

Inter-day Test-retest Reproducibility of the CAT, CCQ, HADS and EQ-5D in Patients With Severe and Very Severe COPD

Henrik Hansen (✉ henrik.hansen.09@regionh.dk)

Hvidovre Hospital Lungemedicinsk Afdeling <https://orcid.org/0000-0001-5464-4088>

Nina Beyer

Institute of Clinical Medicine, University of Copenhagen, Denmark

Anne Frølich

University of Copenhagen Department of Public Health: Københavns Universitet Institut for Folkesundhedsvidenskab

Nina Godtfredsen

Department of Respiratory Medicine, Hvidovre hospital, Denmark

Theresa Bieler

Department of Physical and Occupational Therapy

Research

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

Posted Date: November 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-112096/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

In patients with COPD the COPD Assessment Test (CAT), Clinical COPD Questionnaire (CCQ), Hospital Anxiety and Depression Scale (HADS) and EuroQol 5D (EQ-5D) are widely used patient reported outcome measures (PROMs) of respiratory symptoms, anxiety, depression and quality of life. Despite established responsiveness and minimal important change (MIC), 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 in patients with severe and very severe COPD (FEV₁ <50%) eligible for hospital-based pulmonary rehabilitation.

Patients and Methods

Fifty patients (22 females, mean [SD] age 67 [9] yrs.; FEV₁ 32[9] %; 6-minute walk distance 347 [102] meters; CAT 21 [6] points; BMI: 26 [6] kg/m²) completed the questionnaires (CAT, CCQ, HADS, EQ-5D) 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_{95CI}: 0.80) for CAT; 0.69 (LL_{95CI}: 0.46) for CCQ; 0.86 (LL_{95CI}: 0.75) and 0.90 (LL_{95CI}: 0.82) for HADS-anxiety (A) and depression (D) and 0.87 (LL_{95CI}: 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 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 questionnaires presented moderate to excellent inter-day test-retest reliability and acceptable agreement e.g. SEM and SRD below the established MIC on group level except for the CCQ questionnaire. No floor or ceiling effect of relevance was documented for the questionnaires.

Background

Introduction

In chronic obstructive pulmonary disease (COPD), patient reported outcome measures (PROM) of respiratory symptoms, other symptoms (e.g. anxiety) and health-related quality of life are increasingly used as descriptive instruments or as effect outcome measures¹⁻⁴. Since COPD is an incurable disease with increasing symptoms as the disease progresses, symptom relief is a warranted core outcome in COPD care, including pulmonary rehabilitation (PR)^{1,5}. In addition, the use and importance of PROM as critical effect outcomes are being endorsed by health authorities and scientific societies⁵. An essential requirement of effect outcomes measures is that they are valid and reproducible⁶. Nevertheless, the reproducibility and notably measurement errors of some commonly used PROM to evaluate e.g. PR have only been sparsely reported⁷⁻¹⁴.

Reproducibility concerns the degree to which repeated measurements provide similar results in a specific population⁶. 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^{15,16}. Agreement parameters are preferable when the instrument, e.g. PROM, is used for evaluating changes over time, because they indicate systematic and random errors of patient scores not attributed to true changes in the constructs to be measured¹⁵⁻¹⁷. The COnsensus based Standards for the selection of health Measurement INstruments guideline (COSMIN) recommends that for continuous scores agreement parameters, i.e. the standard error of measurement (SEM), limits of agreement (LOA) or smallest detectable change (SDC) be calculated and reported^{17,18}. A variety of PROMs are being used in all types of study designs related to COPD as well as in clinical practice^{2-4,19-22}. St. George Respiratory Questionnaire (SGRQ) is considered the gold-standard questionnaire covering patients self-reported respiratory symptoms²³. 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 interpret²³. Both CAT and CCQ have proved excellent concurrent validity with SGRQ^{7,12,23-25}. Reliability for CAT and CCQ questionnaires has been reported in several studies concerning patients with COPD, and the 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^{7,8,27,28,9-14,24,26}. Three studies have investigated agreement parameters for CCQ, and reported SEM ranging from 0.10 to 0.21 points for the total score^{24,27,28}, and one study²⁸ reported a 95% LOA from -1.87 to 1.35 points. Regarding CAT only one study has reported agreement parameters, i.e. SEM of 1.92 points, mainly in patients with mild to moderate airflow obstruction, low symptom score and high walking capacity²⁴.

