The COPD assessment test: a systematic review

Nisha Gupta[1,2], Lancelot M. Pinto[1], Andreea Morogan[1] and Jean Bourbeau[1,2]

Affiliations: [1]Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, McGill University, Montréal, QC, Canada. [2]Dept of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, QC, Canada.

Correspondence: Jean Bourbeau, Respiratory Epidemiology and Clinical Research Unit, Montreal Chest Institute, 3650 Rue St-Urbain, Office K1.32, Montréal, QC, H2X 2P4, Canada.
E-mail: jean.bourbeau@mcgill.ca

ABSTRACT The COPD assessment test (CAT) is a self-administered questionnaire that measures health-related quality of life. We aimed to systematically evaluate the literature for reliability, validity, responsiveness and minimum clinically important difference (MCID) of the CAT.

Multiple databases were searched for studies analysing the psychometric properties of the CAT in adults with chronic obstructive pulmonary disease. Two reviewers independently screened, selected and extracted data, and assessed methodological quality of relevant studies using the COSMIN checklist.

From 792 records identified, 36 studies were included. The number of participants ranged from 45 to 6469, mean age from 56 to 73 years, and mean forced expiratory volume in 1 s from 39% to 98% predicted. Internal consistency (reliability) was 0.85–0.98, and test–retest reliability was 0.80–0.96. Convergent and longitudinal validity using Pearson's correlation coefficient were: SGRQ-C 0.69–0.82 and 0.63, CCQ 0.68–0.78 and 0.60, and mMRC 0.29–0.61 and 0.20, respectively. Scores differed with GOLD stages, exacerbation and mMRC grades. Mean scores decreased with pulmonary rehabilitation (2.2–3 units) and increased at exacerbation onset (4.7 units). Only one study with adequate methodology reported an MCID of 2 units and 3.3–3.8 units using the anchor-based approach and distribution-based approach, respectively. Most studies had fair methodological quality.

We conclude that the studies support the reliability and validity of the CAT and that the tool is responsive to interventions, although the MCID remains debatable.

@ERSpublications
Studies support the reliability, validity and responsiveness of the CAT as a HRQoL tool but its MCID remains unclear http://ow.ly/xkVNA

For editorial comments see page 833.

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

Received: Feb 05 2014 | Accepted after revision: May 22 2014 | First published online: July 03 2014

Support statement: This research was supported by a studentship award from the Research Institute of the McGill University Health Centre (Graduate Student Scholarship award to Nisha Gupta), the Collaborative Innovative Research Fund (CIRF) GSK Canada, the Canadian Cohort Obstructive Lung Disease (CanCOLD) Canadian Institutes of Health Research (CIHR)/Rx&D Collaborative Research Program (IRO-93326), and the Respiratory Health Network of the Fonds de recherche du Québec – Santé (FRQS).

Conflict of interest: Disclosures can be found alongside the online version of this article at erj.ersjournals.com

Introduction

Establishing a diagnosis of chronic obstructive pulmonary disease (COPD) requires spirometry; however, recent guidelines suggest that classifying COPD solely by forced expiratory volume in 1 s (FEV_1) % predicted is inadequate in reporting disease severity [1]. Assessing a patient's health-related quality of life (HRQoL) allows clinicians to make individualised patient management decisions; thus, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) strategy document advocates that COPD management no longer be stratified solely by spirometric classification, but through a multidimensional assessment of specific patient attributes [2, 3].

COPD-specific questionnaires assessing HRQoL do exist (*e.g.* the St George's Respiratory Questionnaire (SGRQ) or the Chronic Respiratory Questionnaire (CRQ)), although some are impractical for clinical use as they are time consuming [2]. GOLD consequently proposes using either the modified British Medical Research Council (mMRC) dyspnoea scale or the COPD assessment test (CAT); however, preferential recommendation is given to the CAT since it provides a thorough coverage of the impact of COPD on wellbeing [2].

The CAT was created using COPD patients' input, then developed using modern questionnaire methodology: psychometric analysis and item response theory using Rasch analysis identified items with the best fit to form a unidimensional instrument [4, 5]. The self-administered questionnaire consists of eight items assessing various manifestations of COPD aiming to provide a simple quantified measure of HRQoL [5]. A preliminary evaluation of the CAT's psychometric properties has been promising [5]. Summarising the current knowledge on the performance of this tool as a HRQoL measurement instrument is valuable, as the test could have important roles in COPD clinical practice and research. To the best of our knowledge, a comprehensive review of the psychometric properties of the CAT questionnaire has not been conducted.

Our objectives for this review were to systematically search the literature to evaluate and summarise the psychometric properties of the CAT (reliability, validity, responsiveness and minimum clinically important difference (MCID)) as a HRQoL instrument used in patients with COPD.

Methods

Detailed descriptions of the psychometric properties assessed in the review and the completed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist can be found in the online supplementary material.

Eligibility criteria for study selection

Randomised controlled trials and observational studies (*e.g.* cross-sectional, cohort, *etc.*) with >10 participants were included. Study participants needed to be subjects aged $\geqslant 40$ years diagnosed with COPD (using the GOLD criteria) [2]. Interventions could have been any intervention, placebo, usual care or time. Outcomes evaluated consisted of the CAT reliability, validity, responsiveness or MCID. Studies that reported at least one or more psychometric properties were included. Detailed inclusion and exclusion criteria can be found in the online supplementary material.

Information sources and search

A search was conducted on March 10, 2014 in the Cochrane Database of Systematic Reviews, the Database of Abstracts of Reviews of Effects, and Clinical Query in PubMed, to identify previous systematic reviews on the subject. The solitary review retrieved addressed several HRQoL instruments, had a broad review question and a limited search strategy [6].

