INTERNATIONAL JOURNAL OF BEHAVIORAL MEDICINE, 2(2), 118–134

Reliability and Validity of Dyspnea Measures in Patients With Obstructive Lung Disease

Elizabeth G. Eakin, Dawn E. Sassi-Dambron, Andrew L. Ries, and Robert M. Kaplan

Dyspnea, the clinical term for shortness of breath, is the primary symptom and an important outcome measure in evaluations of patients with lung disease. It is a subjective symptom that has proved difficult to quantify. Many dyspnea measures are available, yet it is difficult, based on the existing literature, to determine the most reliable and valid. In this study, we evaluated 6 measures of dyspnea for reliability and validity: (a) Baseline Dyspnea Index (BDI) and Transition Dyspnea Index, (b) UCSD Shortness of Breath Questionnaire (SOBQ), (c) American Thoracic Society Dyspnea Scale, (d) Oxygen Cost Diagram, (e) Visual Analog Scale, and (f) Borg Scale. Subjects were 143 patients (74 women and 69 men) with obstructive lung disease, ages 40 to 86, $FEV_{1.0}$ 0.36 to 3.53 L, FVC 1.07 to 5.74 L. Dyspnea measures were assessed for test–retest reliability, internal consistency, interrater reliability, and construct validity (i.e., correlations among dyspnea measures and correlations of dyspnea measures with exercise tolerance, health-related quality of life, lung function, anxiety, and depression). Results suggest that the SOBQ and BDI demonstrated the highest levels of reliability and validity among the dyspnea measures examined.

Elizabeth G. Eakin, University of California, San Diego–San Diego State University Joint Doctoral Program in Clinical Psychology, University of California, San Diego, CA, USA, currently at Oregon Research Institute, Eugene, OR, USA; Dawn E. Sassi-Dambron and Andrew L. Ries, Division of Pulmonary and Critical Care Medicine, University of California, San Diego, CA, USA; Robert Kaplan, Department of Family and Preventive Medicine, University of California, San Diego, CA, USA.

This research was supported by University of California Tobacco Related Disease Research Program Grant 2RT0268, National Heart, Lung, and Blood Institute Grant HL 34732 to Robert M. Kaplan, and National Institutes of Health NHLBI Preventive Pulmonary Academic Award No. HL02215 to Andrew L. Ries.

We thank Lela Prewitt, Jan Jasiewicz, and Jay Yancey for their invaluable assistance in this study, and Donald Mahler for his helpful comments on a previous draft of this article.

Correspondence concerning this article should be addressed to Robert M. Kaplan, Department of Family and Preventive Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0622.

Key words: dyspnea, chronic obstructive pulmonary diseases (COPD), obstructive lung disease, reliability, validity, assessment

Obstructive lung diseases are a major cause of death and disability in the United States. The chronic obstructive pulmonary diseases (COPDs), which include emphysema, chronic bronchitis, and nonreversible asthma, are the fourth leading cause of death and account for 4% of all deaths (Higgins, 1989; National Center for Health Statistics, 1994). *Dyspnea,* the subjective sensation of difficult or labored breathing, is one of the most common and disabling symptoms of people with COPD (Kinsman et al., 1983). Numerous authors have described a dyspnea-panic cycle in which the experience of breathlessness leads to anxiety, which creates muscle tension, leading to increased dyspnea and panic (Dudley, Glaser, Jorgenson, & Logan, 1980; Kaplan, Ries, & Atkins, 1985; Renfroe, 1988). The distress caused by dyspnea can become part of a vicious cycle leading to fear of future attacks of shortness of breath. In turn this causes patients with COPD to slowly decrease their activity level resulting in greatly limited independent functioning, quality of life, and furthering the course of deterioration in COPD.

Because there is no cure for COPD, most therapies are aimed at managing symptoms, such as dyspnea, and at increasing independent functioning. Dyspnea is thus an important component in characterizing the disease and evaluating treatment outcomes (Carrieri, Janson-Bjerklie, & Jacobs, 1984; Mahler & Harver, 1992).

Despite its importance, dyspnea is a subjective symptom that has proved difficult to measure. There are no objective criteria against which measures of dyspnea can be compared, and dyspnea is only modestly correlated with measures of pulmonary function ($rs = 0.35$ to 0.43; Burrows, Niden, Barclay & Kasik, 1965; Mahler & Wells, 1988; Mahler & Harver, 1992; Moody, McCormick, & Williams, 1990). Dyspnea is a complex symptom that, like other subjective symptoms, is affected by many factors including past experience, tolerance to discomfort, cultural norms, age, and gender (Pennebaker, 1982). Previous studies have explored the physiological, functional, and psychological components of dyspnea (Burns & Howell, 1969; Dudley, Martin, & Holmes, 1968; Gift, Plaut, & Jacox, 1986; Mahler & Harver, 1992; McGavin, Artvinli, & McHardy, 1978; Moody et al., 1990).

Measurement of dyspnea is further complicated by the number and variation of available measures, as well as by the different approaches to measurement (Mahler & Harver, 1990). Several types of instruments are available including structured interviews, self-report questionnaires, visual analog scales, and numeric ratings. Patients may be asked to give historical reports of dyspnea during a previous period of time or to estimate the amount of dyspnea experienced as they perform various activities. Dyspnea may also be induced in laboratory or research settings via exercise or other physical tasks or by the mechanical addition of breathing loads.

