



Development and first validation of the COPD Assessment Test

P.W. Jones*, G. Harding#, P. Berry[†], I. Wiklund[†], W-H. Chen# and N. Kline Leidy#

ABSTRACT: There is need for a validated short, simple instrument to quantify chronic obstructive pulmonary disease (COPD) impact in routine practice to aid health status assessment and communication between patient and physician. Current health-related quality of life questionnaires provide valid assessment of COPD, but are complex, which limits routine use.

The aim of the present study was to develop a short validated patient-completed questionnaire, the COPD Assessment Test (CAT), assessing the impact of COPD on health status.

21 candidate items identified through qualitative research with COPD patients were used in three prospective international studies (Europe and the USA, n=1,503). Psychometric and Rasch analyses identified eight items fitting a unidimensional model to form the CAT. Items were tested for differential functioning between countries. Internal consistency was excellent: Cronbach's $\alpha=0.88$. Test re-test in stable patients (n=53) was very good (intra-class correlation coefficient 0.8). In the sample from the USA, the correlation with the COPD-specific version of the St George's Respiratory Questionnaire was $r=0.80$. The difference between stable (n=229) and exacerbation patients (n=67) was five units of the 40-point scale (12%; $p<0.0001$).

The CAT is a short, simple questionnaire for assessing and monitoring COPD. It has good measurement properties, is sensitive to differences in state and should provide a valid, reliable and standardised measure of COPD health status with worldwide relevance.

KEYWORDS: Chronic obstructive pulmonary disease, health status assessment, quality of life, questionnaire, reliability, validity

Chronic obstructive pulmonary disease (COPD) is among the leading causes of morbidity and mortality worldwide, with ~80 million people worldwide estimated to have moderate to severe COPD [1]. COPD is characterised by progressive, irreversible limitation of airflow, and a major goal of its treatment is to ensure that the patient's health is optimised. However, despite the availability of clinical guidelines to manage COPD, such as the Global Initiative for Chronic Obstructive Lung Disease (GOLD), there is continued evidence to suggest that a substantial proportion of patients are not achieving the level of treatment success that may be possible [2, 3]. In addition to routine clinical evaluations, a critical step in management is to obtain, from the patient, reliable and valid information on the impact of COPD on their health status. This would include information on daily symptoms, activity limitation and other manifestations of the disease. A standardised patient-centred assessment tool, covering key attributes of COPD health, should facilitate

information gathering and improve communication between patient and clinician. In addition to an overall score, an ideal tool should be able to identify specific areas of greater severity to serve as a focal point for targeted management or the evaluation of management goals, thereby improving both the process and the outcome of care. Available disease-specific health status measures, such as the St George's Respiratory Questionnaire (SGRQ) [4], Chronic Respiratory Disease Questionnaire (CRQ) [5], and the COPD Clinical Questionnaire (CCQ) [6], are reliable, valid, and widely used in clinical trials, or in clinical practice (CCQ). However, some are lengthy and have scoring algorithms that are too complex for routine use in clinical practice. A brief tool that is easy to complete and interpret could be more readily incorporated into routine care.

The studies described here detail the methods used to develop a simple, reliable instrument (the COPD Assessment Test (CAT)) with good measurement properties and with a target number of five to

AFFILIATIONS

*Division of Cardiac and Vascular Science, St George's University of London, and

[†]Global Health Outcomes, GlaxoSmithKline, London, UK.

#Center for Health Outcomes Research, United BioSource Corporation, Bethesda, MD, USA.

For the members of the COPD Assessment Test working group, please see the Acknowledgements section.

CORRESPONDENCE

P.W. Jones

Division of Cardiac and Vascular Science

St George's University of London Cranmer Terrace

London

SW17 0RE

UK

E-mail: pjonas@sgul.ac.uk

Received:

June 30 2009

Accepted after revision:

July 24 2009

This article has online supplementary material available from www.erj.ersjournals.com

seven items from a previously identified pool of 21 items [7]. The first tests of the validity of this new instrument are also presented.

