View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Respiratory Medicine (2007) **101**, 308–316

ELSEVIER

respiratoryMEDICINE

Relative responsiveness of the Chronic Respiratory Questionnaire, St. Georges Respiratory Questionnaire and four other health-related quality of life instruments for patients with chronic lung disease

Milo A. Puhan[a,b], Gordon H. Guyatt[b,c], Roger Goldstein[d], Jeff Mador[e,h], Douglas McKim[f], Elisabeth Stahl[g,h], Lauren Griffith[b], Holger J. Schünemann[b,i,j,*]

[a]*Horten Centre, University of Zurich, Switzerland*
[b]*Department of Clinical Epidemiology and Biostatistics, Hamilton, Ont., Canada*
[c]*Department of Medicine, McMaster University, Hamilton, Ont., Canada*
[d]*Department of Medicine, University of Toronto, Toronto, Ont., Canada*
[e]*Veteran Affairs Medical Center, Buffalo, NY, USA*
[f]*University of Ottawa, Ont., Canada*
[g]*AstraZeneca R & D, Lund, Sweden*
[h]*University of Aberdeen, Aberdeen, UK*
[i]*Department of Medicine and Social and Preventive Medicine, School of Medicine and Biomedical Sciences, State University of New York at Buffalo, Buffalo, NY, USA*
[j]*Clinical Research Development and INFORMAtion Translation Unit, Italian National Cancer Institute Regina Elena, Rome, Italy*

Received 6 July 2005; accepted 26 April 2006

KEYWORDS
Chronic Obstructive Pulmonary Disease;
Quality of Life;
Chronic Respiratory Questionnaire;
St. Georges

Summary
Background: Selection of heath-related quality of life (HRQL) instruments that are most responsive to changes in HRQL prevents investigators from drawing false-negative conclusions about the effectiveness of an intervention. The objective of this study was to compare the responsiveness of the Chronic Respiratory Questionnaire (CRQ), the St. Georges Respiratory Questionnaire (SGRQ) and four other HRQL instruments.

*Corresponding author. Clinical Research Development and Information Translation Unit, Italian National Cancer Institute Regina Elena, Rome, Italy. Tel.: +39 06 52665102.
E-mail address: schuneh@mcmaster.ca (H.J. Schünemann).

0954-6111/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.rmed.2006.04.023

Respiratory Questionnaire; Responsiveness

Methods: We enrolled 177 patients with chronic lung disease, primarily COPD (93%), who completed 8 weeks of respiratory rehabilitation. Patients completed the CRQ, the SGRQ and four generic measures (SF-36, Feeling Thermometer, Standard Gamble and Health Utilities index 3) at the beginning of the rehabilitation program and 12 weeks thereafter. We calculated standardized response means (SRMs) for each instrument, from the change score divided by the standard deviation of the change score.
Results: We observed the largest SRM for the CRQ (0.24–0.66 for the four CRQ domains on the interviewer-administered and 0.56–0.84 for the self-administered format) and the SGRQ (0.33–0.51 for the three SGRQ domains and total score). The CRQ dyspnea domain was statistically significantly more responsive than any other instrument including the SGRQ. For the SGRQ, the total and impacts domain were significantly more responsive than the generic and preference-based instruments.
Conclusions: This study confirms that the CRQ and SGRQ are substantially more responsive than generic measures, and suggests particularly strong responsiveness for the self-administered CRQ.
© 2006 Elsevier Ltd. All rights reserved.

Introduction

Clinicians and investigators recognize the importance of measuring health-related quality of life (HRQL) as a patient-important outcome in clinical trials of patients with chronic obstructive pulmonary disease (COPD) and other chronic lung disease.[1–9] Because there is no single gold standard for the measurement of HRQL, investigators need to select instruments that meet their research goal. If change in HRQL is the outcome of interest, as it is in the evaluation of treatments, responsiveness is a key measurement property. Responsiveness refers to an instrument's ability to detect HRQL change, even if the magnitude of change is small.[10] If instruments are not responsive, investigators and clinicians could draw false-negative conclusions from results of clinical research and monitoring of health programs, including respiratory rehabilitation.[11]

