A systematic overview of the measurement properties of the Chronic Respiratory Questionnaire

Yves Lacasse MD MSc, Eric Wong MDMSc, Gordon Guyatt MD MSc

Design, Measurement and Evaluation Program, Department of Clinical Epidemiology and Biostatistics, McMaster University; Department of Medicine and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario

Y Lacasse, E Wong, G Guyatt. A systematic overview of the measurement properties of the Chronic Respiratory Questionnaire. Can Respir J 1997;4(3):131-139.

BACKGROUND: Assessment of health-related quality of life (HRQL) of patients with chronic lung disease has become an important aspect of many clinical investigations. The authors examined the measurement properties of a disease-specific HRQL questionnaire, the Chronic Respiratory Questionnaire (CRQ), when used by independent investigators in clinical trials and observational studies.
METHODS: All published papers citing the original 1987 CRQ publication were identified using the Science Citation Index, and abstracts presented at international conferences were found by hand search. Clinical trials and observational studies were included if they reported data bearing on the CRQ's measurement properties.
RESULTS: Of 90 papers and 20 abstracts, 32 met the inclusion criterion. CRQ domains of fatigue, mastery and emotional function have high reliability, and face, content and construct validity in differentiating among patients with better and worse HRQL. Because of its self-generated items, the dyspnea domain works less well in discriminating among patients with lesser and greater dyspnea. When CRQ has been used to evaluate treatment, all four domains have performed well in detecting small treatment effects. The minimal important difference in CRQ score (0.5 per item) provides guidance for both planning studies and interpreting results. To maximize CRQ interpretability, investiga-

tors should present results as the mean score per item within each domain on a seven-point scale.
CONCLUSION: The CRQ has proved valid and responsive to change. Its standardization and continued wide use will enhance the understanding of the impact of treatments on patients? HRQL.

Key Words: Chronic lung disease, Quality of life, Questionnaire

Un survol systématique des propriétés de mesure du questionnaire sur les maladies respiratoires chroniques

HISTORIQUE : L'évaluation de la qualité de vie liée à la santé (QVLS) de patients atteints d'une maladie pulmonaire chronique est devenue une partie importante de nombreuses investigations cliniques. Les auteurs ont examiné les propriétés de mesure d'un questionnaire de la qualité de vie liée à la santé et spécifique à une maladie, soit le questionnaire sur les maladies respiratoires chroniques (CRQ), quand il est utilisé par des chercheurs indépendants dans des essais cliniques ou dans des études d'observation.
MÉTHODES : Tous les articles faisant référence à la publication originale du CRQ datant de 1987 ont été identifiés en utilisant le *Science Citation Index* puis, on a procédé à une recherche manuelle de tous les abrégés présentés aux congrès. Les essais cliniques et les études d'observation ont été inclus lorsqu'ils rapportaient des résultats portant sur les propriétés de mesure du CRQ.
RÉSULTATS : Des 90 articles et des 20 abrégés, 32 remplis-

Correspondence and reprints: Dr Gordon Guyatt, Department of Clinical Epidemiology and Biostatistics, McMaster University, Health Sciences Centre, Room 2C12, 1200 Main Street West, Hamilton, Ontario L8N 3Z5. Telephone 905-525-9140 ext 22900, fax 905-577-0017, e-mail guyatt@fhs.csu.mcmaster.ca

saient les critères d'inclusion. Les domaines du CRQ sur la fatigue, la maîtrise de soi, et la fonction émotionnelle démontrent une forte fiabilité et une validité de structure, de contenu et apparente pour différencier parmi les patients ceux dont la QVLS est meilleure ou pire. Parce que ses items sont autogénérés, le domaine de la dyspnée fonctionne moins bien pour opérer une discrimination parmi les patients avec plus ou moins de dyspnée. Quand on a utilisé le CRQ pour évaluer le traitement, tous les 4 domaines ont réussi à détecter les effets minimes du traitement. La différence minimale importante dans les résultats des scores du CRQ (O,5 par

item) fournit un guide pour à la fois planifier des études et interpréter des résultats. Pour maximiser l'interprétation du CRQ, les chercheurs devraient présenter les résultats comme le score moyen par item à l'intérieur de chaque domaine sur une échelle de 7 points.

CONCLUSION : Le CRQ a fait la preuve de sa validité et de sa sensibilité aux changements. Sa normalisation et son usage largement répandu permettront d'améliorer la compréhension de l'impact des traitements sur la QVLS des patients.

Clinicians and investigators have concluded that determining the impact of a treatment requires a measurement of how the patient feels (1-6). This is especially true in diseases, such as chronic lung disease, in which treatment is largely symptomatic and often focused on rehabilitation (7). In response to the need for appropriate measurement instruments, the Chronic Respiratory Questionnaire (CRQ) was developed for patients with functionally limiting chronic lung disease, and initial evidence regarding the usefulness of the instrument was published in 1987 (8). This health-related quality of life (HRQL) measure has now been used by many investigators in diverse settings, and we felt the time was right to determine whether the questionnaire had met its initial promise. In this overview, we examine the measurement properties, including the reliability, validity, responsiveness and interpretability, of the CRQ when used by independent investigators in clinical trials and observational studies.

METHODS

CRQ: The CRQ is an interviewer-administered instrument that measures patients' dyspnea, fatigue, emotional function and mastery (the extent to which they feel they can cope with the disease and its manifestations). The dyspnea domain is individualized; that is, patients choose five activities during which they experience shortness of breath and that they perceive as important in their daily lives. Each domain includes four to seven items, and each item is scored on a scale of 1 to 7 (1 – extremely short of breath to 2 – very short of breath to 7 – not at all short of breath). The questionnaire can be administered with respondents aware (the 'informed' condition) or unaware (the 'blind' condition) of their previous responses (9,10).

Literature search: To retrieve papers citing the original 1987 CRQ publication (8), a computer search of the English-language medical literature from 1987 to 1995 using the Science Citation database was conducted. MEDLINE (1987 to 1995) was also searched using 'CRQ' and 'Chronic Respiratory Questionnaire' as keywords, and abstracts presented at international conferences (American Thoracic Society [1987 to 1995], European Respiratory Society [1987 to 1994] and American College of Chest Physicians [1987 to 1994] meetings) were searched by hand. Additional information concerning the study and use of the questionnaire was requested from the authors of the abstracts as well as from investigators who, it was thought, were conducting clinical

trials using CRQ. From all these sources, clinical trials and observational studies using CRQ to assess health status were selected.