METHODS

Background

The items used to create the CAT were generated in a recent qualitative study [7], based on interviews and focus groups with COPD patients supported by interviews with community physicians and pulmonologists. The study explored the aspects of COPD which were most important in defining patients' health. Identified items included dyspnoea, cough, sputum production and wheeze, as well as systemic symptoms of fatigue and sleep disturbance. Additional indicators included limitations in daily activities, social life, emotional health and feeling in control, along with the use of rescue medication. A draft framework and 21 draft items were developed. Items were formatted as a semantic differential six-point scale, defined with contrasting adjectives. This draft instrument was reviewed by a panel of experts for clinical relevancy. Cognitive debriefing interviews with COPD patients indicated that all items and the format in which they were presented were clear and easy to understand.

Study design

A structured approach to item reduction was used, with a priority placed on reliable measurement properties and an absence of bias due to demographic factors, such as sex and country and/or language. The item reduction process and first evaluation of psychometric properties were based on data from three observational prospective studies in COPD patients from Belgium, France, Germany, the Netherlands, Spain and the USA. Study 1 included stable patients (no exacerbation in the previous 12 weeks) recruited from the USA only. These patients contributed to the item reduction and initial validation studies. Study 2 included patients from the USA with a clinician-confirmed exacerbation at the time of study and contributed to the validation studies. Both studies recruited from primary care and pulmonary clinics. Patients were seen at the clinic twice; normally at baseline and after 2 weeks, although a sub-sample from study 1 completed the second clinic visit after 7 days to assess reproducibility (test-retest). Study 3 included COPD patients from Germany, France, the Netherlands, Spain and Belgium who participated in an European Union (EU) COPD Quality of Life (QoL) Survey Study, a cross-sectional, epidemiological, nonrandomised survey among patients with COPD who had consecutively visited a general practitioner and were invited to participate in a single-visit survey, which included the draft 21 CAT item pool and other measures of disease severity.

The inclusion criteria were: smokers or ex-smokers with a smoking history >10 pack-yrs, aged 40–80 yrs, with a current diagnosis of COPD and forced expiratory volume in 1 s (FEV₁) to forced vital capacity (FVC) ratio ≤70%. Current asthma or significant comorbidity were exclusion criteria. Enrolment for the USA studies was stratified to yield a distribution of patients by GOLD stage: mild (15%), moderate (35%), severe (35%) and very severe (15%). For the EU study, patients with a baseline (post-bronchodilator) FEV₁/FVC ratio ≤70% in the 6 months prior to survey visit were eligible. All three studies excluded patients with asthma as a primary diagnosis or other

active chronic respiratory disease requiring treatment, intervention, or diagnostics. Patients with severe or uncontrolled comorbidities were excluded. All enrolled patients provided written informed consent prior to study procedures.

Statistical analyses

The analysis plan was finalised prior to availability of study data. Analyses conducted were consistent with traditional psychometric theory [8]. In developing the analysis plan, the primary objective was to create a questionnaire made up of the smallest number of items that formed a unidimensional instrument with reliable measurement properties. The process of identifying items for potential deletion was iterative, based on a hierarchical process: 1) age and sex bias; 2) percent of missing responses; 3) floor and ceiling effects; 4) item to total correlation; and 5) tests of redundancy (inter-item correlation). Item response theory (IRT) using Rasch analysis was then used to identify items with the best fit to a unidimensional model and items to be removed due to differential functioning between countries.

Items flagged by item analysis and IRT were potential candidates for item reduction. Item reduction also took into consideration the previously reported qualitative analysis (focus groups and content validity, *etc.*) [7] and consultation from clinical research experts. Once the item reduction process was complete, exploratory factor analysis was conducted and the finalised CAT was examined for reliability and preliminary validation.

Item analysis

Data from the USA (stable) and EU subjects were used to examine the distributional characteristics of the 21 individual items, by country. An item was flagged for potential exclusion if it demonstrated a high missing rate (suggesting that patients had difficulty responding to the item), a floor effect (>25% of patients indicating that they did not experience the symptom or health state), or a ceiling effect (>25% responding to the most severe impact of the item on the scale). Items were also flagged when the item to total score correlation was low, suggesting little contribution to the overall score, or when the inter-item correlation was greater than 0.70, indicating that the items were similar (*i.e.* that one of them was potentially redundant).

Each item was examined for potential bias using the "criterion keying" method [9]. This method examines the association of the items with external criteria, as well as with a bias factor. An item was also flagged for potential exclusion if it had a low association with the external criterion of clinician rating of severity of COPD (mild, moderate, severe and very severe). Bias in responses due to sex and age were tested in the opposite way. An item was flagged for potential exclusion if it had a high association with sex or age, which is some indication of bias.