A review of the literature on dyspnea measurement reveals a lack of consensus and standardization (Eakin, Kaplan, & Ries, 1993). It is difficult to determine the most reliable and valid measures for a given setting. Also, there are few standards for administering the measures or inducing dyspnea.

In this study, we evaluated the reliability and validity of six commonly used dyspnea measures: (a) Baseline Dyspnea Index (BDI) and Transition Dyspnea Index (TDI), (b) UCSD Shortness of Breath Questionnaire (SOBQ), (c) American Thoracic Society Dyspnea Scale (ATS), (d) Oxygen Cost Diagram (OCD), (e) Visual Analog Scale (VAS), and (f) Borg Scale (Borg). This group includes measures using retrospective reports of symptoms during activities of daily living (ADLs), as well as measures during a standardized exercise task to induce dyspnea. Measures were evaluated for ease of administration and scoring, test–retest reliability, internal consistency, interrater reliability, and construct validity.

METHOD

Subjects

The subjects for this study were 143 patients with obstructive lung disease (74 women, 69 men) recruited from newspaper advertisements, local hospitals, and physicians to participate in a clinical trial evaluating a program of dyspnea management strategies (Sassi-Dambron, Eakin, Ries, & Kaplan, 1995). Mean age was 67.7 years (SD = 8.6, range = 40 to 86). A total of 497 individuals were screened during recruitment. Most of those who did not participate either did not have obstructive lung disease or were not interested. The study protocol was approved by the Human Subjects Committee of the University of California, San Diego; all subjects signed an informed consent form prior to participation.

Diagnosis of obstructive lung disease was determined by clinical history and confirmed with pulmonary function evidence of reduced expiratory flows and/or hyperinflation of static lung volumes. Pulmonary function tests included spirometric measurements of vital capacity and expiratory flow rates, lung volumes and airway resistance by body plethysmography, and maximal inspiratory and expiratory pressures to assess respiratory muscle strength. All testing and quality-control procedures followed standard and recommended methods (American Thoracic Society, 1987; Clausen & Zarins, 1982). Subjects with a history of dyspnea due to reversible obstructive lung disease (i.e., asthma) were included even if their expiratory flow rates were within normal limits at the time of the study.

Statistical Analyses

Measures were evaluated for reliability and validity. Reliability was assessed by examining test–retest correlations from dyspnea measures taken at

two time points for the same subjects, and by calculation of coefficier alpha, a statistic used to evaluate internal consistency, or the extent to whic the different items on a dyspnea questionnaire measure the same construc (Cronbach, 1984). Construct validity was assessed by examining correla tions among the dyspnea measures; a valid measure of dyspnea shoul demonstrate a reasonable degree of correlation with other measures of th same construct. The validity of the six dyspnea measures was also evaluate by comparing relations with measures of exercise tolerance, health-relate quality of life, lung function, and selected psychological variables.

Measures

The following measures used in this study were obtained at the baselin evaluation in the clinical trial.

Dyspnea Measures

BDI and TDI. Mahler and colleagues developed a two-part (Baselin and Transition), interviewer-administered dyspnea index that rates dyspne according to three categories: *functional impairment, magnitude of task,* an *magnitude of effort* (Mahler, Weinberg, Wells, & Feinstein, 1984). Th functional impairment category addresses whether the patient has reduced c given up activities or a job due to shortness of breath. The magnitude of tas category rates the type of task that makes the patient breathless (i.e., ligh moderate, or strenuous tasks). The magnitude of effort category assesse how much effort the patient exerts before becoming breathless (e.g., breath less only after extraordinary exertion, or so breathless that she or he has t pause frequently during most tasks). The BDI is used to rate the severity c dyspnea at a single point in time; the TDI is used to assess changes from tha baseline. At baseline, dyspnea in each of the three categories is rated on 5-point scale ranging from 0 (*severe*) to 4 (*unimpaired*). Ratings for each c the three categories are then summed to form a baseline total dyspnea scor (BDI; range = 0 to 12). The TDI is used to rate changes from baseline in eac of the three categories using a 7-point scale ranging from –.3 (*major deter oration*) to +3 (*major improvement*). Ratings from the TDI can be added t form a total transition dyspnea score (TDI; range = –9 to +9).

UCSD SOBQ. The SOBQ was developed and has been used exter sively in the Pulmonary Rehabilitation Program at the University of Califor nia, San Diego to assess shortness of breath with various ADLs (Eakin Prewitt, Ries, & Kaplan, 1994). A modified version of the SOBQ was use in this study. The modification clarified and expanded the dyspnea ratin

scale of the original instrument (Archibald & Guidotti, 1987). The modified questionnaire asks patients to indicate how frequently they experience shortness of breath on a 7-point scale labeled 0 (*never*), 1 (*sometimes*), 2 (*half of the time*), 3 (*most of the time*), 4 (*all of the time*), 5 (*activity given up due to dyspnea*), and NA (*activity not performed, unrelated to dyspnea*) during 21 different ADLs associated with varying levels of exertion. Three additional questions about limitations due to shortness of breath, fear of harm from overexertion, and fear of shortness of breath are included for a total of 24 items.

