

Development of a self-reported Chronic Respiratory Questionnaire (CRQ-SR)

J E A Williams, S J Singh, L Sewell, G H Guyatt, M D L Morgan

Abstract

Background—The Chronic Respiratory Questionnaire (CRQ) is an established measure of health status for chronic obstructive pulmonary disease (COPD). It has been found to be reproducible and sensitive to change, but as an interviewer led questionnaire is very time consuming to administer. A study was undertaken to develop a self-reported version of the CRQ (CRQ-SR) and to compare the results of this questionnaire with the conventional interviewer led CRQ (CRQ-IL). **Methods**—Fifty two patients with moderate to severe COPD participated in the study. Subjects completed the CRQ-SR 1 week after completing the CRQ-IL, and a further CRQ-SR was administered 1 week later. For patients in group A (n=27) the dyspnoea provoking activities that they had previously selected were transcribed onto the second CRQ-SR, while patients in group B (n=25) were not informed of their previous dyspnoea provoking activities when they completed the second CRQ-SR. To assess the short term reproducibility and reliability of the CRQ-SR it was then administered twice at an interval of 7–10 days to a further group of 21 patients. The CRQ-IL was not administered. Longer term reproducibility was examined in 39 stable patients who completed the CRQ-SR at initial assessment and then again 7 weeks later.

Results—Mean scores per dimension, mean differences, and limits of agreement are given for each dimension in the comparison of the two questionnaires. There were no statistically significant differences between the CRQ-IL and CRQ-SR in the mastery and fatigue dimensions ($p>0.05$). A statistically significant difference between the two scores was found in the dyspnoea dimension ($p=0.006$) and the emotional function dimension ($p=0.04$), but these differences were well within the minimum clinically important threshold. No statistically significant difference in the mean dyspnoea score was seen between groups A and B. The CRQ-SR was found to be reproducible both in the short term and after the longer period of 7 weeks, with no statistically or clinically significant differences in any dimension. Test-retest reliability was found to be high in each dimension, both in the short and longer term.

Conclusions—The CRQ-SR is a reproducible, reliable, and stable measure of

health status. It compares well with the CRQ-IL but cannot be used interchangeably. The main advantage of the CRQ-SR over the CRQ-IL is that is quick to administer, reducing assessment time and hence cost.

(Thorax 2001;56:954-959)

Keywords: health status assessment; quality of life; Chronic Respiratory Questionnaire (CRQ); chronic obstructive pulmonary disease

The assessment of health status is an increasingly important outcome measure for chronic obstructive pulmonary disease (COPD). There is a large selection of tools available to measure health status and they can broadly be divided into those measures which are disease specific and those which are generic. A recent study investigating the effect of respiratory rehabilitation concluded that the responsiveness of generic measures to treatment effects in randomised trials in COPD was limited and that it was essential to include disease specific instruments among any measures of outcome.¹

There is a need for health status measures that are practical and easy to use in the clinical setting. The Chronic Respiratory Questionnaire (CRQ) is an established measure of health status which has been widely used for research purposes. It has been found to be reproducible^{2,3} and is sensitive to change.^{1,4-6} However, a major disadvantage of the CRQ is that it is an interviewer led questionnaire making it time consuming to administer,⁶ with the initial interview taking 20–30 minutes and subsequent interviews taking 10–15 minutes. A self-reported version of the CRQ would therefore be attractive. The purpose of this study was to develop a self-reported version of the CRQ and to compare the results of this questionnaire with those of the conventional interviewer led CRQ in subjects with moderate to severe COPD.

Methods

INSTRUMENT DEVELOPMENT

The conventional interviewer led CRQ (CRQ-IL) is divided into four dimensions of dyspnoea, fatigue, emotional function, and mastery (the patient's feeling of control over their disease). The questions covering the dimensions of fatigue, emotional function, and mastery are standardised and the patient is offered an appropriate 7 point scale for each question. The dyspnoea component is not standardised. The patient is required to identify everyday activities which make them breathless and then select, rank, and score the

Department of
Respiratory Medicine,
University Hospitals of
Leicester, Glenfield
Hospital, Leicester
LE3 9QP, UK
J E A Williams
S J Singh
L Sewell
M D L Morgan

Department of Clinical
Epidemiology and
Biostatistics,
McMaster University,
Hamilton, Ontario,
Canada
G H Guyatt

