



This is a repository copy of *Comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/264/>

---

**Article:**

Harper, R., Brazier, J.E., Waterhouse, J.C. et al. (3 more authors) (1997) Comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting. *Thorax*, 52 (10). pp. 879-887. ISSN 0040-6376

---

**Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

## Comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting

R Harper, J E Brazier, J C Waterhouse, S J Walters, N M B Jones, P Howard

### Abstract

**Background** – To assist clinicians and researchers in choosing outcome measures for patients with chronic obstructive pulmonary disease attending routine outpatient clinics, a comparative assessment was undertaken of four questionnaires designed to reflect the patients' perception of their physical and emotional health in terms of their feasibility, validity, reliability, and responsiveness to health change.

**Methods** – Two condition specific questionnaires, the St George's Respiratory Questionnaire (SGRQ) and Guyatt's Chronic Respiratory Questionnaire (CRQ), and two generic questionnaires, the Short Form-36 Health Survey (SF-36) and Euroqol (EQ), were compared for their discriminative and evaluative properties. Spirometric tests and a walking test were also performed. One hundred and fifty six adults who were clinically judged to have COPD and who attended an outpatient chest clinic were assessed at recruitment and six and 12 months later. Patients were also asked whether their health had changed since their last assessment.

**Results** – Completion rates and consistency between items for dimensions of the SGRQ were lower than for dimensions of the other questionnaires. The distributions of responses were skewed for certain dimensions in all questionnaires except the CRQ. Validity was supported for all instruments insofar as patients' scores were associated with differences in disease severity. The generic questionnaires better reflected other health problems. All instruments were reliable over time. The condition specific questionnaires were more responsive between baseline and first follow up visit but this difference did not persist. While certain dimensions of the SF-36 were responsive to patient perceived changes, this did not apply to the derived single index of the EQ. The rating scale of the EQ, however, provided a quick and easy indicator of change.

**Conclusions** – Evidence from this study supports the CRQ and the SF-36 as comprehensive outcome measures for patients with longstanding COPD.

(*Thorax* 1997;52:879-887)

Keywords: chronic obstructive pulmonary disease, quality of life, questionnaires.

The treatment of patients with chronic obstructive pulmonary disease (COPD) is mainly palliative, aimed at relieving symptoms and enhancing quality of life. Despite this, routine outcome measurement for these patients tends to be limited to pathophysiological measures or survival. Methods for assessing health related quality of life are therefore required for two purposes: to enable discrimination between the degrees of severity in a condition, and to evaluate responsiveness to treatment.<sup>1</sup> So far there are few instruments for measuring health related quality of life in patients with chronic bronchitis and emphysema, and none which is satisfactory for use in routine clinical practice.<sup>2</sup> This is surprising since patients experience marked deterioration in their quality of life to the extent that even a limited amount of physical activity results in breathlessness which is debilitating and disturbing.

The problem for the clinician and the researcher is the choice of appropriate instruments. The Chronic Respiratory Questionnaire comprises items drawn from problem areas which have been identified by patients and their relatives, combined into three standardised dimensions – emotional function, fatigue, and mastery – with items for a fourth dimension (dyspnoea) being generated individually by each patient.<sup>3</sup> The St George's Respiratory Questionnaire, another standardised condition specific instrument, covers symptoms, activities and the impact of disease, but without the patient generated items.<sup>4</sup> Both instruments have been designed for a wide range of patients with respiratory disorders and independent validation has been recommended by their respective developers.

In addition to condition specific measures, there is a case for general measures of health status since there may be side effects of treatment which are missed by condition specific instruments. Furthermore, patients with respiratory disease commonly experience comorbidity such as cardiac and musculoskeletal problems, the symptoms of which will not be covered by COPD specific questions. The Short Form-36 Health Survey is a recent generic health status questionnaire which has been shown to be reliable and sensitive to low levels of morbidity.<sup>5</sup> In work with its predecessor, the SF-20, Stewart *et al*<sup>6</sup> demonstrated significantly different profiles of scores for patients with COPD compared with the general population, but the SF-36 has not yet been examined for this group of patients. Finally, given limited health care resources, it is necessary to dem-

School of Health and Related Research (SchHARR)

R Harper  
J E Brazier  
S J Walters  
N M B Jones

Department of Medicine and Pharmacology  
J C Waterhouse  
P Howard

University of Sheffield,  
Sheffield S1 4DA, UK

Correspondence to:  
Dr J E Brazier.

Received 20 December 1996  
Returned to authors  
20 February 1997  
Revised version received  
15 July 1997  
Accepted for publication  
16 July 1997

onstrate cost effectiveness both within and between disease groups, and for this a single index value for health is required. The Euroqol instrument is a recent example which is easy to use and brief, but has yet to be tested on this patient group.<sup>7</sup>

It is possible to use a transition question which asks patients to compare their health “now” with that on a specified earlier occasion in order to assess outcomes. This is well established in the field of arthritis<sup>8-11</sup> where they have been found to correlate with objective clinical variables. The transition question seems particularly relevant to COPD where conventional respiratory measures convey little about the patient’s experience.

In order to assist clinicians and researchers in choosing suitable instruments for discriminative and evaluative purposes in patients with COPD, we have undertaken a comparative assessment of the feasibility of using questionnaires in a routine outpatient clinic environment and have examined their psychometric properties of validity, reliability, and responsiveness to health change. It has been argued that questionnaires are more responsive to health change if they are condition specific, and especially if they are patient generated. This property has also been tested in this study.

## Methods

This observational study, for which ethical approval was granted, was carried out in the chest clinic of a city teaching hospital. Native English speaking patients aged 35 years and over and clinically judged to have COPD were recruited over a four month period. Patients with a clinical diagnosis of asthma, occupational and non-occupational lung fibrosis, and pulmonary malignancy were excluded, as were those whose spirometric tests gave  $FEV_1 > 70\%$  FVC or  $FEV_1 < 70\%$  FVC but with demonstrable reversibility. Main diagnosis and comorbidity together with sociodemographic information were obtained from medical records. Routine spirometric and pulse oximetry measures were obtained on the day of assessment for all patients and a subsample of patients was invited to undertake the six minute walking test (6MWT).