ATS. This self-report respiratory questionnaire consists of five yes or no questions that correspond to a 5-point dyspnea rating scale developed by Fletcher (1952). Based on the response to these questions, dyspnea is rated in terms of one of five grades of severity (Grade 0 = not troubled by breathlessness except on strenuous exertion; Grade 1 = short of breath when hurrying on the level or walking up a slight hill; Grade 2 = walks slower than most people his or her age on the level; Grade 3 = has to stop for breath after walking about 100 yards on the level; Grade 4 = too breathless to leave the house, or breathless after undressing; American Thoracic Society, 1978).

OCD. The OCD is a 10-cm line along which activities are written at intervals that correspond to the metabolic equivalents (or oxygen cost) required to perform them (McGavin et al., 1978). Patients are asked to make a mark on the line indicating the point above which their breathlessness would not allow them to go. Examples of activities on the diagram are "brisk walking uphill," "medium walking on the level," "bedmaking," "standing," and "sleeping."

VAS. The VAS is presented to the subject as a 10-cm vertical or horizontal line, typically accompanied by anchors such as *no breathlessness,* and *greatest breathlessness* (Aitken, 1969). Subjects are instructed to place a mark on the line indicating their level of breathlessness. The reliability and validity of the VAS have been evaluated for use with a dyspnea-producing task, but have not been assessed for administration without such a task (i.e., to obtain historical reports of dyspnea for a given task or for a given time period).

Borg. The modified Borg is an 11-point rating scale ranging from 0 (*none*) to 10 (*severe*) on which patients are asked to rate their level of breathlessness (Borg, 1982). Written descriptors are placed so that a doubling of the numerical rating corresponds to a two-fold increase in sensation intensity. The original Borg ranged from 6 to 20, also with written descrip-

tors, and was revised to give this categorical scale the properties of a ratio scale (Borg, 1970). In this study, the Borg and VAS were administered in two ways: (a) to obtain historical reports of patients' average level of dyspnea during the past week, and (b) to obtain dyspnea ratings following a 6-min walk test.

Exercise Tolerance

Six Minute Walk Test (6MW). The 6-min walk is a standard measure of exercise tolerance used frequently with lung disease patients (Guyatt et al., 1985). The test was conducted in an area free from distractions with standardized encouragement provided by the staff. Subjects were asked to cover as much ground as they could in 6 min. The timed distance walk test has been shown to be highly reliable (Mungall & Hainsworth, 1979), with moderate correlations with tests of pulmonary function and maximum exercise capacity (McGavin et al., 1978). Guyatt and colleagues demonstrated a learning effect across subsequent trials of this test and recommended two practice tests prior to actual test administration (Guyatt et al., 1984). Because of time constraints, we used one practice test, followed by at least 10 min of rest and then a second test. Data from the longer of the two walks were used. Subjects rated their dyspnea using the Borg and VAS at the end of each walk.

Health-Related Quality of Well-Being

Quality of Well-Being Scale (QWB). The QWB is a comprehensive measure of health-related quality of life that includes several components. First, it obtains observable levels of functioning at a point in time from three separate scales: Mobility, Physical Activity, and Social Activity. Second, each patient selects the symptom or problem that he or she found most undesirable from a list of 36 symptoms and problems. Then, the observed level of function and the subjective symptomatic complaint are weighted by preference, or the utility for the state, on a scale ranging from 0 (*dead*) to 1.0 (*optimum function*). The weights are obtained from independent samples of judges who rate the desirability of the observable health status. Using this system, it is possible to place the general health status of any individual on the continuum between death and optimal functioning for any point in time. This system has been used extensively in a variety of medical and health services research studies (Kaplan & Anderson, 1988, 1990). In addition, validity data for patients with COPD have been published. Kaplan and coworkers reported that the QWB was substantially correlated with both performance and physiologic variables relevant to the health status of patients with COPD (Kaplan, Atkins, & Timms, 1984).

Anxiety

Spielberger State–Trait Anxiety Inventory (STAI). The STAI is a self-report measure of both state (used in this study) and trait anxiety that has been used extensively in research. The STAI contains 40 statements to which subjects respond using a 4-point rating scale ranging from 1 (*not at all*) to 4 (*very much so*) indicating the extent to which the statements describe the way they feel. The first 20 statements tap state anxiety by asking subjects to indicate whether the statements reflect the way they feel right now. The second 20 statements tap trait anxiety by asking subjects to indicate whether the statements reflect the way they generally feel. Spielberger and coworkers reported excellent internal consistency (alpha coefficients for State Anxiety scale for working adults ages 50 to 59 = 0.92 for men and 0.90 for women; for the Trait Anxiety scale, alpha = 0.96 for men and 0.89 for women) and validity across a number of groups including a general medical patient population (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983).