A structured search was performed on March 10, 2014 in five electronic general databases: Cochrane Central, PubMed Medline, OvidSP Medline, OvidSP Embase and Thomson Reuters ISI Web of Knowledge Web of Science. The database searches were done from the year 2009 onwards as this was the year that the CAT was developed. We used a variety of search terms, including text words and database-specific subject headings, for articles in English, French or Spanish. The detailed, database-specific search strategies can be found in the online supplementary material. Top-ranked respiratory journals and the ProQuest Dissertations & Theses database were manually searched. Reference lists from existing narrative reviews of the CAT were searched for potential studies, as were bibliographies of all included studies.

Study selection, data collection process and data items

Two reviewers (Nisha Gupta and Lancelot M. Pinto) independently screened the title and abstract of each study identified from the search. All the potentially relevant articles were then retrieved in full-text form and two reviewers (Nisha Gupta and Lancelot M. Pinto) performed the secondary screen after a review of the

DOI: 10.1183/09031936.00025214

full text of the chosen articles. Disagreement on the inclusion or exclusion of a specific study was resolved by reaching a consensus. When necessary, disagreements were resolved by a third reviewer (Jean Bourbeau). A list of excluded studies and reasons for their exclusion was maintained.

Data were electronically extracted from each eligible study using a piloted data extraction form. The form was revised and improved after pilot data extractions were performed to assess concordance between the reviewers.

Two reviewers (Nisha Gupta and Andreea Morogan) independently extracted data from each included study, including study characteristics, population characteristics, interventions and/or events and outcomes studied, along with the corresponding measures of test performance. The data extraction form can be found in the online supplementary material.

Quality assessment of included studies

Two reviewers (Nisha Gupta and Andreea Morogan) independently performed quality assessment for each study using the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist, which is a validated quality assessment tool that evaluates the methodological quality of studies assessing psychometric properties of an instrument; it is the only specific checklist for methodological evaluation of psychometric properties on patient-reported outcomes [7, 8]. The methodological quality of each psychometric property was evaluated through a number of items and was scored using the four-point rating scale of "excellent", "good", "fair" and "poor". An overall score for the methodological quality of a study was given for each psychometric property by taking the lowest rating of an item ("worst score counts" method) [7]. The list of items, scoring rules and psychometric properties to which the checklist applies to can be found in the online supplementary material.

Synthesis of results

A narrative synthesis was employed to summarise the current knowledge on the CAT's reliability, validity, responsiveness and MCID. Data were tabulated through detailed tables that compared the studies with respect to study characteristics, population characteristics and CAT psychometric properties. Data synthesis was based on the provision of appropriate and similar outcomes studied, noting the specific statistical tests used. Computing a range (minimum to maximum) of study results on a particular psychometric property assessed the strength and adequacy of the psychometric property. A discussion of the impact of methodological quality on study results was explored to provide some assessment of quality and heterogeneity between the studies. All analyses were performed using STATA 11 (Stata Corporation, College Station, TX, USA).

Risk of bias across studies

Language bias was assessed by retrieving citations from the search strategy with language filters (English, French and Spanish) and without, and was reported as an average across the five general databases according to the filtered citations as a percentage of the overall citations retrieved.

Results

Study selection

Figure 1 shows the study selection procedure and numbers of studies screened, assessed for eligibility and included in the review. A total of 36 articles were included in the qualitative synthesis.

Study characteristics

Table 1 summarises the study and population characteristics of the included studies according to the number of outcomes assessed: nine (25%) studies assessed reliability (internal consistency and test–retest), 32 (89%) studies assessed validity (concurrent, convergent, longitudinal and known groups validity), ten (28%) studies assessed responsiveness and four (11%) studies assessed MCID [5, 9–43].

The CAT was administered in 32 countries spanning Europe, North America, South America, Asia and Africa, with 17 studies published in 2012 and 11 studies published in 2013. Of the 36 studies, 16 were prospective cohorts and the remaining 20 were cross-sectional. Types of interventions or events evaluated in the prospective cohorts included pulmonary rehabilitation, onset of an acute exacerbation, recovery from an acute exacerbation and usual care; and the duration of follow-up ranged from 2 to 24 weeks. The number of participants ranged between 45 and 6469 with the percentage of female subjects between 0% and 64.4%. The range for mean age was between 55.9 and 73.0 years and mean FEV_1 between 38.7% and 98.0% predicted. 23 studies specified the number of individuals in GOLD grades.

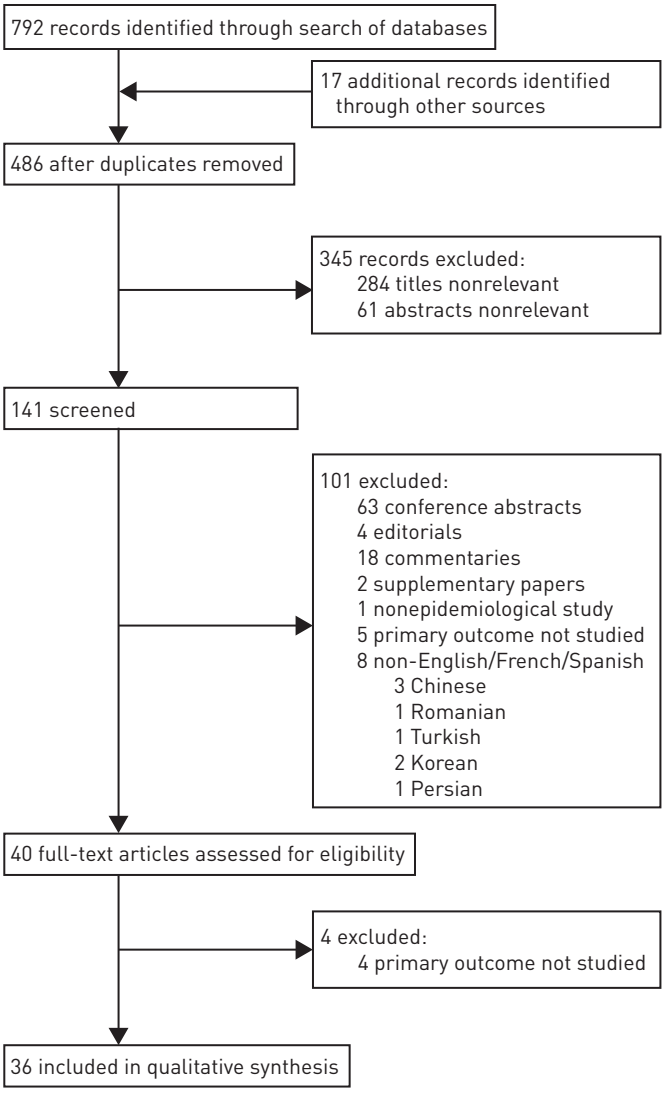

FIGURE 1 Summary of literature search and study selection.