Other symptoms that are frequently reported in COPD are anxiety and depression. The Hospital Anxiety and Depression Scale (HADS) questionnaire is generic and widely used across medical conditions. Among patients with COPD, HADS is used for both symptom screening and evaluation of changes in symptoms following an intervention^{3,22,29,30}. The validity, responsiveness and minimal important change (MIC) for HADS is well established in patients with COPD^{27,31–33}, while no studies have reported reliability and agreement parameters for the HADS in patients with COPD. Likewise, for the widely used generic questionnaire EuroQol 5D (EQ-5D) to assess health related quality of life, we were unable to find any study concerning the reproducibility in patients with COPD.

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 (e.g. respiratory symptoms) is stable^{34,35}. 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 a period might allow patients to recall their earlier responses and a too long period might allow a true change in the status of the patient^{17,34}.

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

Materials 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^{36,37}. We followed the Guideline for Reporting Reliability and Agreement Studies (GRRAS)¹⁶.

Participants

Eligible patients for the RCT were identified and recruited by respiratory nurses during out-patient 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).

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. A consecutive convenience sample size of 50 patients was chosen according to the recommendation from COSMIN (supplement 1 - flowchart)¹⁷.

Inclusion and exclusion criteria³⁶ 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 to FVC < 0.70 ; $FEV_1 < 50\%$; MRC ≥ 2 ; who had not participated in PR within the prior six months³⁶.

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. The patients completed the questionnaires in a pause between two sets of performance tests, i.e. the six-minute walk test and the 30-second sit-to-stand test (Figure 1). Ten raters administered the questionnaires. They were familiar with the questionnaires from clinical practice and had obtained accreditation to be raters. 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 had no influence on the second administration, patients and raters were blinded to the previous response, 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.

Assessment/Test procedures

The raters followed the same procedures (Figure 1), 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. CAT, CCQ, HADS and EQ-5D 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 pre-instruction 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 every-day 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 symptoms¹². It is an 8-item questionnaire where each item scores from 0 to 5 points (0 indicating no impact or symptoms, 5 worst possible impact or symptoms) summing up to a total CAT score range of 0–40 points¹².

Clinical COPD Questionnaire (CCQ) assesses self-reported quality of life⁷. The CCQ consist of 10-items with a total score and 3-domain scores: Symptoms (4-Items), Functional state (4-Items) and Mental state (2-Items). Total- and domain scores range from 0 to 6 (0 = no impairment)⁷.

Hospital Anxiety and Depression Scale (HADS) assesses the level of anxiety and level of depression in medically ill persons³⁸. The scale consists of two sub scales HADS anxiety (HADS-A) and HADS depression (HADS-D), each of which has seven questions with four possible answers (score range 0 to 3). A total subscale scores 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 disorder³⁸.

EuroQol 5-Dimension Questionnaire (EQ-5D), is a generic global questionnaire measuring health-related quality of life³⁹. 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). Each dimension has three levels (no problem, some problem, severe problem), comprising a total of 243 utility scores ranging from -0.624 (worst possible health utility) to 1.0 (best possible health utility). The EQ-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).³⁹

Demographic and descriptive variables

Demographic and descriptive variables, i.e. age, gender, body mass index, smoking status, FEV₁/FVC, FEV₁, GOLD, A/B/C/D stratification⁴⁰, Charlson Comorbidity Index, BODE-index and oxygen supplement were registered at T1⁴¹.