Depression

Center for Epidemiologic Studies Depression Scale (CES–D). The CES–D is a general measure of depression used extensively in epidemiologic studies. The scale includes 20 items and taps dimensions of Depressed Mood, Hopelessness, Appetite Loss, Sleep Disturbance, and Energy Level. Patients are asked to report how often they experienced a particular symptom during the past week on a 4-point scale ranging from rarely or none of the time, 0 (*0 to 1 day*) to most or all of the time, 3 (*5 to 7 days*). Total score on the CES–D ranges from 0 to 60, with scores greater than 15 indicative of clinically significant levels of depressive symptomatology in adults. Radloff (1977) presented extensive data on the reliability and validity of the CES–D. Test–retest correlations are good for a test designed to assess fluctuations in mood (r = 0.57). The CES–D discriminates between clinical and normal populations; the reliability and validity have been replicated across various normal and clinical samples.

Order of Administration

The order of administration of the various instruments was as follows: Subjects were first instructed in the use of the Borg and VAS and were given a number of practice questions to assess their understanding of these scales to rate dyspnea. They were then asked to rate their average level of dyspnea during the past week using the Borg and then the VAS. The QWB was then administered by an interviewer, followed by the BDI. Subjects then com-

test on the Borg and VAS. During a 10-min break between walk tests, subjects completed the ATS, OCD, and UCSD SOBQ. Subjects then completed the second 6MW and rated their dyspnea at the end of the walk on the VAS and Borg. Finally, the STAI and CES–D were completed by subjects. An interviewer remained with subjects during the entire protocol to explain each instrument and to answer any questions.

Test–Retest Procedure

Forty-one patients completed dyspnea measures during the initial assessment and during a telephone retest session 1 to 7 days later (mean retest interval = 2 days). Patients received the retest packet containing the dyspnea measures at the initial visit, and were assisted in completing the packet over the phone. Patients returned the retest packet in a prestamped envelope.

Internal Consistency Procedure for BDI and TDI

The BDI was administered to 24 patients during the baseline assessment by two raters. There were a total of four raters, with one rater administering the BDI to all 24 patients and serving as the standard. The other three raters each administered the BDI to 8 patients along with the standard rater. The correlation of the standard rater with the rating of one of the three other raters was used as a measure of interrater reliability.

The TDI was also administered to 25 subjects by two of the four raters during the follow-up assessment of the treatment–outcome study into which 89 of the 143 patients were randomized. The procedure was the same as for the BDI in which one rater served as the standard.

RESULTS

Descriptive statistics are shown in Table 1. Mean age of study participants was 67.7 years (SD = 8.6). Fifty-one percent were women. Mean $FEV_{1.0}$ was 1.22 L (SD = 0.57 L); mean FVC = 2.54 L (SD = 0.90 L). At the time of study, 17 out of the 143 (12%) subjects had $FEV_{1.0}$ > 80% predicted, indicating low–normal expiratory flow rates. These subjects had a history of dyspnea and reactive airway disease (i.e., asthma).

Test–Retest Reliability and Internal Consistency

Results of the test–retest correlations and internal consistency are presented in Table 2. Test–retest correlations of r = 0.70 and higher are considered

TABLE 1
Subject Characteristics

Variable	*M*	*SD*	*Range*
Age	67.7	8.6	40.0–86.0
Sex	51% female		
Lung function			
VC (L)	3.03	0.96	1.27–5.57
FRC (L)	4.91	1.39	2.19–8.19
FRC (%)	164.00	33.00	82–232
RV (L)	4.05	1.30	1.72–6.90
RV (%)	196.00	55.00	83–330
TLC (L)	7.08	1.70	3.89–10.86
TLC(%)	135.00	21.00	78–178
RV/TLC	57.00	10.00	25–80
FVC (L)	2.54	0.90	1.07–5.74
FVC (%)	77.00	21.00	29–134
$FEV_{1.0}$ (L)	1.22	0.57	0.36–3.52
$FEV_{1.0}$ (%)	53.00	24.00	11–127
$FEV_{1.0}$ / FVC	48.00	13.00	20–87
$FEV_{25-75\%}$ (L/S)	0.56	0.38	0.15–2.57
$FEV_{25-75\%}$ (%)	24.00	18.00	6–140
PEFR (L/S)	4.0	1.8	1.0–9.4
PIFR (L/S)	3.7	1.3	1.4–7.5
MIP (CM H_2O)	67.9	22.4	6.0–125.0
MEP (CM H_2O)	133.4	41.9	60.0–270.0
Dyspnea measures			
BDI	5.0	2.3	0.0–11.0
SOBQ	35.6	17.8	4.0–94.0
ATS	2.3	1.5	0.0–4.0
OCD (cm)	6.5	1.7	2.8–10.0
VAS (post 6MW; cm)	5.5	2.5	0.2–10.0
Borg (post 6MW)	4.3	2.2	0.5–10.0
VAS (average past week; cm)	4.1	2.1	0.0–9.6
Borg (average past week)	3.2	1.3	0.0–8.0
Exercise tolerance			
6MW distance (m)	409	107	33–688
Quality of well-being			
QWB	0.64	0.07	0.50–0.90
Anxiety			
STAI (Form Y-1)	32.8	10.0	20.0–60.0
Depression			
CESD	13.0	8.9	0.0–40.0

Note. $N = 143$. BDI = Mahler's Baseline Dyspnea Index; SOBQ = UCSD Shortness of Breath Questionnaire; ATS = American Thoracic Society Dyspnea Scale; OCD = Oxygen Cost Diagram; VAS = Visual Analog Scale; Borg = Borg Scale; 6MW = Six Minute Walk Test; QWB = Quality of Well-Being Scale; STAI = State-Trait Anxiety Inventory; CESD = Centers for Epidemiologic Studies Depression Scale.