Patient perceived health was assessed using two condition specific and two generic questionnaires. Of the two condition specific questionnaires, the interviewer administered Chronic Respiratory Questionnaire (CRQ) examines four dimensions, with the items relating to the dimension of dyspnoea being identified by each patient individually. Using a seven point scale for responses, the scores for each question for each dimension are simply added together. The self administered St George’s Respiratory Questionnaire (SGRQ) employed in this study comprises 50 items, using weights derived by its developer from a patient population, and from which three dimension scores and a total score may be calculated. The Short-Form-36 Health

Survey (SF-36) is a self administered general health questionnaire of 36 items which generates a profile of scores across eight dimensions of health. Scores are transformed to range from 0 to 100, where 100 indicates good health on each dimension. The Euroqol Classification of Health (EQ) is a recently developed self administered instrument in two sections, the first measuring five dimensions of health from which a single index can be derived (EQ-5D), and the second a visual analogue rating scale measuring global health (EQ rating scale). In addition, patients provided information regarding recent treatments and their use of health and social services.

Three booklets of questionnaires for self administration were prepared – the first with the SF-36, the second with the SGRQ, and the third with the EQ. As a measure of disease severity, the MRC Respiratory Questionnaire<sup>12</sup> was included in the third booklet as were items on recent treatments and the use of resources. These booklets were handed to the patient after their consent to participate in the study had been obtained. Patients completed the booklets while waiting in the clinic, took uncompleted sections home to return promptly, or declined to attempt any further responses. A researcher was on hand in the clinic to answer queries about the questionnaires. The CRQ was administered by interview to a subsample obtained opportunistically in the clinic or at home within a few days of their clinic appointment.

Subsequent follow up assessments of all questionnaires were made after approximately six and 12 months during the patients’ routine appointments at the chest clinic. To examine responsiveness at each follow up the transition question (a modification of item 2 in the SF-36) was used which asked patients to compare their health “now” with that “six months earlier”.

The primary purpose of the analysis was to compare the discriminative and evaluative properties of the four instruments employed in this study. These were assessed, in the baseline and follow up periods respectively, using descriptive statistics as well as more rigorously for completion, internal consistency, validity, reliability, and responsiveness. All dimensions with missing items have been excluded from further analysis except for the CRQ where the recommended substitution procedure was used.

Internal consistency, the relationship between items within a dimension, was examined using the non-parametric correlation coefficient for item-to-dimension score, corrected for overlap, in addition to Cronbach’s alpha.

It is generally recognised that for outcome measures of health there is no “gold standard” for testing validity. However, one method is to examine construct validity where hypotheses or constructs concerning the expected distribution of health between groups may be examined by the measures being validated, with the significance of these differences being tested statistically using

standard *t* tests. The importance of the differences may then be investigated by calculating their effect size, which is the mean difference between groups divided by the pooled standard deviation. This is an indicator of the ability of a measure to discriminate the "signal" from the overall "noise" or variance. In the present analysis the magnitude of effect sizes has been judged against the criteria recommended by Cohen where  $\geq 0.2$  to  $< 0.5$ ,  $\geq 0.5$  to  $< 0.8$ , and  $\geq 0.8$  represent small, moderate, and large changes, respectively.<sup>13</sup> Test-retest reliability in patients who said their health had not changed was examined by intra-class correlations as a measure of agreement, with the mean differences between assessments and their confidence intervals inspected for possible bias.

Responsiveness has been measured in terms of the mean changes in scores between assessments in those who responded to the transition question by saying that their health had improved, worsened, or stayed the same. In order to undertake statistical comparisons of responsiveness, standardised response means (SRMs) have been estimated for each dimension score by dividing the mean change between assessments by the standard deviation of these changes in scores.

## Results

During the four month period 172 patients were invited to participate in the study. One hundred and sixty one patients signed consent forms and 11 patients (six men), who were not significantly older, refused. After signing consent forms five patients failed to answer any questionnaires, leaving a sample of 156 patients. At the initial assessment 152 patients returned the SF-36, 138 the SGRQ, 142 the EQ, and of the 68 patients who were approached all agreed to the administration of the CRQ by interview; 58 patients attempted all four quality of life questionnaires at the initial assessment, and 143 completed the MRC Respiratory Questionnaire.

During the follow up periods 128 patients returned questionnaires at the first follow up visit

after six months, 53 of whom attempted all four quality of life questionnaires, and 100 at the second follow up visit after 12 months, 47 of whom attempted all four questionnaires. Patients were lost to follow up because of death, being too ill, or being discharged from the chest clinic.

The sample comprised 76 men and 80 women whose mean (SD) ages were 67 (10.4) years and 62 (10.3) years, respectively. The mean FEV<sub>1</sub> for the sample was 47% of the predicted value for their age, height, and sex. Substantial morbidity was demonstrated, with 77% having to pause for breath when walking at their own pace on level ground, 83% with persistent coughing and 74% with persistent phlegm. Comorbidity was recorded in the medical records of 47% of cases, of which half was cardiac and a fifth musculoskeletal. Practically all patients used inhalers while 15% of patients were on oxygen therapy. In spite of their health problems, they received little support in their homes from health and social service workers, but made considerable use of health services such as GP consultations and hospital outpatient visits. Seventy seven per cent were retired or permanently unable to work and, compared with the population of the area,<sup>14</sup> they were more likely to be in semi-skilled and less likely to be in professional or managerial occupations.