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. Paired t-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_{1,1} model was used because the assessments were conducted at five centers, and all raters did not instruct each patient^{15,42}. The ICC_{1,1} is a fixed model addressing both systematic and random error. ICCs values between 0–0.49 were considered weak, ≥0.50–0.75 moderate, >0.75–0.90 good and >0.90 excellent reliability⁴³.

Agreement was calculated as standard error of measurement (SEM) and the SEM₉₅ using the equation $SD \cdot \sqrt{1-ICC}$ and respectively $1.96 \times SEM$ (SEM₉₅)^{15,42}. The SEM expresses the measurement error that occur within a single measurement where no real change has occurred and indicates that there is a 68% likelihood that a group of patients' (or a single patient's SEM₉₅) "true" score is within this measurement error^{35,43}.

The corresponding smallest real difference (SRD) was calculated by the equation $\sqrt{2} \times SEM$ (SRD) and $1.96 \times \sqrt{2} \times SEM$ (SRD₉₅) respectively. The SRD represents the smallest real difference to be detected beyond the measurement error of repeated measurement without real change and with a 68% certainty on a group of patients (or a single patient's SRD₉₅)^{42,44,45}. The SEM, SEM₉₅, SRD and SRD₉₅ are presented in actual units. 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).

The minimal important change (MIC) is derived from longitudinal validity studies and preferably determined by anchor-based methods, represents the smallest amount of change in an outcome that might be considered important by the patient or clinician^{35,46}. For evaluative purpose it is important that the MIC can be distinguish from repeated measurement error. Therefore, we determined a questionnaire suitable for evaluative use when the SRD was smaller than the MIC.

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 $mean \pm 1.96 \cdot SD$ and included in the plots^{15,47}. 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) score⁴⁸. 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

Fifty of the 108 eligible patients agreed to participate in the reproducibility study. Twenty-three declined to participate due to the extra testing date, while 35 patients could not be included because they undertook the baseline assessment for the RCT less than one week before the scheduled randomization and intervention. The 58 patients who did not participate in the reproducibility study did not differ significantly from the included patients at baseline (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 ($ICC_{1,1}$) 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. Agreements 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, respectively. The Bland Altman plots with 95% limits of agreement for the questionnaires are shown in Figure 2 A to E. There was no significant difference between scores at T1 and T2 for any of the PROMs (Table 2 and figure 2A-E). 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 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, HADS and EQ-5D-VAS suggesting that they can be used for group evaluative purpose in patients with severe and very severe COPD⁴⁹.

CAT

In line with previous results (ICC ranging from 0.80 to 0.94)^{12-14,24,26} we found good reliability for the CAT in patients with severe end very severe COPD. To our knowledge the study by Tsiligianni et al.²⁴ is the only study that have reported agreement parameters for the CAT in patients with COPD. Although the patients had less symptoms (median CAT score 13 points), less disease severity (65% GOLD group I or II) and milder risk profile (BODE index ≤ 2 points), the agreement parameters (SEM: 1.9 points; LOA 95%: -8.0; 12.0) 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 (with 68% confidence), respectively 5.7 points (with 95% confidence), 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 2 to 3.8 points^{31,50}. We found that the SRD for the CAT is lower than the previously reported MIC, and this suggests that the MIC can only 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_{95} of 5.8 point 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

We found moderate reliability (ICC: 0.69) of the CCQ total score, 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^{7,8,27,28,9-14,24,26}. Similarly, we found SEM (0.5 point) in the higher end than previously reported (SEM respectively reported as 0.2²⁷, 0.4²⁴ and 0.6²⁸). However, it must be noted that the SEM of 0.2 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.²⁸ reported LOA (mean difference of -0.3 with a 95%LOA from -1.9 to 1.4)²⁸, which is very similar to our results. Similar to our study, the study by Berkhoff et al.²⁸ collected data based on comparable routine inclusion criteria, had similar sample size, comparable baseline CCQ scores and involved patients with multimorbidity as

most patients had ≥ 2 comorbidities. The study only differed from ours regarding FEV1% predicted mean, which was 51.0 (15.0) in the Berkhoff study²⁸ and 32.3 (9.0) in our study. There were no floor or ceiling effects in any scores for CCQ.