A number of validated HRQL instruments for patients with chronic lung disease exist and investigators often encounter challenges in selecting the optimal instrument. In general, disease-specific instruments such as the Chronic Respiratory Questionnaire (CRQ)[12] are more responsive than generic instruments such as the Sickness Impact Profile or SF-36. Indeed, earlier studies with COPD patients indicated superior responsiveness of disease-specific instruments.[11,13,14] Thus, when responsiveness is key, investigators need to choose among disease-specific instruments such as the widely used CRQ and St. George Respiratory Questionnaire (SQRQ).[15]

A previous study with COPD patients following respiratory rehabilitation found larger effect sizes for the CRQ,[14] while another study assessing the effects of salmeterol and ipratropium bromide did not clearly favour the CRQ or SQRQ in terms of responsiveness.[16] However, the latter study used a non-validated self-administered CRQ version. We recently validated a self-administered and standardized version of the original CRQ[17–19] that facilitates the administration of the CRQ in a clinical trial or practice setting. Investigators have not yet compared this version of the CRQ with the SGRQ.

Other outcomes have attained greater interest in clinical trials. For example, the comparison of clinical interventions across patient groups or evaluation of costs require generic HRQL instruments and preference-based measures such as the Feeling Thermometer (FT), Standard Gamble (SG), or Health Utilities Index 3 (HUI3).[20] Preference-based instruments yield a single score between 0 and 1.0 and, in theory, allow comparison across different health programs and interventions.[21–24] In addition, preference-based instruments are increasingly used as global measures of HRQL in trials so that responsiveness becomes important for these instruments as well. Indeed scores obtained with generic HRQL instruments have experienced transformation into utility scores because of the importance of a common metric that attempts to measure both HRQL and utilities. However, data addressing the responsiveness of preference-based instruments in relation to other HRQL tools are so far limited.[25,26]

The aim of this study was to compare the responsiveness of the CRQ and other disease-specific, generic and preference-based instruments in patients with chronic lung disease following a respiratory rehabilitation program. There is strong evidence from meta-analyses that respiratory rehabilitation leads to patient-important improvements in HRQL among COPD patients.[27] Therefore,

respiratory rehabilitation provides an ideal context for testing instrument responsiveness.

Methods

Patients and study design

For this analysis, we used data from a multicentre randomized trial that compared the measurement properties of the interviewer- (CRQ-IA) and self-administered CRQ (CRQ-SA) as well as those of individual and standardized dyspnea questions.[19] In brief, we recruited patients with chronic lung diseases (mainly COPD) at four University respiratory rehabilitation programs in Toronto, Ottawa, Hamilton (all Canada) and Buffalo, NY (USA), and randomly assigned them to a group with interviewer- or self-administration of the CRQ. Patients followed an outpatient respiratory rehabilitation program of 8 weeks duration with an emphasis on physical exercise. Patients completed all outcome measures at the beginning of respiratory rehabilitation and approximately 12 weeks thereafter. We previously described that we randomized 281 patients to receive either the CRQ-IA or CRQ-SA,[19] of whom we excluded two before the baseline assessment (inability to read) and of whom 102 did not complete the study for the following reasons: too much work or no time ($n = 25$), too sick to complete rehabilitation program ($n = 21$), SARS outbreak in Toronto ($n = 17$), not available for outcome assessment ($n = 14$), not satisfied with the rehabilitation program ($n = 12$), death of patients ($n = 7$), death of one site investigator ($n = 3$) or move to another place ($n = 3$). The administration of the HRQL instruments was identical for all instruments with the exception of the CRQ (administered as CRQ-IA and CRQ-SA) and the FT (administered with or without prior rating of 3 hypothetical health states) and is described below.

Disease-specific instruments

Chronic Respiratory Questionnaire[12]

The original, interviewer administered CRQ has 20 items asking about the level of impairment in the domains of dyspnea, fatigue, emotional function and mastery. Patients express their degree of impairment from 1 (most severe impairment) to 7 (no impairment). In the present study, we randomized patients to complete either the original interviewer-administered or the self-administered version of the CRQ. All patients completed the original individualized[12] and, in addition, the standardized[17] dyspnea domain. We randomized the mode of administration and the order of individualized and standardized dyspnea questions to prevent order effects. The CRQ served as a model for many methodological studies[13,19,28] and has shown good measurement properties including the self-administered format and the standardized dyspnea questions.[17–19]

St. George's Respiratory Questionnaire[29]