Measurement properties – definitions

Discriminative instruments: A discriminative instrument is one that can distinguish among groups of patients; for instance, to identify patients with more severe dyspnea on daily activities and distinguish them from those with less severe dyspnea. To be useful, a discriminative instrument must be valid and reliable. Validity refers to whether the instrument is measuring what it claims to measure. Investigators demonstrate construct validity by showing that correlations with other measures are in the direction and magnitude that one would expect if the instrument under study is working the way it should (11). Reliability is the ability of the instrument to discriminate consistently among more and less affected patients and is best measured by an intraclass correlation coefficient that compares the variability among patients (the signal) to the variability within patients over time (the noise) (12). Investigators also sometimes look at an instruments' internal consistency, the extent to which different items in an instrument are measuring the same thing. The higher the correlation between items (and thus the higher the chosen statistic, often Cronbach's alpha), the greater the internal consistency (13).

Evaluative instruments: An evaluative instrument measures change over time. Responsiveness is the ability of an evaluative instrument to detect real change, even when it is small and is, along with validity, an important property of an evaluative instrument (14). For an evaluative instrument, validity is demonstrated by showing that longitudinal correlations of change in different measures conform to what one would expect if the instrument under study is measuring what it is supposed to measure.

Interpretability: For a discriminative instrument, a score is interpretable when it tells the reader whether the difference between two patients' function is negligible, small, moderate or large. For an evaluative instrument, a score is interpretable when it tells the reader whether a particular change in score represents a small, moderate or large clinical improvement or deterioration (14). Just as clinicians can distinguish between trivial (50 mL) or large (500 mL) differences in forced expiratory volume in 1 s (FEV_1), they should be able to interpret changes in CRQ scores obtained from trials and reported in the literature. If one knows the minimal important differ-

TABLE 1
Construct validity of Chronic Respiratory Questionnaire (CRQ) as a discriminative instrument: Correlation between CRQ and other health status measures

Reference	Sample size	CRQ correlated with	CRQ domains			
			Dyspnea	Fatigue	Emotional function	Mastery
Wijkstra et al (37)	40	Symptom checklist (SCL-90)				
		Anxiety	–	–	0.50*	−0.55*
		Depression	–	0.53*	0.49*	−0.48*
		Somatization	0.09	0.55*	0.52*	−0.40[†]
		Sensitivity	–	–	0.24	−0.27
Martin (30)	15	Medical Outcome Survey – Short Form 36				
		Physical functioning	0.57[‡]	0.31	0.44	0.14
		Role limitations – emotion	0.08	0.02	0.31	0.05
		Role limitations – physical	−0.05	0.03	0.40	−0.21
		General health perception	0.59[‡]	0.41	0.31	0.49
		Center for Epidemiologic Studies Depression Scale	−0.12	0.66[‡]	0.69[†]	−0.34
		Dyspnea Visual Analog Scale	0.66[‡]	0.35	0.33	0.37
Larson et al (39)	30	American Thoracic Society Breathlessness scale	-0.34[‡]	–	–	–
		Profile of Mood State (POMS)				
		POMS-fatigue	–	−0.74[‡]	–	–
		POMS (total score)	–	–	−0.57[‡]	–
Waterhouse et al (45)[§]	70	Medical Outcome Survey – Short Form 36				
		Physical functioning	0.42	0.31	0.16	0.46
		Role limitations – emotion	0.01	0.39	0.61	0.30
		Role limitations – physical	0.10	0.27	0.38	0.30
		General health perception	0.35	0.58	0.38	0.55
		Vitality	0.18	0.75	0.48	0.44
		Mental health	0.11	0.37	0.63	0.34
		Social functioning	0.23	0.38	0.29	0.34

$P<0.001$; [†]$P<0.01$; [‡]$P<0.05$; [§]Reported Spearman's rank correlation coefficients; statistical significance not provided

ence, the smallest difference in score that patients view as important and that would mandate a change in the patients' management (15), interpretation of changes in score is much easier.

RESULTS

Literature search: Ninety papers were identified through the Science Citation search. The MEDLINE search did not uncover any additional studies. Of the 90, 23 reported results of CRQ use (16-38); the others only mentioned the existence of the instrument. Twenty abstracts were retrieved from 16 different investigators, of which two abstracts contained sufficient information (39,40), and full manuscripts were obtained for another four (41-45). Finally, through personal knowledge and serendipidity, three other recent studies were identified (46-48), giving a total of 32 studies.

Discriminative properties

Face and content validity: Investigators have used the CRQ in at least 32 clinical trials or studies in patients with chronic lung diseases and translated the instrument into different languages for use in clinical trials. The high level of acceptance and use attests to the face validity of the CRQ.

Content validity assesses an instrument's comprehensiveness. The process of generating the original list of items including a review of the literature, suggestions of health professionals and in-depth interviews with patients, makes important omissions unlikely. There were 123 items on the original list: 62 items in the physical domains and 61 items in the emotional domain (49). One hundred patients reported the

frequency and importance of the items in their lives, and the most frequent and important items remained in the final questionnaire. Because only patients with moderate to severe chronic airflow obstruction were included, items particularly relevant to patients with mild disease may have been excluded. This problem is, at least, partly ameliorated by the individualization of the dyspnea domain.

Construct validity: The CRQ was originally designed as an evaluative instrument, and its construct validity as a discriminative instrument has only been examined by four independent groups of investigators (30,37,39,45). Table 1 summarizes the results of these studies. In interpreting the coefficients of correlation, the authors of the current paper qualified the strength of the correlations as follows: coefficients ranging from 0 to 0.20 denoted a negligible correlation; 0.21 to 0.35 a weak correlation; 0.36 to 0.50 a moderate correlation; and larger than 0.50 a strong correlation.

