A randomised trial to evaluate the self-administered standardised chronic respiratory questionnaire

H.J. Schünemann*,#,¶, R. Goldstein+, M.J. Mador*,§, D. McKim/, E. Stahl**,
M. Puhan##, L.E. Griffith¶, B. Grant*,§, P. Austin¶, R. Collins+ and G.H. Guyatt¶,¶¶

ABSTRACT: The original chronic respiratory questionnaire (CRQ), one of the most widely used measures of health-related quality of life (HRQL) in chronic respiratory disease (CRD), is traditionally interviewer administered (IA) and includes an individualised dyspnoea domain. The present authors studied the impact of self-administered (SA) and standardised dyspnoea questions on CRQ measurement properties.

In a factorial design multicentre trial, 177 patients with CRD (mean age 67.7 yrs; mean forced expiratory volume in one second per cent predicted 44.6%) were randomised to CRQ-IA (n=86) or CRQ-SA (n=91), and to initially complete the standardised or individualised items before and after respiratory rehabilitation.

While maintaining validity, the CRQ-SA proved more responsive to changes in HRQL than the CRQ-IA in all domains. Compared with the standardised dyspnoea domain, the individualised dyspnoea domain indicated greater responsiveness. The correlations of baseline scores and change scores with other HRQL instruments indicated good validity of the CRQ-SA.

In conclusion, self-administration and standardisation of the chronic respiratory questionnaire maintains validity and responsiveness relative to the interviewer-administered chronic respiratory questionnaire. These results challenge the assumption that interviewer-administered questionnaires are superior to self-administered questionnaires in older patients with chronic respiratory disease.

KEYWORDS: Chronic obstructive pulmonary disease, health-related quality of life, pulmonary rehabilitation, randomised controlled trial, standardised administration

Clinicians and investigators acknowledge the importance of health-related quality of life (HRQL) in clinical studies of patients with chronic respiratory disease (CRD). The chronic respiratory questionnaire (CRQ), one of the most widely used measures of HRQL in patients with CRD [1–8], has served as a model in many methodological HRQL studies [9–13]. The original CRQ includes 20 items across four domains: dyspnoea (five items), fatigue (four items), emotional functioning (seven items) and mastery (four items). When completing this instrument, patients rate their experience on a seven-point scale, ranging from 1 (maximum impairment) to 7 (no impairment).

While interviewer-administered CRQ (CRQ-IA) may help prevent missing responses and errors in completing the CRQ and other respiratory questionnaires, it limits an instrument's feasibility in large clinical studies [14]. The individualised

dyspnoea domain of the CRQ also potentially increases the complexity and burden on patients and interviewers [14–17]. In the original CRQ, patients choose five important daily activities (individualised dyspnoea items) and report their degree of dyspnoea whilst doing those activities. Individualisation potentially enhances validity and responsiveness (the ability of the CRQ to detect important changes, even if those changes are small) [18].

There is little evidence from randomised studies about the impact of self-administered CRQ (CRQ-SA) *versus* CRQ-IA on measurement properties of HRQL instruments, in particular in patients with CRD. Two observational studies have showed that CRQ-SA scores have no impact on measurement properties of the CRQ [19, 20]. Similarly, while the original CRQ has proved both responsive and valid [21], little is known about the extent to which the individualised

questions improve or impair measurement properties. In one observational study and one randomised study performed in German-speaking countries, standardisation of dyspnoea items reduced the time needed for completion of the CRQ from 16 to 8 min, but impaired CRQ responsiveness [18, 22]. However, previous studies were limited by small sample sizes and observational designs [18–20], and the lack of concomitant administration of other disease-specific instruments needed to test the validity of a standardised CRQ-SA [22].

The aim of the current study was to determine how two modifications of the CRQ, self-administration and standardisation of dyspnoea items, influences CRQ measurement properties. The authors chose the most rigorous study design to address these questions. First, the patients were randomised to CRQ-IA or CRQ-SA before and after a respiratory rehabilitation programme in several rehabilitation centres. Secondly, all patients completed the individualised and standardised versions of the dyspnoea domain, but the order in which they completed them was randomised. Patients also completed a number of other measures related to HRQL, facilitating exploration of CRQ validity. The setting of respiratory rehabilitation was chosen because of its demonstrated effectiveness in improving a patient's HRQL [23].

METHODS
Patients and study design
The current authors conducted the randomised, multicentre trial at four university respiratory rehabilitation programmes, in Toronto, Ottawa, Hamilton (ON, Canada) and Buffalo (NY, USA). Eligible patients included all inpatients and outpatients with CRD starting rehabilitation between September 2001 and March 2003. The rehabilitation programmes were all 8 weeks in duration and included an exercise component. Recruitment was terminated in Hamilton in June 2002 due to the death of the site investigator (D. Stubbing). Follow-up appointments in Toronto proved unfeasible during the severe acute respiratory syndrome (SARS) outbreak, because of temporary closure of the rehabilitation hospital. The hospital housing the Toronto rehabilitation facility served as one of the main care facilities early on during the SARS outbreak.

Patients who were unable to complete the questionnaires due to language limitations were ineligible. The initial interviews and completion of questionnaires took place during the patient's admission to the programme and the follow-up interview took place at a clinic visit ~12 weeks after. At each site, a single research assistant conducted or supervised the interviews at baseline and follow-up. Patients rated their health as it had been in the 2 weeks prior to the interviews.

The current authors randomised 281 patients (Toronto: n=180; Ottawa: n=21; Hamilton: n=51; and Buffalo: n=29) with CRD, most of whom suffered from chronic obstructive pulmonary disease (COPD), to receive either the CRQ-SA or the CRQ-IA. In addition, within each group, all patients who had completed both forms of the dyspnoea domain were randomised to receive either the standardised or individualised form first. Finally, patients were randomised in a factorial design to two different modes of administering the feeling thermometer, another instrument that is useful for measuring changes

in HRQL in patients with CRD. These latter data required separate analysis and fall outside the scope of this report. Randomisation was performed centrally at the method centre, by centre and in blocks of eight. Centres were not aware of block size and received questionnaires in sealed envelopes.