Nonresponse rate and floor and ceiling effect

Only six studies reported the proportions of the population with complete CAT scores (no missing items) and those with missing items [9, 15, 17, 21, 26, 41] (table 2). The percentage of patients with the minimum (floor effect) and maximum (ceiling effect) possible total score was measured in two populations [9, 36] (table 2).

Reliability

The internal consistency of the CAT was reported in eight studies, with the Cronbach's α range from 0.85–0.98 indicating a high correlation between items [5, 9, 10, 14, 15, 18, 19, 21] (table 2). Test–retest was evaluated in five studies to measure reproducibility [5, 9, 10, 14, 30] (table 2), with the CAT administered on two different occasions (at baseline and then either 1 or 2 weeks later) for three studies [5, 10, 30] and on three separate occasions (at baseline, 2 weeks and then 6 weeks later) for one study [14]. The ICC ranged from 0.80–0.96, demonstrating that the CAT is consistent in producing scores when administered repeatedly under stable disease condition.

Validity

Three studies measured concurrent validity by comparing the total CAT score to healthcare utilisation [15, 34, 39]. The number of physician consultations was associated with total CAT scores in one study ($p<0.001$) [34], but not in the others [15, 39]. The number of hospitalisations was directly associated with total CAT scores, irrespective of using the total CAT score ($p<0.001$) [15] or arbitrarily dichotomising the CAT score (*e.g.* <10 and $\geqslant 10$) ($p<0.001$ [34] and $p<0.032$ [39]). Likewise, the number of emergency

DOI: 10.1183/09031936.00025214

TABLE 1 Study and population characteristics of the selected studies of the systematic review

First author [ref.]	Year	Country	Setting	Patient population	Study design	Follow-up weeks	Subjects n	Female %	Age mean±SD	FEV1 % predicted mean±SD
⩾3 measurement properties assessed										
Agusti [9]	2012	Spain	Primary care and hospital	Inpatient and outpatient	Prospective cohort	4	377	7.2	NR	NR
da Silva [10]	2013	Brazil	Hospital	Outpatient	Prospective cohort	1	50[¶]	52.0	62.2±8.4	44.0±13.8
Dodd [11]	2011	England	Primary and secondary care	NR	Prospective cohort	8	297	37.4	69.2±9.3	50.9±18.9
Jones [5]	2009	USA	Primary care and pulmonary clinics	Outpatient	Prospective cohort	2	229	47.0	66.0±8.9	52.30±18.9
Jones [12]	2012	Canada and USA	Primary care and pulmonary clinics	Outpatient	Prospective cohort	6	131	45.0	64.0±9.0[+]	47.0±21.0[+]
Miravitlles [13]	2013	Spain	Primary care and hospital	Inpatient and outpatient	Prospective cohort	6	486	11.8	69.4±9.5	47.7±17.4
Tsiligianni [14]	2012	Greece	Primary and secondary care	NR	Prospective cohort	6	90	10.0	67.0 (58–75)[§]	NR
Tsuda [15]	2012	Japan	Patient database and panel	Database or panel registrants	Cross-sectional		301	22.9	55.9±10.6	NR
2 measurement properties assessed										
Dodd [16]	2012	England	Pulmonary rehabilitation clinic	Outpatient	Prospective cohort	24	118	49.2	73.0±8.0	47.0 (33–67)[§]
Ghobadi [17]	2012	Iran	NR	NR	Cross-sectional		105	0	59.6±11.9	71.0±26.7
Horita [18]	2014	Japan	Hospital	Outpatient	Cross-sectional		85	17.6	71.9±8.9	45.8±14.7
Hwang [19]	2013	South Korea	Primary care and hospital	Outpatient	Cross-sectional		100	3.0	69.2±8.4	63.6±21.1
Jones [20]	2011	Belgium, France, Germany, Italy, Netherlands, Spain and UK	Primary care	Outpatient	Cross-sectional		1817	28.2	64.9±9.6	56.7±20.1
Jones [21][#]	2012	Algeria, Egypt, Jordan, Lebanon, Morocco, Saudi Arabia, Syria, Tunisia, Turkey and UAE	General population survey	Survey participants	Cross-sectional		6469	53.5	NR	NR
Jones [22]	2012	Belgium, France, Germany, Netherlands and Spain	Primary care	Outpatient	Cross-sectional		2294	30.7	64.6±10.0	59.0±21.1
Kelly [23]	2012	England	Hospital clinics	Outpatient	Prospective cohort	NR	224	35.7	63.5±10.3	40.1±17.9
Kon [24]	2013	England	Hospital	Outpatient	Prospective cohort	8	255	39.2	69.5±9.4	45.8±20.3
Kwon [25]	2013	Indonesia, Korea, Vietnam and Hong Kong	Primary and secondary care	Outpatient	Cross-sectional		333	2.4	69.3±9.4	52.1±20.3
Mackay [26]	2012	England	Previously existing COPD cohort study	Outpatient	Prospective cohort	5	161	39.8	71.3±9.4	50.3±16.9
Marchand [27]	2012	Belgium	Tertiary care	Outpatient	Cross-sectional		213	37.0	63.0±10.0	47.0±20.0
Miyazaki [28]	2014	Japan	Hospital	Outpatient	Prospective cohort	NR	403	9.0	72.4±8.0	61.6±21.6
Zhou [29]	2013	China	Pulmonary clinics, secondary and tertiary care	NR	Cross-sectional		6437	25.1	64.9±10.0	NR
1 measurement property assessed										
Al Moamary [30]	2011	Saudi Arabia	Pulmonary rehabilitation clinic and pulmonary clinic	Outpatient	Prospective cohort	2	45	20.0	61.0±8.7	49.7±13.6
Al Moamary [31]	2012	Saudi Arabia, Kuwait, Bahrain, and UAE	Pulmonary clinic	Outpatient	Cross-sectional		120	19.2	63.3±7.3	49.3±13.4
Feliz-Rodriguez [32]	2013	Spain	Hospital	Inpatient	Prospective cohort	12	45	31.1	71.0±10.0	51.0±17.0
Gao [33]	2013	China	Hospital	Inpatient and outpatient	Cross-sectional		263	15.2	66.7±7.4	47.43±14.3