The reported MIC for the CCQ total score is 0.4 in patients with moderate to severe COPD^{50,51}. Our results for the SEM (0.5 point) and SRD (0.7 point) in patients with severe and very severe COPD (Table 2) suggest that the established MIC cannot be distinguished from single and repeated measurement error. Thus, we propose cautiousness for the use of the CCQ to evaluate changes over time on both group and individual level in patients with severe and very severe COPD. Based on our results the CCQ seems the less suitable compared to the CAT questionnaire for assessing self-reported respiratory symptoms.

HADS and EQ-5D

Both HADS and EQ-5D are commonly used outcomes in clinical research^{3,19,22,52–54}, clinical practice⁵⁵ and for public health evaluative purposes⁵⁶.

To our knowledge this is the first study to investigate the reproducibility of the HADS in patients with COPD. We found that the HADS questionnaire showed good reliability in patients with severe and very severe COPD. The agreement parameters SEM of the HADS-A (1.3 point) and HADS-D (0.9 point) and the SRD of the HADS-D (1.3 point) are below the established MIC of 1.5 point³². These results indicate that the questionnaire can discriminate a clinically relevant change from measurement error on group level and thus is suitable for evaluative purposes in a group of patients with severe COPD. The SRD₉₅ of 3.7 point (HADS-A) and 2.5 point (HADS-D) is greater than the MIC suggesting that the HADS questionnaire must be considered less suitable for individual screening and to evaluate 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 severe COPD.

Like for the HADS questionnaire, we could not find any study that have investigated the reproducibility of the EQ-5D questionnaire in patients with COPD. We found that EQ-5D-VAS showed good reliability in patients with severe and very severe COPD. The SEM (6.8 point) was below the established MIC of 8.0 points³³, while the repeated measurement error SRD (9.7 point) exceeded the MIC. This finding indicates some cautiousness for the use of the EQ-5D-3L questionnaire for evaluative purposes, i.e. it is mostly suitable for measuring group changes and in large population-based studies⁵⁶. We found no floor or ceiling effect for the EQ5D 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 on an individual level. Patients with severe and very severe COPD can experience significant fluctuations in daily symptoms without a clinical exacerbation, and it has been suggested that agreement parameters of less stable measurements can be improved if the average of several measurements is used⁴³. Therefore, completion of consecutive questionnaires could be considered in the days or weeks before e.g. control consultations or measurement time-points of a single patient. This could feasible be solved by electronic surveys, although this potentially impacts psychometric properties for the questionnaires⁵⁷. In addition, the agreement parameters of such a measurement procedure have to be investigated in a future study.

Strength And Limitations

This study followed the guideline for reporting reliability and agreement studies (GRRAS), including reports on all relevant reproducibility domains, and a recommended sufficient sample 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 calibrated raters. Furthermore, we reassured that patients were stable and did not have an exacerbation, defined by the Global Initiative for Chronic Obstructive Lung as: *“an acute worsening of respiratory symptoms that results in additional therapy”*⁴⁰ 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^{7,8,27,28,9–14,24,26}.

Conclusion

In conclusion, this is the first reproducibility study to provide full data on all reliability and agreement properties for the commonly used questionnaires CAT, CCQ, HADS and EQ-5D in patients with severe and very severe COPD. The inter-day test-retest reliability of the CAT, CCQ, HADS and EQ-5D were moderate to excellent. In contrast to previous studies this study found the CCQ as the less suitable compared to the CAT questionnaire for assessing self-reported respiratory symptoms, because the SEM and SRD on group level exceeded the previously reported MIC for CCQ total score. Agreement parameters SEM and SRD for CAT, HADS and EQ-5D were smaller than the previously reported MICs indicating that

these PROMS are suitable for evaluating changes over time in a group of patients with severe and very severe COPD. However, they are suboptimal for measuring individual changes over time.