The self-administered standardised chronic respiratory questionnaire
The authors developed the CRQ-SA from the original CRQ-IA using the same wording but omitting all phrases or comments that were included in the CRQ-IA to guide the interviewer [18, 24]. The individualised dyspnoea domain allows the patient to choose five activities that are most important to them in their daily lives. Patients then rate the degree of dyspnoea on these self-selected activities during subsequent administrations of the CRQ.

The current authors chose the five activities for the standardised dyspnoea domain from among those which patients in previous trials had frequently selected as the five most important activities causing shortness of breath [18]. In an observational study and a small randomised trial conducted in German-speaking countries in Europe, the authors found that these activities were highly relevant for patients with COPD [18, 22]. The five activities were: 1) feeling emotional, such as being angry or upset; 2) taking care of your basic needs, such as bathing, showering, eating or dressing; 3) walking; 4) performing chores, such as housework, shopping or groceries; and 5) participating in social activities.

The scoring of standardised questions is identical to individualised questions. The mean score for the individualised and standardised dyspnoea domains was calculated by summing the scores for each dyspnoea question and dividing this by the number of scored dyspnoea questions.

Other health-related qualify of life instruments
To validate and assess responsiveness of the CRQ-SA, a number of other HRQL instruments were used. The current authors selected the following instruments because they are widely used and have proven validity and responsiveness in patients with CRD.

The St George's respiratory questionnaire
The St George's respiratory questionnaire (SGRQ) [25] measures the impact of respiratory disease on HRQL and well-being. The responses to its 53 items can be aggregated into an overall score and three domain scores for symptoms, activity and impact. Weighting each item and dividing the summed weights by the maximum possible weight and expressing the result as a percentage provides the patient's score. The scores range from 0 (best) to 100% (worst). Previous results have provided evidence that the SGRQ is valid and responsive, and the minimal important difference (MID) has been found to be ~4% [26, 27]. The current authors hypothesised that CRQ scores would correlate highly with the scores on the SGRQ.

In addition, several generic HRQL instruments were administered. The authors hypothesised that correlations would be moderate with the generic instruments on corresponding domains unless noted otherwise.

The medical outcomes short form 36

The medical outcomes short form 36 (SF-36) [28] assesses a wide range of health problems. It consists of eight domains, including physical functioning, role limitations-physical, bodily pain, general health, vitality, social functioning, role limitations-emotional and mental health. The SF-36 scores range 0–100, with higher scores indicating better functioning and well-being. The SF-36 is valid and responsive to change in a variety of patient groups, including patients with CRD [22].

The standard gamble

The standard gamble (SG) [29] offers the patient two alternatives from which a choice must be made. Choice A is the certain outcome that the patient will stay in a healthy state (their own health state or a marker state) for t yrs until death. The alternative, choice B, is a hypothetical treatment with two possible outcomes: 1) returning to full health (probability p) for t yrs, at the end of which they die; or 2) immediate death (probability 1–p). The indifference probability, p*, is the utility value for the health state in choice A in the interval from dead (0) to full health (1). The greater a respondent's willingness to accept the risk of a worse outcome (*e.g.* death) to avoid the health state in choice A, then the lower the utility of the state in choice A. Validity and responsiveness of the SG in patients with CRD are poor compared to other generic instruments, and low correlations between the SG and the CRQ were expected [30, 31].

The feeling thermometer

The feeling thermometer (FT) [29, 31] is a visual analogue scale presented in the form of a thermometer in which the best state is full health (100) and the worst state is dead (0) [29]. The FT has demonstrated good measurement properties in patients with CRD and the MID for the FT is ~5–7 [27, 31].

Health utilities index 3

Health utilities index (HUI)3 [32] collects data using a 15-item self-administered questionnaire, designed to ask the minimum number of questions required to classify a patient's health status [32]. Patients are asked to recall their health status over the week prior to the interview. The eight attributes in the HUI3 multi-attributed utility measure are vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain. Each item has five or six response options. Based on the questionnaire responses, a utility score on a 0–1 scale was calculated, where 0 represents dead and 1 full health. Previous studies of patients with CRD suggest that it is valid and responsive to change in these patients [27].

At the follow-up visit, the same interviewer administered or supervised the administration of the instruments to each patient in the same order as at baseline. Patients self administered the SGRQ, HUI3, FT and SF-36 under supervision, and interviewers administered the SG.

Statistical analysis

Analysis focused on the comparison of the SA and IA versions, and the standardised and individualised dyspnoea items. The mean scores were calculated on each domain by excluding missing responses, *e.g.* if patients answered only three of the five dyspnoea questions, the mean score for these three answers provided the basis for analysis.

Test for interaction

The present authors tested for the presence of interaction between the CRQ administration technique (self administration *versus* interviewer administration) and the dyspnoea-domain method (individualised or standardised). A repeated-measure regression was performed, with the change in the dyspnoea domain score as the dependent variable, and the type of dyspnoea question (self administration *versus* interviewer administration), the type (individualised or standardised) and the order (standardised *versus* individualised dyspnoea questions first) of administration and their two- and three-way interaction terms as the independent variables. No evidence of an interaction (lowest p for the interaction terms $p=0.31$) was found.

Discriminative properties

Cross-sectional validity of the standardised and the individualised dyspnoea domains and the other domains were evaluated using data from the baseline and the follow-up visit. For this analysis, Pearson's correlation coefficients of the CRQ domains were calculated with the other CRQ domains, as well as the SGRQ, FT, SG, HUI3 and SF-36.

Evaluative properties

To assess the relative responsiveness of the CRQ–IA and CRQ-SA, the mean change was calculated between baseline and follow-up scores, as were the corresponding 95% confidence intervals (CI) for each administration format. To determine if any apparent differences in responsiveness could be explained by chance, change in CRQ scores were compared with the CRQ versions using an unpaired t-test. For the comparison of the individualised and standardised dyspnoea domains, mean scores were compared using a paired t-test. To evaluate the longitudinal validity of the CRQ-SA compared with the CRQ-IA, Pearson's correlation coefficients of the change scores of the two administration modes with the change in the validation instruments were calculated.