Correspondence to:
J E A Williams
johanne.williams@
uhl-tr.nhs.uk

Received 20 December 2000
Returned to authors
30 April 2001
Revised version received
13 July 2001
Accepted for publication
1 August 2001

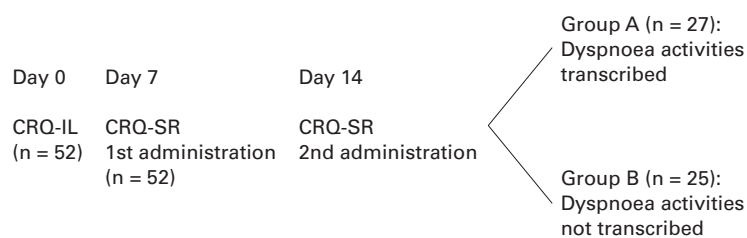


Figure 1 Flow chart showing study timetable for the repeatability of the dyspnoea dimension.

five most important activities on a 7 point scale which spans from 1 (extremely short of breath) to 7 (not at all short of breath). Every patient will have a unique list of activities. In each dimension the lower the score, the greater the degree of dysfunction.

The self-reported CRQ (CRQ-SR) was developed in conjunction with the original author (GG). The basic structure, content, and scoring of the CRQ-SR is exactly the same as the CRQ-IL. For the dimensions of fatigue, emotional function, and mastery the wording of the questions and answers has not been altered. The only difference is the format of the questionnaire—that is, the patient ticks an appropriate answer on a questionnaire rather than being asked a question by an interviewer. In the dyspnoea section patients select activities which make them breathless from a list of activities on the questionnaire (these are the same activities as the CRQ-IL) or they can volunteer any additional activities. They are then required to select, rank, and score the five most important activities in the conventional manner.

To evaluate the CRQ-SR we compared it with the CRQ-IL, then examined its reproducibility, stability, and reliability in subjects with moderate to severe COPD. The study was approved by the Leicestershire ethics committee.

COMPARISON WITH CRQ-IL

Fifty two patients with moderate to severe COPD who had been referred for pulmonary rehabilitation were recruited after written informed consent was obtained. At the initial assessment patients completed the CRQ-IL as part of the pre-rehabilitation assessment. At the end of the assessment patients were given the CRQ-SR to complete at home 1 week later. This was before starting rehabilitation. Patients were given simple instructions on how to fill in the questionnaire and it was completed unsupervised. Patients were not informed of the activities they had selected previously in the dyspnoea dimension of the CRQ-IL when they completed the CRQ-SR. In the CRQ-IL patients are asked questions about their health status relating to the previous 2 weeks. It was therefore important that patients were given the CRQ-SR within this time period in order to compare the two questionnaires directly. However, the CRQ-SR needed to be administered a sufficient time from the CRQ-IL to reduce the likelihood that patients could remember their previous responses. Seven days was considered to be a reasonable time frame in which to

administer the CRQ-SR. The questionnaires were given in the same order to the entire cohort—that is, CRQ-IL followed by CRQ-SR. This was necessary for pragmatic reasons based on service delivery.

REPEATABILITY OF DYSPNOEA DIMENSION

Unlike the dimensions of fatigue, emotion, and mastery, the dyspnoea dimension cannot be standardised, with each individual selecting his or her own unique list of activities that make them breathless. Some investigators have found the dyspnoea dimension to be less reliable than the other dimensions, with a low internal consistency.^{6,7} In the CRQ-IL the activities that are initially selected are used each time the questionnaire is administered. We wished to evaluate further the repeatability of the dyspnoea dimension if dyspnoea provoking activities were not transcribed onto a subsequent administration of the CRQ-SR—that is, the patient was allowed to select a new list of activities.

The original cohort of 52 patients who completed both questionnaires were asked to complete another CRQ-SR 7 days later. For the first consecutive 27 patients (group A) the dyspnoea provoking activities that were chosen initially were transcribed by an administrator onto the second administration of the CRQ-SR. The patients were then required to score those activities in the conventional manner. The next 25 patients (group B) were not informed of the dyspnoea provoking activities they had selected previously when they completed the second CRQ-SR—that is, they selected a second list of activities which they ranked and scored in the conventional manner (fig 1).