TABLE 1 Continued

First author [ref.]	Year	Country	Setting	Patient population	Study design	Follow-up weeks	Subjects n	Female %	Age mean±SD	FEV1 % predicted mean±SD
Idrees [34]#	2012	Algeria, Egypt, Jordan, Lebanon, Morocco, Saudi Arabia, Syria, Tunisia, Turkey, UAE and Pakistan	General population survey	Survey participants	Cross-sectional		1014	23.6	NR	NR
Kon [35]	2014	England	Hospital and pulmonary clinics	Inpatient and outpatient	Prospective cohort	8	565[i]	42.1	70.0±9.0	47.6 (45.9–49.3)##
Nishimura [36]	2013	Japan	Secondary care	Healthy industrial workers	Cross-sectional		1333	34.7	56.0±8.2	95.8±14.8
Okutan [37]	2013	Turkey	Hospital	Outpatient	Cross-sectional		90	NR	68.5±10.9	NR
Papaioannou [38]	2013	Greece	Tertiary care	Inpatient	Prospective cohort	6	230	11.7	71.2±8.8	52.8±20.1
Polatli [39]#	2012	Algeria, Egypt, Jordan, Lebanon, Morocco, Saudi Arabia, Syria, Tunisia, Turkey, UAE and Pakistan	General population survey	Survey participants	Cross-sectional		1392	24.4	NR	NR
Raghavan [40]	2012	Canada	Random sample of general population	Random sample participants	Cross-sectional		532	53.0	60.1±11.4	98.0±17.0
Ringbaek [41]	2012	Denmark	Secondary care	Outpatient	Cross-sectional		90	64.4	69.5±8.7	38.7±12.9
Uzaslan [42]#	2012	Algeria, Egypt, Jordan, Lebanon, Morocco, Saudi Arabia, Syria, Tunisia, Turkey, UAE and Pakistan	General population survey	Survey participants	Cross-sectional		1392	24.4	NR	NR
Varol [43]	2014	Turkey	Hospital	NR	Cross-sectional		165	9.7	65.1±9.9	43.7±14.8

FEV1: forced expiratory volume in 1 s; NR: not reported; COPD: chronic obstructive pulmonary disease. #: references [21, 34, 39, 42] come from the same cross-sectional cohort (the BREATHE study); however, each reference evaluates different psychometric properties with varying subcohorts. ¶: reported population characteristics for face-to-face questionnaire group only, where the CAT was completed in a face-to-face interview. +: reported age and FEV1 for 67 subjects participating in study 1 of Jones [12]. §: data are presented as median (interquartile range). ƒ: reported population characteristics for 565 subjects participating in study 1 of Kon [35]. ##: data are presented as mean (95% CI).

DOI: 10.1183/09031936.00025214

TABLE 2 COPD assessment test (CAT) nonresponse rate, floor and ceiling effect, internal consistency and test–retest reliability

Completed CATs and missing items	0 missing items %	1 missing item %	Subjects n
Agusti [9]	99.7	0.3	377
Ghobadi [17]	100	0.0	105
Jones [21]#	99.3	0.7	6469
Mackay [26]	98.3	1.7	161
Ringbaek [41]	98.9	1.1	90
Tsuda [15]	100	0.0	301
Floor¶ and ceiling+ effect	**Floor effect %**	**Ceiling effect %**	**Subjects n**
Agusti [9]	0.5	0.3	377
Nishimura [36]	0.0	7.6	145##
Internal consistency§	**Cronbach's α**	**Subjects n**	**p-value**
Agusti [9]	0.86	NR	NR
da Silva [10]	0.98	50	0.001
Horita [18]	0.85	85	NR
Hwang [19]	0.85	100	NR
Jones [5]	0.88	1490	NR
Jones [21] Arabic version	0.85	4807	NR
Jones [21] Turkish version	0.86	1590	NR
Tsiligianni [14]	0.86	90	NR
Tsuda [15]	0.89	301	NR

Test–retest^{f}	ICC	n	Administration duration weeks			p-value
			1	2	3	
Agusti [9]	0.83	NR	NR	NR	NR	NR
Al Moamary [30]	0.90	45	0	2	NA	0.00008
da Silva [10]	0.96	50	0	1	NA	0.93–0.97¶¶
Jones [5]	0.80	53	0	1	NA	NR
Tsiligianni [14]	0.94	90	0	2	6	0.92–0.96¶¶

COPD: chronic obstructive pulmonary disease; ICC: intra-class correlation coefficient; NR: not reported; NA: not applicable. #: missing defined as more than one item not completed so the questionnaire was considered unusable; if the score was missing for only one item, this was replaced by the mean value of the remaining seven items. ¶: percentage of patients with the minimum possible CAT total score (0/40). +: percentage of patients with the maximum possible (40/40). §: interrelatedness of the items within the CAT. ^{f}: ability of the CAT to provide consistent scores over time when administered repeatedly under stable disease condition. ##: floor and ceiling effect of the CAT for COPD participants defined by fixed ratio. ¶¶: 95% CI for ICC estimate.

room visits was associated with the total CAT scores in two studies ($p<0.001$ [15] and $p<0.001$ [34]), but not in the other [39].