Rasch analysis

Rasch analysis was conducted to examine whether each item exhibited Guttman scaling properties; for an item of given severity, the patient is more likely to respond to other less severe items and less likely to respond to items of greater severity. To form a reliable measurement instrument, all items

should form a unidimensional construct (*i.e.* the response scale to all the items should measure COPD severity but represent different degrees of severity). Rasch models also assume that all items have uniform discrimination power between high and low severity.

The results were assessed in order to test both the quality of fit of the individual items to a unidimensional model and the quality of the fit of all of the items taken together to the model. The fit statistics approximate to the Chi-squared distribution when the data show a good fit to the model. Within Rasch analysis, the severity of an item response and of the patients is measured using logits, which is the log odds of a 50% response of a patient of a given severity responding positively to that item. Within the model used for developing the instrument, the mean severity for the items and patient should approximate zero logits with a standard deviation of 1.0. More details of the statistical tests used are given in the online supplementary material and in the online supplementary material to the article by MEGURO *et al.* [10].

Reliability

Internal consistency and reliability of the finalised instrument was assessed using Cronbach's formula for coefficient α using pooled data from all patients. Values >0.70 are generally considered acceptable for aggregate data [8, 11, 12]. Reproducibility was tested using paired *t*-tests, Pearson correlation and intraclass correlation coefficient (ICCC).

Initial tests of construct and discriminant validity

In all studies, physicians rated the patient's overall level of COPD severity using a global scale with the categories mild, moderate, severe and very severe. In the USA patients, FEV₁ and scores for the SGRQ-C (a shorter version of the SGRQ specific for COPD that produces directly equivalent scores to the original [10]) were available for tests of construct and discriminant ability.

Analysis

SAS statistical software version 8.2 (SAS Institute, Cary, NC, USA) was used for all analyses except for those in the Rasch analysis, for which RUMM2020 (RUMM Laboratory, Perth, Australia) software was used. All statistical tests used a significance level of 0.05 unless otherwise noted. Results are presented as mean \pm SD, unless otherwise stated.

RESULTS

Patient demographics

A total of 1,503 subjects from six countries were included in the item reduction phase of the CAT, comprising Belgium (n=71), France (n=294), Germany (n=431), the Netherlands (n=109), Spain (n=369) and the USA (n=229). Demographic and clinical data are presented in table 1. Mean \pm SD age ranged from 63.7 \pm 9.6 to 68.0 \pm 9.0 yrs, with ~60% males in all countries, except Spain (88% males). USA patients had more severe obstruction than those in the EU, with FEV₁ in the USA 52.3 \pm 18.9% predicted and 57.8 \pm 19.9% pred in EU.

Item reduction

Patients used the full range of the response scale (0–5) for all 21 items, with mean item responses on the 0–5 scale ranging from 1.0 \pm 1.3 to 3.4 \pm 1.4. All items had minimal ($\leq 1\%$) rates of missing data. None showed evidence of age or sex effects. Four items demonstrated a high rate of floor effects (*i.e.* patients indicated that they did not experience the symptom or health state at all), ranging from 27% to 47% of subjects. Two items demonstrated ceiling effects; for both, 26% of subjects responded with the highest possible score. The item-to-total correlations ranged from 0.52 to 0.77. At this stage in the reduction process, four items were therefore deleted: three due to substantial floor effects and one because of a low item-to-total correlation. One item demonstrating a high floor effect (42%), "I am confident leaving my home despite my lung condition", was retained as a potential marker for mild patients whose health may get worse. Two items with ceiling effects were retained at this stage as potential indicators for