Bland-Altman plots were used to assess whether the difference between the standardised and individualised dyspnoea domain ratings were influenced by the mean score of the standardised and individualised scores or whether they differed by administration mode.

Multiple testing

For the evaluation of validity in the current study, a number of HRQL instruments were used, some with several dimensions. As a result, a multitude of statistical tests were performed, thereby increasing the chances of statistically significant findings in the absence of adjustment of the p-values. However, in the present report, no adjustments for multiple comparisons have been made, as validation instruments were used with correlated scores and because there is no consensus on how adjustments should be made for this type of analysis.

RESULTS

Patient characteristics

A total of 281 patients were enrolled for this study, of whom 104 (37.0%) did not complete the follow-up visit (Toronto: n=72, 40.0%; Ottawa: n=5, 22.7%; Hamilton: n=22, 43.1%; Buffalo: n=5, 17.2%). Figure 1 describes the flow of participants
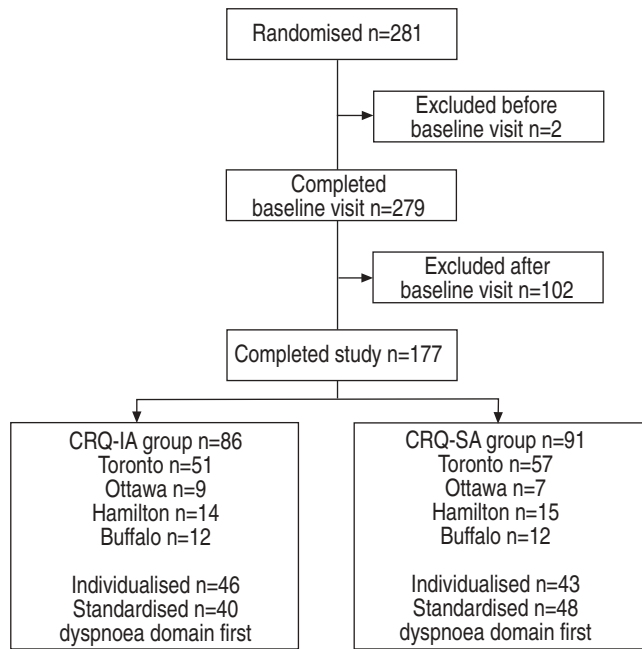
FIGURE 1. Flow diagram describing patient enrolment and completion of the study. CRQ-IA: interviewer-administered chronic respiratory questionnaire; CRQ-SA: self-administered chronic respiratory questionnaire.

through each stage of the randomised study and the reasons for exclusion from the study. A total of 281 patients were randomised, of whom two were withdrawn because of their inability to read (prior to completion of the baseline visit). Of the remaining 279 patients, 102 were excluded: 25 patients stated too much work, no time or uncomfortable with the

questions; 21 patients were too sick to continue the rehabilitation programme; 17 patients could not be followed-up because of the SARS outbreak in Toronto; 12 were not satisfied with the rehabilitation programme; six could not be reached for follow-up; seven patients died; three patients stopped because of the death of the investigator responsible for the Hamilton rehabilitation programme; three patients moved; and eight patients did not complete the CRQ or the SGRQ at both visits. The participation rate did not differ within the four centres for all randomised groups. Patients included in the analysis proved similar to those not included, with respect to sex, age, length and type of respiratory diagnosis, smoking history, and employment status in the patients (the lowest p-value for differences between these groups was p=0.16 for forced expiratory volume in one second (FEV1) % predicted).

Table 1 demonstrates the similarity of patients' characteristics in those randomised to CRQ-SA and CRQ-AI groups. There were no apparent differences between the groups and, in general, patients were older and almost all had COPD. When the baseline characteristics of the patients were compared according to whether they were randomised to complete the standardised (n=88) versus the individualised (n=89) CRQ dyspnoea domain first, no important differences were observed. However, patients in the group that received the standardised dyspnoea domain first were more likely to live alone and had a shorter duration of illness.

Instrument responsiveness
Investigator-administered versus self-administered chronic respiratory questionnaire
Table 2 shows the baseline, follow-up and mean change scores of the CRQ domains by randomisation group. Compared with

TABLE 1	Demographic information for patients completing the interviewer-administered (IA) and self-administered (SA) version of the chronic respiratory questionnaire (CRQ)			
Patient characteristic	CRQ-IA	CRQ-SA	Standardised dyspnoea domain first	Individualised dyspnoea domain first
Patients n	86	91	88	89
Sex F	39 (45.3)	34 (37.4)	37 (42.0)	36 (40.4)
Age yrs mean ± SD	68.4 ± 7.9	67.0 ± 8.3	67.8 ± 7.6	67.6 ± 8.7
Diagnosis				
COPD	80 (93.0)	84 (92.3)	80 (90.9)	84 (94.4)
Other#	5 (5.8)	7 (7.7)	8 (9.1)	4 (4.5)
FEV1 % pred mean ± SD	42.9 ± 15.6¶	45.4 ± 21.8+	44.7 ± 20.6§	43.6 ± 17.1ƒ
Time since diagnosis yrs				
Mean ± SD	10.9 ± 11.8	10.3 ± 12.5,	8.7 ± 10.3	12.5 ± 13.5
Median (IQR)	8 (2–15)	6 (3–13)	5.5 (2–12)	8 (4–18.5)
Smoking history				
Never	5 (5.8)	4 (4.4)	5 (5.7)	4 (4.5)
Current smoker	1 (1.2)	2 (2.2)	1 (1.4)	2 (2.2)
Quit	80 (93.0)	85 (93.4)	82 (93.2)	83 (93.3)
Living alone	23 (26.7)	32 (35.2)	35 (39.8)	20 (22.5)
Employed	7 (8.1)	10 (11.0)	7 (8.0)	10 (11.2)

Data are presented as n (%) unless otherwise stated. F: female; COPD: chronic obstructive pulmonary disease; FEV1 % pred: forced expiratory volume in one second per cent predicted; IQR: interquartile range. #: other diagnoses include idiopathic pulmonary fibrosis, chronic pulmonary aspergillosis, post-pulmonary resection, bronchiectasis or information missing (n=1); ¶: n=74; +: n=74; §: n=76; ƒ: n=72.