The numbers of items completed varied considerably between instruments and between dimensions within each instrument. Completion rates of 98% were achieved for the administered CRQ except for the patient generated dimension of Dyspnoea (87%). Of the self completed questionnaires completion rates for dimensions varied from 92–96% (EQ) to 30–76% for the SGRQ. Completion of the Impact dimension of the SGRQ, and consequently the Total which is a composite of three dimensions, was the lowest (43% and 30%, respectively). Two items

Table 1 Descriptive statistics: dimension scores for patients at initial assessment

	<i>n</i>	Mean (SD)	Median	% patients on "floor" (worst health score)	% patients at "ceiling" (best health score)
SF-36 <sup>1</sup>					
Physical functioning	124	29.0 (25.0)	25.0	14.2	0.8
Social functioning	139	45.0 (27.5)	44.4	5.6	7.0
Role limitations (physical)	127	18.1 (32.8)	00.0	68.7	11.5
Role limitations (emotional)	122	44.8 (43.9)	33.3	40.5	34.1
Pain	138	53.1 (28.3)	55.6	4.3	13.5
Mental health	138	64.7 (21.0)	68.0	1.4	3.5
Vitality	133	34.5 (20.1)	35.0	6.6	0.0
General health perception	127	29.2 (18.4)	25.0	6.1	0.0
SGRQ <sup>2</sup>					
Symptoms	105	77.7 (17.1)	80.2	7.4	0.0
Activity	79	80.9 (18.9)	86.5	25.9	0.0
Impact	60	52.5 (18.5)	52.5	0.0	0.0
Total	42	65.4 (15.9)	64.8	0.0	0.0
CRQ <sup>3</sup>					
Dyspnoea	60	14.5 (4.5)	14.0	0.0	0.0
Fatigue	68	13.0 (4.3)	13.0	0.0	0.0
Emotional function	68	29.2 (7.5)	29.0	0.0	0.0
Mastery	68	17.9 (4.9)	19.0	0.0	0.0
EQ <sup>1</sup>					
EQ-5D	125	52.4 (15.7)	53.1	0.0	3.2
Rating scale	132	50.9 (16.4)	50.0	0.8	0.0

<sup>1</sup> For SF-36 and EQ: dimensions range from 0 (worst health) to 100 (best health).

<sup>2</sup> For SGRQ: dimensions range from 100 (worst health) to 0 (best health).

<sup>3</sup> For CRQ: each dimension has its own range.

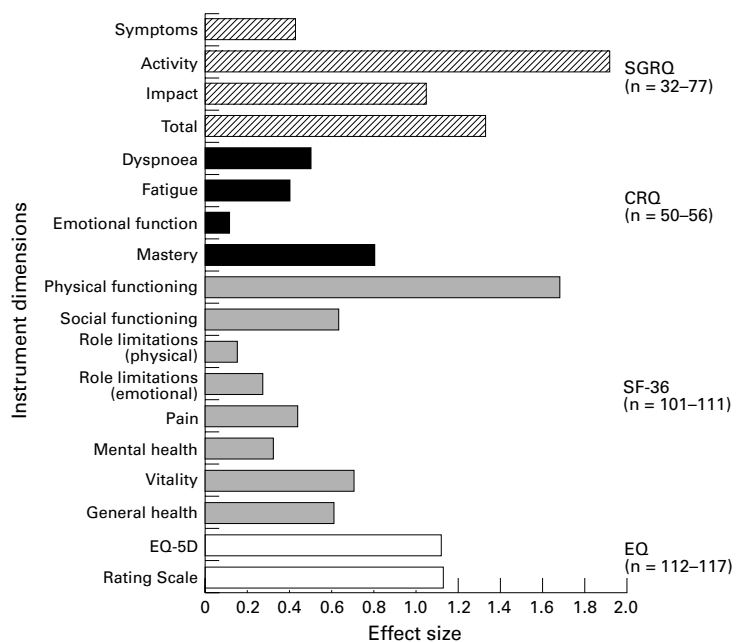


Figure 1 Effect sizes (mean difference between groups divided by pooled standard deviation: small  $\geq 0.2$  to  $<0.5$ , moderate  $\geq 0.5$  to  $<0.8$ , large  $\geq 0.8$ ) for the dimensions of the four instruments in relation to breathlessness (cut-off value: having to stop for breath when walking on level ground at own pace).

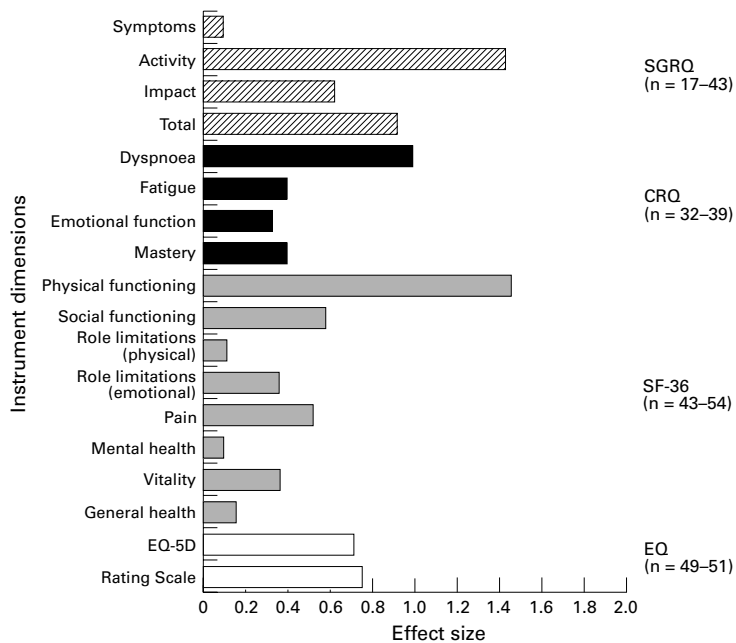


Figure 2 Effect sizes (mean difference between groups divided by pooled standard deviation: small  $\geq 0.2$  to  $<0.5$ , moderate  $\geq 0.5$  to  $<0.8$ , large  $\geq 0.8$ ) for dimensions of the four instruments in relation to distance walked (metres) on the six minute walking test (cut-off value: median average for the distance walked ( $\leq 302$  m;  $>302$  m)).