TABLE 2
Reliability: Test-Retest Correlations and Internal Consistency

Measure	*Test-Retest Correlation*[a]	*Internal Consistency Alpha*
BDI	0.76	0.80
SOBQ	0.94	0.91
ATS	0.72	0.76
OCD	0.64	N/A
VAS[b]	0.54	N/A
Borg[b]	0.45	N/A

Note. BDI = Mahler's Baseline Dyspnea Index; SOBQ = UCSD Shortness of Breath Questionnaire; ATS = American Thoracic Society Dyspnea Scale; OCD = Oxygen Cost Diagram; VAS = Visual Analog Scale; Borg = Borg Scale.

[a]All $p < .001$. [b]Patients were asked to estimate their average level of dyspnea within the past week.

reasonable evidence of reliability for research studies (Nunnally, 1978). The SOBQ had the highest test–retest correlation, followed by the BDI and ATS. The OCD, Borg, and VAS (average rating for past week) all demonstrated lower test–retest correlations. Repeatability of Borg and VAS ratings during an exercise task was not evaluated in this study, but has been discussed elsewhere (Belman, Brooks, Ross, & Mohsenifar, 1991; Wilson & Jones, 1989).

Internal consistency was calculated for the dyspnea measures with more than one item. A criterion of 0.70 was chosen as good evidence for reliability (Nunnally, 1978). As seen in Table 2, all three dyspnea measures with more than one item have good levels of internal consistency, with the SOBQ having the highest alpha coefficient.

Interrater Reliability for the BDI and TDI

As seen in Table 3, we were able to achieve a good level of interrater reliability for the BDI ($r = 0.88$ for total score). Interrater reliability for the TDI was $r = 0.83$ for total score.

Validity of Dyspnea Measures

The results of the construct validity evaluation using correlations among the dyspnea measures are presented in Table 4. Correlations among the measures that use ADLs as the stimulus for dyspnea (i.e., BDI, SOBQ, ATS, OCD) range from $r = 0.48$ to $r = 0.70$, indicating a reasonable degree of convergence on the construct of dyspnea. Overall, the Borg and VAS (average dyspnea in past week and post 6-min walk) demonstrated smaller correlations with the BDI, SOBQ, ATS, and OCD, $r = 0.27$ to $r = 0.50$.

The construct validity evaluation using correlations between dyspnea measures and other variables with which they are assumed to be related is presented in Table 5. The dyspnea measures demonstrated small to moderate correlations with the 6MW and the QWB. Correlations between dyspnea measures and lung function variables (i.e., FEV_1 and FVC) are small to moderate, consistent with previous reports in the literature. Correlations of dyspnea measures with measures of depression and anxiety are generally weak, but often statistically significant. Overall, the Borg and VAS demonstrated lower correlations with the other variables than did the BDI, SOBQ, ATS, and OCD.

TABLE 3
Interrater Reliability for Mahler's Baseline Dyspnea Index (BDI) and Transition Dyspnea Index (TDI)

	Interrater Correlation[a]
BDI category	
Functional impairment	0.72
Magnitude of task	0.77
Magnitude of effort	0.76
Total score	0.88
TDI category	
Functional impairment	0.93
Magnitude of task	0.52
Magnitude of effort	0.74
Total score	0.83

[a]All $p < .001$.

TABLE 4
Validity: Correlations Among Dyspnea Measures

Measure	*BDI*	*SOBQ*	*ATS*	*OCD*	*VAS*[a]	*BORG*[a]	*VAS*[b]	*BORG*[b]
BDI	—	−.70**	−.59**	.52**	−.50**	−.40**	−.42**	−.48**
SOBQ		—	.52**	−.48**	.45**	.39**	.35**	.41**
ATS			—	−.52**	.33**	.36**	.35**	.38**
OCD				—	−.28**	−.27**	−.27**	−.31**
VAS (past week)					—	.71**	.22**	.22*
BORG (past week)						—	.09	.11
VAS (post 6MW)							—	.79**
BORG (post 6MW)								—

Note. BDI = Mahler's Baseline Dyspnea Index; SOBQ = USCD Shortness of Breath Questionnaire; ATS = American Thoracic Society Dyspnea Scale; OCD = Oxygen Cost Diagram; VAS = Visual Analog Scale; Borg = Borg Scale; 6MW = Six Minute Walk Test.
[a]*Past week.* [b]*Post 6-min walk.*

TABLE 5
Validity: Correlations of Dyspnea Measures With Exercise Tolerance, Quality of Life, Lung Function, Anxiety, and Depression

Measure	*6MW*	*QWB*	*FEV_1*	*FVC*	*CESD*	*STAI*
BDI	.62**	.50**	.40**	.28**	−.32**	−.37**
SOBQ	−.47**	−.45**	−.28**	−.20*	.25**	.36**
ATS	−.53**	−.29**	−.38**	−.32**	.15	.20*
OCD	.54**	.33**	.41**	.29**	−.12	−.16
VAS (past week)	−.32**	−.37**	−.26**	.23**	.31**	.31**
Borg (past week)	−.28**	.20**	−.22*	−.13	.21*	.22**
VAS (post 6MW)	−.33**	−.35**	−.22*	−.14	.15	.28**
Borg (post 6MW)	−.37**	−.27**	−.26**	−.21*	.08	.23**

Note. BDI = Mahler's Baseline Dyspnea Index; SOBQ = UCSD Shortness of Breath Questionnaire; ATS = American Thoracic Society Dyspnea Scale; OCD = Oxygen Cost Diagram; VAS = Visual Analog Scale; Borg = Borg Scale; 6MW = Six Minute Walk Test. $*p < .05$. $**p < .01$.