The self-administered SGRQ consists of 76 items addressing the effect of respiratory disease on HRQL. A total score and three domain scores for symptoms, activity and impact summarize the responses and range from 0% (best) to 100% (worst). Previous studies showed good reliability, cross-sectional and longitudinal validity of the SGRQ.[13–15]

Generic instruments

The Medical Outcomes Short Form 36 (SF-36)[30]: The SF-36 consists of 8 domains including physical functioning, role limitations-physical, bodily pain, general health, vitality, social functioning, role limitations-emotional, and mental health. The domains scores can be aggregated into two composite scores for physical and mental health. The SF-36 scores range from 0 to 100, with higher scores indicating better functioning and well-being. Both composite scores are standardized to have a mean score of 50 and standard deviation of 10 in the general US population. The SF-36 is valid and responsive to change in a variety of patient groups, including patients with COPD.[3]

Preference-based instruments

Feeling Thermometer[31]

The FT is a visual analogue scale presented in the form of a thermometer with 100 intervals, in which the best state is full health (equal to a score of 100) and the worst state is dead (a score of 0). We asked patients to reflect the health status in their score as it was during the last 7 days. We had administered the FT with or without clinical marker states.[32] The latter occurred in random order in a 2×2 factorial design. The FT has demonstrated good measurement properties in patients with COPD.[24,33]

Standard Gamble[31]

The SG offers the patient two alternatives from which a choice must be made: Choice A is a hypothetical treatment with two possible outcomes: (1) returning to full health (probability p)

for t years, at the end of which they die or (2) immediate death (probability $1-P$). The alternative (choice B) is the certain outcome that he/she will stay in their own health state for t years until death. 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. dead) to avoid the health state in choice A, then the lower is the utility of the state in choice A to them.

Health Utilities Index 3[20]

We used the self-administered Health Utilities Index 3 questionnaire with 15 items and asked patients to recall their health status over the week prior to the interview. The 8 attributes in the Health Utilities Index 3 multi-attribute utility measure are vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain. Each item has 5 or 6 response options. Based on the questionnaire responses we calculated a utility score from 0 to 1.0 where 0 represents dead and 1.0 full health.[24]

Statistical analysis

To compare the responsiveness across instruments, we calculated standardized response means (SRMs) for each individual patient by dividing the change scores of each instrument (follow-up minus baseline score) by the standard deviations (SDs) of change scores of the whole study group. We compared SRMs among all instruments using paired t-tests and calculated 95% confidence intervals for SRMs using the approach proposed by Beaton et al.[34] We performed all comparisons across instruments within one group including all patients. For the comparison of the SRMs, we combined both CRQ (CRQ-IA and CRQ-SA) and both FT groups to facilitate comparisons with the other HRQL instruments because the SRMs did not differ significantly between the CRQ-IA and CRQ-SA and the FT with or without clinical marker states. The larger sample size provided greater precision to estimate differences in relative responsiveness between the HRQL instruments.

The SRM, a widely used measure for responsiveness,[35] differs from change scores in that it is corrected for the SD of the change score.[34] We chose SRMs because it removes dependency on sample size as opposed to t-tests that depend on sample size. The SRM is superior to Cohen's effect size (change scores divided by SDs of baseline scores) because the signal (change score) relates to the noise term (SD of change score) directly. Using the SD of the change scores in the SRM rather than the SD of the baseline scores is an advantage over Cohen's effect size because baseline scores may be unrelated to change score variability. We performed the analyses using SAS 8.2 (SAS Institute Inc., Cary, NC, USA).

Results

Mean age of the 177 included patients was 69.0 years (SD 8.7), 59% were males, mean FEV_1 was 42.8 (19.2) % predicted and the proportion of patients with COPD was 93%. 7% of enrolled patients had other chronic lung diseases including idiopathic pulmonary fibrosis, chronic pulmonary aspergillosis, post-pulmonary resection and bronchiectasis.