(“breathless when sitting or lying still” and “unpleasant side effects from my medication”) were missed by 28% and 27% of respondents, respectively, and five items (“breathless when playing sports and games”, “medication interferes with my life”, “chest trouble is a nuisance to my family”, “cannot move far from my bed”, and “my cough hurts”) were missed by 20% or more of respondents. Differences in age, sex, FEV<sub>1</sub>% predicted, and recent use of health services be-

tween patients who completed or did not complete the Impact dimension, which was the most incomplete, were small and not consistent.<sup>15</sup> Completion rates for the SGRQ were found to be slightly higher in the clinic than at home but they were still substantially below those of the SF-36 or EQ.<sup>15</sup> Completion rates for the SF-36 varied from 80% for the Role limitations (emotional) and 82% for Physical functioning to 91% for Pain, Mental health, and Social functioning.

By design, all items addressing a dimension of a questionnaire will be reasonably well correlated. Items were found to correlate with their hypothesised dimensions by 0.4 or more in 33 of 35 (94%) SF-36 items, 17 of 20 (85%) CRQ items, and 27 of 50 (54%) SGRQ items. Calculations of Cronbach’s alpha statistic ranged from 0.75 to 0.99 for the SF-36, 0.71 to 0.84 for SGRQ, and 0.80 to 0.85 for three dimensions of the CRQ, with the dimension measuring Dyspnoea reaching 0.64. This statistic is not relevant to the EQ where there was only one item per dimension.

Dimension scores were least skewed in the case of the CRQ, for which no respondents were either on the “floor” (the worst health score) or the “ceiling” (the best health score) (table 1). For the SF-36 the majority of respondents’ scores were on the “floor” or “ceiling” for the Role dimensions, while for Physical functioning 35% of respondents scored in the lowest decile, all of which indicate constraints on measuring changes in health from baseline. Only the Activity score distribution of SGRQ demonstrated an obvious “floor” effect of 26%. Substantial “ceiling” effects were observed on all dimensions of the EQ, although this was not evident in the distribution of the Rating scale nor the EQ-5D.

Support for construct validity of the SF-36 and the EQ was provided by a comparison of the scores profile for COPD patients with an age-matched and sex-matched sample of the general population drawn from the same city. This was markedly different for all dimensions of the SF-36 except Mental health. Similarly, patients’ scores on the EQ indicated poorer health compared with the same matched sample of the general population.<sup>16</sup>

All questionnaires were found to include dimensions which differentiated significantly between patients on the basis of their experience of breathlessness, where this was defined as severe if respondents stated that they stopped for breath when walking on level ground at their own pace (fig 1).<sup>15</sup> (As the numbers of questionnaires available for comparison varied, a comparison was made between patients who attempted all four questionnaires (n = 58) and those who attempted three or less (n = 98). The results for patients in these two groups revealed no significant differ-

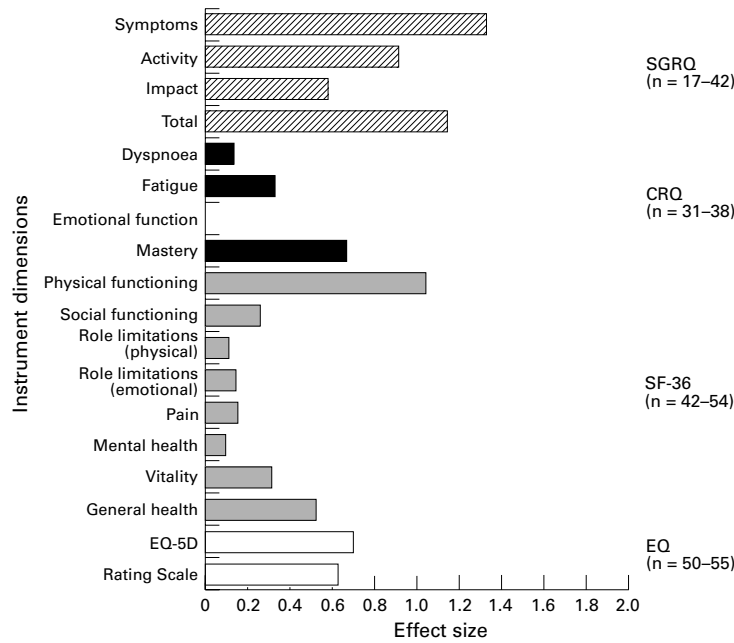


Figure 3 Effect sizes (mean difference between groups divided by pooled standard deviation; small  $\geq 0.2$  to  $< 0.5$ , moderate  $\geq 0.5$  to  $< 0.8$ , large  $\geq 0.8$ ) for dimensions of the four instruments in relation to breathlessness as measured by a visual analogue scale (VAS; 0 = not breathless; 100 = severely breathless) at the end of the six minute walking test (cut-off value: median average for VAS ( $\leq 65$ ;  $> 65$ )).