SHORT TERM REPRODUCIBILITY OF THE CRQ-SR

The short term reproducibility of the CRQ-SR was then examined in a further 21 patients. It was administered twice at an interval of 7–10 days to patients before starting pulmonary rehabilitation. For this cohort the CRQ-IL was not administered. An administrator transcribed the activities selected in the dyspnoea dimension from the initial CRQ-SR onto the second CRQ-SR. The activities were then scored in the conventional manner.

LONGER TERM STABILITY OF THE CRQ-SR

The longer term stability of the CRQ-SR was examined in a further 39 subjects. At the initial assessment patients were given the questionnaire. Exercise tolerance was assessed using the shuttle walk test (SWT)⁸ and forced expiratory volume in 1 second (FEV₁) was measured. The patients then entered a control period of 7 weeks, at the end of which they were given a second CRQ-SR with the dyspnoea activities initially selected transcribed by an administrator. Exercise tolerance and FEV₁ were re-tested to ensure clinical stability. Again, the CRQ-IL was not administered to this cohort.

STATISTICAL ANALYSIS

The method of scoring is identical for both formats of the questionnaire and the results are

Table 1 Comparison of mean (SD) scores per dimension between the CRQ-IL and CRQ-SR (n=52)

Dimension	CRQ-IL	CRQ-SR	Mean (SD) difference*	Limits of agreement†	p value‡
Dyspnoea	2.79 (0.88)	2.43 (0.86)	0.36 (0.89)	-1.38 and 2.10	0.006
Fatigue	3.14 (1.06)	3.30 (1.20)	-0.15 (0.95)	-2.01 and 1.71	0.25
Emotion	4.42 (1.24)	4.20 (1.14)	0.22 (0.77)	-1.29 and 1.73	0.04
Mastery	4.22 (1.40)	4.36 (1.35)	-0.15 (1.03)	-2.17 and 1.87	0.30

*Mean difference between CRQ-IL and CRQ-SR.

†Mean difference \pm 1.96 SD.

‡p = p value on difference between CRQ-IL and CRQ-SR from paired *t* tests.

presented as mean score per dimension, obtained by dividing the total score in each dimension by the number of questions in that dimension. This was on advice from the original author (GG) and produces an average score on the 7 point Likert scale, enabling comparisons to be made between dimensions—that is, all four dimensions have a range from 1 to 7.

For the comparison of the CRQ-IL and CRQ-SR the difference between the mean values was analysed using paired *t* tests and a *p* value of <0.05 was considered to be statistically significant. Bland and Altman have described the use of limits of agreement to evaluate how a new method of measurement agrees with an established technique.⁹ The mean difference (bias), standard deviation of the bias (SD), and limits of agreement are presented for each dimension.

To examine the reliability of the CRQ-SR (both in the short and longer term) the intraclass correlation coefficient (ICC) was used to assess the test-retest reliability of each dimension. The 95% confidence intervals for the ICC and the repeatability coefficient for each dimension⁹ are also presented. For the ICC a value of >0.7 was taken to be reliable.^{10 11}

All analyses were performed using the Statistical Package for Social Sciences (SPSS) version 9.0.

Results

COMPARISON WITH CRQ-IL

Fifty two patients (30 men) with moderate to severe COPD of mean (SD) age 66.5 (7.9) years and mean FEV₁ 1.13 (0.58) l completed both the CRQ-IL and the CRQ-SR. Table 1 presents a comparison of the mean scores per dimension and the limits of agreement for each dimension between the two questionnaires. No statistically significant difference in the mean score per dimension between the CRQ-IL and the CRQ-SR was seen in the fatigue and mastery dimensions ($p>0.05$), but a small significant difference was seen between the two scores in the emotional function ($p=0.04$) and dyspnoea dimensions ($p=0.006$). However, the mean difference between the two questionnaires in each dimension was well below the minimum clinically important threshold of 0.5 as described by Redelmeier *et al.*¹² The limits of agreement for each dimension are also presented graphically in fig 2.