DISCUSSION

This study evaluated the reliability and validity of six dyspnea measures commonly used in research and practice. One notable exception is the dyspnea component of the Chronic Respiratory Disease Questionnaire (CRQ; Guyatt, Berman, Townsend, Pugsley, & Chambers, 1987). This interviewer-administered questionnaire was not included in this study due to time constraints and the estimated 15 to 25 min needed to administer the dyspnea component of the CRQ. The instruments evaluated in this study include measures that assess the functional state resulting from dyspnea during ADLs (i.e., BDI/TDI, SOBQ, ATS, OCD), as well as measures of dyspnea intensity following a standardized exercise task (i.e., VAS and Borg). The Borg and VAS were also used to obtain historical reports of the average level of dyspnea experienced during the past week.

The results demonstrate some interesting findings about the usefulness and comparative performance of several different approaches to the measurement of dyspnea in patients with obstructive lung disease.

Practical Issues

BDI and TDI. The BDI can be used to collect reliable and valid information. However, as an interviewer-administered instrument with few specific instructions, its optimal use depends on the clinical experience and expertise of the raters and may involve additional work on the part of raters to standardize its administration. In our experience, to use this instrument reliably, it was necessary for our four raters to discuss and standardize

questions and to come to some consensus as to how ratings should be made on each one of the three scales. Ongoing assessment of interrater reliability to check for tendencies of each rater to stray from initial standardization was also needed.

The TDI may be affected by bias on the part of the patient and interviewer because it asks both individuals to make judgments about improvement versus deterioration in the patient's status since baseline. Like the BDI, the TDI lacks standardized questions for raters. In addition, administering the TDI requires that the interviewer have on hand the patient's baseline scores as well as a copy of the BDI. The TDI refers back to the patient's state at baseline, and the wording of the magnitude of task category refers the interviewer back to the levels described in the BDI, which are not described on the TDI.

UCSD SOBQ. The SOBQ is a simple, self-administered instrument requiring little subject instruction and taking approximately 5 min to complete. Although good levels of reliability and validity were observed, modifications were indicated. The rating scale contains a "not applicable" category that resulted in frequent unscored responses. The scale has subsequently been modified so that respondents are asked to make ratings for all items; the revised SOBQ has recently been evaluated (Eakin et al., 1994).

ATS. The ATS can be easily administered by an interviewer or can be filled out by the patient with little instruction. However, it does not appear to capture some of the complexity of the dyspnea construct, given its low and nonsignificant correlation with the psychological variables.

OCD. The OCD was moderately related to other dyspnea measures. As a one-item measure, the OCD demonstrated lower test–retest reliability than the other instruments ($r = 0.64$). The instructions for the OCD were confusing for some patients; thus, additional explanation for which there are no standards was often required. Given the difficulty our patients had in understanding this instrument and its low reliability, the OCD seems somewhat less useful for research or practice.

Borg and VAS. The Borg was administered easily and quickly. The majority of subjects had no difficulty understanding how to rate their dyspnea using this numerical scale. The VAS was difficult for some subjects who had trouble understanding the concept of using a line to rate their dyspnea, even after explanation.

In this study, the Borg and VAS were used in two ways: (a) to obtain historical reports of dyspnea (i.e., average dyspnea during the past week),

and (b) in conjunction with an exercise task (i.e., the 6-min walk distance). We were not able to evaluate the reliability of the Borg and VAS ratings during exercise in this study. The literature on their reproducibility with various exercise tasks suggests that the reliability of Borg and VAS ratings depends on the length of time between tests (e.g., same day vs. different days) and also on the exercise task (e.g., steady state vs. incremental exercise; Belman et al., 1991; Wilson & Jones, 1989). Our weak test–retest reliability correlations for the Borg and VAS ratings of average dyspnea in the past week likely result from both measurement error and the true variability of the symptom of dyspnea.

The Borg and VAS, which obtain historical reports of dyspnea experienced during the past week, do not substitute for the more extensive dyspnea measures. However, the Borg and VAS are the only instruments available for measuring dyspnea intensity during exercise or with other physical tasks. Further research is needed to elucidate the effects of varying exercise tasks and perhaps to develop a standard exercise task that maximizes both the reliability and validity of Borg and VAS dyspnea ratings.