Abbreviations

COPD: chronic obstructive pulmonary disease; PROM: patient reported outcome measures; PR: pulmonary rehabilitation; COSMIN: COnsensus based Standards for the selection of health Measurement INstruments guideline; SEM: standard error of measurement; LOA: limits of agreement; SDC: smallest detectable change; SRD: smallest real difference; SGQR: St. George Respiratory Questionnaire; CAT: COPD Assessment Test; CCQ: COPD Clinical Questionnaire; HADS: Hospital Anxiety and Depression Scale; EQ-5D: EuroQol 5D; ICC: intraclass correlation coefficient; MIC: minimal important change; FVC: forced vital capacity; FEV1: 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; SpO₂: arterial oxygen saturation as measured by pulse oximetry; dyspnea: perceived dyspnea (Borg cr-10); 30sec-STS, 30 seconds sit-to-stand chair test; GRRAS: Guideline for Reporting Reliability and Agreement Studies;

Declarations

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).

Consent for publication

Not applicable

Availability of data and materials

Proposal for data use should be addressed to henrik.hansen.09@regionh.dk. Data access in Denmark are under very strict juristic data protection law. Any possible access or sharing demands a part application to; (1) Danish Data Protection agency, (2) ethics committee of the capital region, (3) national health Data authorities. Only if the applications are approved data will be considered available for sharing. The authors will not be able to support this process and a prolonged process must be expected.

Funding

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

Competing interest

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.

Author Contribution

Concept and Design of study: all authors; Acquisition of Data: HH and blinded personnel; Analysis of data: HH, TB, NG; Drafting of Manuscript: HH; Revision of manuscript critically for important intellectual content: all authors; Approval of final manuscript: all authors.

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 Kalleose, Clinical Research Center, Copenhagen University Hospital Hvidovre for analytical support.

References

1. 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). doi:10.1164/rccm.201309-1634ST
2. McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic obstructive pulmonary disease (Review). *Cochrane Libr* 2015. 2015;(2). doi:10.1002/14651858.CD003793.pub3.
3. 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

4. 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*. January 2017;thoraxjnl-2016-209026. doi:10.1136/thoraxjnl-2016-209026
5. 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
6. 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
7. 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
8. Damato S, Bonatti C, Frigo V, et al. Validation of the Clinical COPD questionnaire in Italian language. *Health Qual Life Outcomes*. 2005;3(1):9. doi:10.1186/1477-7525-3-9
9. Ställberg B, Nokela M, Ehre P-O, Hjemdal P, Wikström Jonsson E. Health and Quality of Life Outcomes Validation of the Clinical COPD Questionnaire (CCQ) in primary care. 2009. doi:10.1186/1477-7525-7-26
10. 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
11. Antoniu SA, Puiu A, Zaharia B, Azoicai D, Antoniu SA, Puiu A. Health status during hospitalisations for chronic obstructive pulmonary disease exacerbations: the validity of the Clinical COPD Questionnaire Health status during hospitalisations for chronic obstructive pulmonary disease exacerbations: the validity of. 2014;7167(November 2015). doi:10.1586/14737167.2014.887446
12. 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
13. 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). www.smj.org.sa. Accessed August 13, 2019.
14. 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
15. 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
16. 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
17. Mokkink LB, De Vet · H C W, Prinsen · C A C, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. 2018;27:1171-1179. doi:10.1007/s11136-017-1765-4
18. Mokkink LB, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. 2019;(July).
19. 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
20. 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 WD-C, Higginson IJ. *Series Chronic Obstructive Pulmonary Disease 2 Palliative Care and Management of Troublesome Symptoms for People with Chronic Obstructive Pulmonary Disease*. Vol 390.; 2017. www.thelancet.com. Accessed August 12, 2019.
22. Hansen H, Bieler T, Beyer N, et al. Supervised pulmonary tele-rehabilitation versus pulmonary rehabilitation in severe COPD: a randomised multicentre trial. *Thorax*. March 2020;thoraxjnl-2019-214246. doi:10.1136/thoraxjnl-2019-214246
23. Ringbaek T, Martinez G, Lange P. A 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. 2014:873-884. 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. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4075870/pdf/1806-3713-jbpneu-39-04-0402.pdf>. Accessed August 13, 2019.