REPEATABILITY OF DYSPNOEA DIMENSION

Table 2 presents the mean dyspnoea scores for groups A (n=27, 19 men, mean (SD) age 66.8 (8.2) years, mean (SD) FEV₁ 1.02 (0.56) l) and B (n=25, 11 men, mean (SD) age 66.2 (7.6) years, mean (SD) FEV₁ 1.3 (0.59) l). There were no statistically significant differences between the two administrations of the CRQ-SR in the two groups. The percentage agreement in the dyspnoea items selected between the two administrations of the

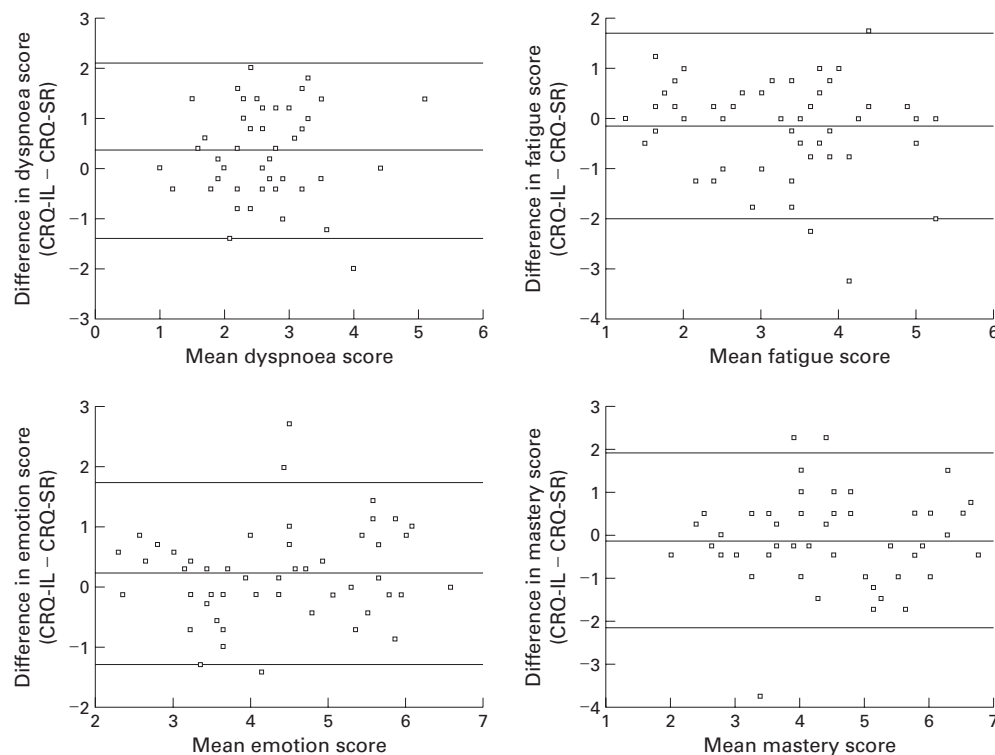


Figure 2 Difference in mean values for CRQ-IL and CRQ-SR data for each dimension.

Table 2 Comparison of mean (SD) dyspnoea scores between group A (dyspnoea responses transcribed, n=27) and group B (dyspnoea responses not transcribed, n=25) (two administrations of CRQ-SR 7 days apart)

	1st CRQ-SR	2nd CRQ-SR	Mean (SD) difference*	95% CI†	p value‡
Group A	2.5 (0.75)	2.59 (0.87)	-0.08 (0.61)	-0.33 to 0.15	0.45
Group B	2.35 (0.97)	2.56 (0.97)	-0.21 (0.57)	-0.44 to 0.03	0.08

*Mean difference between 1st and 2nd CRQ-SR.

†95% confidence interval around the difference between 1st and 2nd CRQ-SR.

‡p value on difference between 1st and 2nd CRQ-SR from paired *t* tests.

Table 3 Percentage of subjects in group B selecting different dyspnoea activities on the second administration of the CRQ-SR

Number of same dyspnoea activities selected on both administrations of CRQ-SR	Percentage of group B (n=25)
0/5 (no agreement of dyspnoea activities)	21%
1/5	10.5%
2/5	10.5%
3/5	21%
4/5	21%
5/5 (all 5 dyspnoea activities selected on both administrations)	16%

CRQ-SR for patients in group B is shown in table 3. Only 16% of patients selected exactly the same list of dyspnoea activities on both occasions while 21% selected none of the activities that were on their original list.