TABLE 2	Baseline, follow-up and change scores of the investigator-administered (IA) and self-administered (SA) version of the chronic respiratory questionnaire (CRQ)				
Domain	**Visit**	**CRQ-IA[#]**	**CRQ-SA[¶]**	**Difference**	**p-value[+]**
Individualised dyspnoea	Baseline	3.29 (3.07–3.50)	3.00 (2.78–3.22)	0.29 (-0.01–0.60)	0.06
	Follow-up	3.94 (3.68–4.20)	4.00 (3.73–4.28)	-0.06 (-0.44–0.32)	0.74
	Change	0.65 (0.44–0.87)	1.01 (0.76–1.26)	-0.36 (-0.68– -0.03)	0.03
Standardised dyspnoea	Baseline	4.29 (4.02–4.56)	4.11 (3.86–4.37)	0.17 (-0.19–0.54)	0.35
	Follow-up	4.82 (4.54–5.10)	4.88 (4.61–5.15)	-0.06 (-0.45–0.32)	0.75
	Change	0.53 (0.30–0.76)	0.77 (0.54–1.00)	-0.24 (-0.56–0.09)	0.15
Fatigue	Baseline	4.04 (3.73–4.35)	3.85 (3.62–4.08)	0.19 (-0.19–0.58)	0.32
	Follow-up	4.33 (4.02–4.63)	4.48 (4.24–4.73)	-0.16 (-0.54–0.23)	0.43
	Change	0.28 (0.04–0.53)	0.63 (0.41–0.86)	-0.35 (-0.68– -0.02)	0.04
Emotional function	Baseline	4.86 (4.58–5.14)	4.71 (4.48–4.94)	0.15 (-0.21–0.50)	0.41
	Follow-up	5.07 (4.82–5.33)	5.21 (4.98–5.43)	-0.13 (-0.47–0.20)	0.43
	Change	0.22 (0.02–0.42)	0.50 (0.31–0.69)	-0.28 (-0.55– -0.01)	0.04
Mastery	Baseline	5.03 (4.76–5.29)	4.60 (4.34–4.87)	0.42 (0.05, 0.80)	0.03
	Follow-up	5.37 (5.11–5.64)	5.33 (5.07–5.59)	0.04 (-0.33–0.41)	0.82
	Change	0.34 (0.15–0.54)	0.72 (0.51–0.94)	-0.38 (-0.67– -0.09)	0.01

Data are presented as mean (95% confidence interval) unless otherwise stated. [#]: n=86; [¶]: n=91; [+]: p-values from unpaired t-tests for independent samples comparing the CRQ-IA with the CRQ-SA.

the CRQ-IA, baseline scores on the CRQ-SA were consistently lower (more severe dysfunction) for each of the domains. The difference was statistically significant for the mastery domain only (p=0.03). The differences between the CRQ-IA and CRQ-SA in the follow-up scores were small and not statistically significant.

Respiratory rehabilitation led to statistically significant improvements in all domains, independent of the administration format, but the CRQ-SA demonstrated larger score changes than the CRQ-IA for each of the domains and most of the differences were statistically significant.

Table 3 shows the baseline, follow-up and mean change scores of the other HRQL instruments by randomisation group. Compared with the CRQ-IA, baseline scores in the CRQ-SA group indicated more severe dysfunction for each of the domains. The difference was statistically significant only for the FT (p=0.03). At follow-up, each of the validation instruments were similar and not statistically significant. Respiratory rehabilitation led to statistically significant improvements in all domains, independent of the randomisation group. CRQ-SA group change scores tended to be larger than CRQ-IA group change scores, but none of these differences were statistically significant.

Standardised *versus* individualised dyspnoea
Table 2 also shows that patients who completed the CRQ-IA had a mean improvement in individualised dyspnoea of 0.65 and in standardised dyspnoea of 0.53 (mean (95% CI) 0.12 (-0.04–0.28); p=0.13). Patients who completed the CRQ-SA had a mean improvement in individualised dyspnoea of 1.01 and in standardised dyspnoea of 0.77 (0.24 (0.09–0.39); p=0.002). Because there was no interaction between the mode of administration and whether patients received the individualised or standardised dyspnoea questions first, these data were pooled. The mean improvement in the pooled analysis was

0.84 for the individualised and 0.65 for the standardised dyspnoea domain (0.18 (0.07–0.29); p<0.001).

The Bland-Altman plot in figure 2 shows that the differences between change scores on the standardised and individualised dyspnoea domains were relatively constant across the range of dyspnoea scores and independent of the mode of administration.

Validity
Interviewer-administered chronic respiratory questionnaire *versus* self-administered chronic respiratory questionnaire
Table 4 addresses the cross-sectional validity of the CRQ-IA and CRQ-SA at baseline. In general, moderate-to-high correlations of the CRQ with other HRQL instruments were observed, independent of the format of administration. There were no clear trends towards higher correlations for the CRQ-IA or CRQ-SA (28 out of 40 correlation coefficients were higher for the CRQ-IA). Correlations at follow-up were similar to those at baseline (data not shown).

Table 5 addresses the correlations of the change scores between the baseline visit and a follow-up visit of the CRQ-IA and CRQ-SA domains with the other measures (longitudinal validity). Most of these correlations were moderate to strong and statistically significant. Although three of the correlations on the CRQ-IA were significantly lower (all for the HUI3), overall there was no clear trend toward lower correlations with the CRQ-IA (18 out of 40 coefficients were stronger for the CRQ-IA).