Study Limitations

This study was intended to provide information on the selection and use of dyspnea measures. Several methodological issues must be considered in interpreting the results. Seventeen of the 143 subjects had a diagnosis of asthma, based on clinical history, and did not meet the pulmonary function test criteria for COPD. It was felt that inclusion of subjects with obstructive lung disease (i.e., asthma) would allow us to assess the reliability and validity of the dyspnea measures across a broad range of disease severity. However, we are unable to say whether certain measures are more appropriate for patients with different levels of disease severity (i.e., asthma vs. COPD). Categorizing the sample into levels of disease severity and calculating the reliability and validity coefficients for each group would result in attenuated correlations due to the restriction in range. In addition, the order of administration of the dyspnea measures was not randomized and thus order effects cannot be ruled out. It was felt that the advantage gained by ruling out order effects was outweighed by the need to maintain standardization in this complex protocol. It is possible that the accuracy of the data was adversely affected for the more compromised participants due to the length of the protocol. Interviewers made every effort to maximize the comfort of participants and to allow adequate rest periods. Finally, to assess the validity of the dyspnea measures evaluated in this study, we chose a number of psychosocial, physical function, and lung function variables against which to compare the dyspnea measures. These variables were chosen following a detailed review of the literature on the correlates of dyspnea (Eakin et al., 1993). The results of this study by no means provide an exhaustive evalua-

tion of the complex construct of dyspnea. The range of correlations among the dyspnea measures may indicate that the instruments measure different dimensions of the dyspnea construct. We encourage continued study of the factors underlying this complex construct.

Summary

The choice of dyspnea measures depends on the purpose of the application and the reliability and validity of the measures. For historical reports of the extent to which dyspnea affects functional ability, the results of this study suggest that the UCSD SOBQ and the BDI demonstrated the highest levels of reliability and validity among this group of measures. The VAS and Borg were less reliable and valid for obtaining historical reports of dyspnea (i.e., average dyspnea during the past week). They appear to be most useful in assessing dyspnea during or following a physical task such as exercise or an added breathing load. They are the only instruments that provide information about symptom intensity at a given point in time and may be most useful for repeated measurement in individual subjects. However, their potential usefulness to evaluate changes resulting from an intervention may be affected by several extraneous factors, including interval between tests, type of task, desensitization to dyspnea unrelated to treatment, or changing disease status.

The appropriate choice of a dyspnea measure also depends on the setting in which it is to be used. Clinicians experienced in evaluating patients with lung diseases may find an interviewer administered scale like the BDI/TDI easy to complete. On the other hand, for self-reported patient assessment, either for clinical or research application, the SOBQ may be most appropriate because it can be completed easily by patients with little instruction or supervision. We hope to report additional studies on the SOBQ in the near future.

REFERENCES

Aitken, R. B. (1969). A growing edge of measurement of feelings. *Proceedings of the Royal Society of Medicine, 62,* 989–993.

American Thoracic Society. (1978). Recommended respiratory disease questionnaires for use with adults and children in epidemiological research. *American Review of Respiratory Disease, 118,* 7–35.

American Thoracic Society. (1987). Standardization of spirometry—update. *American Review of Respiratory Disease, 136,* 1285–1296.

Archibald, C. J., & Guidotti, T. L. (1987). Degree of objectivity measured impairment and perceived shortness of breath with activities of daily living in patients with chronic obstructive pulmonary disease. *Canadian Journal of Rehabilitation, 1*(1), 45–54.

Belman, M. J., Brooks, L. R., Ross, D. J., & Mohsenifar, Z. (1991). Variability of breathlessness measurement in patients with chronic obstructive pulmonary disease. *Chest, 99,* 566–571.

Borg, G. (1970). Perceived exertion as an indicator of somatic stress. *Scandinavian Journal of Rehabilitation Medicine, 2*(3), 92–98.

Borg, G. (1982). Psychophysical bases of perceived exertion. *Medicine and Science in Sports and Exercise, 14*(5), 377–381.

Burns, B. H., & Howell, J. B. L. (1969). Disproportionately severe breathlessness in chronic bronchitis. *Quarterly Journal of Medicine, 151,* 277–294.

Burrows, B., Niden, A. H., Barclay, W. R., & Kasik, J. E. (1965). Chronic obstructive lung disease: II. Relationship of clinical and physiologic findings to the severity of airways obstruction. *American Review of Respiratory Disease, 91,* 665–678.

Carrieri, V. K., Janson-Bjerklie, S., & Jacobs, S. (1984). The sensation of dyspnea: A review. *Heart & Lung, 13*(4), 436–447.

Clausen, J. L., & Zarins, L. P. (1982). *Pulmonary function testing guidelines and controversies: Equipment, methods, and normal values.* New York: Academic.

Cronbach, L. J. (1984). *Essential of psychological testing* (4th ed.). New York: Harper & Row.

Dudley, D. L., Glaser, E. M., Jorgenson, B. N., & Logan, D. L. (1980). Psychosocial concomitants to rehabilitation in chronic obstructive pulmonary disease: Part 1. Psychosocial and psychological considerations. *Chest, 77,* 413–420.

Dudley, D. L., Martin, C. J., & Holmes, T. H. (1968). Dyspnea: Psychologic and physiologic observations. *Journal of Psychosomatic Research, 11,* 325–339.

Eakin, E. G., Kaplan, R. M., & Ries, A. L. (1993). Measurement of dyspnea in chronic obstructive pulmonary disease. *Quality of Life Research, 2,* 181–191.