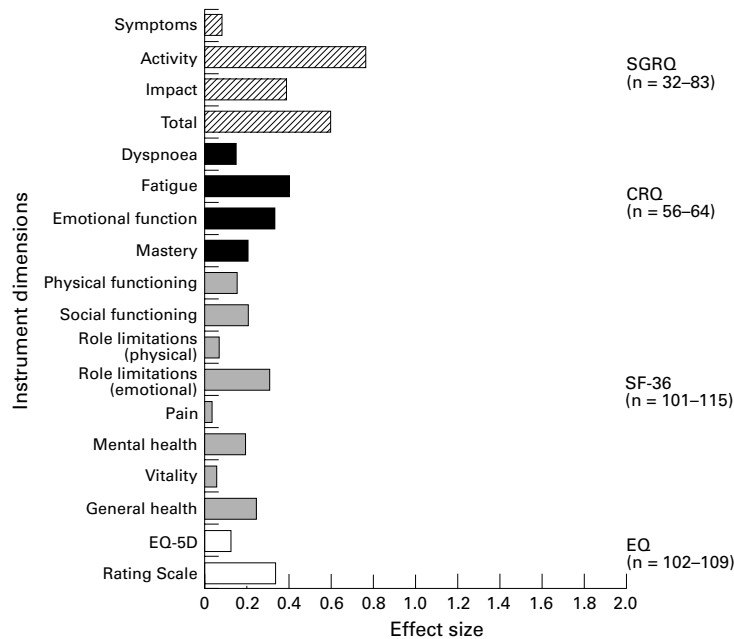


Figure 4 Effect sizes (mean difference between groups divided by pooled standard deviation; small  $\geq 0.2$  to  $< 0.5$ , moderate  $\geq 0.5$  to  $< 0.8$ , large  $\geq 0.8$ ) for dimensions of the four instruments in relation to FEV<sub>1</sub>% predicted (cut-off value: median average for FEV<sub>1</sub>% predicted ( $\leq 41$ ;  $> 41$ )).

ences with regard to age, sex, physical and respiratory function characteristics and questionnaire dimensions with one exception – patients who completed all four questionnaires walked, on average, further ( $p=0.01$ .) Differentiation on the basis of exercise tolerance, though in the expected direction, was less impressive (figs 2 and 3).<sup>15</sup> The physiological measure of FEV<sub>1</sub>%

predicted was not systematically related to any dimensions of any questionnaire.<sup>15</sup>

In addition to examining intra-questionnaire characteristics, inter-questionnaire comparisons may be made. The four questionnaires were compared directly by calculating effect sizes for the dimension scores in relation to breathlessness, distance walked in the 6MWT, breathlessness measured at the end of the 6MWT, and FEV<sub>1</sub>% predicted (figs 1–4; directions of signs have been ignored), and recent hospital admission and comorbidity.<sup>15</sup> For this purpose, patients were divided into groups representing severe and less severe breathlessness (fig 1), into two groups around the median for respiratory indicators (figs 2–4), and according to the presence or absence of other indicators.<sup>15</sup>

With regard to breathlessness on exertion, large effect sizes were observed for Activity, Impact and Total (SGRQ), Mastery (CRQ), Physical functioning (SF-36), and the two EQ dimensions; moderate effect sizes were observed for Dyspnoea (CRQ), Social functioning (SF-36); and small effect sizes were seen for Fatigue (CRQ), Symptoms (SGRQ), Pain, Mental health and Role limitations (emotional) (SF-36) (fig 1).

Tests of exercise tolerance produced some large or moderate effect sizes for both condition specific and generic questionnaires (figs 2 and 3). For FEV<sub>1</sub>% predicted effect sizes were moderate to small (fig 4).

Six dimensions of the SF-36 questionnaire identified patients who had been admitted to hospital within the last six months. The Pain and Physical functioning dimensions of the SF-36 and Activity, Impact and Total dimensions of the SGRQ distinguished the presence of comorbidity.<sup>15</sup>

Reliability over the first interval of six months was examined for all four instruments, revealing no evidence of bias between assessments (table 2). The 95% confidence intervals around the mean differences included zero for all dimensions except General health perception (SF-36), but exceeded 10 points (on the 100 point scales) for five dimensions of the SF-36. Reliability over the second six month period was similar but with slightly lower levels of correlation.<sup>15</sup>

Responsiveness for groups of patients was first examined by relating the differences in mean scores to patient perceived health change between initial assessment and first follow up, and between first and second follow up visits, and for this purpose the five categories have been combined into three – worse, same, and better.

At the first follow up visit the mean score differences were associated with the transition question for the two condition specific questionnaires (SGRQ and CRQ) with results for patients feeling “better” being associated with positive changes in scores and those for patients feeling “worse” being associated with negative changes in scores (table 3). These mean di-er-

Table 2 Intraclass ( $r_i$ ) correlation coefficients and mean differences between scores for those who said their health had not changed between initial assessment and first follow up ( $n = 58$ )

	<i>n</i>	<i>r<sub>i</sub></i>	Mean difference (SD)	95% CI
<b>SF-36</b>				
Physical functioning	35	0.86	-0.4 (11.7)	(-4.4 to 3.5)
Social functioning	49	0.42	-1.6 (26.7)	(-9.2 to 6.1)
Role limitations (physical)	39	0.21	8.3 (44.9)	(-6.1 to 22.7)
Role limitations (emotional)	39	0.18	6.0 (56.1)	(-12.0 to 24.0)
Pain	47	0.48	-4.3 (27.2)	(-3.7 to 12.2)
Mental health	46	0.74	0.4 (15.1)	(-4.0 to 4.9)
Vitality	47	0.60	1.6 (16.3)	(-3.1 to 6.3)
General health perception	41	0.44	7.7 (20.2)	(1.4 to 14.0)
<b>SGRQ</b>				
Symptoms	34	0.76	1.6 (13.4)	(-3.0 to 6.2)
Activity	21	0.74	-1.9 (9.8)	(-6.2 to 2.4)
Impact	16	0.46	-3.3 (16.0)	(-7.6 to 0.9)
Total	10	0.84	1.2 (8.7)	(-4.3 to 6.6)
<b>CRQ†</b>				
Dyspnoea	26	0.77	1.1 (12.2)	(-4.2 to 6.4)
Fatigue	25	0.92	0.3 (7.9)	(-5.0 to 5.7)
Emotional function	21	0.79	2.8 (13.8)	(-2.6 to 8.3)
Mastery	26	0.84	0.3 (11.9)	(-4.4 to 5.1)
<b>EQ</b>				
EQ-5D	42	0.67	0.8 (11.0)	(-3.2 to 4.8)
Rating scale	49	0.65	-0.2 (13.3)	(-4.0 to 3.6)

† The CRQ dimensions have been transformed onto a 0 (worst function) to 100 (best function) scale.