Table 1 shows the SRMs of the disease-specific, generic and preference-based instruments. In general, SRMs were higher for the CRQ-SA compared to the CRQ-IA. The disease-specific CRQ and SGRQ showed larger SRMs compared to similar domains on the generic and preference-based instruments. However, the physical composite score of the SF-36 showed similar responsiveness (SRM = 0.37) was similar to the fatigue, emotions and mastery domains of the CRQ-IA and the activities and impact domain on the SGRQ. The CRQ dyspnea domains tended to have larger SRMs compared to the CRQ fatigue, emotions and mastery domains. The SGRQ total score showed a larger SRM than the SGRQ domains scores. While the signal of the SGRQ total score (mean change score of 5.4) was similar compared to the signal on the three domains (4.7–6.2), the standard deviation of change scores was lower (11.2 for total score versus 15.4 for the activities, 12.5 for the impacts and 17.6 for the symptoms domain).

Table 2 shows the comparisons of SRMs across instruments. The individualized dyspnea domain of the CRQ showed significantly higher responsiveness that all other instruments while the standardized dyspnea domain was significantly more responsive than the generic and preference-based instruments and the SGRQ domains. The mastery domain of the CRQ was more responsive compared to the generic and preference-based instruments, but did not differ from the SGRQ total and impacts score. The CRQ fatigue and emotions domain were significantly more responsive than the SF-36 mental composite score and the Health Utilities Index 3.

The SGRQ total score was significantly more responsive than the generic and preference-based instruments. However, only the impacts domain was significantly more responsive than the generic and preference-based instruments.

Table 1 Standardized response means of disease-specific, generic and preference-based instruments in patients with COPD after respiratory rehabilitation.

Instrument	Standardized response mean (95% confidence interval)
Disease-specific instruments	
CRQ-IA ($n = 86$)	
Dyspnea (Individualized)	0.66 (0.45, 0.87)
Dyspnea (Standardized)	0.50 (0.29, 0.71)
Fatigue	0.25 (0.04, 0.46)
Emotional function	0.24 (0.03, 0.45)
Mastery	0.38 (0.17, 0.59)
CRQ-SA ($n = 91$)	
Dyspnea (Individualized)	0.84 (0.63, 1.05)
Dyspnea (Standardized)	0.69 (0.48, 0.90)
Fatigue	0.60 (0.39, 0.81)
Emotional function	0.56 (0.35, 0.77)
Mastery	0.70 (0.49, 0.91)
SGRQ ($n = 177$)	
Total Score	0.51 (0.36, 0.66)
Activities	0.33 (0.18, 0.48)
Impacts	0.46 (0.31, 0.61)
Symptoms	0.34 (0.19, 0.49)
Generic instrument	
SF-36 ($n = 177$)	
Physical composite score	0.37 (0.22, 0.52)
Mental composite score	0.17 (0.02, 0.32)
Physical functioning domain	0.30 (0.15, 0.45)
Role physical domain	0.21 (0.06, 0.36)
Bodily pain domain	0.22 (0.07, 0.37)
General health domain	0.20 (0.06, 0.35)
Vitality domain	0.37 (0.22, 0.52)
Social functioning domain	0.21 (0.06, 0.36)
Role emotional domain	0.07 (-0.08, 0.22)
Mental health domain	0.15 (0.002, 0.30)
Preference-based instruments	
Feeling Thermometer ($n = 177$)	0.28 (0.13, 0.43)
Standard Gamble ($n = 177$)	0.21 (0.06, 0.36)
Health Utilities Index 3 ($n = 177$)	0.20 (0.05, 0.35)

Among the generic and preference-based instruments, the SF-36 physical composite score was most responsive, but the differences with other instruments did not reach statistical significance. The FT showed a larger SRM compared to the SG and Health Utilities Index 3, but these differences were not statistically significant.

Discussion

This comparison of six HRQL and preference-based instruments showed that the self-administered CRQ, in particular the individualized and standardized dyspnea domains, are most responsive to changes in patients with chronic lung disease undergoing respiratory rehabilitation. The generic and preference-based instruments showed inferior relative responsiveness and did not differ significantly from each other.

Strengths of this study include the standardized administration or supervision of instruments by trained interviewers present in all centers. In addition we included a battery of the most widely used disease-specific, generic and preference-based instruments available for patients with chronic lung disease and randomized the order of administration for the individualized and standardized CRQ dyspnea domains to eliminate order effects.

Table 2 Comparison of standardized response means ($n = 177$).