Eakin, E. G., Prewitt, L. M., Ries, A. L., & Kaplan, R. M. (1994). Validation of the UCSD Shortness of Breath Questionniare [abstract]. *Journal of Cardiopulmonary Rehabilitation, 14*(5), 322.

Fletcher, C. M. (1952). The clinical diagnosis of pulmonary emphysema—an experimental study. *Proceedings of the Royal Society of Medicine, 45,* 577–584.

Gift, A. G., Plaut, S. M., & Jacox, A. (1986). Psychologic and physiologic factors related to dyspnea in subjects with chronic obstructive pulmonary disease. *Heart & Lung, 15*(6), 595–601.

Guyatt, G. H., Berman, L. B., Townsend, M. Pugsley, S. O., & Chambers, L. W. (1987). A measure of quality of life for clinical trials in chronic lung disease. *Thorax, 42,* 773–778.

Guyatt, G. H., Pugsley, S. O., Sullivan, M. J., Thompson, P. J., Berman, L. B., Jones, N. L., Fallen, E. L., & Taylor, D. W. (1984). Walking test performance: The effect of encouragement. *Thorax, 39,* 818–822.

Guyatt, G. H., Thompson, P. J., Berman, L. B., Sullivan, M. J., Townsend, M., Jones, N. L., & Pugsley, S. O. (1985). How should we measure function in patients with chronic heart and lung disease? *Journal of Chronic Disease, 38*(6), 517–524.

Higgins, M. W. (1989). Chronic airways disease in the United States: Trends and determinants. *Chest, 96*(Suppl.), 328s–334s.

Kaplan, R. M., & Anderson, J. P. (1988). The quality of well-being scale: Rationale for a single quality of life index. In S. R. Walker & R. M. Rosser (Eds.), *Quality of life: Assessment and application* (pp. 51–77). London: MTP.

Kaplan, R. M., & Anderson, J. P. (1990). The general health policy model: An integrated approach. In B. Spiker (Ed.), *Quality of life assessments in clinical trials* (pp. 131–49). New York: Raven.

Kaplan, R. M., Atkins, C. J., & Timms, R. (1984). Validity of a quality of well-being scale as an outcome measure in chronic obstructive pulmonary disease. *Journal of Chronic Disease, 37*(2), 85–95.

Kaplan, R. M., Ries, A., & Atkins, C. J. (1985). Behavioral issues in the management of chronic obstructive pulmonary disease. *Annals of Behavioral Medicine, 7*(4), 5–9.

Kinsman, R. A., Yaroush, R. A., Fernandez, R. A., Dirks, J. F., Schnocket, M., & Fukuhara, J. (1983). Symptoms and experiences in chronic bronchitis and emphysema. *Chest, 5,* 755–761.

Mahler, D. A., & Harver, A. (1990). Clinical Measurement of Dyspnea. In D. A. Mahler (Ed.), *Dyspnea* (pp. 75–126). New York: Futura.

Mahler, D. A., & Harver, A. (1992). A factor analysis of dyspnea ratings, respiratory muscle strength, and lung function in patients with chronic obstructive pulmonary disease. *American Review of Respiratory Disease, 145,* 467–470.

Mahler, D. A., Weinberg, D. H., Wells, C. K., & Feinstein, A. R. (1984). The measurement of dyspnea: Contents, interobserver agreement, and physiologic correlates of two new clinical indexes. *Chest, 85,* 751–758.

Mahler, D. A., & Wells, C. K. (1988). Evaluation of clinical methods for rating dyspnea. *Chest, 93,* 580–586.

McGavin, C. R., Artvinli, M., & McHardy, G. J. R. (1978). Dyspnoea, disability, and distance walked: Comparison of estimates of exercise performance in respiratory disease. *British Medical Journal, 2,* 241–243.

Moody, L., McCormick, K., & Williams, A. (1990). Disease and symptom severity, functional status, and quality of life in chronic bronchitis and emphysema (CBE). *Journal of Behavioral Medicine, 13*(3), 297–306.

Mungall, I. P. F., & Hainsworth, R. (1979). Assessment of respiratory function in chronic obstructive airway disease. *Thorax, 34,* 254–258.

National Center for Health Statistics, Centers for Disease Control (1994). Mortality Patterns—United States, 1992. *Morbidity and Mortality Weekly Report, 43,* 916–921.

Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.

Pennebaker, J. W. (1982). *The psychology of physical symptoms.* New York: Springer-Verlag.

Radloff, L. S. (1977). The CES–D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385–401.

Renfroe, K. L. (1988). Effect on progressive relaxation on dyspnea and state anxiety in patients with chronic obstructive pulmonary disease. *Heart & Lung, 17,* 403–413.

Sassi-Dambron, D. E., Eakin, E. G., Ries, A. L., & Kaplan, R. M. (1995). Treatment of dyspnea in chronic obstructive pulmonary disease: A controlled clinical trial of dyspnea management strategies. *Chest, 107,* 724–729.

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *State–trait anxiety inventory (Form Y).* Palo Alto, CA: Consulting Psychologists.

Wilson, R. C., & Jones, P. W. (1989). A comparison of the visual analogue scale and modified Borg scale for the measurement of dyspnea during exercise. *Clinical Science, 76,* 277–282.