VALUE IN HEALTH 14 (2011) 307–315

available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/jval

Brief, Valid Measures of Dyspnea and Related Functional Limitations in Chronic Obstructive Pulmonary Disease (COPD)

Susan E. Yount, PhD[a,], Seung W. Choi, PhD[a], David Victorson, PhD[a], Bernice Ruo, MD, MAS[b], David Cella, PhD[a], Susan Anton, MPH, DrPH[c], Alan Hamilton, PhD[d]*

[a] *Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA*
[b] *Division of General Internal Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA*
[c] *Boehringer-Ingelheim Pharmaceuticals, Inc., Ridgefield, CT, USA*
[d] *Boehringer Ingelheim (Canada) Ltd., Burlington, ON, Canada*

Keywords:
Computer adaptive testing
Chronic obstructive pulmonary disease
Dyspnea
Measurement

A B S T R A C T

Objective: Chronic obstructive pulmonary disease (COPD) is a progressive disease with functional decline leading to disability. Dyspnea, the prominent symptom, can be measured using existing measures, but a lack of consensus about standardization of dyspnea measurement remains. We examined the psychometric performance of two item-response theory–based (IRT) measures of dyspnea and related functional limitations (FLs) in patients with COPD and simulated computerized adaptive testing (CAT) of the banks to determine the number of questions required to achieve high precision.

Methods: A total of 102 patients completed banks measuring dyspnea and FLs (33 items), from which the 10-item dyspnea and FL short forms were scored as well as other self-report measures of respiratory and physical function and emotional distress. A subset of patients completed the banks 7 to 10 days later. Pulmonary function test results were obtained from medical charts.

Results: The 33-item banks and 10-item short forms had excellent internal consistency (alphas >0.9) and test-retest reliability (intraclass correlation coefficients >0.89). Patients sorted by severity level on the Medical Research Council scale were differentiated by item banks ($P < 0.001$) and the short forms ($P < 0.01$). The banks and short forms were also associated with related measures of dyspnea (e.g., Baseline Dyspnea Index, $r = 0.47–0.53$), physical function (e.g., 36-Item Short Form Health Survey, $r = -0.83$ to -0.86) and forced expiratory volume in 1 second ($r = -0.32$ to -0.35). On average, CAT required 4 and 5 items for accurate measurement of dyspnea and FLs, respectively.

Conclusion: The Functional Assessment of Chronic Illness Therapy–Dyspnea short forms and banks provide options for brief, psychometrically sound measures of dyspnea and/or FLs in COPD.

Funding: Susan Anton, MPH, DrPH and Alan Hamilton, PhD are employed by Boehringer-Ingelheim Pharmaceuticals. This work was supported in part by a grant from Boehringer-Ingelheim Pharmaceuticals. Susan E. Yount, PhD, Seung W. Choi, PhD, David Victorson, PhD, Bernice Ruo, MD, MAS have nothing to disclose.

* *Address correspondence to:* Susan Yount, Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, 625 North Michigan Avenue, Suite 2700, Chicago, IL 60611, USA.
E-mail: s-yount@northwestern.edu.

1098-3015/$36.00 – see front matter Copyright © 2011, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.
doi:10.1016/j.jval.2010.11.009

Introduction

Chronic obstructive pulmonary disease (COPD) is a major public health problem in the United States, affecting approximately 24 million people and millions more around the world [1]. It is the fourth leading cause of death in the world and is projected to be ranked fifth in burden of disease worldwide by the year 2020 [2]. The National Heart, Lung, and Blood Institute of the National Institutes of Health estimates that the economic burden of COPD in the United States was $49.9 billion in direct and indirect costs in 2010 [3]. COPD is also a major source of disability and impaired health-related quality of life (HRQL) [4].

COPD includes chronic lung diseases (i.e., chronic bronchitis, emphysema) that are characterized by progressive obstruction of the airflow into and out of the lungs and increased dyspnea [5]. It can be extremely disabling and limit an individual's function across several important areas including upper extremity function, balance, strength, exercise performance, and self-reported daily activities such as vacuuming, climbing stairs, and engaging in recreation [6]. Clinical evaluations typically include both pulmonary function tests (PFTs) (e.g., spirometry) and patient-reported symptoms.

The diagnosis of COPD is often confirmed by pulmonary functioning (e.g., PFTs), including spirometry, which measures airflow out of fully inflated lungs over time in liters [7]. However, objective measures of pulmonary functioning, such as PFTs, do not correlate particularly well with patient reports of COPD-associated symptoms [8]. One of the most prominent patient-reported symptoms of COPD is dyspnea, a subjective experience that can only be measured from the patient's perspective. Different people will have different thresholds for noticing, reporting, and rating the severity of the symptom. Other factors, such as the level of activity, presence and severity of comorbid disease, environmental conditions, and anxiety, may influence dyspnea ratings as well. The lack of association with PFTs is, therefore, not surprising. It does, however, underscore the need for psychometrically sound measures of dyspnea and related functional limitations (FLs) for use in clinical trials.