TABLE 1 Clinical characteristics by country

Characteristic	Belgium	France	Germany	The Netherlands	Spain	USA
Subjects	71	294	431	109	369	229
Age yrs	66 \pm 10.3	64 \pm 10.6	65 \pm 9.9	64 \pm 9.6	68 \pm 9.0	66 \pm 8.9
Sex						
Male	46 (65)	190 (65)	274 (64)	67 (61)	323 (88)	122 (53)
Female	25 (35)	104 (35)	156 (36)	42 (39)	46 (12)	107 (47)
Missing	0	0	1	0	0	0
Clinician-rated severity						
Mild	17 (24)	39 (13)	70 (16)	14 (13)	89 (24)	26 (11)
Moderate	27 (38)	171 (58)	215 (50)	60 (55)	180 (49)	95 (42)
Severe	19 (27)	75 (26)	114 (26)	28 (26)	85 (23)	80 (35)
Very severe	8 (11)	9 (3)	29 (7)	7 (6)	14 (4)	28 (12)
Missing	0	0	3 (1)	0	1 (0)	0
FEV₁ % pred	66 \pm 17	62 \pm 20	56 \pm 20	56 \pm 17	59 \pm 20	52 \pm 19

Data are presented as n, n (%) or mean \pm SD. FEV₁: forced expiratory volume in 1 s; % pred: % predicted.

patients whose health may improve; both pertained to breathing while walking up a hill or flight of stairs.

The item-to-item correlations were moderate to high, ranging from 0.30 to 0.89. Correlations ≥ 0.70 were noted between four pairs of items (indicating that one of each pair may be redundant). Based on these findings, three additional items were deleted. Two breathing difficulty/wheeze items qualified as “easy or very difficult” were deleted due to potential problems related to the translatability of this concept, and one “activity limitation” item was deleted because a similar item demonstrated a wider range of responses and was also deemed to be better worded. Items measuring “tired” and “energy” were correlated ($r=0.71$) but were maintained during this round of analysis, since they measured different concepts and the correlation coefficient was only slightly over the threshold.

Seven iterations of Rasch analysis were conducted to identify items with the best fit to a unidimensional model and items to be removed due to differential functioning between countries. A total of six items were deleted, with the model improving at each iteration in terms of overall fit (*i.e.* reduction in Chi-squared value) and/or better distribution of item responses evidenced by the standard deviation approaching one. An eight-item model solution was found. Testing the removal of further items to create a smaller (seven-item) instrument resulted in either worse measurement properties or development of differential item functioning between countries. For this reason, the eight-item solution was accepted. The final eight items demonstrated a good model fit (Chi-squared 124.7 ± 0.601 ; $p=0.00012$) with no country bias and covered cough, phlegm, chest tightness, breathlessness going up a hill/stairs, activity limitation at home, confidence in leaving home, sleep and energy.

A total of 13 items were deleted owing to: high floor effects (three), low item-to-total correlations (one), high item-to-item correlations (three) and poor performance on the IRT analyses (six); see online supplementary material for more details. The remaining eight items that form the CAT cover a wide range of COPD severity (fig. 1) and can be scored as a single scale, with scores ranging from 0 to 40 (Appendix). Higher scores represent worse health. Most items are distributed evenly across the severity range, but the item concerned with breathlessness on stairs/hills has greatest discriminant power for milder patients, whereas the one concerning confidence leaving the home discriminates better in more severe patients. The frequency distribution of the patients' scores shows good matching of the items to the severity of this patient population (fig. 1).

Preliminary psychometric properties of the finalised CAT

The data used for item reduction also allowed a number of initial tests of the consistency and reliability of the eight items that formed the CAT as an instrument. Internal consistency ($n=1,490$) was excellent with Cronbach's $\alpha=0.88$ and test-retest in stable patients ($n=53$) was equally good (ICCC=0.8). The cumulative frequency distribution of scores (USA stable and EU patients) shows that the entire scaling range is used (fig. 2). CAT scores by country are presented in table 2; there was a significant difference ($p<0.0001$), mainly due to

Belgium, where the scores were higher but in a smaller sample of patients than the other countries.

In the patients from the USA, Rasch analysis was used to test whether the measurement properties of the eight CAT items differed between stable ($n=229$) and acute states of exacerbation ($n=67$). No items showed evidence of this, showing that the CAT should provide a reliable measurement of differences in COPD severity between these states.

SGRQ-C data were only available for the USA patients for these validity tests. The correlation between the CAT and SGRQ-C in stable patients was very good: $r=0.8$, $n=227$ (fig. 3) and equally good ($r=0.78$, $n=67$) in acute patients with an exacerbation. The CAT scores were significantly different between patients with acute and stable disease (fig. 4); the mean difference was 4.7 units (95% CI 2.7–6.7 units) (paired *t*-test $p<0.0001$). This equates to 12% of the scaling range. In the same patients, the difference in SGRQ was 12.0 units (95% CI 6.6–17.4), *i.e.* 12% of its scaling range. The effect size (mean difference divided by the standard deviation of the stable patients) for the CAT was 0.63 and for the SGRQ-C it was 0.62.