Because the sample size of Martin's study (30) was so small and, thus, the estimates of the correlations very imprecise, her results were of questionable significance. Overall, these studies of the discriminative validity of the CRQ (30,37,39,45) were interpreted as providing strong evidence of the construct validity of the fatigue, emotional function and mastery domains of CRQ in their discriminative function. It was not expected that the dyspnea domain, in which patients rate their shortness of breath on different activities, would discriminate among patients. Because patients are asked to choose activities that make them short of breath, it is likely that all will choose items that make them moderately

TABLE 2
Studies of the internal consistency (Cronbach's alpha) of Chronic Respiratory Questionnaire (CRQ)

Reference	Sample size	CRQ domains			
		Dyspnea	Fatigue	Emotional function	Mastery
Wijkstra et al (37)	40	0.51, 0.53*	0.78, 0.71*	0.81, 0.87*	0.83, 0.88*
Waterhouse et al (45)	61	0.64	0.85	0.84	0.80
Güell et al (46)	60	Not calculated	0.80	0.86	0.84

*CRQ administered twice

to severely dyspneic. Consequently, the variability between patients is small, and correlation between items or correlations between the dyspnea domain and other instruments are be attenuated. This does not, however, compromise the instrument in its evaluative function. Inferences regarding the discriminative validity of the other three domains would be strengthened if subsequent investigators make a priori predictions regarding the magnitude of the correlations they would expect if CRQ is measuring disease-specific HRQL and then test these predictions.

Reliability: In its initial development, CRQ was tested in 25 patients with stable chronic airflow limitation by a single interviewer six times at two-week intervals (8). Mean scores were similar in all four dimensions at each administration with no significant trend towards improvement or deterioration. The coefficient of variation (within person standard deviation divided by the mean) was 6% for the dyspnea domain, 9% for both fatigue and emotional function, and 12% for mastery. Güell and colleagues (46) obtained similar results during the validation process of a Spanish translation of CRQ in 30 stable patients with chronic obstructive pulmonary disease (COPD), the coefficients of variation ranging from 6% to 9%. These are comparable with other measures of respiratory function such as FEV_1 (8%) (50).

The reliability of CRQ has since been formally assessed by four groups of investigators. To examine the test-retest reliability of the CRQ before using it as a primary outcome in a respiratory rehabilitation trial, Wijkstra et al (37) administered a Dutch translation of the CRQ to 40 patients with stable COPD two days apart. Test-retest reliability, using the Spearman-Brown reliability coefficient, was good (P>0.7) in all four domains.

Martin (30) used a self-administered form of CRQ in 15 patients with COPD after completion of a pulmonary rehabilitation program. Eight of 15 patients returned nine days later for a retest, and three patients filled out the questionnaire at home on the day of the retest and returned it by mail. High correlations (r=0.68 to r=0.90) were found in three of the four domains. The fatigue domain did not show a significant correlation. Problems with this work included a very small sample size and the absence of any attempt to ensure patients were stable. For instance, the author mentioned that one patient arrived at the testing room by wheelchair for the first test, but walked to the room for the second test.

Larson et al (39) administered CRQ twice to 30 stable COPD patients one week apart. During the second administration, patients were blinded to their responses from the first test. Strong correlations (r=0.64 to r=0.76) were seen in all four domains. Finally, during a formal validation process of a Spanish translation of CRQ, Güell et al (46) administered the questionnaire twice to 60 patients with stable COPD three months apart. Intraclass correlations over the three months were 0.80, 0.68 and 0.67 for the fatigue, emotional function and mastery domains, respectively.

Internal consistency: Table 2 summarizes the findings of three studies (37,45,46) regarding the internal consistency of the CRQ. For the dyspnea domain, the investigators found a lower Cronbach's alpha. Again, this was attributed the lower internal consistency of the CRQ dyspnea dimension because of the individualization of the items in this domain.

Evaluative properties

Construct validity: Direct evidence of construct validity of the CRQ as an evaluative instrument was provided during its development (8). A number of a priori predictions about how closely changes in different measures (walk distance, global rating of symptoms by the patients) should correlate with changes in CRQ domains were generated. Weak to moderate correlations were found. Overall, good agreement was observed between the predicted and observed correlations (weighted Kappa 0.51; P<0.05).

Other investigators have calculated correlations between changes in CRQ scores and other measures. In a controlled trial of inhaled salbutamol and oral theophylline in 24 patients with COPD (23), the correlations between changes in CRQ-dyspnea scores and changes in FEV_1, 6 min walk distance and a global rating of dyspnea were 0.55, 0.52 and 0.84, respectively (P<0.01 for all correlations). Guimont and Bourbeau (44), using a French translation of the CRQ in a randomized controlled trial of inhaled steroids in 16 patients with COPD, found similar correlations. These authors also found significant correlations between changes in FEV_1 and CRQ-dyspnea scores (r=0.62; P<0.01) but no significant correlation between the changes in 6 min walk distance and the changes in CRQ-dyspnea scores. Finally, Güell et al (46) examined the correlations between the changes in each domain of CRQ and the change in pulmonary function and effort variables in a group of 30 patients admitted into a respiratory rehabilitation program lasting for six months. The only statistical significant correlations found were between the total lung capacity changes and the changes in the CRQ-fatigue scores (r=0.37; P=0.043), and between the changes in maximal work load in the cycle ergometer test and the changes in the CRQ-mastery scores (r=0.39; P=0.035).