Standardised *versus* individualised dyspnoea
The comparison of the individualised and standardised dyspnoea domains revealed consistently higher correlations at baseline for the standardised domain on both the CRQ-IA and CRQ-SA (16 out of 16 coefficients; p<0.05 for 11 comparisons; table 4). It did not show a clear trend for higher
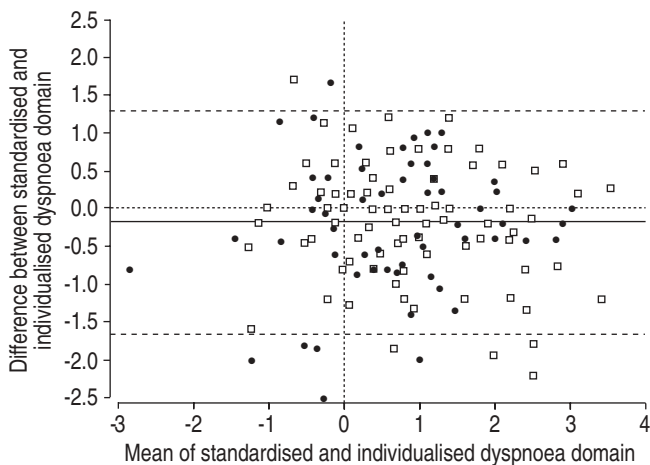
TABLE 3	Baseline, follow-up and change scores for other instruments by chronic respiratory questionnaire (CRQ) administration group				
Instrument	**Visit**	**CRQ-IA[#]**	**CRQ-SA[¶]**	**Difference**	**p-value[+]**
Feeling thermometer	Baseline	63.1 (59.2–67.0)	57.3 (53.7–60.8)	5.8 (0.6–11.1)	0.03
	Follow-up	66.0 (61.8–70.2)	64.3 (60.8–67.7)	1.7 (-3.7–7.1)	0.53
	Change	2.8 (-0.5–6.2)	7.0 (3.0–11.0)	-4.1 (-9.3–1.1)	0.12
SF36 physical aggregate	Baseline	34.1 (32.2–35.9)	32.7 (31.1–34.3)	1.4 (-1.0–3.8)	0.26
	Follow-up	36.1 (34.1–38.1)	35.4 (33.7–37.1)	0.6 (-2.0–3.3)	0.63
	Change	2.0 (0.6–3.4)	2.8 (1.4–4.1)	-0.8 (-2.7–1.2)	0.45
SF36 mental aggregate	Baseline	49.9 (47.3–52.5)	49.9 (47.3–52.4)	0.1 (-3.5–3.7)	0.96
	Follow-up	50.2 (47.6–52.8)	52.1 (49.9–54.3)	-1.9 (-5.3–1.5)	0.27
	Change	0.3 (-1.6–2.2)	2.2 (0.6–3.9)	-2.0 (-4.4–0.5)	0.11
St George's activities	Baseline	73.2 (69.1–77.2)	74.2 (70.9–77.6)	-1.1 (-6.3–4.1)	0.68
	Follow-up	70.3 (66.1–74.5)	67.8 (64.4–71.2)	2.4 (-2.9–7.8)	0.37
	Change	-2.9 (-6.7–0.9)	-6.4 (-9.1– -3.7)	3.5 (-1.0–8.1)	0.13
St George's impacts	Baseline	34.6 (31.1–38.2)	39.5 (36.0–42.9)	-4.9 (-9.8–0.1)	0.05
	Follow-up	30.6 (27.0–34.1)	32.4 (28.6–36.1)	-1.8 (-6.9–3.3)	0.48
	Change	-4.1 (-6.8– -1.3)	-7.1 (-9.6– -4.6)	3.0 (-0.7–6.7)	0.11
St George's symptoms	Baseline	60.2 (55.9–64.4)	60.8 (56.3–65.2)	-0.6 (-6.7–5.5)	0.84
	Follow-up	53.2 (48.6–57.7)	55.8 (51.3–60.3)	-2.6 (-9.0–3.8)	0.42
	Change	-7.0 (-10.7– -3.3)	-5.0 (-8.7– -1.3)	-2.0 (-7.2–3.2)	0.44
Standard gamble	Baseline	0.68 (0.62–0.73)	0.65 (0.60–0.70)	0.03 (-0.05–0.10)	0.47
	Follow-up	0.71 (0.66–0.76)	0.72 (0.68–0.77)	-0.01 (-0.08–0.06)	0.74
	Change	0.03 (-0.03–0.09)	0.07 (0.02–0.12)	-0.04 (-0.12–0.04)	0.31
HUI3	Baseline	0.59 (0.54–0.65)	0.58 (0.52–0.65)	0.01 (-0.08–0.09)	0.84
	Follow-up	0.60 (0.54–0.66)	0.66 (0.62–0.71)	-0.06 (-0.14–0.02)	0.12
	Change	0.01 (-0.04–0.06)	0.08 (0.03–0.13)	-0.07 (-0.14–0.001)	0.05

Data are presented as mean (95% confidence interval) unless otherwise stated. Higher scores on the St George's questionnaire represent worse health-related quality of life. IA: interviewer administered; SA: self administered; SF36: the medical outcomes short form 36; HUI3: health utilities index 3. [#]: n=86; [¶]: n=91; [+]: p-values from unpaired t-tests for independent samples comparing the CRQ-IA with the CRQ-SA.

longitudinal validity of the standardised dyspnoea domain independent of whether the questionnaire was self

FIGURE 2. Bland-Altman plot showing the difference between the standardised and individualised dyspnoea domain ratings against the mean of the standardised and individualised scores. ●: self-administered chronic respiratory questionnaire (CRQ); □: interviewer-administered (CRQ); ——: mean; -----: the mean ± 1.96 SD of the change scores.

administered or interviewer administered (nine out of 16 correlation coefficients were higher for the standardised domain and one out of two significantly different correlations coefficients were higher for the standardised domain; table 5).