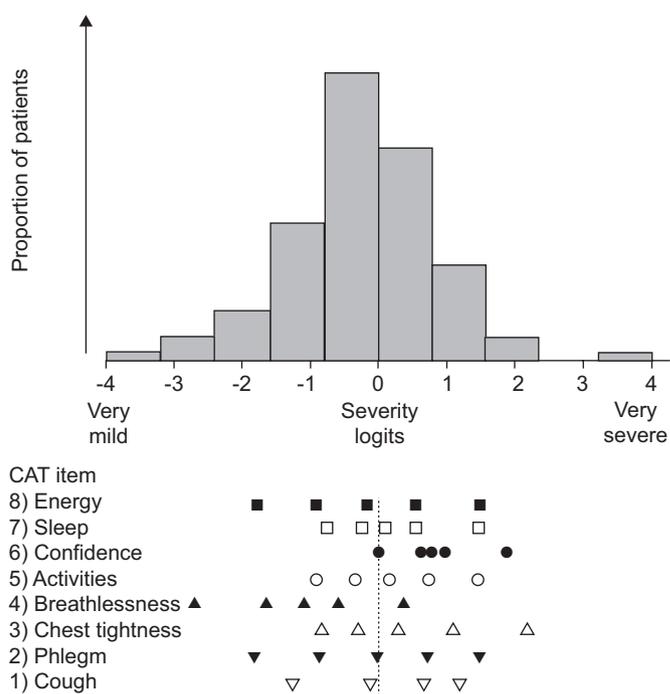


FIGURE 1. Rasch “item map” showing the severity of each item. The units are logits (log odds) of a 50% probability that a patient of a given level of severity will affirm a given response category. Each symbol shows the level of severity at the boundary between two adjacent response categories (between 0 and 1, 1 and 2, etc.), *i.e.* when the probability of a positive response in the adjacent categories is 50%. The frequency distribution of the patient's severity (measured in logits) is plotted above the x-axis. The mean item severity for each item is tabulated in the online supplement. CAT: COPD Assessment Test; COPD: chronic obstructive pulmonary disease.

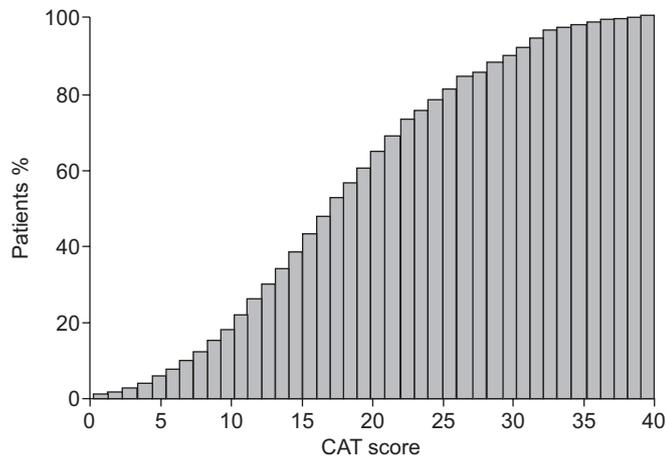


FIGURE 2. Cumulative frequency distribution of COPD Assessment Test (CAT) score in 1,503 patients with stable chronic obstructive pulmonary disease (COPD). 80% of patients used 63% of the scaling range; 10% of patients had a score <7 units and 10% had a score >32.

DISCUSSION

This study has created a short, simple patient-completed questionnaire for COPD with very good measurement properties. It covers a broad range of effects of COPD on patients' health, despite the small number of component items. The quality of fit to a Rasch unidimensional model suggests that it has true interval scaling properties. Based on data from six countries, tests of internal consistency show that it provides a reliable measure of overall COPD severity from the patient's perspective, independent of language. This should ensure that it is relevant to an international COPD population and applicable for global use. Preliminary tests of validity show the entire scaling range of the instrument is used by a COPD population.