Some of these correlations may seem disappointing. However, explanations for lower correlations may include small changes in pulmonary function and exercise testing results

TABLE 3
Chronic Respiratory Questionnaire (CRQ) in trials of respiratory rehabilitation in chronic obstructive lung disease

Reference	Design	Sample size	Data collection and presentation	Results
Busch and McClements (17)	Randomized controlled trial; rehab 18 weeks	Treatment: 7 Control: 7	Dyspnea domain only	Data not reported; "No change in dyspnea in either group"
Simpson et al (33)	Randomized controlled trial; rehab* 8 weeks	Treatment: 14 Control: 14	All four domains; total score within domains; graphical display (histograms) only	Significant effect on dyspnea and mastery (P<0.01) and fatigue (P<0.05). No change in the control group. The resolution of the histogram does not allow a precise estimate of the results on a 7-point scale
Vale et al (35)	Case series; six week rehab + 11 months' follow-up	29	All four domains of CRQ; results presented as a global score	Baseline 4.1±1.1; post-rehab: 5.2±1.1 (compared with baseline P<0.0001); follow-up 4.8±1.2 (compared with post-rehab P<0.0005; compared with baseline P<0.0001)
Wijkstra et al (38)	Randomized controlled trial; rehab 12 weeks	Treatment: 28 Control: 15	All four domains of the CRQ; total score within domains	Significant change in three domains (dyspnea, emotional, mastery). One example dyspnea: rehab baseline 3.9±0.9; 12 week 4.8±1.0 (P<0.001); control baseline 3.9±1.2; 12 week 3.8±1.3; differences in changes between groups P<0.01
Goldstein et al (20)	Randomized controlled trial; rehab eight weeks; measures done at 24 weeks	Treatment: 40 Control: 40	All four domains of the CRQ; scores within domains reported on a 7-point scale	Mean differences: dyspnea 0.6 (0.1 to 1.1), P=0.0061; fatigue 0.5 (0 to 0.9), P=0.0507; emotional function 0.4 (0.1 to 0.7), P=0.015; mastery 0.7 (0.4 to 1.1); P=0.0002
Güell et al (47)	Randomized controlled trial; rehab six months	Treatment: 27 Control: 29	All four domains of the CRQ; scores within domains reported on a 7-point scale	One example dyspnea rehab baseline 3.1±1.0; at 6 months 4.3±1.4 (P<0.001); control baseline: 3.2±1.0; at 6 months 3.1±1.1

*Exercise training: weight lifting. Rehab Rehabilitation

over time attenuating the correlations, and a true underlying weak relation among quality of life and pulmonary function test results and measures of exercise capacity (3-5). Inferences from all these correlations would have been strengthened if a priori predictions had been made regarding their magnitude.

The way the results obtained from CRQ are reported also influences their validity. At least one author (35) reported CRQ scores as a total encompassing all four domains. Patients who completed a six-week out-patient pulmonary rehabilitation program were followed to evaluate whether the gains in exercise capacity and quality of life could be sustained over time. Global CRQ scores were 81.2±21.4 at baseline, 104.7±22.2 after the program (a mean improvement of 1.2 per item) and 96.4±23.3 at 11 months' follow-up. While providing an overall assessment of HRQL, a single number may obscure important information, particularly when only small changes are observed. For instance, imagine an intervention that improves dyspnea but, for whatever reason, makes the patients lose control over the disease. Changes in both dyspnea and mastery will not be captured by an overall score.

Responsiveness of CRQ to respiratory rehabilitation trials: Comprehensive respiratory rehabilitation programs that include exercise therapy and any form of psychological support are now thought to improve exercise capacity without any change in expiratory flow rates. HRQL has been measured in some trials using a variety of health status indexes, including CRQ. Table 3 presents the cumulative experience with CRQ in trials of respiratory rehabilitation.

In general, the CRQ has proved responsive in detecting rehabilitation effects, not only in before-after studies but also in randomized control trials. Rehabilitation programs have generally led to small to moderate, but important, changes in HRQL. Even with the small sample sizes, CRQ was able to exclude the play of chance as an explanation of the differences. In only one study did the CRQ fail to detect a difference between treatment and control (17). In this study of a six-week home-based rehabilitation program, investigators used the dyspnea domain of CRQ as the sole health status index. Seven patients in each of rehabilitation and conventional community care groups completed the trial. As with any negative trial, explanations include the treatment not working, the outcome measure not being responsive, or the study, by virtue of its sample size, not being powerful enough to detect an underlying treatment effect. Given the positive results of all the other trials, the very small sample size seems to be the most likely explanation in this study.

Responsiveness of CRQ to other interventions: Waterhouse et al (45) directly assessed the responsiveness of the CRQ by comparing the self-perceived change in health status after various interventions in about 50 patients with chronic lung diseases with their CRQ changes in scores over a 12-month period. The self-perceived change was measured using a transition question that asked: "Compared to the last time you completed the questionnaire, how would you rate your health in general now?" The patients were then divided in three groups: those who improved, those who stayed the same and those who deteriorated. Mean differences in CRQ scores followed the direction of perceived health changes, supporting the ability of the CRQ to detect change following intervention.

Table 4 summarizes the cumulative experience with the CRQ in studies of specific interventions other than rehabili-

TABLE 4
Chronic Respiratory Questionnaire (CRQ) as an evaluative instrument in trials of interventions other than rehabilitation in patients with chronic obstructive pulmonary disease