SHORT TERM REPRODUCIBILITY OF THE CRQ-SR

Short term reproducibility of the CRQ-SR was examined in 21 patients (12 men) of mean age 65.8 (7.8) years and mean FEV₁ 0.88 (0.47) l. Table 4 presents a comparison of the mean scores per dimension between the two administrations of the CRQ-SR 7–10 days apart. There were no statistically significant or clinically important differences in the mean score per dimension between the two administrations of the CRQ-SR in any dimension (p>0.05). The repeatability coefficient and the intraclass correlation for each dimension are also shown. All four dimensions had high test-retest reliability (ICC 0.83–0.95).

LONGER TERM STABILITY OF THE CRQ-SR

The stability of the CRQ-SR was examined in a further 39 patients (21 men) of mean age 69.7 (7.3) years, mean FEV₁ 0.82 (0.27) l,

Table 4 Short term reproducibility of the CRQ-SR: comparison of mean (SD) scores per dimension between two administrations of the CRQ-SR 7 days apart, repeatability coefficient, and test-retest reliability (ICC) within each dimension (n=21)

Dimension	1st CRQ-SR	2nd CRQ-SR	Mean (SD) difference*	95% CI†	p value‡	Repeatability coefficient	ICC (95% CI)
Dyspnoea	2.23 (0.86)	2.51 (1.24)	-0.29 (0.79)	-0.65 to 0.08	0.16	1.58	0.83 (0.58 to 0.93)
Fatigue	3.34 (1.41)	3.50 (1.40)	-0.15 (0.59)	-0.42 to 0.11	0.21	1.18	0.95 (0.89 to 0.98)
Emotion	4.31 (1.13)	4.45 (1.32)	-0.14 (0.69)	-0.46 to 0.17	0.66	1.38	0.91 (0.79 to 0.97)
Mastery	4.46 (1.27)	4.56 (1.39)	-0.09 (0.78)	-0.45 to 0.26	0.73	1.56	0.91 (0.78 to 0.96)

*Mean difference between 1st and 2nd CRQ-SR. †95% confidence interval around the difference between 1st and 2nd CRQ-SR. ‡p = p value on difference between 1st and 2nd CRQ-SR from paired *t* tests.

Table 5 Longer term stability of the CRQ-SR: comparison of mean (SD) scores per dimension between two administrations of the CRQ-SR 7 weeks apart, repeatability coefficient, and test-retest reliability (ICC) (n=39)

Dimension	1st CRQ-SR	2nd CRQ-SR	Mean (SD) difference*	95% CI†	p value‡	Repeatability coefficient	ICC (95% CI)
Dyspnoea	2.51 (0.87)	2.65 (1.07)	-0.13 (0.75)	-0.37 to 0.11	0.27	1.50	0.83 (0.67 to 0.91)
Fatigue	3.43 (1.13)	3.42 (1.12)	0.001 (0.77)	-0.24 to 0.26	0.92	1.54	0.87 (0.75 to 0.93)
Emotion	4.23 (1.15)	4.27 (1.34)	-0.004 (0.77)	-0.29 to 0.21	0.72	1.54	0.90 (0.80 to 0.94)
Mastery	4.35 (1.38)	4.22 (1.45)	0.13 (0.95)	-0.18 to 0.43	0.40	1.90	0.88 (0.76 to 0.93)

*Mean difference between 1st and 2nd CRQ-SR.

†95% confidence interval around the difference between 1st and 2nd CRQ-SR.

‡p = p value on difference between 1st and 2nd CRQ-SR from paired *t* tests.

mean SWT 186 (111) m. At the end of the 7 week control period FEV₁ and exercise tolerance were re-tested to ensure clinical stability: mean FEV₁ 0.84 (0.33) l, mean difference 0.02 l (95% CI -0.07 to 0.03), p=0.48; mean SWT 197 (128) m, mean difference -11 m (95% CI -29 to 8.0), p=0.24. Table 5 shows a comparison of the mean scores per dimension between the two administrations of the CRQ-SR 7 weeks apart. No statistically significant or clinically important differences between the two administrations of the CRQ-SR were seen in any dimension (p>0.05). The repeatability coefficient and the intraclass correlation for each dimension are also presented in table 5. All four dimensions demonstrated high test-retest reliability (ICC 0.83–0.90).

Discussion

The results of this study show that, in comparison with the “gold standard” CRQ-IL, the CRQ-SR detects similar levels of fatigue, mastery, and emotional function. Cook *et al.*¹³ in their comparison of the interviewer and self-reported formats of the Asthma Quality of Life Questionnaire, concluded that questionnaires which are designed to ascertain problems and dysfunction in patients with chronic illness produce similar results even when using different questionnaire formats. The results of our study are consistent with this finding, particularly for the standardised dimensions of emotional function, fatigue, and mastery.