Patient-reported dyspnea, however, has proved difficult to measure in a comprehensive and psychometrically sound manner. Several general and disease-specific self-report instruments have previously been developed, but criticism in the research and clinical literature has highlighted an overall lack of consensus and standardization [9].

Ideally, a measure of dyspnea for use in clinical trials should be brief, self-administered, standardized, created with cross-cultural input, easily understood by patients and physicians, and responsive to changes in severity across the disease spectrum. In response to the noted limitations of existing measures, we undertook the task of developing a new and psychometrically improved self-report measure of COPD-related dyspnea and related FLs. The first task was to construct a patient-driven conceptual model to elucidate the relationships between dyspnea and other important factors that not only influence it, but are also influenced by it [10]. This model served as the basis of our creation of a new self-reported tool to measure dyspnea and related FLs.

We then developed a patient-derived conceptual framework and an item-response theory (IRT)–based item bank of dyspnea severity and related FLs [11]. To maximize geographic diversity as well as efficiency of data collection, the framework was tested using an online Internet panel of COPD patients across the United States, 608 respondents at baseline and a subset of 236 respondents at 7-10 days' follow up [11]. By taking advantage of modern test development approaches, including IRT [12,13]. we were able to "calibrate" patient item responses along a severity continuum that represents a given latent trait, such as dyspnea. A calibrated item bank allows one to estimate a person's location along a continuum of increasingly difficult items and determine which items provide maximum information on a given concept [14,15]. This enables the creation of fixed-length short forms from a pool of calibrated items as well as computerized adaptive testing (CAT), a dynamic measurement platform that allows items to be selected based on an individual's previous responses. CAT administration reduces test length dramatically while improving measurement quality [16–18]. This step resulted in the development of two 33-item item banks for dyspnea and FLs, and from those banks, a 20-item short-form measure of dyspnea (10 items) and its FLs (10 items) was created, using a methodology described elsewhere [11].

As the final step in the development and testing of a new measure of COPD-related dyspnea and FLs, we now report on the validity of the 20-item Functional Assessment of Chronic Illness Therapy (FACIT)–Dyspnea scale, tested with a different sample of 102 individuals with a diagnosis of COPD who were obtained from outpatient pulmonary medicine and internal medicine clinics. The 20 items include 10 that assess dyspnea severity (Dyspnea-10) and 10 that assess dyspnea related FLs (FL-10). We compare and contrast the performance of these two short forms with the item banks for dyspnea and FLs (33 items each) and provide information from a simulated CAT for consideration by researchers/clinicians when making choices about method of administration.

Methods

Participants

Patients were eligible for the study if they had a diagnosis of COPD, were fluent in English, were at least 18 years of age, and had no current diagnosis of psychosis or dementia. To balance recruitment between patients with mild versus more severe dyspnea, patients were recruited from general internal medicine clinics, where those with milder cases of COPD tend to be seen, as well as specialized pulmonology clinics and a PFT laboratory, where individuals with more severe COPD tend to receive their treatment.

Procedure

After the informed consent process, patients completed a sociodemographic questionnaire to obtain basic background information. Clinical information was obtained from patients (e.g., diagnosis, smoking history, previous/existing comorbid

conditions, medication use, recency/severity of exacerbations), and thorough medical chart review (e.g., PFT results). The most recent PFT results for each patient were obtained from patients' medical records. Patients then completed a battery of questionnaires (described later) administered via touch-screen laptop computer. Seven to 10 days after completion of the baseline questionnaires, patients completed a brief clinical form to capture any change in clinical status since baseline and repeated the dyspnea and FL item banks. If patients were scheduled to return to the clinic within the assessment window, they completed the assessments via touch-screen laptop computer; if not, the questionnaires were completed via telephone interview.

Measures

In addition to completing sociodemographic questionnaires, patients completed a baseline assessment battery that consisted of the following measures.

FACIT-Dyspnea Item Banks
Patients completed item banks for dyspnea and FLs, each with 33 items [11]. Included in these 33-item banks were the items that comprise the FACIT-Dyspnea short forms (Dyspnea-10 and FL-10). All dyspnea items include a 4-point rating scale (no shortness of breath = 0; mildly short of breath = 1; moderately short of breath = 2; severely short of breath = 3; or I did not do this in the past 7 days = 4).

Although respondents in this study did not complete short forms as stand-alone measures, the forms are constructed so that respondents are presented with 10 common tasks and asked to rate the severity of their dyspnea when completing these tasks over the past 7 days using the same response scale as the banks. Those who report not doing the task are asked to report whether it was attributable to dyspnea (shortness of breath) or simply because they did not have an opportunity to do the task in the past week. If the response is because of dyspnea (i.e., "stopped trying" or "knew could not do it because of shortness of breath"), the response is treated the same as the response "severely short of breath." Otherwise, the response is treated as missing (i.e., not included for scoring). Next, using the same 10 items, respondents are asked to rate the amount of difficulty they experienced when doing these tasks on a 4-point Likert scale (no difficulty = 0; a little difficulty = 1; some difficulty = 2; much difficulty = 3). Respondents are not presented with items that were rated as "I did not do" on the dyspnea severity subscale; however, those who report dyspnea-related task avoidance are scored as a 3 (much difficulty) on the FL subscale, whereas nonexposure task avoiders are not scored. The FACIT-Dyspnea short forms are scored such that a high score represents high levels of dyspnea or FLs.