Reference	Intervention design	Sample size	Data collection and presentation	Results reported on a 7-point scale
Jaeschke et al (26)	Multiple crossover trial of theophylline and/or salbutamol versus placebo	24	Physical = dyspnea + fatigue; emotional = emotion + mastery	Difference in scores between salbutamol and placebo: physical 0.6 (P=0.006); emotional 0.6 (P=0.015). Difference in scores between theophylline and placebo: physical 0.5 (P=0.027); emotional 0.3 (P=0.048)
Elliott et al (18)	Case series: nocturnal positive pressure ventilation	12; only six completed six months of therapy	All four domains; results expressed as % of maximal score; graphical display (histograms) only	"Trend towards improvement; no significant differences"; the resolution of the histogram does not allow a precise estimate of the results on a 7-point scale
Guyatt et al (21)	Randomized controlled trial of inspiratory muscle training (IMT)	IMT: 43; control: 39	Physical = dyspnea + fatigue; emotional = emotion + mastery	95% CI around the difference in changes in scores between groups included 0 and therefore were not significant: physical −0.6 to 0.2; emotional −0.6 to 0.1
Levine et al (28)	Case series: negative pressure ventilation	5	Dyspnea domain only	No data presented; "no change"
Lock et al (29)*	Crossover trial of liquid versus aqueous oxygen	15	All four domains	No data presented; "no consistent change in any of its domains"
McKay et al (31)	Crossover trial of high dose versus low dose of theophylline versus placebo	15	All four domains; scores reported on 0-100 scales	Within-patient differences: dyspnea difference =1.1, 95% CI 0.4 to 1.6; fatigue difference = 0.6, 95% CI 0.2 to 1.0
Weir and Burge (36)	Sequential placebo controlled parallel groups trial of high versus low dose inhaled steroids ± oral steroids	High dose: 51 Low dose: 47	All four domains; original scaling	Significant improvement in dyspnea and mastery domains; minimal important difference not reached. Example dyspnea domain – high dose baseline 3.6; after steroids 4.1; low dose baseline 3.5; after steroids 4.3; treatment effect P<0.0001
Crockett et al (43)	Case series: domiciliary oxygen therapy	57	Three domains (excluded the dyspnea domain); scores reported on 0 to 10 scales	No data presented; statistically significant improvement in all three domains in women only; MID reached for the emotional function only
Hansen et al (24)	Crossover trial of terbutaline dry powder versus nebulizer	25	All four domains. Scores reported in three different ways (1) original scaling; (2) physical = dyspnea + fatigue, psychological = emotion + mastery; (3) total score	Median differences (95% CI around the differences): dyspnea 0.0 (−0.6 to 0.2); fatigue 0.1 (0.0 to 0.5); physical 0.1 (−0.4 to 0.3); emotional 0.0 (−0.3 to 0.3); mastery 0.1 (0.0 to 0.5); psychological 0.1 (−0.1 to 0.4); total score 0.2 (−0.2 to 0.3).
Jaeschke et al (25)	Multiple crossover trial of inhaled terbutaline 500 versus 1000 versus 1500 µg	25	All four domains; original scaling	Differences in scores 500 versus 1500 µg: dyspnea: −0.1 (−0.3 to 0.1); fatigue: −0.04 (−0.3 to 0.2); emotional function: −0.01 (−0.2 to 0.2) mastery: −0.05 (−0.2 to 0.1)
Singh et al (34)	CRQ used to ascertain stability of the patients participating in the development of a new walk test	7	All four domains	No data presented; "no significant difference between study days"

*Study population patients with oxygen-dependent chronic lung disease. MID Minimally important difference

tation and shows that the CRQ has consistently shown theophylline's benefit in improving HRQL. The small sample sizes in the trials of Elliott et al (18) and Levine et al (28) result in an extremely low power to detect even large treatment effects, and thus fail to exclude treatment effects or raise concern about the unresponsiveness of the CRQ. For larger studies with negative results, possible explanations include the CRQ's unresponsiveness or the interventions' ineffectiveness. Given the CRQ's responsiveness to both rehabilitation and theophylline effects, the latter explanation seems more likely. This is especially true when one considers that there have been a number of other negative studies of inspiratory muscle training (50), and that, in both the inspiratory muscle training and high dose bronchodilator studies,

other measures of functional status also failed to show any differences in favour of the experimental treatment.

Interpretability: Clinicians will not find the results of health status measurement useful if they cannot understand the magnitude of the differences between treatment and control groups, and particularly if they cannot distinguish between trivial and important differences. Tables 3 and 4 include columns that present the size of CRQ score differences between experimental and control interventions, as reported by the investigators and on a common seven-point scale. Elliott et al (18) expressed results of the CRQ as the percentage of the maximal possible score while McKay et al (31) transformed CRQ scores to a 0 to 100 scale. In the initial work on the CRQ, the scores of each domain were presented by

simply adding the items, resulting in different maximum scores in each domain. The different ways of presenting the results make interpretation extremely difficult. It is suggested that dividing the score of each domain by the number of items, leading to possible scores of 1 to 7 on all domains, becomes the standard method of reporting.

Even if one standardizes the presentation of results, one still must decide what differences are trivial, small, medium and large. This issue has been addressed in a number of studies in which the CRQ was administered repeatedly, and the patients were asked to make global ratings of their changes in dyspnea, fatigue and emotional function. Changes in CRQ scores were then compared with the global ratings. The results were similar for all domains: the minimal clinically important difference was approximately 0.5; 1.0 represented moderate change; and changes greater than 1.5 represented large change (15). Using this methodology, Juniper at al (51,52) found similar ranges of small, medium and large effects in instruments using comparable seven-point scales for measuring quality of life in adult and childhood asthma.

Subsequently, Redelmeier et al (48) have used another method in which patients were asked to judge themselves relative to others with respect to their fatigue and emotional function (48). When the investigators used mean between-patient difference in CRQ scores for patients to start rating themselves as either 'a little bit better' or 'a little bit worse' to estimate the minimal clinically important difference, it once again was approximately 0.5. The method is not appropriate for the dyspnea domain because each patient has different activities.

Given this information, one can better interpret the results presented in Tables 3 and 4. Both rehabilitation and pharmacological interventions, when effective, demonstrate small- to medium-sized important effects on HRQL.

DISCUSSION

We have examined the measurement properties of the CRQ when used by independent investigators in clinical trials and observational studies. In this discussion, we will review a number of issues relevant to the use of the questionnaire.

Target population – what is 'chronic lung disease'? Review of the process of item generation (49) showed that CRQ has been directed primarily towards an adult population of patients suffering from COPD. Indeed, unstructured interviews with ambulatory patients with COPD (defined as a FEV_1 less than 70% of the predicted value associated with a ratio of FEV_1 to vital capacity less than 0.7 on all out-patient clinic visits during the previous two years) were used to draw items for the questionnaire. Validity of the CRQ when it is used in different populations may be limited. The questionnaire is not likely to remain valid in young patients with asthma – indeed, quality of life questionnaires developed for asthmatic patients look very different from CRQ (51) – but may remain useful in elderly patients with obstructive lung disease secondary to alpha-1 antitrypsin deficiency (40) or severe restrictive lung disease secondary to kyphoscoliosis.

Ideally, the measurement properties of CRQ need to be confirmed before use in a new population.