Number of missing items
There were more answered items in the individualised than in the standardised dyspnoea domain at baseline (96.5 versus 70.9% on the CRQ-SA and 98.9 versus 75.8% on the CRQ-IA, respectively). At least three standardised dyspnoea items were completed by 97.7 and 98.9% of patients on the CRQ-IA and CRQ-SA, respectively. The proportion of patients who responded to standardised dyspnoea items did not differ by item indicating that patients perform the activities with approximately equal frequency. The proportions of missing items were similar at baseline and follow-up. There were no missing items on the fatigue, emotional function and mastery domain on either the CRQ-SA or the CRQ-IA.

DISCUSSION
This large multicentre randomised controlled study addressed how a widely used instrument to assess HRQL in CRD functioned in self-administered versus interviewer-administered formats, and using standardised versus individualised dyspnoea items. Despite the older age and considerable morbidity of the patients involved in the study, self administration did not

TABLE 4	Cross-sectional validity of the interviewer-administered (IA) and self-administered (SA) versions of the chronic respiratory questionnaire (CRQ): correlation[#] for baseline scores									
Instrument	**CRQ-IA domains**					**CRQ-SA domains**				
	Individualised dyspnoea	**Standardised dyspnoea**	**Fatigue**	**Emotional function**	**Mastery**	**Individualised dyspnoea**	**Standardised dyspnoea**	**Fatigue**	**Emotional function**	**Mastery**
Feeling thermometer	0.26[¶]	0.44[¶]	0.44[+]	0.38[+]	0.44	0.37[¶]	0.60[¶]	0.70[+]	0.66[+]	0.55
SF-36 physical aggregate score	0.41[¶]	0.66[¶]	0.58	0.37	0.43	0.43	0.50	0.37	0.23	0.21
SF-36 mental aggregate score	0.25[¶]	0.52[¶]	0.61	0.77	0.67	0.22[¶]	0.45[¶]	0.67	0.76	0.62
St George's activities	-0.34[¶]	-0.69[¶]	-0.50	-0.34	-0.40	-0.52	-0.67	-0.44	-0.33	-0.33
St George's impacts	-0.41[¶]	-0.64[¶]	-0.66	-0.67	-0.62	-0.37[¶]	-0.60[¶]	-0.61	-0.60	-0.63
St George's symptoms	-0.30[¶]	-0.49[¶]	-0.50	-0.42	-0.48	-0.26[¶]	-0.46[¶]	-0.32	-0.38	-0.44
Standard gamble	0.05	0.17	0.30	0.32	0.25	0.16	0.26	0.27	0.37	0.23
HUI3	0.39[¶]	0.64[¶]	0.72[+]	0.67	0.58	0.32	0.46	0.52[+]	0.63	0.41

SF-36: the medical outcomes short form 36; HUI3: health utilities index 3. [#]: Pearson's correlation coefficient, $r>0.21$ or $r<-0.21$ significant at $p<0.05$, lower scores on the St George's respiratory questionnaire indicate a higher health-related quality of life. [¶]: $p<0.05$ for the comparison of Pearson's correlation coefficient between the individualised and standardised dyspnoea domain; [+]: $p<0.05$ for the comparison of Pearson's correlation coefficient between the CRQ-IA and the CRQ-SA.

impair validity and it enhanced responsiveness. The individualised dyspnoea format demonstrated superior responsiveness to the standardised dyspnoea domain, while the standardised dyspnoea format showed superior cross-sectional validity.

The enhanced responsiveness of the CRQ-SA relative to the CRQ-IA appears largely due to lower baseline scores, indicating greater dysfunction and leaving room for greater improvement (table 2). The increase in responsiveness may be related

TABLE 5	Longitudinal validity of the interviewer-administered (IA) and the self-administered (SA) formats of the chronic respiratory questionnaire (CRQ): correlation[#] for change scores									
Instrument and domain	**CRQ-IA domains**					**CRQ-SA domains**				
	Individualised dyspnoea	**Standardised dyspnoea**	**Fatigue**	**Emotional function**	**Mastery**	**Individualised dyspnoea**	**Standardised dyspnoea**	**Fatigue**	**Emotional function**	**Mastery**
Feeling thermometer	0.43	0.31	0.28	0.20	0.19	0.38	0.40	0.50	0.32	0.29
SF36 physical aggregate score	0.41	0.46	0.31	0.17	0.13	0.46[¶]	0.33[¶]	0.30	0.23	0.26
SF36 mental aggregate score	0.33	0.46	0.50	0.47	0.40	0.26	0.38	0.52	0.45	0.33
St Georges's activities	-0.47	-0.49	-0.41	-0.21	-0.32	-0.45	-0.44	-0.35	-0.25	-0.23
St George's impacts	-0.55	-0.52	-0.49	-0.46	-0.40	-0.48	-0.50	-0.43	-0.41	-0.51
St George's symptoms	-0.29	-0.29	-0.15	-0.14	-0.22	-0.19	-0.18	-0.21	-0.31	-0.33
Standard gamble	0.11	0.16	0.13	0.03	0.002	0.12	0.09	0.13	0.09	-0.01
HUI3	0.14[¶,+]	0.35[¶]	0.25[+]	0.38	0.17[+]	0.46[+]	0.53	0.52*	0.53	0.46*

SF-36: the medical outcomes short form 36; HUI3: health utilities index 3. [#]: Pearson's correlation coefficient; $r>0.21$ or $r<-0.21$ significant at $p<0.05$; lower scores on the St George's respiratory questionnaire indicate higher health-related quality of life. [¶]: $p<0.05$ for the comparison of Pearson's correlation coefficient between the individualised and standardised dyspnoea domain; [+]: $p<0.05$ for the comparison of Pearson's correlation coefficient between the CRQ-IA and the CRQ-SA.

to patients being more willing to acknowledge severe dysfunction in the absence of an interviewer [28, 33]. However, this greater willingness to acknowledge impairment did not affect the relative validity of the self-administered and interviewer-administered formats, which showed similar correlations with other questionnaires, in both the cross-sectional (table 4) and longitudinal analysis (table 5). Three out of 40 longitudinal correlations were higher on the self-administered compared with the interviewer-administered version, and all of these differences were on the HUI3. One explanation of the difference between self-administered and interviewer-administered longitudinal validity on the HUI3 is that the HUI3 is itself self-administered. However, other self-administered questionnaires did not show the same difference. These isolated results are puzzling and could represented a chance phenomenon.