Convergent validity was assessed in 21 studies in which the CAT was compared to various questionnaires [5, 9–15, 17–20, 22, 23, 25, 27–29, 31, 41, 42] (table 3). The patients were in stable state when measuring this property unless it was unreported in the studies. CAT longitudinal validity was reported in six studies and the interventions or events consisted of pulmonary rehabilitation or recovery from an acute exacerbation [9, 11–13, 16, 24] (table 3).

19 studies reported known groups validity and the categories that could differ in HRQoL varied [5, 9, 11, 14, 17, 20–23, 25–29, 33, 36, 37, 40, 43] (table 4). The CAT score was statistically different ($p<0.05$) in the following categories: COPD GOLD grades [9, 14, 17, 20, 22, 25, 27, 37]; primary care physician-rated COPD GOLD grades [20, 22]; healthy individuals *versus* individuals diagnosed with COPD [21, 28, 33, 36, 40]; infrequent exacerbators *versus* frequent exacerbators (defined as no acute exacerbation in the last 6 months *versus* acute exacerbation in the last 6 months; 0–1, 2–4 or >4 exacerbations per year; and <2 or ⩾2 exacerbations per year) [9, 23, 26, 33, 43]; exacerbation state *versus* stable state [5, 9, 20]; body mass index (BMI) (defined as BMI <18.5 $kg \cdot m^{-2}$, BMI ⩾18.5 and <23 $kg \cdot m^{-2}$, or BMI ⩾23 $kg \cdot m^{-2}$) [25]; and mMRC score [9, 27, 37]. The CAT score was not statistically different ($p \geqslant 0.05$) in the following categories: sex [11, 20, 21, 29]; age (defined as ⩽65 years *versus* >65 years) [20, 29]; current smokers *versus* nonsmokers [23]; and comorbidities (defined as 0, 1–2 or ⩾3 comorbidities) [20, 25].

TABLE 3 COPD assessment test (CAT) convergent and longitudinal validity

Instrument	Studies n	Convergent validity[#]		Longitudinal validity[¶] with study details		
		Pearson's correlation range	Spearman's correlation range	Pearson's correlation range	Type of intervention/event	Duration weeks
Disease-specific questionnaires						
SGRQ-C	7	0.69–0.82	0.64	0.63	Exacerbation recovery	4
CCQ	4	NA	NA	0.60	Exacerbation recovery	6
		0.68–0.78	0.64–0.76	0.13	Pulmonary rehabilitation	8
SGRQ	5	0.72–0.74	0.65–0.84	0.36	Pulmonary rehabilitation	8
CRQ	5[§]	-0.48– -0.33	NA	-0.50– -0.38	Pulmonary rehabilitation	6–8
General quality of life questionnaires						
SF-36 (general health)	1	0.58	NA			
SF-12 (physical component)	2	-0.60	-0.65			
SF-12 (mental component)	2	-0.34	-0.58			
SF-6D	1	-0.53	NA			
Clinical and physiological measures						
mMRC	11	0.29–0.61	0.42–0.61	0.20	Pulmonary rehabilitation	8
6MWT	4	-0.37– -0.27	-0.37– -0.24			
FEV_1 % predicted	10	-0.55– -0.17	-0.56– -0.23			

COPD: chronic obstructive pulmonary disease; SGRQ: St George's Respiratory Questionnaire; SGRQ-C: SGRQ for COPD; CCQ: Clinical COPD Questionnaire; CRQ: Chronic Respiratory Questionnaire; SF: Short Form; mMRC: modified Medical Research Council dyspnoea scale; 6MWT: 6-min walk test; FEV_1: forced expiratory volume in 1 s; NA: not applicable. [#]: correlation between the total CAT score and the score of another instrument that measures a similar construct. [¶]: correlation between the change in total CAT score and the change in score of another instrument over time with an intervention/event. [§]: JONES [12], KON [24] and AL MOAMARY [31] had individual CRQ domain scores correlated to the total CAT score; therefore, individual correlations of the different domains were averaged to give a total CRQ correlation to the total CAT score.

Responsiveness

10 studies examined responsiveness of the CAT (table 5) [9, 11–14, 16, 24, 26, 32, 38]. The CAT was responsive to pulmonary rehabilitation in four studies and the range of mean change in CAT score was -3.0– -2.2 units at the end of the intervention [11, 12, 16, 24]. The majority of patients improved with pulmonary rehabilitation, whether it lasted for 8 weeks [11, 16, 24] or 6 weeks [12]. Responsiveness was reassessed from the end of an 8-week intervention to 6 months later, and it was determined that the total CAT score deteriorated slightly from the end of rehabilitation [16].

The majority of patients' total CAT score deteriorated with onset of an exacerbation: mean CAT score increased by 4.7 units ($p<0.001$) [26]. However, with exacerbation recovery on treatment, patients' CAT score improved over 2, 4, 6 or 12 weeks [9, 12, 13, 32, 38] (table 5). Patients with and without depressive symptoms improved their CAT scores with exacerbation recovery, but those patients without depressive symptoms had greater improvement in CAT scores over 6 weeks [38].

Minimum clinically important difference

Four studies attempted to determine the MCID of the CAT [11, 12, 14, 35]. Three studies employed the anchor-based approach to calculate MCID. Of those, two studies found that their range of external responses (*e.g.* much better, a little better, no different or a little worse; responders or nonresponders) was not used by an equal proportion of patients; therefore, it was not possible to determine the MCID [11, 12]. The other study identified a decrease of 2 units as an MCID estimate [35]. Two studies used the distribution-based approach and determined an MCID for the CAT of 3.76 units [14] and a decrease that ranged from 3.3 to 3.8 units [35].