In the comparison of the two different formats of the CRQ, patients were not informed of their dyspnoea responses from the CRQ-IL when they completed the CRQ-SR. It is therefore not surprising that we found a significant difference between the mean scores for the dyspnoea dimension (although the difference did not reach the threshold of the minimum clinically important difference¹²). The limits of agreement⁹ also suggest that, although the two formats of the CRQ may detect similar levels of dysfunction, they cannot be used interchangeably.

Some studies have suggested that self-reported questionnaires may yield more information, particularly about sensitive items, and that, when compared with an interviewer administered format, respondents tend to report higher levels of dysfunction.^{13 14} We did not find an overall trend for answering the CRQ-SR either more positively or negatively than the interviewer led questionnaire. The dimensions of dyspnoea and emotion showed higher levels of dysfunction in the self-reported questionnaire, but the mastery and fatigue dimensions showed lower levels of dysfunction. From their work on the Asthma Quality of Life Questionnaire, Cook *et al*¹³ recommended that, when an absolute level of dysfunction is required, a self-reported version of a questionnaire (if it exists) should be chosen to optimise the probability that subjects will report all problem areas.

Previous research has found the dyspnoea dimension of the CRQ-IL to be less reliable than the other dimensions.^{6 7} Wijkstra *et al*⁷ found the dimension to have a low internal consistency and recommended that it should not be included in the overall score in comparative research. However, these findings were not confirmed by Hajiro *et al*¹⁵ who found the internal consistency of the dyspnoea dimension to be as high as the mastery and fatigue dimensions. In the Spanish translation of the CRQ the internal consistency of the dyspnoea dimension was not analysed as the researchers felt the individual nature of the domain made it inappropriate for discriminating between patients.¹¹

We set out to explore how repeatable the dyspnoea dimension was if the activities that were chosen initially were not transcribed onto a subsequent administration of the CRQ-SR. No statistically significant differences were found in mean dyspnoea scores in a comparison of subjects whose dyspnoea responses either were or were not transcribed. It was therefore somewhat surprising that the agreement in the dyspnoea items chosen was so low, given that the two administrations of the CRQ-SR were only 1 week apart (only 16% selected exactly the same list on each occasion). However, Wijkstra *et al*⁷ found the reproducibility of the dyspnoea dimension to be lower than the other dimensions, even when the CRQ-IL was given only 1 day apart. Our results appear to indicate that the dimension reflects a more general perceived level of severity of dyspnoea rather than being activity specific. However, it is still recommended that the activities that are selected initially should be transcribed onto any subsequent administration of the CRQ-SR, as is the case for the CRQ-IL.

When evaluating test re-test reliability the usual approach is to administer the measures on two separate occasions separated by a time interval sufficiently short that it can be assumed that the underlying process is unlikely to have changed. The difficulty is in choosing the appropriate time interval—too long and things may have changed; too short and patients may remember their first response. In choosing a time interval of 7 days between the

two administrations of the CRQ-SR (or between the CRQ-IL and CRQ-SR), there could be a change in a subject's condition that may be reflected in changes to the CRQ-SR scores. However, this was a stable group of patients about to commence rehabilitation, thus reducing the likelihood that their condition had altered in the intervening week. Also, the questionnaire asks patients to reflect on their condition in the previous 2 weeks so there is a crossover period of at least 1 week.

The CRQ-IL has been extensively examined by previous researchers and found to be reproducible.^{2 6 7 11} We have shown that the CRQ-SR is reproducible, reliable, and stable. There were no statistically or clinically significant differences between the mean scores of the CRQ-SR in any dimension, either after an interval of 7 days or after a period of documented clinical stability of 7 weeks. The CRQ-SR also had high test-retest reliability. Our results compare well with other researchers who have studied the reliability of the CRQ-IL.^{2 6 7 11 12} Guell *et al*¹¹ in the Spanish translation of the CRQ found the intraclass correlation coefficient to be 0.80, 0.68, and 0.67 for the dimensions of fatigue, emotion, and mastery, respectively. Our figures were higher with results ranging from 0.83 to 0.95 for the short term reliability and 0.83–0.90 for the longer term reliability. These figures are similar in magnitude to those found by Harper *et al*⁶ in their analysis of the CRQ-IL in subjects who said their health had not changed over a 6 month study period.