Baseline Dyspnea Index (BDI) [19,20]
The BDI is a multidimensional instrument for measuring breathlessness based on three aspects of dyspnea: 1) functional impairment, 2) magnitude of task, and 3) magnitude of effort. The interviewer-administered version of the BDI used in this study grades the three categories from 0 (severe) to 4 (unimpaired); a baseline focal score ranging from 0 to 12 is obtained by adding the three individual scores together.

Medical Research Council (MRC) Dyspnea Scale [21]
The MRC dyspnea scale consists of five statements about perceived dyspnea, and patients select the grade that applies to the impact of dyspnea on their mobility: grade 1, "I only get breathless with strenuous exercise"; grade 2, "I get short of breath when hurrying on the level or up a slight hill"; grade 3, "I walk slower than people of the same age on the level because of breathlessness or have to stop for breath when walking at my own pace on the level"; grade 4, "I stop for breath after walking 100 yards or after a few minutes on the level"; grade 5, " I am too breathless to leave the house."

36-Item Short Form Health Survey (SF-36) [22]
The SF-36 is a 36-item questionnaire with eight subscales (physical functioning, social functioning, role limitations because of a physical problem, role limitations because of an emotional problem, mental health, vitality, pain, and general health perception) and one single item to measure health change. The eight subscales form two distinct higher order summary scales: the physical component summary scale, and the mental summary component scale. Transformed scores on all scales range from 0 to 100, with higher scores indicating better HRQL.

Hospital Anxiety and Depression Scale (HADS) [23]
The HADS is a 14-item self-report inventory to measure anxiety and depression. The scale consists of seven items for anxiety (HADS-A) and seven for depression (HADS-D). The items are scored on a 4-point scale from 0 (not present) to 3 (considerable). The item scores are summed, yielding subscale scores on the HADS-A and the HADS-D ranging from 0 to 21.

Chronic Respiratory Questionnaire–Self-Administered Standardized (CRQ-SAS) Version (CRQ-SAS) [24,25]
The Chronic Respiratory Questionnaire (CRQ) is a respiratory-specific health status instrument originally developed for patients with COPD that assesses four aspects of COPD-related quality of life: dyspnea, fatigue, emotional function, and mastery, each of which is scored on a 7-point Likert-type scale. In this study, analyses focused on the five-item dyspnea subscale only because that was most relevant to the validation of the FACIT-Dyspnea scale.

Data analysis

The analyses described in the following were conducted on the 33-item dyspnea and FL banks and the 10-item short forms. Three respondents with excessive missing responses (>90%) on either of the banks were removed from the analyses.

Reliability
Internal consistency analyses based on baseline data were conducted for the item banks and short forms of the new dyspnea measure using Cronbach's alpha coefficients. A Cronbach's alpha of 0.70 or greater was considered necessary to minimally meet the recommended standards. Test-retest reliability was evaluated by calculating an ICC for the item banks and short forms at baseline and 7 to 10 days post-baseline using data from

the full item bank testing. Both Cronbach's alpha and the ICC have a range of possible values from 0.0 to 1.0.

Discriminant validity

A frequency distribution was calculated on the MRC dyspnea scale at baseline. Cross-sectional analyses of item bank and the short form scores at baseline focused on differentiating definable groups defined according to the MRC dyspnea scale. We used analysis of variance and/or pairwise *t* tests to compare mean dyspnea and FL bank and short form scores and MRC dyspnea scale categories, with the expectation that patients with higher MRC dyspnea scale scores would report greater dyspnea/FLs. Effect sizes were calculated for group comparisons to provide an indication of the clinical significance of group differences. We used eta squared (η^2), which is the proportion of the total variance in scores that is explained by MRC dyspnea scale categories. For η^2, an effect size of 0.14 is considered a large effect [26]. For pairwise comparisons of adjacent groups, we used Cohen's *d* effect size measure. For Cohen's *d*, an effect size of 0.8 and higher is considered a large effect.

Convergent validity

The associations between the item bank and short form scores, CRQ-SAS dyspnea subscale, HADS, and subscales of the SF-36 were evaluated using correlational analyses. The FACIT-Dyspnea scores (banks and short forms) were expected to be most highly correlated with the CRQ-SAS dyspnea subscale and physical functioning subscales of the SF-36 and correlated to a lesser extent with the socioemotional subscales of the SF-36 and HADS.

Concurrent validity

The association between item bank scores and short form scores and the BDI scores were evaluated using correlations. Because the well-established BDI has been extensively used to assess dyspnea, the dyspnea scores and the BDI scores were expected to correlate.

Post hoc CAT simulations

CAT simulations can provide useful information regarding the potential effectiveness of an item bank under CAT administration for a population of interest. Thus, we simulated CAT based on baseline ratings of dyspnea and FLs from the FACIT-Dyspnea scale (n = 99) using the Firestar CAT simulator [27]. The minimum and maximum numbers of items to administer were set at 3 and 12, respectively, with a minimum standard error criterion set at 0.3.