Risk of bias within and across studies

The methodological quality of the studies was mostly rated fair (30 (83%) studies), with one rated poor, four rated good and one rated excellent, according to the COSMIN checklist. Internal consistency, test–retest and convergent and known groups validity were evaluated in studies of fair and good methodological

DOI: 10.1183/09031936.00025214

TABLE 4 COPD assessment test (CAT) known groups validity[#]

COPD GOLD grades	GOLD I	GOLD II	GOLD III	GOLD IV		p-value
Agusti [9]	12.4±NR	14.4±NR	20.9±NR	20.9±NR		0.01
Ghobadi [17]	14.6±NR	21.1±NR	25.1±NR	28.4±NR		0.001
Jones [20]	16.2±8.8	16.3±7.9	19.3±8.2	22.3±8.7		0.0001[¶]
Jones [22]	15.9±8.6	16.5±8.0	19.2±8.1	22.4±9.0		0.0001[¶]
Kwon [25]	16.5±NR	16.0±NR	19.0±NR	21.3±NR		0.001
Marchand [27]	14.3±NR	16.6±NR	19.1±NR	23.7±NR		0.0001[¶]
Okutan [37]	9.2±6.2	18.2±8.1	20.9±8.2	22.5±0.79		0.001[+]
Tsiligianni [14]	10.1±NR	10.5±NR	13.0±NR	26.0±NR		0.0023[¶]
Healthy individuals *versus* COPD	**No COPD**	**COPD**				**p-value**
Gao [33]	4.0±2.1	10.0±5.3				0.001
Jones [21] Turkey	8.1 (7.6–8.6)[§]	20.9 (19.6–22.2)[§]				0.0001
Jones [21] Arabic countries	5.4 (5.2–5.6)[§]	16.6 (15.5–16.8)[§]				0.0001
Miyazaki [28]	9.4±6.6	12.4±8.3				0.01
Nishimura [36]	5.8±4.4	7.3±5.2				0.001
Raghavan [40]	6.9±6.2	9.2±6.6				0.01
COPD status state	**Stable**	**Exacerbation**				**p-value**
Agusti [9]	15.8±8.1	22.4±8.4				0.01
Jones [5]	16.0±NR	20.7±NR				0.0001
Jones [20]	17.2±8.3	21.3±8.4				0.0001
mMRC score	**0**	**1**	**2**	**3**	**4**	**p-value**
Agusti [9]	8.7±NR	14.8±NR	18.3±NR	24.9±NR	28.1±NR	0.01
Marchand [27]	NA	12.7±NR[f]	19.9±NR	25.5±NR	25.2±NR	0.0001
Okutan [37]	7.7±2.2	13.9±6.1	21.7±5.9	27.5±2.2	27.3±9.7	0.021[##]

Data are presented as subcategory total CAT score mean±SD, when available from the study, unless otherwise stated. COPD: chronic obstructive pulmonary disease; GOLD: Global Initiative for Chronic Obstructive Lung Disease; mMRC: modified Medical Research Council dyspnoea scale; NR: not reported; NA: not applicable. [#]: ability of the CAT to differentiate scores between subcategories known to vary on health status; [¶]: GOLD I and GOLD II are not significant; [+]: GOLD II, GOLD III and GOLD IV are not significant; [§]: data are presented as mean (95% CI); [f]: total CAT score for 0–1 mMRC score subcategory; [##]: mMRC scores 2, 3 and 4 are not significant.

quality; longitudinal validity and responsiveness were examined in studies of fair methodological quality; MCID was evaluated in studies of fair and excellent methodological quality; and concurrent validity was assessed in studies of poor and fair methodological quality. Language bias was minimal across the five general databases; English, French and Spanish filtered citations as a percentage of the overall citations retrieved were 95.4%.

Discussion

The goal in designing a HRQoL tool is for it to accurately and reliably measure HRQoL, and this review identifies the CAT's adequacy as a HRQoL instrument. Although several articles have been published on the CAT, this is the first study to systematically review the available literature evaluating the CAT's psychometric properties in a defined population of patients with COPD.

The psychometric properties of the CAT are both acceptable and favourable. The CAT is reliable: the interrelatedness of the eight items within the questionnaire indicate high internal consistency, while the stability of CAT total scores after repeated administrations confirms its reproducibility over time. Furthermore, the CAT demonstrates good construct validity through convergent, longitudinal and known groups validity. It is evident that the CAT is responsive and able to detect a change in score over time: the CAT score improved with pulmonary rehabilitation and exacerbation recovery on treatment, and the CAT score deteriorated with the onset of an exacerbation. Only one study reliably identified a decrease of 2 units as the MCID estimate through an anchor-based approach.

The majority of the studies (83%) did not report missing data on the CAT, and in some studies only subjects with complete data were analysed, leading to a significant proportion of patients being excluded. Future research needs to examine the floor and ceiling effects of the CAT, as they have only been addressed in two populations. Overall, the methodological quality of the studies was rated fair. All of the

TABLE 5 COPD assessment test (CAT) responsiveness[#] to pulmonary rehabilitation and exacerbation recovery on treatment

Study	Duration weeks	Study defined groups	CAT score			p-value
			Pre	Post	Change[¶]	
Pulmonary rehabilitation						
Dodd [11]	8	NA	20.5±7.4	17.5±7.7	-2.9±NR	0.001
Dodd [16]	8	NA	22.1±7.5	19.2±7.6	-2.9±NR	0.001
Jones [12]	6	NA	17.9±6.5	15.7±6.9	-2.2±NR	0.002
Kon [24]	8	NA	21.8±7.2	18.8±7.4	-3.0±NR	0.001
Exacerbation recovery on treatment						
Agusti [9][+]	2	Much better	NR	NR	-8.9±9.1	NR
		Considerably improved	NR	NR	-4.8±6.0	NR
		Slightly improved	NR	NR	-4.6±4.7	NR
		No change	NR	NR	0	NR
		Slightly worse	NR	NR	1.6±NR	NR
		Quite a lot worse	NR	NR	4.7±NR	NR
Feliz-Rodriguez [32]	12	NA	22.8±4.9	15.6±4.5	-6.5±3.9	0.001
Jones [12]	4	NA	21.4±7.7	19.9±7.7	-1.4±5.3	0.03
Miravitlles [13]	6	NA	22.0±7.0	12.1±5.9	-9.9±5.1	0.001
Papaioannou [38]	6	Depressive symptoms	NR	NR	-7.0±NR	0.012
		No depressive symptoms	NR	NR	-11.0±NR	0.012