	CRQ dyspnea (individualized)	CRQ dyspnea (standardized)	CRQ fatigue	CRQ emotions	CRQ mastery	SQRQ total score	SQRQ activities	SQRQ impacts	SQRQ symptoms	SF36-physical	SF36-mental	Feeling Thermometer	Standard Gamble
SQRQ total score	**0.27****	0.12	−0.07	−0.09	0.06								
SQRQ activities	**0.45****	**0.30****	0.11	0.09	**0.24****								
SQRQ impacts	**0.30****	**0.15***	−0.03	−0.05	0.10								
SQRQ symptoms	**0.41****	**0.26****	0.07	0.05	**0.20***								
SF36-physical	**0.40****	**0.26****	0.07	0.06	**0.21***	**0.15***	−0.04	0.11	−0.01				
SF36-mental	**0.61****	**0.47****	**0.28****	**0.27****	**0.42****	**0.36****	0.17	**0.32****	**0.20***				
Feeling Thermometer	**0.47****	**0.32****	0.13	0.12	**0.26****	**0.20***	0.02	0.17	0.06	0.06	−0.15		
Standard Gamble	**0.54****	**0.39****	**0.21***	0.19	**0.33****	**0.28****	0.10	**0.24***	0.14	0.13	−0.10	0.05	
Health Utilities Index 3	**0.55****	**0.40****	**0.22****	**0.21****	**0.36****	**0.27****	0.08	**0.25****	0.15	0.12	−0.08	0.08	0.002

Differences in standardized response means between variables in columns and rows (positive scores indicating higher SRM in columns)
*Differences significant at $P \leqslant 0.05$.
**Differences significant at $P \leqslant 0.01$.

A recent systematic review comparing the responsiveness of specific and generic HRQL instruments in RCTs showed superior responsiveness of specific instruments across a wide range of patients and interventions.[36] To appraise studies that included comparisons of responsiveness between specific, generic and preference-based instruments in COPD patients (on which our study focused), we performed a PubMed search. We used the following terms "QUALITY OF LIFE [MESH]", "instrument* OR scale* OR tool* or questionnaire*", "COMPARATIVE STUDY [MESH]" and "Lung Diseases, Obstructive [MESH] OR COPD [MESH]". Of 119 potentially relevant articles, 14 reported on comparisons of relative responsiveness.[11,14,16,25,26,37–45] The majority of studies reported superior responsiveness of disease-specific instruments (CRQ, SGRQ, Respiratory Quality of Life Questionnaire and Pulmonary Functional Status Instrument) compared to generic instruments (SF-36, Sickness Impact Profile, Nottingham Health Profile, Quality of Well Being),[11,37,38,42–45] while two studies did not find differences.[14,41] One study found superior responsiveness of the Nottingham Health Profile compared to the SGRQ.[39] Two studies[14,38] comparing the CRQ and SGRQ concluded that the CRQ is more responsive, while two other studies[16,40] did not favour one over the other instrument in terms of responsiveness.

However, only one study formally compared responsiveness within patients across instruments[11] while most studies assessed responsiveness for each instrument separately. For example, the studies assessed whether the difference between baseline and follow-up scores was statistically significant for each instrument using one-sample *t*-tests[14,16,25,26,38,40,45] and declared similar responsiveness when instruments detected significant differences. Other studies used Cohen's effect sizes14 or SRMs11. Only one study used, in addition to *t*-tests, the minimal important difference to define responsiveness as the ability of an instrument to detect patient-important changes.[16]

Consensus does not exist about the most appropriate method to assess and compare responsiveness.[46,47] Unlike most earlier studies, we did not rely on *t*-tests[14,16,25,26,38,40,45] or Cohen's effect sizes[14] to assess and compare responsiveness. Significance testing using *t*-tests depends partly on the observed change scores and variability in change scores. However, this test also depends on sample size, which should not influence judgments about responsiveness.[46,48] Cohen's effect size is calculated by dividing the change score by the standard deviation of the baseline score. In contrast to variability in change, there is no

necessary or logical relationship between between-person variability at baseline and instrument responsiveness. We used the SRM to compare responsiveness within patients across several instruments, because the SRM deals with the aforementioned limitations in that the denominator captures variability in change (the noise component of the signal-to-noise ratio) and removes the dependency on sample size.[46,49,50] However, the multiple comparison we used in this analysis do raise the possibility of a type I error. One solution to this problem is reliance on lower *P*-values and readers may focus on the *P*-values that are less than 0.01 in our analysis. Other methods of correction for multiple testing have limitations because the comparison we conducted were not independent.