One surprising use of the CRQ is in studies of allergic (32) and nonallergic (19) rhinitis. In the former population, Meltzer et al (32) used the dyspnea and emotional function domains of CRQ in a double-blind, multicentre, randomized, placebo controlled, parallel group study of four weeks. A total of 123 patients assessed the effectiveness of one or two doses of ipratropium bromide aqueous nasal spray (IPBR) or placebo. The investigators chose these two domains because more than 95% of patients thought that rhinorrhea interfered with their daily activities and more than 80% thought that rhinorrhea affected them emotionally at least some of the time. The investigators found that both dosages of IPBR were statistically significantly better than placebo in reducing the mean duration and severity of rhinorrhea. The patients noted an improvement in their emotional function in CRQ with both dosages, but the improvement was only statistically significant with the higher dosage. There was no significant change in their dyspnea score.

Georgitis et al (19) used an identical design in patients with nonallergic rhinitis and found that a similar proportion of patients reported their rhinorrhea interfered with activities and emotional function. In this group of 174 patients, they found a statistically significant improvement in rhinorrhea and in CRQ-measured emotional function in patients receiving IPBR. Based on the evidence of these two trials, face validity was established because patients in both trials indicated their rhinorrhea influenced their HRQL in these two domains. However, a priori hypotheses should have been set for correlations of these two domains with other measurements for a more rigorous testing of construct validity. While the CRQ did well in these studies, instruments developed specifically for rhinitis are likely to do better (53).

Mode of administration: The CRQ is designed to be administered by a trained interviewer, and we provide a training manual and training tape for anyone interested in the questionnaire. Interviewer administration ensures compliance and minimizes misunderstanding. Limitations of self-administration include difficulties with both individualized items and informed administration of the questionnaire. However, self-administration may still be possible in a subset of patients, and the CRQ has been used as a self-administered questionnaire by some investigators (30).

Through personal communication with the authors of 12 studies included in Tables 3 and 4 (17-20,24,27,29-33,38), we learned that the CRQ has been administered as often with the respondents unaware (the blind condition) as aware (the informed condition) of their previous responses. The results from one observational study (9) and one randomized trial (10) suggested that informed administration enhanced the questionnaire's measurement properties. Therefore, we recommend that the investigators adopt the informed condition when administering the CRQ, though its performance appears satisfactory with either mode of administration.

Sample size and interpretation of CRQ results: Our knowledge of the differences in CRQ scores that correspond

TABLE 5
Sample size requirements to detect large and small treatment effects in randomized controlled trials and observational studies using Chronic Respiratory Questionnaire as an outcomeaccording to various levels of type I and II errors*

		Type I error							
	One-tailed	0.05		0.025		0.01		0.005	
	Two-tailed	0.1		0.05		0.02		0.01	
		RCT	Obs	RCT	Obs	RCT	Obs	RCT	Obs
	80%	8	6	9	7	11	8	13	9
		15	10	18	11	23	14	26	15
		50	28	63	32	81	41	94	47
Power: 1 – type II error	**85%**	9	7	10	7	13	9	14	9
		17	11	20	12	25	15	29	17
		58	32	72	36	91	46	105	53
	90%	10	7	12	8	14	9	16	10
		20	12	24	14	29	17	30	18
		69	35	85	43	105	53	120	60

*Upper number in each cell is the number of patients per group needed to detect a large treatment effect ($\Delta=1.5$); the middle number in each cell is the number of patients per group needed to detect a moderate treatment effect ($\Delta=1.0$); the lower number in each cell is the number of patients per group needed to detect a small treatment effect corresponding to the minimal clinically important difference ($\Delta=0.5$); σ assumed to be 1.0; n is increased according to Lachin's method (56) when the sample size is less than 30. Obs Observational studies; RCT Randomized controlled trial

to small but important, moderate and large treatment effects (0.5, 1.0 and 1.5, respectively) (15) allows investigators planning a clinical trial to compute sample sizes according to the magnitude of the treatment effect that they would like to detect (Table 5). Once a study is completed, readers can look to the size of the treatment effect that was observed (the point estimate of the treatment effect) and the confidence intervals around that point estimate to interpret the results. McKay et al (31) found a moderate difference of 1.1 in dyspnea as a result of theophylline administration (Table 4). However, the lower bound of the confidence interval around the difference, 0.4, represents an average effect that is less than the minimal important difference.

On the other hand, Busch and McClements (17) failed to exclude the role of chance as an explanation for the small differences in the CRQ that they found in patients receiving or not receiving rehabilitation. Their omission of the confidence interval around the difference that they found makes interpretation of their results more difficult. Because of the study's small sample size, it is likely that the upper boundary of the confidence interval, had it been calculated, would have included an important treatment-mediated difference. In another trial with a 'negative' result, Jaeschke et al (25) found that the mean difference in dyspnea between high and low dose terbutaline in a multiple crossover trial including 25 patients with COPD was –0.1, and the 95% CI was –0.3 to 0.1. The upper boundary of the confidence interval, 0.1, is considerably less than the minimal important difference. Therefore, we can be confident that higher doses of terbutaline do not have an important impact on patients' dyspnea in daily living.

Such clear and specific interpretation of CRQ results requires investigators to not only include both the point estimate and the confidence intervals, but to present results in a consistent fashion. We strongly recommend presenting a change in the average score per item on a 1 to 7 scale.

Cultural adaptability: The CRQ has been translated and used in clinical trials in Dutch (37), Danish (24), French (44), Spanish (46), German (27), Italian (42), Swedish and Chinese (personal communications). Details of most of these translations are not available. Studies using translations of CRQ have not reported major problems, and some have presented data suggesting strong measurement properties (37,46). Potential problems with translation and different cultures are manifold (54,55), and data regarding validity are a requirement for confident use of a new translation.