Data are presented as mean±SD, when available from the study, unless otherwise stated. COPD: chronic obstructive pulmonary disease; NA: not applicable; NR: not reported. [#]: ability of the CAT to detect a change in score over time between interventions. [¶]: negative change in CAT score is an improvement in CAT score (or health-related quality of life (HRQoL)); positive change in CAT score is a deterioration in CAT score (or HRQoL). [+]: categories of change determined based on the HRQoL transition.

psychometric properties, except concurrent validity, were assessed in studies rated fair, good or excellent methodological quality.

Although the diversity of the studies retrieved resulted in examination of psychometric properties in many COPD populations, allowing for generalisation of the results, further assessment of validity and responsiveness needs to be completed in specific patient populations (*e.g.* females, younger age groups, mild disease) to assess the CAT's capability of discriminating between these groups. It would be of great utility to evaluate the predictive validity of the CAT to determine whether it can predict future clinical outcomes (*e.g.* mortality, hospital admission, disease progression or exacerbation). Similarly, while there is no correct manner to determine the MCID, several studies must attempt to provide estimates, so that multiple results can be combined to provide a true value (see online supplementary material for methods of calculating MCID). Moreover, linked to the development of the questionnaire was a grading system based on the CAT score, for which the development group proposed potential management considerations according to each scenario. An investigation of the impact of the CAT on the quality of the primary care consultations in patients with COPD has been conducted in a randomised controlled trial, although research needs to be advanced in this area due to the study's methodological limitations [44].

Strengths of this review include exhaustive search strategies across multiple databases, independent study retrieval, screening, data extraction and assessment of study quality. Data were insufficient to perform a meta-analysis: the variety of outcomes studied, methodological heterogeneity and diverse study populations prevented the generation of a common summary effect of a specific psychometric property, so a meta-analysis was deemed inappropriate.

There were, however, limitations. Investigating heterogeneity was not possible, but heterogeneity between studies is to still be expected. Likewise, formal assessments evaluating publication bias through a funnel plot could not be conducted. Although no language filters were applied in the search strategy, an assessment of the language bias indicated that a minor language bias could be of concern; however, the findings appear appropriate given that the CAT's development was studied in certain languages (*e.g.* English) but not in others. Limitations of the data from the included studies must also be considered (*e.g.* standard deviations for known groups validity were not presented in the majority of studies), although they reflect the dearth of literature available on the CAT.

DOI: 10.1183/09031936.00025214

Conclusion

This review employed rigorous methodology to provide a comprehensive overview of the CAT's psychometric properties in patients with COPD. The studies support the reliability and validity of the CAT and that the tool is responsive to interventions, although the MCID remains debatable. Since the CAT demonstrates good performance and is a simple and quick tool that assesses the HRQoL in patients with COPD, there is a growing interest in its use in clinical practice. Studies are needed to evaluate the use of this questionnaire for the symptomatic assessment of patients with COPD in the new GOLD classification. It cannot be assumed that the CAT behaves similarly with different patient population characteristics; thus, studies must also attempt to determine the validity of the CAT in females, patients with mild disease or individuals at risk, and younger and older patients.