The superior responsiveness of the SGRQ total score compared to its domain scores deserves attention. The increased number of items in the total score resulted in decreased variability in change, and thus an increased signal (change score) to noise (SD of change scores) ratio. Thus by calculating total scores, the SRM becomes larger compared to the domain scores. There are, however, limitations to using a total score. If effects of an intervention vary across domains of HRQL, the total score obscures potentially important information from HRQL domains and may result in misleading interpretations. In the present study, the effect of respiratory rehabilitation was similar in all domains and the SGRQ total score may adequately represent the overall treatment effect. In the present article, we did not calculate CRQ total scores because use of CRQ total scores requires additional research.

Investigators should not base the selection of the instrument on responsiveness alone, but also on longitudinal validity. In order to understand if an instrument measures changes in HRQL that one intends to measure, one needs to consider correlations of change scores with those of other instruments. We have previously described the longitudinal construct validity for the CRQ-IA and CRQ-SA and SGRQ.[19] The results of the latter and other studies suggest that the CRQ and SGRQ measure similar constructs.[13,14,17,28]

In recent years, investigators suggested that HQRL instruments can be used in clinical practice for evaluation purposes but also to facilitate the discussion of HQRL issues between patients and clinicians.[51] Clinicians should also choose instruments based on evidence about the measurement properties. But there are practical aspects that are important for a successful introduction of HQRL instruments into clinical practice. First, the duration and mode of administration (interviewer versus self-administration) should require as few resources as possible. Second, calculation of domain or summary scores should be straightforward and not involve complicated algorithms. Third, there should be evidence about the minimal important difference to guide interpretation of domain or summary scores. Considering these practical aspects, the CRQ-SA is likely to be the most prudent choice as a HQRL instrument for clinical practice.

This study confirms that the disease-specific HRQL instruments (CRQ and SGRQ), in particular the CRQ dyspnea domain, are substantially more responsive than generic measures when used for the evaluation of respiratory rehabilitation. It is possible that the relative responsiveness of HRQL instruments may depend on the population studied and contexts other than respiratory rehabilitation.[48] Studies using fully validated instruments evaluating the extent to which responsiveness depends on patients or interventions would enhance our understanding of the measurement of change in HRQL.

Acknowledgments

The work was in part supported by an unrestricted educational grant from AstraZeneca, Lund, Sweden. The sponsor was not involved in the study design, conduction of the trial and analysis of data. One of the authors is an employee of AstraZeneca and participated in the study through discussion and she reviewed drafts of the original manuscript making wording and presentation suggestions. Pfizer Limited, UK, provided additional funding. The CRQ-IA and CRQ-SA are copyrighted by McMaster University, Hamilton, Canada; Principal Authors Dr. Gordon Guyatt and Dr. Holger Schünemann. Use of the CRQ requires permission by McMaster University and the authors.

References

1. Calverley P, Pauwels R, Vestbo J, Jones P, Pride N, Gulsvik A, et al. Combined salmeterol and fluticasone in the treatment of chronic obstructive pulmonary disease: a randomised controlled trial. *Lancet* 2003;**361**(9356):449–56.
2. Goldstein RS, Gort EH, Stubbing D, Avendano MA, Guyatt GH. Randomised controlled trial of respiratory rehabilitation. *Lancet* 1994;**344**(8934):1394–7.
3. Griffiths TL, Burr ML, Campbell IA, Lewis-Jenkins V, Mullins J, Shiels K, et al. Results at 1 year of outpatient multidisciplinary pulmonary rehabilitation: a randomised controlled trial. *Lancet* 2000;**355**(9201):362–8.