Results

Sample

Of the 103 consented participants, 102 reported sociodemographic data. Three respondents were removed because of excessive missing responses (fewer than five valid responses). As described in Table 1, this sample was nearly equally divided in terms of sex (53% female), it was overwhelmingly non-Hispanic (99%), and the majority were white (85%). The mean (SD) age was 70 (10) years, consistent with 63% of the sample being retired. The participants were overall well educated, with 28% having some college, 25% college graduates, and 29% having earned a graduate degree. Twenty percent of the participants reported an all-source family income of greater than $60,000, and 35% reported a family income of $100,000 or more.

Table 1 – Sample sociodemographic characteristics (N = 102).

Variable	Value (%)
Age, mean ± SD (range)	70 ± 10.3 (40–92)
Sex (male)	48 (47%)
Hispanic/Spanish/Latino origin	1 (1%)
Race/ethnicity	
White	87 (85%)
Black	9 (9%)
Other	6 (6%)
Marital status	
Married	54 (53%)
Divorced	22 (22%)
Widowed	16 (16%)
Single	10 (10%)
Living arrangement	
Alone	34 (33%)
With other adult(s) and/or dependents	68 (67%)
Highest grade in school completed	
Less than 12th grade	4 (4%)
High school graduate or equivalent	14 (14%)
Some college	29 (28%)
College grad or advanced degree	55 (54%)
Current occupational status	
Homemaker	6 (6%)
Unemployed	2 (2%)
Retired	64 (63%)
On disability	5 (5%)
Full-time or part-time employed	25 (25%)
Family income	
≤$19,999	13 (14%)
$20,000–$39,999	14 (15%)
$40,000–$59,999	15 (16%)
$60,000–$99,999	18 (20%)
≥$100,000	32 (35%)

As detailed in Table 2, which describes the sample's clinical characteristics, participants reported having a diagnosis of COPD for a median of 3 years (range: 0–37 years). Fifteen percent of the sample was smoking at the time of their participation, and 72% had smoked tobacco products in the past. Participants reported a range of current and previous comorbid conditions. Ninety percent were using an inhaler, 59% a steroid inhaler, and 22% were taking oral steroids (45% intermittently, 55% continuously). Forty-two percent of the sample reported experiencing an exacerbation of their COPD less than a month before their study participation, 30% had an exacerbation between 1 and 3 months previously, and 27% more than 3 months before their participation. The severity of the exacerbations was rated as mild by 56%, moderate by 26%, and severe by 18%.

Internal consistency reliability

The internal consistency reliability, as measured by Cronbach's alpha (n = 99, after removing cases with fewer than five valid responses on the bank), was 0.973 for the Dyspnea item

Table 2 – Sample clinical characteristics (N = 102).

Variable	Value (%)	
Diagnosis (in medical chart)		
COPD	67 (66%)	
Chronic bronchitis	8 (8%)	
Emphysema	29 (28%)	
Bronchiectasis	21 (21%)	
Median no. of years since diagnosis (range)	3 (0–37)	
Smoking history		
Currently smoke tobacco	15 (15%)	
Mean ± SD number of years smoked (range)	43 ± 10 (23–60)	
Smoked tobacco products in the past	73 (72%)	
Mean ± SD number of years smoked (range)	31 ± 14 (2–58)	
	Current	Previous
Comorbid medical conditions (top 8 current)		
Hypertension	41 (40%)	44 (44%)
Acid reflux (heartburn)	36 (35%)	46 (46%)
Arthritis	37 (36%)	41 (41%)
Back pain	32 (31%)	45 (45%)
Asthma	28 (27%)	32 (32%)
Depression	17 (17%)	21 (21%)
Insomnia	14 (14%)	15 (15%)
Coronary artery disease	13 (13%)	20 (20%)
Medications		
Use of any inhalers	92 (90%)	
Use of any steroid inhaler	60 (59%)	
Taking any oral steroids	22 (22%)	
Frequency of steroid use		
Intermittent	10 (45%)	
Continuous	12 (55%)	
Other respiratory medications		
Theophylline	3 (3%)	
Montelukast	8 (8%)	
Other	13 (13%)	
Pulmonary function test results (n = 58)		
Median FEV_1/FVC	0.63	
Median FEV_1 percentage predicted	58	
Most recent exacerbation		
<1 mo ago	42 (42%)	
Between 1 and 3 mo ago	30 (30%)	
>3 mo ago	27 (27%)	
Severity of most recent exacerbation		
Mild	56 (56%)	
Moderate	26 (26%)	
Severe	18 (18%)	
Use of any assistive devices	46 (46%)	

COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity.

bank as well as the FL item bank. The alpha coefficients for the Dyspnea-10 and FL-10 were 0.922 and 0.928, respectively.

Test-retest reliability

Test-retest reliability, evaluated by an ICC, showed good measurement stability of the full-length banks and short forms over 7 to 10 days (n = 82). The ICCs were 0.916 and 0.933 for the 33-item dyspnea and FL banks, respectively, and 0.896 and 0.912 for Dyspnea-10 and FL-10, respectively.