The comparison of standardised and individualised dyspnoea formats is consistent with two smaller previous studies [18, 22]. Compared with the standardised questions, the individualised items produced a narrower distribution of scores and a lower mean score. It is likely that the opportunity to focus on their most important activities led patients to choose individualised items for which they experience the most troublesome dyspnoea. The standardised approach did not permit this focus and, when each patient addresses the same items, a much wider distribution of degree of dyspnoea resulted. Thus, standardising the dyspnoea domain improves the discriminative properties of the CRQ dyspnoea domain across patients by increasing the between-person variance.

These differences in mean and distribution are associated with differences in measurement properties. The wider distributions are likely to explain the higher cross-sectional correlations between standardised items and other HRQL instruments than between individualised items and those same instruments (table 4). The higher correlations suggest improved discriminative validity of the standardised approach, an important measurement property for differentiating severity of HRQL impairment between patients.

The individualised questions tended to be slightly more responsive than the standardised questions in both the CRQ-SA and CRQ-IA, and the difference reached conventional levels of statistical significance in the former. A probable explanation for the increased responsiveness of the individualised domain is the likelihood that patients will be more attentive to changes on activities they consider most important. Another possibility is that changes in activity pattern over time may add random error to the standardised dyspnoea items.

Since the differences in responsiveness between individualised and standardised dyspnoea domains are small, the relative impact on sample size requirements will also be small. However, if sample size is a critical issue, investigators planning a new study should consider using the individualised CRQ dyspnoea domain.

At the same time, changing to standardised items did not attenuate the correlation between change in the CRQ dyspnoea domain and change in other HRQL measures. This indicates that standardisation of the dyspnoea domain does not compromise longitudinal validity. Thus, if increases in sample size are feasible and investigators feel the greater convenience of the standardised items is important in their study, they can choose the standardised approach without fearing they are compromising the CRQ validity.

It was observed here that the standardised approach led to a larger number of missing items on the dyspnoea domain. This finding is not surprising because patients with the individualised approach ensured that the patients would choose items that are both important to them and that they have undertaken in the last week. This restriction does not apply to the standardised dyspnoea items, making it far more likely that respondents would not have undertaken the activity during the time frame specified in the questionnaire.

At the same time, patients almost invariably completed at least three standardised dyspnoea items. Because prior results suggested that reducing the number of items from five to two had only a small impact on responsiveness and validity of the CRQ, it is not unexpected that the standardised CRQ maintained validity in the current study [34].

The strengths of the current trial include the randomised design and the use of a wide variety of validation instruments. It was ensured that all study procedures were consistent across centres and all interviewers were trained to ensure adherence to the study protocol. The multicentre trial increases the generalisability of the current findings.

Limitations of the study included the relatively high loss to follow-up. The loss to follow-up was, to some extent, due to unusual events (the SARS epidemic and the death of an investigator) and is unlikely to have introduced bias into the current study, as there was no difference in the participation rate between the groups. In addition, systematic differences in patients who completed the trial compared with those who did not complete the trial was not observed. Another limitation prevented the authors from including an exercise performance test, as an added part of the study protocol, which would have provided additional validation data. However, several other studies evaluated exercise performance in relation to the original CRQ and found moderate correlations.

Patients in the CRQ-SA group tended to show slightly greater improvements on other HRQL instruments and, thus, the increased responsiveness observed with the CRQ-SA could have resulted from a slight imbalance in randomisation (table 3). However, changes in the scores were significantly different between the CRQ-IA and CRQ-SA group only for the CRQ, but not for other HRQL instruments. Furthermore, even if the enhanced responsiveness of the CRQ-SA was a result of bias, it doesn't compromise the conclusion that measurement properties are at least maintained with a CRQ-SA.

In conclusion, the present authors have demonstrated the validity and responsiveness of a standardised self-administered version of the CRQ. Investigators will lose neither responsiveness nor validity by choosing self administration over interviewer administration and the loss in responsiveness by choosing standardised *versus* individualised items is small. Investigators can deal with the lower responsiveness of the standardised dyspnoea questions by adjusting the sample size, making the CRQ-SA highly suitable for use in large

scale clinical trials. Thus, under most circumstances, the resource- and time-consuming interviewer-administration method using individualised dyspnoea questions or interviewer administration need no longer hinder use of the CRQ.

Because there are few randomised studies comparing self administration to interviewer administration of health-related quality of life measures, not only in chronic respiratory disease but also in other illnesses, this study adds important information to the field of patient-reported outcome assessment. Because of the extensive use of the chronic respiratory questionnaire, the current findings also have important implications for measuring health-related quality of life in chronic respiratory disease, both in clinical settings, such as rehabilitation programmes, and in investigational settings, such as randomised intervention trials. The reduced resource and time requirements of the self-administered standardised chronic respiratory questionnaire makes it an attractive instrument for the assessment of health-related quality of life in chronic respiratory disease.

REFERENCES

1 Guell R, Casan P, Sangenis M, Morante F, Belda J, Guyatt GH. 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.

2 van den Boom G, Rutten-van Molken MP, Molema J, Tirimanna PR, van Weel C, van Schayck CP. The cost effectiveness of early treatment with fluticasone propionate 250 microg twice a day in subjects with obstructive airway disease. Results of the DIMCA program. *Am J Respir Crit Care Med* 2001; 164: 2057–2066.

3 Bendstrup KE, Ingemann Jensen J, Holm S, Bengtsson B. Out-patient rehabilitation improves activities of daily living, quality of life and exercise tolerance in chronic obstructive pulmonary disease. *Eur Respir J* 1997; 10: 2801–2806.