ences, which systematically reflected the direction of perceived health change, reached a high degree of significance for the dimensions of Dyspnoea and Fatigue in the CRQ and for Symptoms, Activity and Total in the case of SGRQ. For the generic questionnaires the mean score differences reached significance for Physical functioning and Social functioning while the mean difference in the EQ rating scale was significant, but not for the EQ-5D. Mean differences in respiratory function or distance walked in the 6MWT were not significantly related to perceived health change.<sup>15</sup>

These mean changes have been translated into standardised response means (SRMs) for those who reported a health change, in order to permit a more formal comparison of the responsiveness of these instruments (figs 5 and 6). For the first follow up period the SRMs observed for the symptom dimensions of each condition specific instrument – that is, Dyspnoea (CRQ) and Total

(SGRQ) – were large, while Fatigue (CRQ) and Symptoms and Activity (SGRQ) were moderate (fig 5). Of the generic instruments, Physical functioning and Social functioning (SF-36) and the Rating Scale (EQ) showed moderate SRMs. The SRM for the EQ-5D, however, failed to reach significance.

The picture was not as clear for the second follow up period (fig 6). The sizes of the SRMs for the symptom dimensions were considerably reduced to within the moderate to small range, and for other dimensions were inconsistent, with the SRM for the EQ-5D again below the criteria.

In the above calculations of responsiveness the results for all patients were included, which might have served to distort the changes. A more critical comparison of questionnaires was provided by calculating score changes and SRMs for the same patients completing the same questionnaires. In this stringent analysis patient groups were inevitably small, but the results for the complete sample were confirmed.

## Discussion

The results of the study have confirmed the feasibility of use and the acceptability of three of the four outcome measures for patients with COPD in an outpatient clinic setting, with high response rates being achieved. The numbers of patients attempting the four questionnaires were not identical, but the results for those attempting all four were similar to those who attempted three or less (data available on request).

The SGRQ can be self administered but, with the minimal supervision provided in this study, levels of item completion were relatively lower than for the other self administered questionnaires, suggesting low comprehensibility and/or irrelevance of some items. Furthermore, there were low correlations between some items and their dimensions, inferring a lack of homogeneity. Dimensions of the SGRQ which are obviously

Table 3 Mean score differences<sup>1</sup> between initial assessment and first follow up in relation to patient perceived health change

	Perceived health change by first follow up						<i>p</i> values <sup>2</sup>
	Worse		Same		Better		
	Mean (SD) difference	<i>n</i>	Mean (SD) difference	<i>n</i>	Mean (SD) difference	<i>n</i>	
<b>SF-36</b>							
Physical functioning	-13.0 (21.8)	30	0.4 (11.7)	35	6.6 (10.9)	16	0.001
Social functioning	-10.5 (20.1)	34	1.6 (26.7)	49	13.2 (24.4)	16	0.005
Role limitations (physical)	-4.2 (27.1)	30	-8.3 (44.9)	39	13.3 (28.1)	15	0.154
Role limitations (emotion)	7.1 (52.4)	28	-6.0 (56.1)	39	4.4 (45.2)	15	0.58
Pain	-1.3 (26.6)	34	4.3 (27.2)	47	-2.5 (17.3)	18	0.51
Mental health	-6.8 (20.5)	33	-0.4 (15.1)	46	-2.9 (15.5)	19	0.27
Vitality	-6.0 (17.6)	31	-1.6 (16.3)	47	2.1 (14.7)	19	0.23
General health perception	-5.7 (14.8)	32	-7.1 (20.2)	41	2.9 (10.9)	18	0.11
<b>SGRQ</b>							
Symptoms	-6.7 (17.0)	23	1.6 (13.4)	34	16.8 (11.4)	16	0.001
Activity	-3.7 (11.7)	19	-1.9 (9.8)	21	10.4 (10.7)	13	0.001
Impact	-2.0 (17.3)	13	-3.3 (16.0)	16	12.1 (14.8)	7	0.100
Total	-6.8 (12.0)	8	1.2 (8.7)	10	15.1 (9.0)	4	0.008
<b>CRQ</b>							
Dyspnoea	-9.5 (6.5)	14	-1.1 (12.2)	21	14.9 (16.2)	13	0.001
Fatigue	-3.9 (11.4)	16	-0.2 (7.9)	26	11.3 (14.2)	14	0.001
Emotional function	-3.4 (13.8)	16	-2.8 (13.8)	26	4.3 (7.7)	14	0.12
Mastery	-4.9 (12.0)	16	-0.3 (11.9)	25	4.2 (8.2)	14	0.09
<b>EQ</b>							
EQ-5D	-5.1 (12.1)	30	-0.8 (11.0)	42	-3.2 (10.7)	22	0.28
Rating scale	-10.5 (14.6)	33	0.2 (13.3)	49	3.6 (13.6)	21	0.001

<sup>1</sup> Where a mean difference >0 indicates a health improvement.

<sup>2</sup> *p* values from one-way ANOVA to compare mean score differences by perceived health change group.