CRQ use and study design: The excellent measurement properties of the CRQ cannot compensate for studies with flawed designs. Randomization is required to avoid bias caused by noncomparability of treatment groups, and double-blinding is required to avoid placebo effects, cointervention and bias in assessment of outcomes. No matter how responsive an instrument, if a treatment is not adequately tested, the result is likely to be negative. For example, Kirsten et al (27) conducted a trial of theophylline therapy withdrawal in 38 clinically stable patients with severe COPD over six days. The CRQ was adapted by replacing 'two weeks' by 'two days' in all questions. Final results were reported as a global score, with the potential limitations already underlined. No significant change in the global score was observed. It may well be that even had the treatment been effective, two days was too short a time to observe changes in emotional function and mastery or to provide a sufficient sample of observations to judge dyspnea in day-to-day living.
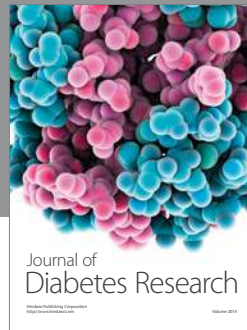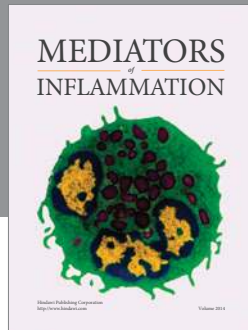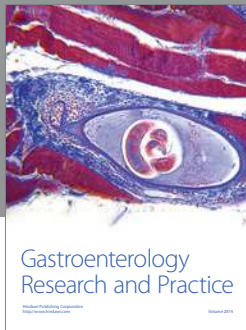
CONCLUSION

The CRQ is a disease-specific HRQL instrument that has, in the hands of a number of investigators, performed well in clinical trials concerning chronic lung disease. While its primary use is likely to remain as an outcome measure for clinical trials, three of the four domains can also be used to discriminate among patients with more versus less severe quality of life impairment. Use of the CRQ in chronic lung disease provides an example in which adequate standardization and widespread use of an instrument enhances our understanding of the impact of treatments on the way patients feel.