4 Green RH, Singh SJ, Williams J, Morgan MD. A randomised controlled trial of four weeks *versus* seven weeks of pulmonary rehabilitation in chronic obstructive pulmonary disease. *Thorax* 2001; 56: 143–145.

5 Neder JA, Sword D, Ward SA, Mackay E, Cochrane LM, Clark CJ. Home based neuromuscular electrical stimulation as a new rehabilitative strategy for severely disabled patients with chronic obstructive pulmonary disease (COPD). *Thorax* 2002; 57: 333–337.

6 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: 773–778.

7 van den Boom G, Rutten-van Molken MP, Tirimanna PR, van Schayck CP, Folgering H, van Weel C. Association between health-related quality of life and consultation for respiratory symptoms: results from the DIMCA programme. *Eur Respir J* 1998; 11: 67–72.

8 Brightling CE, Monteiro W, Ward R, *et al.* Sputum eosinophilia and short-term response to prednisolone in chronic obstructive pulmonary disease: a randomised controlled trial. *Lancet* 2000; 356: 1480–1485.

9 Jones PW. Testing health status (''quality of life'') questionnaires for asthma and COPD. *Eur Respir J* 1998; 11: 5–6.

10 Wyrwich KW, Fihn SD, Tierney WM, Kroenke K, Babu AN, Wolinsky FD. Clinically important changes in health-related quality of life for patients with chronic obstructive pulmonary disease. An expert consensus panel report. *J Gen Intern Med* 2003; 18: 196–202.

11 Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999; 52: 861–873.

12 Hajiro T, Nishimura K, Tsukino M, Ikeda A, Koyama H, Izumi T. 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: 785–790.

13 Hajiro T, Nishimura K, Tsukino M, Ikeda A, Koyama H, Izumi T. Analysis of clinical methods used to evaluate dyspnoea in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998; 158: 1185–1189.

14 Jones PW. Issues concerning health-related quality of life in COPD. *Chest* 1995; 107: 187S–193S.

15 Wijkstra PJ, Van Altena R, Kraan J, Otten V, Postma DS, Koeter GH. Quality of life in patients with chronic obstructive pulmonary disease improves after rehabilitation at home. *Eur Respir J* 1994; 7: 269–273.

16 Morgan MD. Experience of using the CRQ (Chronic Respiratory Questionnaire). *Respir Med* 1991; 85: 23–24; discussion 33–37.

17 Tu SP, McDonell MB, Spertus JA, Steele BG, Fihn SD. A new self-administered questionnaire to monitor health-related quality of life in patients with COPD. Ambulatory Care Quality Improvement Project (ACQUIP) Investigators. *Chest* 1997; 112: 614–622.

18 Schünemann HJ, Griffith L, Jaeschke R, *et al.* A Comparison of the original chronic respiratory questionnaire with a standardized version. *Chest* 2003; 124: 1421–1429.

19 Williams JE, Singh SJ, Sewell L, Guyatt GH, Morgan MD. Development of a self-reported Chronic Respiratory Questionnaire (CRQ-SR). *Thorax* 2001; 56: 954–959.

20 Williams JE, Singh SJ, Sewell L, Morgan MD. Health status measurement: sensitivity of the self-reported Chronic Respiratory Questionnaire (CRQ-SR) in pulmonary rehabilitation. *Thorax* 2003; 58: 515–518.

21 Guyatt GH, Townsend M, Keller J, Singer J, Nogradi S. Measuring functional status in chronic lung disease: conclusions from a randomized control trial. *Respir Med* 1991; 85: 17–21; discussion 33–37.

22 Puhan M, Behnke M, Laschke M, *et al.* Self-administration and interviewer-administration of the chronic respiratory questionnaire: A randomized trial in three German-speaking countries. *Respir Med* 2004; 98: 342–350.

23 Lacasse Y, Brosseau L, Milne S, *et al.* Pulmonary rehabilitation for chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2002; 3: CD003793.

24 Schünemann HJ, Guyatt GH, Ståhl E, *et al.* Validation of a new version of the chronic respiratory questionnaire

(CRQ): the self-administered standardized CRQ. *Eur Respir J* 2002; 20: 156S.

25 Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med* 1991; 85: 25–31, discussion 33–37.

26 Jones PW DL. Relationship between change on the St. George's Respiratory Questionnaire score and patient's perception of treatment efficacy after one year therapy with nedocromil sodium. *Am J Respir Crit Care Med* 1994; 149: A211.

27 Schünemann H, Griffith L, Jaeschke R, Stubbing D, Goldstein R, Guyatt GH. Evaluation of the minimal important difference for the feeling thermometer and St. Georges Respiratory questionnaire in patients with chronic airflow limitation. *J Clin Epidemiol* 2003; 56: 1170–1176.

28 Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30: 473–483.

29 Bennett KJ. Measuring health state preferences and utilities: rating scale, time trade-off, and standard gamble techniques. *In:* Spilker B, ed. Quality of Life and Pharmacoeconomics in Clinical Trials. 2nd Edn. Philadelphia, Lippincott-Raven, 1996; p. 259.

30 Schünemann H, Stahl E, Mador J, *et al.* Validation of a self-administered, standardized version of the chronic respiratory questionnaire (CRQ). *Am J Respir Crit Care Med* 2003; 167: A226.

31 Schünemann HJ, Griffith L, Stubbing D, Goldstein R, Guyatt GH. A clinical trial to evaluate the measurement properties of 2 direct preference instruments administered with and without hypothetical marker states. *Med Decis Making* 2003; 23: 140–149.

32 Feeny D, Furlong W, Torrance GW, *et al.* Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medi Care* 2002; 40: 113–128.

33 McHorney CA, Kosinski M, Ware JE Jr. Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail *versus* telephone interview: results from a national survey. *Med Care* 1994; 32: 551–567.

34 Moran LA, Guyatt GH, Norman GR. Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *J Clin Epidemiol* 2001; 54: 571–579.