The item reduction process followed a rigorous methodology that combined classical test theory and IRT, together with

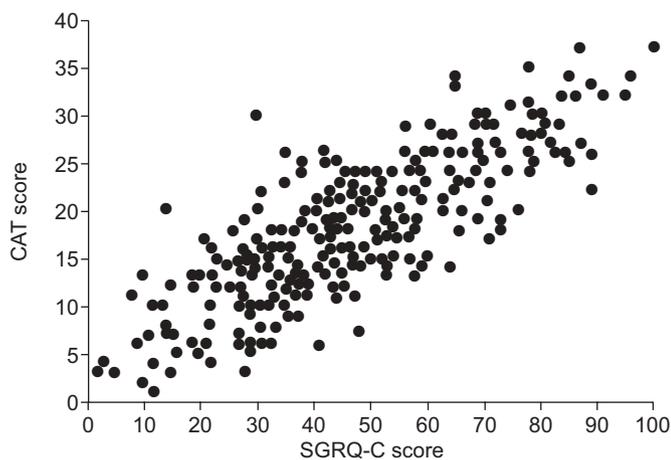


FIGURE 3. Pearson correlation between scores in the chronic obstructive pulmonary disease (COPD)-specific version of the St George's Respiratory Questionnaire (SGRQ-C) and COPD Assessment Test (CAT) in 229 stable patients from the USA. $r=0.80$, $p<0.0001$.

Country	Subjects	CAT score
Belgium	71	21.5±9.9
France	294	18.5±8.8
Germany	431	18.2±8.1
The Netherlands	109	16.0±7.4
Spain	369	16.4±8.9
USA	229	17.8±7.5

Data are presented as n or mean±sd. COPD: chronic obstructive pulmonary disease.

careful monitoring of content. Items with poor measurement properties were removed, while preserving broad coverage of the different effects of COPD. When deciding whether to include or exclude an item during questionnaire development, it is necessary to balance its weaknesses and strengths against its overall contribution. We did not exclude an item on the basis of one criterion alone, but because of its overall performance compared with the others. Developer and clinical judgment were required in this process, but since Rasch methodology provides an objective test of the quality of each item's fit to a unidimensional model that contains all the items, the composition of the final item set was driven more by the patient's responses than is the case when questionnaire development is based solely on classical test methodology.

The final content covers: cough, phlegm, chest tightness, breathlessness going up hills/stairs, activity limitation at home, confidence leaving home, sleep and energy. The principle that the instrument should have reliable measurement properties with all items meeting tight statistical requirements was achieved. Item content was used as a guide to decision making, particularly when choosing between items to remove. The original intention was to produce an instrument with five to

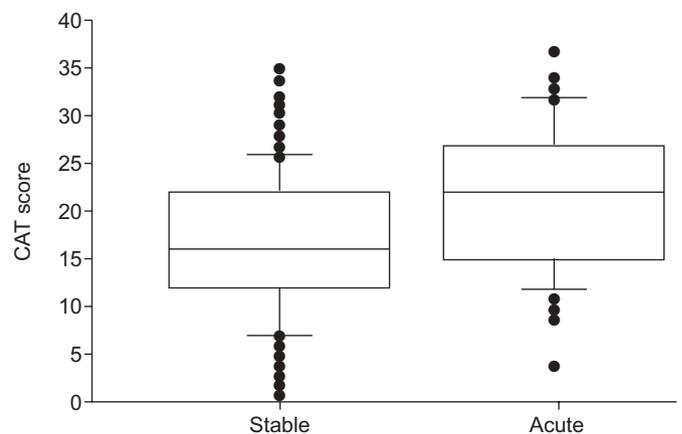


FIGURE 4. Box and whisker plot of COPD Assessment Test (CAT) scores of 229 stable patients and 67 patients measured on the day of presentation with an acute exacerbation. COPD: chronic obstructive pulmonary disease. Boxes represent medians and interquartile ranges, whiskers represent 10% and 90% limits. ●: individual patients who lie outside the 10% and 90% limits. $p<0.0001$ (unpaired t-test).

seven items; however, the item reduction process dictated eight items. Removing items to produce a seven-item instrument reduced content coverage and worsened measurement properties, principally through the emergence of differential functioning between countries in some items.

