inoperable lung cancer undergoing chemotherapy: a randomized clinical trial. *Lung Cancer*. 2020;145:76–82. doi:10.1016/j.lungcan.2020.05.003
31. Davidson M, Keating J. Patient-reported outcome measures (PROMs): how should I interpret reports of measurement properties? A practical guide for clinicians and researchers who are not biostatisticians. *Br J Sports Med*. 2014;48(9):792–796. doi:10.1136/bjsports-2012-091704
32. Mokkink LB, De Vet HCW, Prinsen CA, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171–1179. doi:10.1007/s11136-017-1765-4
33. Hansen H, Bieler T, Beyer N, Godtfredsen N, Kallelose T, Frølich A. COPD online-rehabilitation versus conventional COPD rehabilitation – rationale and design for a multicenter randomized controlled trial study protocol (CORE trial). *BMC Pulm Med*. 2017;17(1):140. doi:10.1186/s12890-017-0488-1
34. Hansen H, Beyer N, Frølich A, Godtfredsen N, Bieler T. Intra- and inter-rater reproducibility of the 6-minute walk test and the 30-second sit-to-stand test in patients with severe and very severe COPD. *Int J Chron Obstruct Pulmon Dis*. 2018;13:3447–3457. doi:10.2147/COPD.S174248
35. Kon SSC, Canavan JL, Jones SE, et al. Minimum clinically important difference for the COPD assessment Test: a prospective analysis. *Lancet Respir Med*. 2014;2(3):195–203. doi:10.1016/S2213-2600(14)70001-3
36. Smid DE, Franssen FME, Houben-Wilke S, et al. Responsiveness and MCID estimates for CAT, CCQ, and HADS in patients with COPD undergoing pulmonary rehabilitation: a prospective analysis. *J Am Med Dir Assoc*. 2017;18(1):53–58. doi:10.1016/j.jamda.2016.08.002
37. Kon SSC, Dilaver D, Mittal M, et al. The clinical COPD questionnaire: response to pulmonary rehabilitation and minimal clinically important difference. *Thorax*. 2014;69(9):793–798. doi:10.1136/thoraxjnl-2013-204119
38. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the hospital anxiety and depression scale. *J Psychosom Res*. 2002;52:69–77. doi:10.1016/S0022-3999(01)00296-3
39. Puhan MA, Frey M, Büchi S, Schünemann HJ. The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease. *Health Qual Life Outcomes*. 2008;6:46. doi:10.1186/1477-7525-6-46
40. Brooks R, Rabin R, De Charro F. *The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective: Evidence from the EuroQol BIOMED Research Programme*. Netherlands: Springer; 2003.
41. Zanini A, Aiello M, Adamo D, et al. Estimation of minimal clinically important difference in EQ-5D visual analog scale score after pulmonary rehabilitation in subjects with COPD. *Respir Care*. 2015;60(1):88–95. doi:10.4187/respcare.03272
42. Nolan CM, Longworth L, Lord J, et al. The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. *Thorax*. 2016;71(6):493–500. doi:10.1136/thoraxjnl-2015-207782
43. Agusti A, Hurd S, Jones P et al. Global initiative for chronic obstructive lung; 2017. Available from: <http://goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/>. Accessed May 10, 2021.
44. Danish Society of Respiratory Medicine. Lungefunktionsstandard Spirometri Og Peakflow; 2007. Available from: <https://www.lungemedicin.dk/fagligt/klaringsrapporter/5-lfu-standard/file.html>. Accessed March 28, 2019.
45. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231–240. doi:10.1519/15184.1
46. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. Upper Saddle River, NJ: Prentice Hall; 2009.
47. Hopkins WG. Measures of reliability in sports medicine and science. *Sport Med*. 2000;30(1):1–15. doi:10.2165/00007256-200030010-00001
48. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med*. 1998;26(4):217–238. doi:10.2165/00007256-199826040-00002
49. Terwee CB, Roorda LD, Knol DL, De Boer MR, De Vet HCW. Linking measurement error to minimal important change of patient-reported outcomes. *J Clin Epidemiol*. 2009;62(10):1062–1067. doi:10.1016/j.jclinepi.2008.10.011
50. Comins JD, Brodersen J, Christensen KB, Jensen J, Hansen CF, Krogsgaard MR. Responsiveness, minimal important difference, minimal relevant difference, and optimal number of patients for a study. *Scand J Med Sci Sport*. 2020.
51. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307–310. doi:10.1016/S0140-6736(86)90837-8
52. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4(4):293–307. doi:10.1007/BF01593882
53. de Vet HCW, Terwee CB. The minimal detectable change should not replace the minimal important difference. *J Clin Epidemiol*. 2010;63(7):804–805. doi:10.1016/j.jclinepi.2009.12.015
54. Holland AE, Mahal A, Hill CJ, et al. Home-based rehabilitation for COPD using minimal resources: a randomised, controlled equivalence trial. *Thorax*. 2017;72(1):57–65. doi:10.1136/thoraxjnl-2016-208514

55. Chaplin E, Hewitt S, Apps L, et al. Interactive web-based pulmonary rehabilitation programme: a randomised controlled feasibility trial. *BMJ Open*. 2017;7(3):e013682. doi:10.1136/bmjopen-2016-013682
56. Spruit MA, Augustin IM, Vanfleteren LE, et al. Differential response to pulmonary rehabilitation in COPD: multidimensional profiling on behalf of the CIRO+ rehabilitation network. *Eur Respir J*. 2015;46:1625–1635. doi:10.1183/13993003.00350-2015
57. Sørensen J, Gudex C, Davidsen M, Brønnum-Hansen H, Pedersen KM. Danish EQ-5D population norms. *Scand J Public Health*. 2009;37(5):467–474. doi:10.1177/1403494809105286
58. White MK, Maher SM, Rizio AA, Bjorner JB. A meta-analytic review of measurement equivalence study findings of the SF-36<sup>®</sup> and SF-12<sup>®</sup> health surveys across electronic modes compared to paper administration. *Qual Life Res*. 2018;27(7):1757–1767. doi:10.1007/s11136-018-1851-2

### Patient Related Outcome Measures

Dovepress

### Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups.

The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>