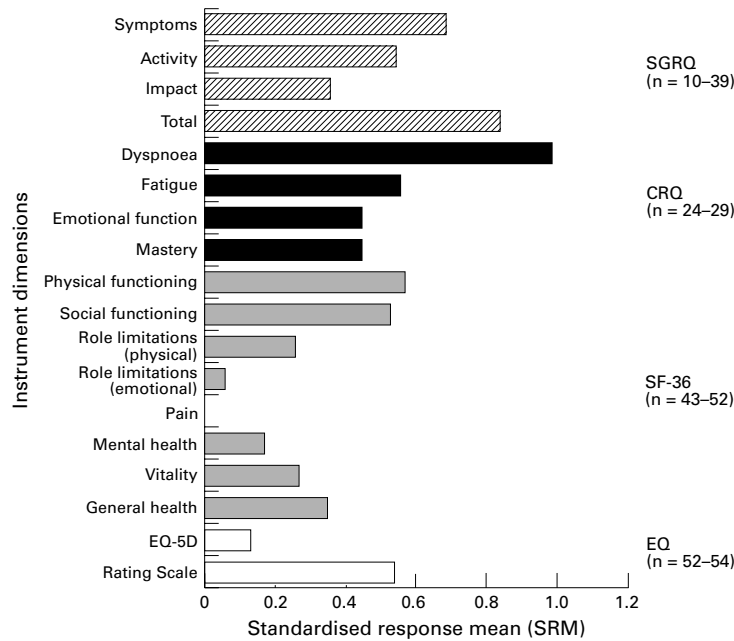


Figure 5 Standardised response means (SRM = mean change between assessments divided by the standard deviation of change: small  $\geq 0.2$  to  $< 0.5$ ; moderate  $\geq 0.5$  to  $< 0.8$ ; large  $\geq 0.8$ ) for dimensions of the four instruments for the first follow up period (between initial assessment and first follow up period six months later) for those who reported a change.

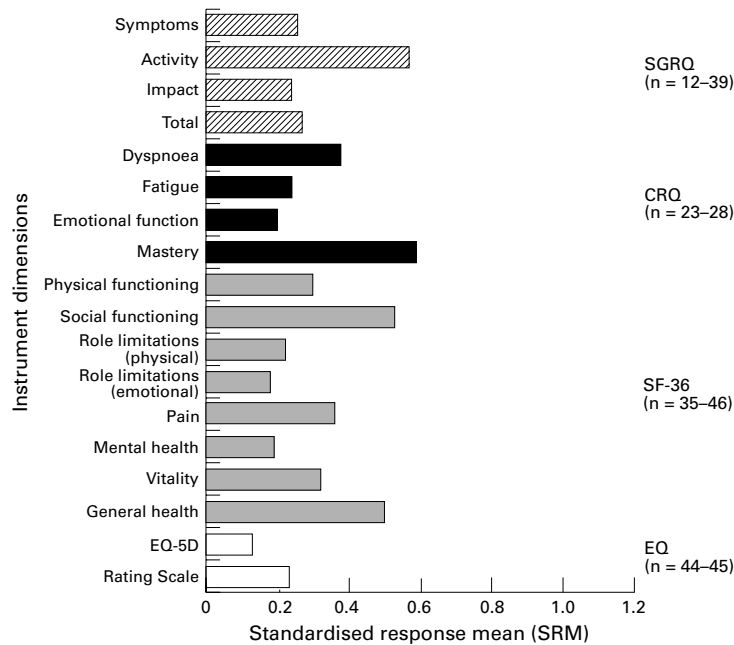


Figure 6 Standardised response means (SRM = mean change between assessments divided by the standard deviation of change: small  $\geq 0.2$  to  $< 0.5$ ; moderate  $\geq 0.5$  to  $< 0.8$ ; large  $\geq 0.8$ ) for dimensions of the four instruments for the second follow up period (between first follow up at six months and second follow up at 12 months) for those who reported a change.

related to mobility, such as Activity and Total, were found to have large and moderate effect sizes, clearly discriminating between patients on the severity of their disease. In addition, the more general dimensions of Activity, Impact and Total identified the presence of comorbidity. However, results obtained with this questionnaire in this study suggest that some revision may be required

before choosing it for an elderly patient group with long established chest disease.

The CRQ was used in a subsample of patients. In generating items which cause dyspnoea, patients provide the clinician with insight into the restrictions which the disease inflicts on their activities and life style. Being interviewer administered, completion levels were high. In addition, impressive properties of internal consistency, validity, and responsiveness were achieved. The widespread use of the CRQ in a clinic setting, however, is limited by the need for administration by interview and any version for self completion would need to be validated on a national population.

The design of the SF-36 as a measure of general health status was reflected in its wider scope to detect comorbidity and recent hospital admission and thereby indicates its potential use in assessing side effects of treatment and the impact of complications. It is clear that the Role limitations dimensions require re-designing, for example, to provide five point scales in place of their Yes/No response format as well as instructions to cover more clearly people who are no longer employed in the workplace<sup>17,18</sup> and in whom the disease onset is later in life. We would not propose to discard the dimension as a whole, since the fulfilment of a person's role is an important consideration in the field of rehabilitation medicine. Furthermore, we would add extra items to the Physical functioning dimension – such as walking within the home or getting up from a chair – in order to overcome the “floor” effect observed in our study.

A single index measure of health is an important tool for policy makers, applying both to choices between cost effective treatments for different groups of patients and to choices between treatments for the same group, and was designed for use in economic evaluations. The EQ in this study has demonstrated acceptability to patients, high rates of item completion, a reasonable distribution of responses for the Rating scale, and clear differentiation of patients according to the severity of their disease. The Rating scale, but not the EQ-5D, has been shown to be responsive to the minor changes in health which are typical of patients with chronic disease. Rating scales have been used with considerable success for other conditions and patient groups,<sup>19</sup> their disadvantages being that they provide less information about the nature of the change and are inadequate for cross-sectional comparisons. The results from this study suggest that the EQ-5D is inappropriate for use in this patient group.

In addition to the comparison of appropriate questionnaires for patients with COPD in an outpatient setting, this study has contributed to the discussion of more general issues.

With regard to exercise tolerance, the relationship observed between the walking test and physical factors measured by the questionnaires – such as physical function, activity and general health status – is not surprising and leads us to suggest that, if verified in other studies, questionnaires could replace the



walking test which many COPD patients find onerous. Again, the lack of association between health related quality of life and objective measures of respiratory function has illustrated the importance of obtaining subjective health measures for this group of patients.