Discriminant validity

The mean scaled score for each MRC dyspnea scale category showed a clear linear pattern of escalation across the MRC dyspnea scale categories. Figure 1 graphically displays the between-group mean differences across the categories and the within-group variations within each category based on the four measures. The most severe MRC dyspnea scale category (category 5) contained only four cases and, hence, was combined with the next lower category. Patients in different MRC categories were characterized by substantially and systematically different dyspnea/FL item bank and short form scores (Dyspnea-33 bank [$F_{3,95} = 40.02$, $P < 0.001$], FL-33 bank [$F_{3,95} = 40.88$, $P < 0.001$], Dypsnea-10 short form [$F_{3,95} = 35.28$, $P < 0.01$], and FL-10 short form ($F_{3,95} = 37.44$, $P < 0.01$]). Subsequent pairwise comparisons using the Bonferroni step-down (Holm) correction revealed that the dyspnea/FL measures distinctly separated each higher category as having a substantially higher mean score ($P < 0.01$). The effect size estimates (using pooled SDs) ranged from 0.811 to 1.356, indicating mean differences between adjacent cate-

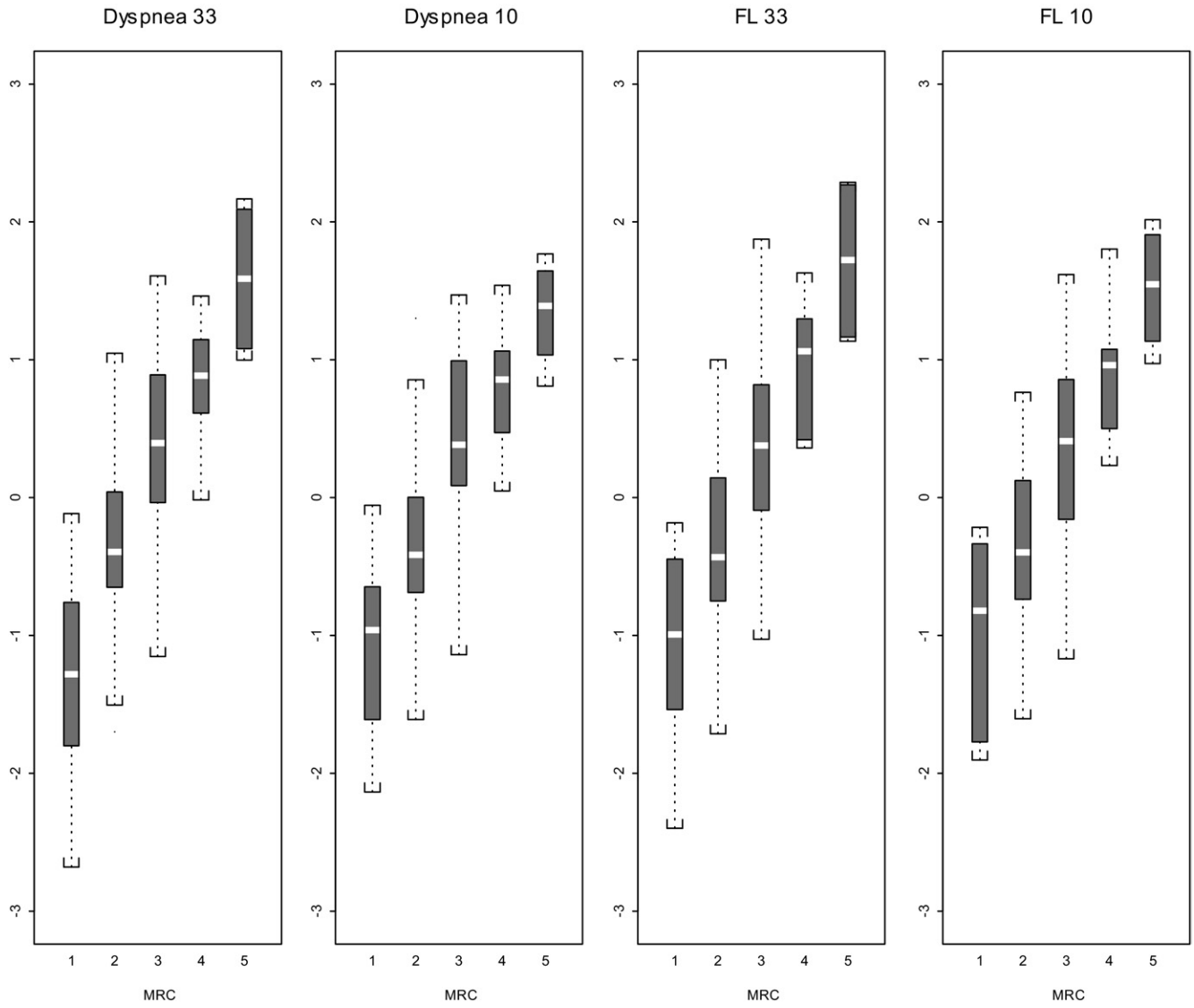

Fig. 1 – The distribution of dyspnea and FL scores by MRC dyspnea scale categories are graphically displayed. The shaded boxes represent the middle 50% of the scores for each MRC dyspnea scale category, the band near the middle of the box the median, and the upper and lower brackets the maximum and minimum scores. All scores are expressed as scaled scores with a mean of 0 and an SD of 1.

gories that were more than a full SD apart in all but two comparisons.