The final CAT consists of eight items, each formatted as a semantic six-point differential scale (Appendix), making the tool easy to administer and easy for patients to complete. The items were selected to cover a wide range of disease severity, with the intention that the greatest discriminant power would be in the mild to moderate range. Based on our findings, as shown by the item map in figure 1, the items related to cough and phlegm have greater discriminant power for milder disease; items concerning chest tightness and confidence leaving home are more discriminative in severe COPD and the remaining items capture moderate health status impairment.

A limitation of this study is that the reliability and validation findings are based on data from the USA only, owing to the current availability of such data. However, it is clear that the CAT has very similar discriminative properties to the much more complex SGRQ-C, showing that it will be able to measure the impact of COPD on individual patient's health. Validation of an instrument is a continuous process and international studies will be performed to further test its psychometric properties. Use of standardised techniques will ensure linguistic and cultural validity in all languages.

The CAT will provide clinicians and patients with a simple and reliable measure of overall COPD-related health status for the assessment and long-term follow-up of individual patients. It is not a diagnostic tool; its role is to supplement information obtained from lung function measurement and assessment of exacerbation risk. The content and layout of the CAT will allow identification of key areas of health impairment that the clinician can then explore further in the consultation. It has good repeatability and its discriminative properties suggest that it is likely to be sensitive to treatment effects at a group level. In common with all other measurements suggested for routine use in individual patients, including FEV₁ and the Medical Research Council dyspnoea scale [13], we do not expect that its signal/repeatability ratio will enable it to determine reliably, in every patient, whether they have had a clinically worthwhile response to a specific treatment. Despite this limitation, the CAT should improve communication between clinician and patient, enabling a common understanding of the severity and impact of the patient's disease. This, in turn, should enable COPD treatment to be better targeted and management optimised.

APPENDIX

For Appendix, see following page.

SUPPORT STATEMENT

This study and the development of the COPD Assessment Test (CAT) were funded by GlaxoSmithKline.

STATEMENT OF INTEREST

Statements of interest for all authors, and for the study itself, can be found at www.erj.ersjournals.com/misc/statements.dtl

ACKNOWLEDGEMENTS

The authors would like to thank M. Tabberer, N. Banik, H. McDowell and L. Adamek of GlaxoSmithKline (London, UK) and the CAT working group members for their important contributions to this work.

The members of this working group are as follows. A. Agusti, Hospital Clinic, University of Barcelona, Barcelona, Spain; W. Bailey, University of Alabama Lung Health Center, Birmingham, AL, USA; O. Bauerle, Centro Medico Las Americas, Merida, Mexico; D. Halpin, Royal Devon and Exeter Hospital, Exeter, UK; C. Jenkins, Woolcock Institute of Medical Research, Camperdown, Australia; P. Kardos, Maingau Hospital, Frankfurt, Germany; M. Levy, University of Edinburgh, Edinburgh, UK; F. Martinez, University of Michigan, Ann Arbor, MI, USA; M. Miravittles, Hospital Clinic, University of Barcelona, Barcelona, Spain; S. Molitor, University of Hanover, Hanover, Germany; D. Price and M. Thomas, University of Aberdeen, Aberdeen, UK; N. Roche, University of Paris 5, Paris, France; M. Salapatras, European Federation of Allergy and Airway Diseases Patients Association, Greece; T. van der Molen, University Medical Center Groningen, Groningen, the Netherlands; and J. Walsh, COPD Foundation, Miami, FL, USA.

In addition, we acknowledge the support of Innovex Medical Communications (Bracknell, UK) in the administration of the manuscript submission. This support was funded by GlaxoSmithKline.