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. doi:10.2147/COPD.S66028
29. Sibillitz 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, Lillelund 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. 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
32. 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
33. 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
34. Mokkink Cecilia AC Prinsen Donald L Patrick Jordi Alonso Lex M Bouter LB, Mokkink CL. *COSMIN Study Design Checklist for Patient-Reported Outcome Measurement Instruments*; 2019. www.cosmin.nl. Accessed May 14, 2020.
35. 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
36. Hansen H, Bieler T, Beyer N, Godtfredsen N, Kallemsø T, Frølich A. COPD online-rehabilitation versus conventional COPD rehabilitation – rationale and design for a multicenter randomized controlled trial study protocol (COPRe trial). *BMC Pulm Med.* 2017;17(1):140. doi:10.1186/s12890-017-0488-1
37. 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;Volume 13:3447-3457. doi:10.2147/COPD.S174248
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. Brooks R, Rabin R CF. *Health Measurement and Valuation of Health Status Using EQ-5D: A European Perspective: Evidence from the EuroQol BIOMED Research Programme.* Springer Netherlands; 2003.
40. Agusti A, Hurd S, Jones P, Fabbri LM, Martinez F VC et al. *Global Initiative for Chronic Obstructive Lung Disease*; 2017. <http://goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/>.
41. Danish Society of Respiratory Medicine. *Lungefunktionsstandard Spirometri Og Peakflow*; 2007. <https://www.lungemedicin.dk/fagligt/klaringsrapporter/5-lfu-standard/file.html>. Accessed March 28, 2019.
42. Weir JP. Quantifying Test-Retest Reliability Using the Intraclass Correlation Coefficient and the SEM. *J Strength Cond Res.* 2005;19(1):231-240. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.457.2590&rep=rep1&type=pdf>. Accessed July 24, 2018.
43. Portney LG WM. *Foundations of Clinical Research: Applications to Practice*, 3rd editio. New Jersey: Prentice Hall; 2008. doi:10.1016/s0039-6257(02)00362-4
44. Hopkins WG. Measures of Reliability in Sports Medicine and Science. *Sport Med.* 2000;30(1):1-15.
45. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 1998;26(4):217-238. <http://www.ncbi.nlm.nih.gov/pubmed/9820922>. Accessed July 24, 2018.
46. 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
47. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (London, England).* 1986;1(8476):307-310. <http://www.ncbi.nlm.nih.gov/pubmed/2868172>. Accessed July 24, 2018.
48. 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
49. 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
50. 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

51. 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
52. 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
53. 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
54. 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
55. 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
56. 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
57. White MK, Maher SM, Rizio AA, Bjorner JB. A meta-analytic review of measurement equivalence study findings of the SF-36® and SF-12® Health Surveys across electronic modes compared to paper administration. *Qual Life Res*. 2018;27(7):1757-1767. doi:10.1007/s11136-018-1851-2