4. Troosters T, Gosselink R, Decramer M. Short- and long-term effects of outpatient rehabilitation in patients with chronic obstructive pulmonary disease: a randomized trial. *Am J Med* 2000;**109**(3):207–12.
5. Normandin EA, McCusker C, Connors M, Vale F, Gerardi D, ZuWallack RL. An evaluation of two approaches to exercise conditioning in pulmonary rehabilitation. *Chest* 2002; **121**(4):1085–91.
6. Brightling CE, Monteiro W, Ward R, Parker D, Morgan MD, Wardlaw AJ, et al. Sputum eosinophilia and short-term response to prednisolone in chronic obstructive pulmonary disease: a randomised controlled trial. *Lancet* 2000; **356**(9240):1480–5.
7. Mahler DA, Wire P, Horstman D, Chang CN, Yates J, Fischer T, et al. Effectiveness of fluticasone propionate and salmeterol combination delivered via the Diskus device in the treatment of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2002;**166**(8):1084–91.
8. Poole PJ, Veale AG, Black PN. The effect of sustained-release morphine on breathlessness and quality of life in severe chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998;**157**(6 Pt 1):1877–80.
9. Guyatt G, Montori V, Devereaux PJ, Schunemann H, Bhandari M. Patients at the center: in our practice, and in our use of language. *ACP J Club* 2004;**140**(1):A11–2.
10. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993;**118**(8):622–9.
11. Guyatt GH, King DR, Feeny DH, Stubbing D, Goldstein RS. Generic and specific measurement of health-related quality of life in a clinical trial of respiratory rehabilitation. *J Clin Epidemiol* 1999;**52**(3):187–92.
12. 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**(10):773–8.
13. Harper R, Brazier JE, Waterhouse JC, Walters SJ, Jones NM, Howard P. Comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting. *Thorax* 1997;**52**(10):879–87.
14. Singh SJ, Sodergren SC, Hyland ME, Williams J, Morgan MD. A comparison of three disease-specific and two generic health-status measures to evaluate the outcome of pulmonary rehabilitation in COPD. *Respir Med* 2001;**95**(1):71–7.
15. Jones PW, 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**(6):1321–7.
16. Rutten-van Molken M, Roos B, Van Noord JA. An empirical comparison of the St. George's Respiratory Questionnaire (SGRQ) and the Chronic Respiratory Disease Questionnaire (CRQ) in a clinical trial setting. *Thorax* 1999;**54**(11): 995–1003.
17. Schunemann HJ, Griffith L, Jaeschke R, Goldstein R, Stubbing D, Austin P, et al. A comparison of the original chronic respiratory questionnaire with a standardized version. *Chest* 2003;**124**(4):1421–9.
18. Puhan MA, Behnke M, Laschke M, Lichtenschopf A, Brandli O, Guyatt GH, et al. Self-administration and standardisation of the chronic respiratory questionnaire: a randomised trial in three German-speaking countries. *Respir Med* 2004;**98**(4): 342–50.
19. Schunemann HJ, Goldstein R, Mador MJ, McKim D, Stahl E, Puhan M, et al. A randomised trial to evaluate the self-administered standardised chronic respiratory questionnaire. *Eur Respir J* 2005;**25**(1):31–40.
20. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care* 2002;**40**(2):113–28.
21. Chakravorty I, Cayton RM, Szczepura A. Health utilities in evaluating intervention in the sleep apnoea/hypopnoea syndrome. *Eur Respir J* 2002;**20**(5):1233–8.
22. Hurskainen R, Teperi J, Rissanen P, Aalto AM, Grenman S, Kivela A, et al. Clinical outcomes and costs with the levonorgestrel-releasing intrauterine system or hysterectomy for treatment of menorrhagia: randomized trial 5-year follow-up. *J Am Med Assoc* 2004;**291**(12):1456–63.
23. Kalra L, Evans A, Perez I, Melbourn A, Patel A, Knapp M, et al. Training carers of stroke patients: randomised controlled trial. *Br Med J* 2004;**328**(7448):1099.
24. Schunemann 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**(2):140–9.
25. Katsura H, Yamada K, Kida K. Usefulness of a linear analog scale questionnaire to measure health-related quality of life in elderly patients with chronic obstructive pulmonary disease. *J Am Geriatr Soc* 2003;**51**(8):1131–5.
26. Nishiyama O, Taniguchi H, Kondoh Y, Nishimura K, Suzuki R, Takagi K, et al. The effectiveness of the visual analogue scale 8 in measuring health-related quality of life for COPD patients. *Respir Med* 2000;**94**(12):1192–9.
27. Lacasse Y, Brosseau L, Milne S, Martin S, Wong E, Guyatt GH, et al. Pulmonary rehabilitation for chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2004(4): CD003793.
28. Wijkstra PJ, TenVergert EM, van Altena R, Otten V, Postma DS, Kraan J, et al. Reliability and validity of the chronic respiratory questionnaire (CRQ). *Thorax* 1994;**49**(5):465–7.
29. Jones PW, Quirk FH, Baveystock CM. The St. George's Respiratory Questionnaire. *Respir Med* 1991;**85**(Suppl B): 25–31.
30. Ware Jr. JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;**30**(6):473–83.
31. Bennet KJ. *Measuring health state preferences and utilities: rating scale, time trade-off, and standard gamble techniques. Quality of life and pharmacoeconomics in clinical trials*. Philadelphia: Lippincott-Raven; 1996. p. 259.
32. Schünemann HJ, Goldstein R, Mador J, McKim D, Stahl E, Griffith L, et al. Do marker states improve measurement properties of utility instruments: a randomized multi-center trial in patients with chronic respiratory disease. *Qual Life Res* 2006;**15**(1):1–14.
33. Puhan MA, Behnke M, Devereaux PJ, Montori VM, Braendli O, Frey M, et al. Measurement of agreement on health-related quality of life changes in response to respiratory rehabilitation by patients and physicians—a prospective study. *Respir Med* 2004;**98**(12):1195–202.
34. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *J Clin Epidemiol* 1997;**50**(1): 79–93.
35. Zou GY. Quantifying responsiveness of quality of life measures without an external criterion. *Qual Life Res* 2005; **14**(6):1545–52.
36. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. *J Clin Epidemiol* 2003;**56**(1):52–60.
37. Borson S, McDonald GJ, Gayle T, Deffebach M, Lakshminarayan S, VanTuinen C. Improvement in mood, physical