Convergent validity

Table 3 displays the ICCs between dyspnea item bank and short form scores and the validity measures (i.e., CRQ-SAS dyspnea, HADS and its subscales, SF-36 subscales and two component scales, and BDI). The correlations are very consistent between the respective banks and short forms. In most comparisons, the ICC for the 33-item bank was slightly higher than that of the short form in each domain of dyspnea or FLs.

The dyspnea and FL measures were most highly correlated with the SF-36 Physical Function scale (−0.86 to −0.82) followed by the SF-36 Physical Component Scale (−0.83 to −0.76) and the CRQ-SAS dyspnea (−0.77 to −0.80). The lowest correlations were with the SF-36 Emotional Well-being scale (−0.35 to −0.39).

The dyspnea bank and short form were significantly correlated with FEV_1 percentage predicted ($r = -0.30$, $P < 0.05$; $r = -0.33$, $P < 0.05$, respectively).

CAT simulation

On average, CAT administered 4.39 and 4.43 items for the dyspnea and FL banks, respectively. People at the lower (healthier) end of the spectrum were given relatively more items, but only the four to five most extreme cases exceeded seven items. With fewer than five items, on average, the CAT generated scores that were highly correlated with the bank scores; the correlations between the full bank and CAT were 0.953 for both dyspnea and FLs. The top five most frequently adminis-

Table 3 – Validity coefficients based on pearson correlation coefficients (N=99).

Validity measures	Mean	SD	Dyspnea-33	Dyspnea-10	FL-33	FL-10
Baseline dyspnea index	6.0	2.3	−0.51	−0.46	−0.52	−0.52
CRQ-SAS						
Dyspnea	27.3	6.3	−0.80	−0.78	−0.79	−0.77
Fatigue	17.3	5.4	−0.53	−0.52	−0.53	−0.50
Emotional	37.1	8.3	−0.43	−0.43	−0.43	−0.40
Mastery	20.9	5.2	−0.56	−0.56	−0.54	−0.50
HADS	8.9	6.5	0.54	0.53	0.57	0.53
HADS-anxiety	4.9	3.8	0.44	0.44	0.46	0.45
HADS-depression	4.0	3.4	0.53	0.51	0.57	0.52
SF-36						
Physical function	34.3	11.5	−0.85	−0.82	−0.86	−0.84
Role physical	37.4	11.4	−0.58	−0.56	−0.60	−0.56
Bodily pain	47.0	11.1	−0.45	−0.42	−0.49	−0.49
General health	38.9	10.7	−0.59	−0.57	−0.60	−0.62
Vitality	44.7	10.1	−0.59	−0.56	−0.61	−0.58
Social function	45.0	12.3	−0.61	−0.59	−0.64	−0.61
Role emotional	38.0	12.1	−0.55	−0.55	−0.57	−0.52
Mental health	49.7	9.3	−0.39	−0.37	−0.37	−0.35
Physical component scale	37.4	9.6	−0.79	−0.76	−0.83	−0.82
Mental component scale	47.7	9.5	−0.47	−0.46	−0.47	−0.43

CRQ-SAS, Chronic Respiratory Questionnaire–Self-Administered Standardized version; FL, functional limitation; HADS, Hospital Anxiety and Depression Scale; SF-36, 36-Item Short Form Health Survey.

tered tasks by the CAT were carrying something weighing 10 to 20 lb (about 4.5–9 kg, such as a large bag of groceries) from one room to another, walking 50 steps/paces on flat ground at a normal speed without stopping, walking up 20 stairs (2 flights) without stopping, walking (faster than your usual speed) for 0.5 mile (almost 1 km) without stopping, and preparing meals.

Discussion

This study evaluated the reliability and validity of the FACIT-Dyspnea scale in people with a diagnosis of COPD. The measures included both the item banks for dyspnea and FLs (33 items each) and the short forms for the same two constructs (10 items each). In this validation study, patients completed items from the two banks, and their scores on the 10-item short forms were calculated as summed subsets of the banks.