Tables

Table 1 Characteristics		
	Included	Not included
Sex, men/women (n)	28/22	21/37
Age, years (SD)	66.6±9.0	69.4±9.1
Body mass index, mean kg ·m ⁻² , (SD)	25.4±5.6	25.8±5.6
FEV ₁ % predicted, mean (SD)	32.3±9.0	35.1±9.4
FEV ₁ /FVC, 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 (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
<i>CCQ-symptoms, mean, (SD)</i>	<i>2.88±1.00</i>	<i>2.75±1.02</i>
<i>CCQ-functional, mean (SD)</i>	<i>2.88±1.16</i>	<i>2.71±1.19</i>
<i>CCQ-Mental, mean (SD)</i>	<i>2.96±1.48</i>	<i>2.48±1.56</i>
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
<p>Notes: Data are presented as mean ±standard deviation and median (range) or percent in non-normally distributed variables. Any statistically significant difference between included vs. not-included participants denoted *<i>p</i><0.05</p> <p>Abbreviations: FEV₁, 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.</p>		

Table 2 Inter-day test-retest reproducibility									
Variables	Test-day 1 (T1)	Test-day 2 (T2)	Difference	Floor/Ceiling (n)	ICC _{1,1} (LL ₉₅)	SEM (SEM%)	SEM ₉₅ (SEM ₉₅ %)	SRD (SRD%)	SRD ₉₅ (SRD ₉₅ %)
CAT total, <i>points</i>	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, <i>points</i>	2.90±0.92	2.72±0.82	-0.18 [-0.41; 0.06]	0/0 0/0 2/3 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			0.71 (0.49)		1.00 (36)	0.72 (26)	1.44 (51)
CCQ-functional	2.88±1.16	2.78±1.08	-0.18 [-0.45; 0.08]		0.78 (0.62)	0.51 (19)	1.02 (36)	0.74 (26)	2.47 (88)
CCQ-Mental	2.96±1.48	2.65±1.22	-0.10 [-0.37; 0.17]		0.57 (0.25)	0.52 (19)	1.76 (63)	1.27 (45)	1.33 (47)
			-0.31 [-0.73; 0.11]			0.90 (32)			
HADS-A, <i>points</i>	5.72±3.63	5.80±3.52	-0.10 [-0.64; 0.80]	1/0 0/0	0.86 (0.75)	1.33 (23)	2.61 (45)	1.88 (32)	3.69 (64)
HADS-D, <i>points</i>	4.62±3.00	4.36±2.69	-0.20 [-0.76; 0.24]		0.90 (0.82)	0.90 (20)	1.76 (39)	1.27 (28)	2.49 (55)
EQ-5D VAS, <i>score</i>	49.22±19.5	50.28±18.64	1.10 [-2.69; 4.81]	0/1 0/3	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]		0.77 (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%]. Significant difference between test-days is denoted as *P<0.05.

Abbreviations: Floor, lowest score; Ceiling, highest score; n, number; ICC_{1,1}, 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; SEM₉₅, standard error of the measurement at the 95% confidence level; SEM₉₅%, 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; SRD₉₅, smallest real difference at the 95% confidence level; SRD₉₅%, smallest real difference as a percentage of the mean; CAT, COPD Assessment Test; ; CCQ, Clinical COPD Questionnaire; HADS-A and -D, Hospital Anxiety and Depressions Scale (HADS); EQ-5D VAS, EuroQol 5-Dimension Questionnaire Visual Analogue Scale; EQ-5D-3L Utility, EuroQol 5-Dimension 3-likert utility score.

Figures



Assessment and progression:

1. Subject history/introduction, while seated: resting blood pressure, resting heart rate, resting SpO₂, resting dyspnea. Standing: anthropometric measures, i.e. weight and height (30 minutes)
2. Instruction and performing 6MWT, end-heart rate, end-SpO₂, 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 SpO₂, resting dyspnea (5 minutes)
7. Instruction and performing 6MWT, end-heart rate, end-SpO₂, 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: SpO₂, 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 Depressions Scale (HADS); EQ-5D-3L, EuroQol 5-Dimension 3-likert utility score and VAS score.