symptoms, and function with nortriptyline for depression in patients with chronic obstructive pulmonary disease. *Psychosomatics* 1992;**33**(2):190–201.
38. de Torres JP, Pinto-Plata V, Ingenito E, Bagley P, Gray A, Berger R, et al. Power of outcome measurements to detect clinically significant changes in pulmonary rehabilitation of patients with COPD. *Chest* 2002;**121**(4):1092–8.
39. Doll H, Duprat-Lomon I, Ammerman E, Sagnier PP. Validity of the St. George's respiratory questionnaire at acute exacerbation of chronic bronchitis: comparison with the Nottingham health profile. *Qual Life Res* 2003;**12**(2): 117–32.
40. Hajiro T, Nishimura K, Jones PW, Tsukino M, Ikeda A, Koyama H, et al. A novel, short, and simple questionnaire to measure health-related quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1999;**159**(6):1874–8.
41. Jaeschke R, Guyatt GH, Willan A, Cook D, Harper S, Morris J, et al. Effect of increasing doses of beta agonists on spirometric parameters, exercise capacity, and quality of life in patients with chronic airflow limitation. *Thorax* 1994;**49**(5):479–84.
42. Jones PW, Bosh TK. Quality of life changes in COPD patients treated with salmeterol. *Am J Respir Crit Care Med* 1997; **155**(4):1283–9.
43. Ries AL, Kaplan RM, Limberg TM, Prewitt LM. Effects of pulmonary rehabilitation on physiologic and psychosocial outcomes in patients with chronic obstructive pulmonary disease. *Ann Intern Med* 1995;**122**(11):823–32.
44. Stavem K, Erikssen J, Boe J. Performance of a short lung-specific health status measure in outpatients with chronic obstructive pulmonary disease. *Respir Med* 1999;**93**(7):467–75.
45. Tsukino M, Nishimura K, McKenna SP, Ikeda A, Hajiro T, Zhang M, et al. Change in generic and disease-specific health-related quality of life during a one-year period in patients with newly detected chronic obstructive pulmonary disease. *Respiration* 2002;**69**(6):513–20.
46. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;**53**(5):459–68.
47. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. *Med Care* 2000;**38**(9 Suppl):II84–90.
48. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 2003;**12**(4):349–62.
49. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990;**28**(7):632–42.
50. El Moussaoui R, Opmeer BC, Bossuyt PM, Speelman P, de Borgie CA, Prins JM. Development and validation of a short questionnaire in community acquired pneumonia. *Thorax* 2004;**59**(7):591–5.
51. Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. *J Am Med Assoc* 2002;**288**(23):3027–34.