REFERENCES

1. Agle DP, Baum GL. Psychological aspects of chronic obstructive pulmonary disease. Med Clin North Am 1977; 61:749-58.
2. McSweeny AJ, Heaton RK, Grant IN, et al. Chronic obstructive pulmonary disease: socioemotional adjustment and life quality. Chest 1980;77:309-11.
3. Guyatt GH, Thompson PJ, Berman LB, et al. How should we measure function in patients with chronic heart and lung disease? J Chron Dis 1985;38:517-24.
4. Wegner RE, Jörres RA, Kirsten DK, Magnussen H. Factor analysis of exercise capacity, dyspnoea ratings and lung function in patients with severe COPD. Eur Respir J 1994;7:725-9.
5. Wijkstra PJ, Ten Vergert EM, van der Mark TW, et al. Relation of lung function, maximal inspiratory pressure, dyspnoea, and quality of life with exercise capacity in patients with chronic obstructive pulmonary disease. Thorax 1994;49:468-72.
6. Mahler DA, Harver A. A factor analysis of dyspnea ratings, respiratory muscle strength, and lung function in patients with chronic obstructive pulmonary disease. Am Rev Respir Dis 1992;145:467-70.
7. Canadian Thoracic Society Workshop Group. Guidelines for the assessment and management of chronic obstructive pulmonary disease. Can Med Assoc J 1992;147:420-8.
8. Guyatt GH, Berman LB, Townsend M, Pugsley SO, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. Thorax 1987;42:773-8.
9. Guyatt GH, Berman LB, Townsend M, Taylor DW. Should study subjects see their previous responses? J Chron Dis 1985;38:1003-7.
10. Guyatt GH, Townsend M, Keller JL, Singer J. Should study subjects see their previous responses: data from a randomized control trial. J Clin Epidemiol 1989;42:913-20.
11. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chron Dis 1985;38:27-36.
12. Streiner DL, Norman GR. Health Measurement Scales. Oxford: Oxford University Press, 1989:79-96.
13. Kramer MS, Feinstein AR. Clinical biostatistics. LIV. The biostatistics of concordance. Clin Pharmacol Ther 1981;29:111-23.
14. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993;118:622-9.
15. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Controlled Clin Trials 1989;10:407-15.
16. Addis GJ. What use is steady-state theophylline in severe chronic obstructive pulmonary disease – A preliminary report. Post Grad Med J 1991;67:S34-5.
17. Busch AJ, McClements JD. Effects of a supervised home exercise program for patients with chronic obstructive pulmonary disease. Phys Therapy 1988;68:469-74.
18. Elliott MW, Simonds AK, Carroll MP, Wedzicha JA, Branthwaite MA. Domiciliary nocturnal nasal intermittent positive pressure ventilation in hypercapnic respiratory failure due to chronic obstructive lung disease – Effects on sleep and quality of life. Thorax 1992;47:342-8.
19. Georgitis JW, Banov C, Boggs PB, et al. Ipratropium bromide nasal spray in nonallergic rhinitis: Efficacy, nasal cytological response and patient evaluation on quality of life. J Allergy Clin Immunol 1994;24:1049-55.
20. Goldstein RS, Gort EH, Stubbing D, Avendano MA, Guyatt GH. Randomised controlled trial of respiratory rehabilitation. Lancet 1994;344:1394-7.
21. Guyatt GH, Keller J, Singer J, Halcrow S, Newhouse MT. Controlled trial of respiratory muscle training in chronic airflow limitation. Thorax 1992;47:598-602.
22. Guyatt GH, Townsend M, Nogradi S, Pugsley SO, Keller JL, Newhouse MT. Acute response to bronchodilator: An imperfect guide to bronchodilator therapy in chronic airflow limitation. Arch Intern Med 1988;148:1949-52.
23. Guyatt GH, Townsend M, Keller JL, Singer J, Nogradi S. Measuring functional status in chronic lung disease: Conclusions from a randomized control trial. Respir Med 1989;83:293-7.
24. Hansen NCG, Evald T, Ibsen TB. Terbutaline inhalations by the Turbuhaler as replacement of domiciliary nebulizer therapy in severe chronic obstructive pulmonary disease. Respir Med 1994;88:267-71.
25. Jaeschke R, Guyatt GH, Willan A, et al. Effect of increasing dose of beta-agonists on spirometric parameters, exercise capacity, and quality of life in patients with chronic airflow limitation. Thorax 1994;49:479-84.
26. Jeaschke R, Guyatt GH, Singer J, Keller J, Newhouse MT. Mechanism of bronchodilator effect in chronic airflow limitation. Can Med Assoc J 1991;144:35-9.
27. Kirsten DK, Wegner RE, Jörres RA, Magnussen H. Effects of theophylline withdrawal in severe chronic obstructive pulmonary disease. Chest 1993;104:1101-7.
28. Levine S, Levy SF, Henson DJ. Effect of negative-pressure ventilation on ventilatory muscle endurance in patients with severe chronic obstructive pulmonary disease. Am Rev Respir Dis 1992;146:722-9.
29. Lock SH, Blower G, Prynne M, Wedzicha JA. Comparison of liquid and gaseous oxygen for domiciliary portable use. Thorax 1992;47:98-100.
30. Martin LL. Validity and reliability of a quality-of-life instrument. The Chronic Respiratory Disease Questionnaire. Clin Nurs Res 1994;3:146-56.
31. McKay SE, Howie CA, Thomson AH, Whiting B, Addis GJ. Value of theophylline treatment in patients handicapped by chronic obstructive lung disease. Thorax 1993;48:227-32.
32. Meltzer EO, Orgel HA, Bronsky EA, et al. Ipratropium bromide aqueous nasal spray for patients with perennial allergic rhinitis: a study of its effect on their symptoms, quality of life, and nasal cytology. J Allergy Clin Immunol 1992;90:242-9.
33. Simpson K, Killian K, McCartney N, Stubbing DG, Jones NL. Randomised controlled trial of weightlifting exercise in patients with chronic airflow limitation. Thorax 1992;47:70-5.
34. Singh SJ, Morgan MDL, Hardman AE, Rowe C, Bardsley PA. Comparison of oxygen uptake during a conventional treadmill test and the shuttle walking test in chronic airflow limitation. Eur Respir J 1994;7:2016-20.
35. Vale F, Reardon JZ, ZuWallack RL. The long-term benefits of out-patient pulmonary rehabilitation on exercise endurance and quality of life. Chest 1993;103:42-5.
36. Weir DC, Burge PS. Effects of high-dose inhaled beclomethasone dipropionate, 750 mg and 1500 mg twice daily, and 40-mg per day oral prednisone on lung function, symptoms, and bronchial hyperresponsiveness in patients with nonasthmatic chronic airflow obstruction. Thorax 1993;48:309-16.
37. Wijkstra PJ, Tenvergert EM, VanAltena R, et al. Reliability and validity of the Chronic Respiratory Questionnaire (CRQ). Thorax 1994;49:465-7.
38. Wijkstra PJ, van Altena R, Kraan J, Otten V, Postma DS, Koëter GH. Quality of life in patients with chronic obstructive pulmonary disease improves after rehabilitation at home. Eur Respir J 1994;7:269-73.
39. Larson JL, Covey MK, Berry JK, Wirtz S, Kim MJ. Reliability and validity of the Chronic Respiratory Disease Questionnaire. Am Rev Respir Dis 1993;147:A530.
40. Sherman S, Brown B, Wallace S, Sachs G, Knebel A. Severity of illness and quality of life in alpha-1 antitrypsin deficiency. Am J Respir Crit Care Med 1995;151:A449.
41. Clini E, Scalvini S, Simoni P, Bruletti G, Rocchi S, Quadri A. Measurement of health status in patients with chronic obstructive lung disease. Eur Respir J 1992;5(Suppl):217s.
42. Clini EM, Scalvini S, Simoni P, et al. Cycloergometer effort test evaluation and quality of life in chronic obstructive lung disease. Ital J Chest Dis 1993;47:71-6.
43. Crockett AJ, Cranston JM, Moss JR, Alpers JH. Survival and quality of life with chronic lung disease on long-term oxygen therapy. Am J Respir Crit Care Med 1994;149:A181.
44. Guimont C, Bourbeau J. Dyspnea assessed by the Chronic Respiratory Disease Questionnaire in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 1995;151:A466.
45. Waterhouse JC, Harper R, Marsh F, Jones NMB, Brazier JE, Howard P. Measuring quality of life in chronic obstructive pulmonary disease. Can we assess change? Eur Respir J 1994;7(Suppl):419s.
46. Güell R, Casan P, Sangenis M, Morante F, Belda J, Guyatt GH. The Spanish translation and evaluation of a quality-of-life questionnaire in patients with chronic obstructive pulmonary disease. Arch Bronconeumol 1995;31:202-10.
47. Güell R, Morante F, Sangenis M, Casan P. Effects of respiratory rehabilitation on the effort capacity and on the health-related quality of life of patients with chronic obstructive pulmonary disease. Eur Respir J 1995;8(Suppl):356s. (Abst)
48. Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. J Clin Epidemiol 1996;49:1215-9.
49. Guyatt GH, Townsend M, Berman LB, Pugsley SO. Quality of life in patients with chronic airflow limitation. Br J Dis Chest 1987;81:45-54.
50. Smith K, Cook D, Guyatt GH, Madhavan J, Oxman AD. Respiratory muscle training in chronic airflow limitation: a meta-analysis. Am Rev Respir Dis 1992;145:533-9.
51. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. J Clin Epidemiol 1994;47:81-7.
52. Juniper EF, Guyatt GH, Ferrie PJ, Griffith LE. Measuring quality of life in asthma. Am Rev Respir Dis 1993;147:832-8.
53. Juniper EF, Guyatt GH. Development and testing of a new measure of health status for clinical trials in rhinoconjunctivitis. Clin Exp Allergy 1991;21:77-83.
54. Guyatt GH. The philosophy of health-related quality of life translation. Qual Life Res 1993;2:461-5.
55. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46:1417-32.
56. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. Controlled Clin Trials 1981;2:93-113.

The Scientific World Journal

Gastroenterology
Research and Practice

MEDIATORS of INFLAMMATION

Journal of
Diabetes Research

Disease Markers

Journal of
Immunology Research

International Journal of
Endocrinology

PPAR Research

Hindawi

Submit your manuscripts at
http://www.hindawi.com

BioMed
Research International

Journal of
Ophthalmology

Stem Cells
International

Evidence-Based
Complementary and
Alternative Medicine

eCAM

Journal of
Obesity

Journal of
Oncology

Computational and
Mathematical Methods
in Medicine

Behavioural
Neurology

Parkinson's
Disease

AIDS
Research and Treatment

Oxidative Medicine and
Cellular Longevity