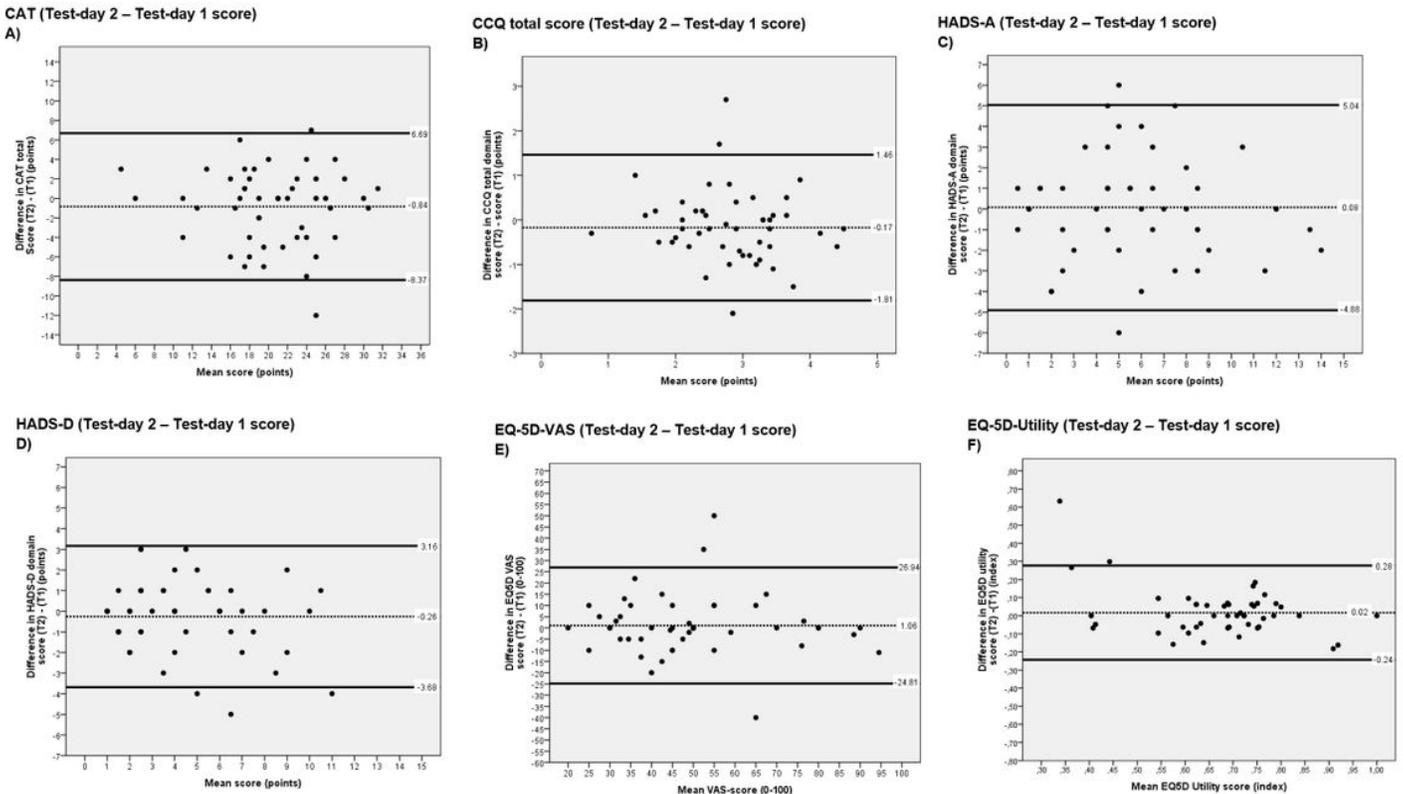


Figure 2

Bland and Altman plots of the CAT, CCQ-total, HADS-A, HADS-D, EQ5D-VAS and EQ5D-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 Depression Scale (HADS); EQ-5D VAS score, EuroQol 5-Dimension Questionnaire Visual Analogue Scale; EQ-5D-3L Utility, EuroQol 5-Dimension 3-likert utility score.