The measurement properties of the banks and short forms were very good. The internal consistency of the banks and short forms were quite high (>0.90), as was the test-retest reliability over 7 to 10 days (also >0.90). The bank and short form scores successfully differentiated patients by MRC dyspnea scale categories. For both banks and short forms, all scores across MRC dyspnea scale groups were in the expected direction, i.e., patients with the lowest MRC dyspnea scale scores (i.e., better dyspnea status) had the lowest dyspnea/FLs scores (i.e., better dyspnea/functional status), and those with higher MRC dyspnea scale scores had FACIT-Dyspnea scores reflecting poorer dyspnea/FL status. Further, the dyspnea banks and short forms were correlated in the expected direction and magnitude with other study measures. Concurrent validity was demonstrated by the FACIT-Dyspnea measures' significant correlations with prominent existing measures of dyspnea, the BDI ($r = 0.47–0.52$). Evidence of convergent validity was suggested by higher correlations with measures expected to be more highly related to the FACIT-Dyspnea measures (e.g., CRQ-SAS dyspnea, SF-36 Physical Function) and lower correlations with measures of less related constructs (e.g., HADS-A or -D, SF-36 Social Functioning). Consistent with the literature [8], the FACIT-Dyspnea measures were more weakly associated with objective measures of lung function ($r = 0.30–0.35$ with PFTs). These findings closely mirror those reported previously with the Internet-based calibration sample of 608 individuals with COPD [11].

We demonstrated that CAT provides an efficient measurement of dyspnea and FLs in a clinical population of patients with mild to severe dyspnea. Our post hoc simulations demonstrated that the amount of reduction in test length was substantial compared to the full-bank and even to the static short form. This should be attractive to clinical researchers whose major concern is to minimize burden, especially with patients who have physical limitations.

Although a number of self-report dyspnea and respiratory health measures currently exist, they have some important limitations. We developed a measure that includes aspects of dyspnea that are highly rated by both patients and clinicians; was developed using a conceptual model and methodology consistent with the US Food and Drug Administration (FDA) PRO Guidance [28]; that is patient-completed versus requiring interviewer administration or clinician rating; is very brief, requiring an average of four to five items for each CAT and 20 by static short form; has strong psychometric qualities, including reflecting differences in severity across the disease spectrum; and is consistent in its format, context, and response options; thereby addressing the majority the limitations of current instruments [9,19,21,24,29–39]. In addition, the FACIT-Dyspnea measures were created with cross-cultural input, facilitating the eventual development and validation in languages other than English. The FACIT-Dys-

pnea also offers administration options, depending on preference and logistical constraints, which include fixed-length short forms as well as CAT.

This study is not without limitations. One factor that was largely related to the geographic location of the participating hospitals was the sex and age of the participants who were recruited using convenience sampling at one suburban and one urban health care setting. The sample was largely white, well-educated, and relatively affluent, which is likely not representative of the US population with COPD. A second potential limitation is the sample size of 102 participants with COPD. This limits the degree to which the results that we report are generalizable to the larger population of individuals with COPD. In addition, a relatively small subsample had available PFT data in their medical charts (n = 58), which could have resulted in inadequate power to detect a significant association with the FACIT-Dyspnea measures. Even though one of the strengths of this new tool is its flexibility in application as CAT, the simulations that were conducted are only the first step toward this tool's validated use, and future studies should prospectively seek to validate this CAT in different clinical samples. Finally, this study reported on cross-sectional results only; future studies will be needed to assess the measures' ability to detect responsiveness longitudinally.

The set of FACIT-Dyspnea measures now offers clinical trialists, clinical researchers, and clinicians options for selecting brief, psychometrically sound measures of dyspnea and/or FLs and was developed in accordance with the FDA PRO Guidance to satisfy regulatory concerns. Future studies will be important to demonstrate the measures' psychometric performance in larger, more diverse populations of individuals with COPD.

Acknowledgments

The authors acknowledge Mary Gabb, MS, a medical writer working with American Healthcare Communications, for editorial input and formatting the article for journal submission. Support for Ms. Gabb was provided by Boehringer-Ingelheim Pharmaceuticals.