The CRQ-IL has been shown to be sensitive to change in numerous clinical trials^{4 5 16} and further work has been undertaken to confirm whether the CRQ-SR is a similarly sensitive tool. A potential weakness of our study was that, because of service delivery constraints, all patients completed the questionnaires in the same order—that is, CRQ-IL followed by CRQ-SR. However, the 7 days between each administration makes the possibility of patients remembering their responses less likely and the same conditions applied to the entire cohort. The advantage of the CRQ-SR over the CRQ-IL is that it is quick to use, significantly reducing the time it takes to administer the questionnaire. Patients reported that the CRQ-SR took approximately 5–10 minutes to complete for the initial administration and 5 minutes for the second. After being given simple instructions, patients are able to fill in the questionnaire at home, thus increasing their privacy and the chance that they will answer questions more frankly.¹⁴ The response rate was found to be good and any sections not filled in correctly can be easily amended when the patient returns the questionnaire. The use of the CRQ-SR also has a cost implication, with the reduction of interviewer time helping to reduce the cost of research trials or clinical services.

In conclusion, the CRQ-SR compares well with the CRQ-IL although the two questionnaires cannot be used interchangeably. The CRQ-IL has been used extensively in clinical trials but is time consuming to administer. There is also the possibility that patients may not reveal

the true extent of their problems to an interviewer. The CRQ-SR has the advantage of greater perceived privacy for the patient and has been found to have a significant impact on assessment time and hence cost. It is a reproducible, stable, and reliable measure of health status in subjects with moderate to severe COPD.

Copies of the CRQ-SR can be obtained from the first author.

- 1 Guyatt GH, King DR, Feeny DH, *et al.* Generic and specific measurement of health-related quality of life in a clinical trial of respiratory rehabilitation. *J Clin Epidemiol* 1999; 52:187–92.
- 2 Guyatt GH, Berman LB, Townsend M, *et al.* A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987;42:773–8.
- 3 Guyatt GH, Townsend M, Keller J, *et al.* Measuring functional status in chronic lung disease: conclusions from a randomised control trial. *Respir Med* 1989;83:293–7.
- 4 Singh SJ, Smith DL, Hyland ME, *et al.* A short outpatient pulmonary rehabilitation programme: immediate and longer term effects on exercise and quality of life. *Respir Med* 1998;92:1146–54.
- 5 Griffiths TL, Burr ML, Campbell IA, *et al.* Results at 1 year of outpatient multidisciplinary pulmonary rehabilitation: a randomised controlled trial. *Lancet* 2000;355:362–8.
- 6 Harper R, Brazier JE, Waterhouse JC, *et al.* Comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting. *Thorax* 1997;52:879–87.
- 7 Wijkstra PJ, Ten Vergert EM, Van Altena R, *et al.* Reliability and validity of the chronic respiratory questionnaire (CRQ). *Thorax* 1994;49:465–7.
- 8 Singh SJ, Morgan MDL, Scott S, *et al.* Development of a shuttle walking test of disability in patients with chronic airway obstruction. *Thorax* 1992;47:1019–24.
- 9 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307–10.
- 10 Nunnally J. *Psychometric theory*. 2nd ed. New York: McGraw-Hill, 1978: 7.
- 11 Guell R, Casan P, Sangenis M, *et al.* Quality of life in patients with chronic respiratory disease: the Spanish version of the Chronic Respiratory Questionnaire (CRQ). *Eur Respir J* 1998;11:55–60.
- 12 Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimum important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol* 1996; 49:1215–9.
- 13 Cook DJ, Guyatt GH, Juniper E, *et al.* Interviewer versus self-administered questionnaires in developing a disease specific, health-related quality of life instrument for asthma. *J Clin Epidemiol* 1993;46:529–34.
- 14 Bergner M, Bobbit RA, Carter W, *et al.* The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981;19:787–805.
- 15 Hajiro T, Nishimura K, Tsukino M, *et al.* Comparison of discriminative properties among disease-specific questionnaires for measuring health-related quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998;157:785–90.
- 16 Guell R, Casan P, Belda J, *et al.* Long term effects of outpatient rehabilitation of COPD: a randomised trial. *Chest* 2000;117:976–83.