Dimensions related to mental health in both condition specific and generic questionnaires appeared to be less relevant and showed little or no differences between patient groups. Perceived mental health in older people, including those with chronic ill health, has been reported to be good in other studies because of patients' lower expectations and coming to terms with limitations.<sup>20 21</sup>

This study has provided the first opportunity to undertake comparisons of responsiveness between the two leading condition specific instruments for COPD, both of which have been shown to be responsive to change. The condition specific dimensions were most responsive in the first follow up period, but this was not repeated in the second follow up period. A similar result was found by Fitzpatrick *et al*<sup>10</sup> with respect to attenuated responsiveness in the second follow up period of rheumatology patients. The instability in the degree of responsiveness, particularly for the key dimensions of Symptoms and Dyspnoea, is an interesting finding and clearly requires exploration in future. It could have implications for the repeated use of questionnaires in clinical trials.

Over and above the use of questionnaires to evaluate changes in health between groups of patients, clinicians may have an interest in the use of these instruments at the individual patient level. In this study the measurement of patient perceived health has provided important additional information to conventional measures such as spirometric tests. For a busy clinician the Rating scale of the EQ and the transition question have been shown to detect change, but neither provides insight into the nature and full extent of a patient's perceived health problem. Moreover, the transition question provides only crude categorical information about the extent of any improvement. The SGRQ and the SF-36 are in need of some revision. The CRQ is relatively expensive in the interview format and the self completed version would need to be validated. There remains, therefore, some caution in measuring and interpreting changes over time in individual patients.

Further development in appropriate outcome measures for patients with COPD is necessary. The SGRQ could be improved by modification and the SF-36 by the addition of items specific to COPD, while the self completed version of the CRQ would need to be validated. Another option is the addition of dimensions to a condition specific instrument to cover domains important to patients with COPD but which are currently excluded, such as pain. These and other developments would enhance the assessment of health and improve the quality of data. The measurement of patient perceived health in COPD is sufficiently important to warrant such effort.

In conclusion, the feasibility of using condition specific (CRQ and SGRQ) and generic (SF-36 and EQ) questionnaires with COPD patients in an outpatient setting has been demonstrated, with high response rates being achieved. The discriminative and evaluative properties of the condition specific CRQ and SGRQ have been confirmed in those patients who completed them. Some revision of the SGRQ would improve the internal consistency and completion of items. Both generic measures exhibited valid patterns in terms of breathlessness and exercise tolerance and comparisons on the basis of disease severity, while the SF-36 provided wider scope. Certain dimensions of the SF-36 and the Rating scale of the EQ, but not the EQ-5D, were responsive to health change.

Results from this study and elsewhere indicate that the optimum strategy in outcome measurement is to use a condition specific together with a generic instrument. This combines the advantages of greater responsiveness with the monitoring of those aspects of health especially affected by COPD, as well as providing a broader picture of a patient group with considerable comorbidity. With certain reservations outlined in the text, the present evidence suggests that the CRQ, SF-36 and Rating scale of the EQ are the instruments of choice for patients with chronic obstructive pulmonary disease.

The authors acknowledge the assistance of their colleagues in the Respiratory Function Unit and Chest Clinic, Royal Hallamshire Hospital, Sheffield, and the Medical Care Research Unit, University of Sheffield Medical School, and thank Dr Paul Jones and Dr Gordon Guyatt for the use of their questionnaires. The study was funded by the Medical Research Council of the United Kingdom.

- Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993;118:622-9.
- Jones PW. Quality of life measurement for patients with diseases of the airways. *Thorax* 1991;46:676-82.
- 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.
- Jones PQ, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992;145:1321-7.
- Brazier JE, Harper R, Jones NMB, O' Cathain A, Thomas KJ, Usherwood T, *et al*. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992;305:160-4.
- Stewart AL, Greenfield S, Hays RD, Wells K, Rogers WH, Berry SD, *et al*. Functional status and well-being of patients with chronic conditions. *JAMA* 1989;262:907-13.
- The Euroqol Group. Euroqol - a facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-208.
- Garratt AM, Ruta DA, Abdalla MI, Russell IT. SF-36 health survey questionnaire: II. Responsiveness to changes in health status in four common clinical conditions. *Quality Health Care* 1994;3:186-92.
- Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. Transition questions to assess outcomes in rheumatoid arthritis. *Br J Rheumatol* 1993;32:807-11.
- Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. *J Rheumatol* 1993;20:429-36.
- Ziebland S, Fitzpatrick R, Jenkinson C, Mowat A, Mowat A. Comparison of two approaches to measuring change in health status in rheumatoid arthritis: the Health Assessment Questionnaire (HAQ) and Modified HAQ. *Ann Rheum Dis* 1992;51:1202-5.
- Medical Research Council. *BMJ* 1960;1:1665.
- Cohen J. *Statistical power analysis for the behavioural sciences*. New York, Academic Press: 1978.
- General Household Survey. London, HMSO, 1988.
- Brazier JE, Howard P, Williams BT. *Comparison of outcome measures for patients with COPD*. Report to Medical Research Council, London (Ref: G9203850).

- 16 Brazier J, Jones N, Kind P. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. *Quality of Life Res* 1993;2:169-80.
- 17 Hayes V, Morris J, Wolfe C, Morgan M. The SF-36 health survey questionnaire: is it suitable for use with older adults? *Age Ageing* 1995;24:120-5.
- 18 Jenkinson C, Lawrence K, McWhinnie D, Gordon J. Sensitivity to change of health status measures in a randomized controlled trial: comparison of the COOP charts and the SF-36. *Quality of Life Res* 1995;4:47-52.
- 19 Nicholl JP, Brazier JE, Milner PC, Westlake L, Kohler B, Williams BT, *et al.* Randomised controlled trial of cost-effectiveness of lithotripsy and open cholecystectomy as treatments for gallbladder stones. *Lancet* 1992;340:801-7.
- 20 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.
- 21 Cassileth BR, Lusk EJ, Strouse TB, Miller DS, Brown LL, Cross PA, *et al.* Psychosocial status in chronic illness: a comparative analysis of six diagnostic groups. *N Engl J Med* 1984;311:506-10.