REFERENCES

[1] Mannino DM, Homa DM, Akinbami LJ, Ford ES, Redd SC. Chronic obstructive pulmonary disease surveillance—United States, 1971-2000. Respir Care 2002;47:1184–99.
[2] Global Initiative for Chronic Obstructive Pulmonary Disease. Global Strategy for the Diagnosis, Management, and Prevention of COPD, Updated 2008. Available from: www.goldcopd.com. [Accessed June 15, 2010].
[3] National Heart, Lung and Blood Institute. Morbidity and Mortality: 2009 Chartbook on Cardiovascular, Lung and Blood Diseases. Bethesda, MD: US Department of Health and Human Services, Public Health Service, National Institutes of Health. Available from: www.nhlbi.nih.gov/resources/docs/2009_ChartBook.pdf. [Accessed June 15, 2010].
[4] Ferrer M, Alonso J, Morera J, et al. Chronic obstructive pulmonary disease stage and health-related quality of life. The quality of life of chronic obstructive pulmonary disease study group. Ann Intern Med 1997;127:1072–9.
[5] Edelman NH, Kaplan RM, Buist AS, et al. Chronic obstructive pulmonary disease. Task Force on Research and Education for the Prevention and Control of Respiratory Diseases. Chest 1992;102(3 Suppl.):243S–56S.
[6] Eisner MD, Blanc PD, Yelin EH, et al. COPD as a systemic disease: impact on physical functional limitations. Am J Med 2008;121:789–96.
[7] Petty TL. Benefits of and barriers to the widespread use of spirometry. Curr Opin Pulm Med 2005;11:115–20.
[8] Eltayara L, Becklake MR, Volta CA, Milic-Emili, J. Relationship of chronic dyspnea and flow limitation in COPD patients. Am J Respir Crit Care Med 1996;154:1726–34.
[9] Eakin EG, Sassi-Dambron DE, Ries AL, Kaplan RM. Reliability and validity of dyspnea measures in patients with obstructive lung disease. Int J Behav Med 1995;2:118–34.
[10] Victorson DE, Anton S, Hamilton A, et al. A conceptual model of the experience of dyspnea and functional limitations in chronic obstructive pulmonary disease. Value Health 2009;12:1018–25.
[11] Choi SW, Victorson DE, Yount S, et al. Development of a conceptual framework and calibrated item banks to measure patient reported dyspnea severity and related functional limitations. Value Health 2011;14:xx–xx.
[12] Chang C-H, Reeve BB. Item Response Theory and its applications to patient-reported outcomes measurement. Eval Health Prof 2005;28:264–82.
[13] Cella D, Chang CH. A discussion of Item Response Theory (IRT) and its applications in health status assessment. Med Care 2000;38(9 Suppl.):1166–72.
[14] Stansfeld SA, Roberts R, Foot SP. Assessing the validity of the SF-36 General Health Survey. Qual Life Res 1997;6:217–24.
[15] Lord FM, Novick MR. Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley, 1968.
[16] Bjorner JB, Chang CH, Thissen D, Reeve BB. Developing tailored instruments: item banking and computerized adaptive assessment. Qual Life Res 2007;16(Suppl. 1):95–108.
[17] Cella D, Gershon R, Lai J-S, Choi SW. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Qual Life Res 2007;16:133–41.
[18] Choi SW, Swartz RJ. Comparison of CAT item selection criteria for polytomous items. Appl Psychol Meas 2009;33:419–40.
[19] Mahler DA, Weinberg DH, Wells CK, Feinstein AR. The measurement of dyspnea: contents, interobserver agreement, and physiologic correlates of two new clinical indexes. Chest 1984;85:751–8.
[20] Mahler DA, Ward J, Fierro-Carrion G, et al. Development of self-administered versions of modified baseline and transition dyspnea indexes in COPD. COPD 2004;1:1–8.
[21] Medical Research Council Committee. Standardised questionnaire on respiratory symptoms: a statement prepared and approved by the MRC Committee on the Aetiology of Chronic Bronchitis (MRC breathlessness score). BMJ 1960;2:1665.
[22] Ware JE Jr, Sherbourne CD. The MOS 36 Item Short Form Health Survey (SF 36). 1. Conceptual framework and item selection. Med Care 1992;30:473–83.
[23] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67:361–70.
[24] Guyatt GH, Berman LB, Townsend M, et al. A measure of quality of life for clinical trials in chronic lung disease. Thorax 1987;42:773–8.

[25] Schunemann HR, Goldstein R, Mador MJ, et al. A randomised trial to evaluate the self-administered standardised chronic respiratory questionnaire. Eur Respir J 2005;25:31–40.
[26] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;10:37–46.
[27] Choi SW. Firestar: computerized adaptive testing simulation program for polytomous IRT models. Appl Psychol Meas 2009;33:644–5.
[28] US Food And Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for industry. Patient-Reported Outcome measures: Use in medical product development to support labeling claims. Available from: www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. [Accessed December 2009].
[29] Ferris BG. Recommended respiratory disease questionnaires for use with adults and children in epidemiological research. Am Rev Respir Dis 1978;118(Suppl.):S1–53.
[30] Stoller JK, Ferranti R, Feinstein AR. Further specification and evaluation of a new clinical index for dyspnea. Am Rev Respir Dis 1986;34:1129–34.
[31] McGavin CR, Artvinli M, McHardy GJR. Dyspnoea, disability and distance walked: comparison of estimates of exercise performance in respiratory disease. BMJ 1978;2:241–3.
[32] Eakin EG, Resnikoff PM, Prewitt LM, et al. Validation of the USCD Shortness of Breath Questionnaire [abstract]. J Cardiopulm Rehab 1994;14:322.
[33] Hyland ME, Bott J, Singh S, Kenyon CAP. Domains, constructs, and the development of the breathing problems questionnaire. Qual Life Res 1994;3:245–6.
[34] Borg G. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982;14:377–81.
[35] Jones PW, Quirk FH, Baveystock CM. The St. George's Respiratory Questionnaire. Respir Med 1991;85:25–31.
[36] Mahler DA. Mechanisms and measurement of dyspnea in chronic obstructive pulmonary disease. Proc Am Thorac Soc 2006;3,234–8.
[37] Fletcher CM, Elmes PC, Wood CH. The significance of respiratory symptoms and the diagnosis of chronic bronchitis in the working population. BMJ 1959;2:257.
[38] Aitken RB. A growing edge of measurement of feelings. Proc R Soc Med 1969;62:989–93.
[39] Quirk FH, Jones PW. Repeatability of two new short airways questionnaires. Thorax 1994;49:1075–9.