References

1 Vestbo J, Hurd SS, Agusti AG, *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2013; 187: 347–365.
2 Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management, and Prevention of COPD. Revised 2011.
3 Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993; 118: 622–629.
4 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; 18: 208–215.
5 Jones PW, Harding G, Berry P, *et al.* Development and first validation of the COPD assessment test. *Eur Respir J* 2009; 34: 648–654.
6 Weldam SW, Schuurmans MJ, Liu R, *et al.* Evaluation of quality of life instruments for use in COPD care and research: a systematic review. *Int J Nurs Stud* 2013; 50: 688–707.
7 Terwee CB, Mokkink LB, Knol DL, *et al.* Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012; 21: 651–657.
8 Mokkink LB, Terwee CB, Patrick DL, *et al.* The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; 63: 737–745.
9 Agusti A, Soler JJ, Molina J, *et al.* Is the CAT questionnaire sensitive to changes in health status in patients with severe COPD exacerbations? *COPD* 2012; 9: 492–498.
10 da Silva GP, Morano MT, Viana CM, *et al.* Portuguese-language version of the COPD assessment test: validation for use in Brazil. *J Bras Pneumol* 2013; 39: 402–408.
11 Dodd JW, Hogg L, Nolan J, *et al.* The COPD assessment test (CAT): response to pulmonary rehabilitation. A multicentre, prospective study. *Thorax* 2011; 66: 425–429.
12 Jones PW, Harding G, Wiklund I, *et al.* Tests of the responsiveness of the COPD assessment test following acute exacerbation and pulmonary rehabilitation. *Chest* 2012; 142: 134–140.
13 Miravitlles M, Garcia-Sidro P, Fernandez-Nistal A, *et al.* Course of COPD assessment test (CAT) and clinical COPD questionnaire (CCQ) scores during recovery from exacerbations of chronic obstructive pulmonary disease. *Health Qual Life Outcomes* 2013; 11: 147.
14 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: 20.
15 Tsuda T, Suematsu R, Kamohara K, *et al.* Development of the Japanese version of the COPD assessment test. *Respir Investig* 2012; 50: 34–39.
16 Dodd JW, Marns PL, Clark AL, *et al.* The COPD assessment test (CAT): short- and medium-term response to pulmonary rehabilitation. *COPD* 2012; 9: 390–394.
17 Ghobadi H, Ahari SS, Kameli A, *et al.* The relationship between COPD assessment test (CAT) scores and severity of airflow obstruction in stable COPD patients. *Tanaffos* 2012; 11: 22–26.
18 Horita N, Yomota M, Sasaki M, *et al.* Evaluation of the chronic obstructive pulmonary disease assessment test in Japanese outpatients. *Clin Respir J* 2014; 8: 213–219.
19 Hwang YI, Jung KS, Lim SY, *et al.* A validation study for the Korean version of chronic obstructive pulmonary disease assessment test (CAT). *Tuberc Respir Dis* 2013; 74: 256–263.
20 Jones PW, Brusselle G, Dal Negro RW, *et al.* Properties of the COPD assessment test in a cross-sectional European study. *Eur Respir J* 2011; 38: 29–35.
21 Jones PW, Shahrour N, Nejjari C, *et al.* Psychometric evaluation of the COPD assessment test: data from the BREATHE study in the Middle East and North Africa region. *Respir Med* 2012; 106: Suppl. 2, S86–S99.
22 Jones PW, Brusselle G, Dal Negro RW, *et al.* Patient-centred assessment of COPD in primary care: experience from a cross-sectional study of health-related quality of life in Europe. *Prim Care Respir J* 2012; 21: 329–336.
23 Kelly JL, Bamsey O, Smith C, *et al.* Health status assessment in routine clinical practice: the chronic obstructive pulmonary disease assessment test score in outpatients. *Respiration* 2012; 84: 193–199.
24 Kon SS, Clark AL, Dilaver D, *et al.* Response of the COPD assessment test to pulmonary rehabilitation in unselected chronic respiratory disease. *Respirology* 2013; 18: 974–977.
25 Kwon N, Amin M, Hui DS, *et al.* Validity of the COPD assessment test translated into local languages for Asian patients. *Chest* 2013; 143: 703–710.
26 Mackay AJ, Donaldson GC, Patel AR, *et al.* Usefulness of the chronic obstructive pulmonary disease assessment test to evaluate severity of COPD exacerbations. *Am J Respir Crit Care Med* 2012; 185: 1218–1224.
27 Marchand E, Maury G. Évaluation du COPD assessment test (CAT) chez des patients BPCO en état stable [Evaluation of the COPD assessment test in patients with stable COPD]. *Rev Mal Respir* 2012; 29: 391–397.
28 Miyazaki M, Nakamura H, Chubachi S, *et al.* Analysis of comorbid factors that increase the COPD assessment test scores. *Respir Res* 2014; 15: 13.
29 Zhou QT, Mei JJ, He B, *et al.* Chronic obstructive pulmonary disease assessment test score correlated with dyspnea score in a large sample of Chinese patients. *Chin Med J* 2013; 126: 11–15.

30 Al Moamary MS, Al-Hajjaj MS, Tamim HM, *et al.* The reliability of an Arabic translation of the chronic obstructive pulmonary disease assessment test. *Saudi Med J* 2011; 32: 1028–1033.
31 Al Moamary MS, Tamim HM, Al-Mutairi SS, *et al.* Quality of life of patients with chronic obstructive pulmonary disease in the Gulf Cooperation Council countries. *Saudi Med J* 2012; 33: 1111–1117.
32 Feliz-Rodriguez D, Zudaire S, Carpio C, *et al.* Evolution of the COPD assessment test score during chronic obstructive pulmonary disease exacerbations: determinants and prognostic value. *Can Respir J* 2013; 20: e92–e97.
33 Gao Y, Hou Q, Wang H. Assessment of health status in patients with newly diagnosed chronic obstructive pulmonary disease. *PLoS One* 2013; 8: e82782.
34 Idrees M, Koniski ML, Taright S, *et al.* Management of chronic obstructive pulmonary disease in the Middle East and North Africa: results of the BREATHE study. *Respir Med* 2012; 106: Suppl. 2, S33–S44.
35 Kon SS, Canavan JL, Jones SE, *et al.* Minimum clinically important difference for the COPD assessment test: a prospective analysis. *Lancet Respir Med* 2014; 2: 195–203.
36 Nishimura K, Mitsuma S, Kobayashi A, *et al.* COPD and disease-specific health status in a working population. *Respir Res* 2013; 14: 61.
37 Okutan O, Tas D, Demirer E, *et al.* Evaluation of quality of life with the chronic obstructive pulmonary disease assessment test in chronic obstructive pulmonary disease and the effect of dyspnea on disease-specific quality of life in these patients. *Yonsei Med J* 2013; 54: 1214–1219.
38 Papaioannou AI, Bartziokas K, Tsikrika S, *et al.* The impact of depressive symptoms on recovery and outcome of hospitalised COPD exacerbations. *Eur Respir J* 2013; 41: 815–823.
39 Polatli M, Ben Kheder A, Wali S, *et al.* Chronic obstructive pulmonary disease and associated healthcare resource consumption in the Middle East and North Africa: the BREATHE study. *Respir Med* 2012; 106: Suppl. 2, S75–S85.
40 Raghavan N, Lam YM, Webb KA, *et al.* Components of the COPD assessment test (CAT) associated with a diagnosis of COPD in a random population sample. *COPD* 2012; 9: 175–183.
41 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* 2012; 9: 12–15.
42 Uzaslan E, Mahboub B, Beji M, *et al.* The burden of chronic obstructive pulmonary disease in the Middle East and North Africa: results of the BREATHE study. *Respir Med* 2012; 106: Suppl. 2, S45–S59.
43 Varol Y, Ozacar R, Balci G, *et al.* Assessing the effectiveness of the COPD assessment test (CAT) to evaluate COPD severity and exacerbation rates. *COPD* 2014; 11: 221–225.
44 Gruffydd-Jones K, Marsden HC, Holmes S, *et al.* Utility of COPD assessment test (CAT) in primary care consultations: a randomised controlled trial. *Prim Care Respir J* 2013; 22: 37–43.

DOI: 10.1183/09031936.00025214