REFERENCES

- Mathers CD, Stein C, Ma Fat D, *et al.* Global Burden of Disease 2000: Version 2 Methods and Results. Geneva, World Health Organization, 2000; pp. 1–108.
- Mannino DM, Homa DM, Akinbami LJ, *et al.* Chronic obstructive pulmonary disease surveillance – United States, 1971–2000. *MMWR Surveill Summ* 2002; 51: 1–16.
- Rennard S, Decramer M, Calverley PM, *et al.* Impact of COPD in North America and Europe in 2000: subjects' perspective of Confronting COPD International Survey. *Eur Respir J* 2002; 20: 799–805.
- Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med* 1991; 85: Suppl. B, 25–31.
- Larson JL, Covey MK, Berry JK, *et al.* Reliability and validity of the Chronic Respiratory Disease Questionnaire. *Am Rev Respir Dis* 1993; 147: A530.
- van der Molen T, Willemsse BW, Schokker S, *et al.* Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health Qual Life Outcomes* 2003; 1: 13.
- Jones P, Harding G, Wiklund I, *et al.* Improving the process and outcome of care in COPD: development of a standardised assessment tool. *Prim Care Respir J* 2009; in press.
- Psychometric Theory. 3rd Edn. New York, McGraw-Hill, 1994.
- O'Leary CJ, Jones PW. The influence of decisions made by developers on health status questionnaire content. *Qual Life Res* 1998; 7: 545–550.
- Meguro M, Barley EA, Spencer S, *et al.* Development and validation of an improved, COPD-specific version of the St. George Respiratory Questionnaire. *Chest* 2007; 132: 456–463.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297–334.
- Hays RD, Revicki DA. Reliability and validity, including responsiveness. In: Fayers P, Hays RD, eds. *Assessing Quality of Life in Clinical Trials*. New York, Oxford University Press, 2005; pp. 25–39.
- O'Donnell DE, Aaron S, Bourbeau J, *et al.* Canadian Thoracic Society recommendations for management of chronic obstructive pulmonary disease – 2007 update. *Can Respir J* 2007; 14: Suppl. B, 5B–32B.

APPENDIX: CONTENT AND STRUCTURE OF THE FINAL CAT QUESTIONNAIRE

How is your COPD?

For each item below, place a mark (√) in the box that best describes your experience.

Example: I am very happy

0	1	2	3	4	5
---	---	---	---	---	---

 I am very sad

		SCORE						
I never cough	<table border="1" style="display: inline-table;"><tr><td style="width: 20px; text-align: center;">0</td><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table>	0	1	2	3	4	5	I cough all the time
0	1	2	3	4	5			
I have no phlegm (mucus) in my chest at all	<table border="1" style="display: inline-table;"><tr><td style="width: 20px; text-align: center;">0</td><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table>	0	1	2	3	4	5	My chest is completely full of phlegm (mucus)
0	1	2	3	4	5			
My chest does not feel tight at all	<table border="1" style="display: inline-table;"><tr><td style="width: 20px; text-align: center;">0</td><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table>	0	1	2	3	4	5	My chest feels very tight
0	1	2	3	4	5			
When I walk up a hill or one flight of stairs I am not breathless	<table border="1" style="display: inline-table;"><tr><td style="width: 20px; text-align: center;">0</td><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table>	0	1	2	3	4	5	When I walk up a hill or one flight of stairs I am very breathless
0	1	2	3	4	5			
I am not limited doing any activities at home	<table border="1" style="display: inline-table;"><tr><td style="width: 20px; text-align: center;">0</td><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table>	0	1	2	3	4	5	I am very limited doing activities at home
0	1	2	3	4	5			
I am confident leaving my home despite my lung condition	<table border="1" style="display: inline-table;"><tr><td style="width: 20px; text-align: center;">0</td><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table>	0	1	2	3	4	5	I am not at all confident leaving my home because of my lung condition
0	1	2	3	4	5			
I sleep soundly	<table border="1" style="display: inline-table;"><tr><td style="width: 20px; text-align: center;">0</td><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table>	0	1	2	3	4	5	I don't sleep soundly because of my lung condition
0	1	2	3	4	5			
I have lots of energy	<table border="1" style="display: inline-table;"><tr><td style="width: 20px; text-align: center;">0</td><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table>	0	1	2	3	4	5	I have no energy at all
0	1	2	3	4	5			

SCORE

--	--

Reproduced with permission from GlaxoSmithKline. GlaxoSmithKline is the copyright owner of the COPD Assessment Test (CAT). However, third parties will be allowed to use the CAT free of charge. The CAT must always be used in its entirety. Except for limited reformatting the CAT may not be modified or combined with other instruments without prior written approval. The eight questions of the CAT must appear verbatim, in order, and together as they are presented and not divided on separate pages. All trademark and copyright information must be maintained as they appear on the bottom of the CAT and on all copies. The final layout of the final authorised CAT questionnaire may differ slightly but the item wording will not change. The CAT score is calculated as the sum of the responses present. If more than two responses are missing, a score cannot be calculated; when one or two items are missing their scores can be set to the average of the non-missing item scores.