Respiratory Medicine (2004) **98**, 1195–1202

respiratoryMEDICINE

Measurement of agreement on health-related quality of life changes in response to respiratory rehabilitation by patients and physicians—A prospective study

Milo A. Puhan[a,*], Michaela Behnke[b], P.J. Devereaux[c], Victor M. Montori[d], Otto Braendli[e], Martin Frey[f], Holger J. Schünemann[g,h]

[a]*Horten Centre, University Hospital of Zurich, Bolleystrasse 40, Postfach Nord, CH-8091 Zurich, Switzerland*
[b]*Pulmoresearch Institute, Grosshansdorf, Germany*
[c]*Department of Medicine, McMaster University, Hamilton, Ont., Canada*
[d]*Mayo Clinic, Division of Endocrinology, Diabetes, Metabolism and Internal Medicine, Rochester, MN, USA*
[e]*Zuercher Hoehenklinik Wald, Faltigberg-Wald, Switzerland*
[f]*Klinik Barmelweid, Barmelweid, Switzerland*
[g]*Departments of Medicine and of Social & Preventive Medicine, University at Buffalo, Buffalo, New York, USA*
[h]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont., Canada*

Received 22 December 2003; accepted 13 April 2004

KEYWORDS
Quality of life;
Physician–patient relations;
COPD;
Respiratory rehabilitation

Summary *Introduction*: To provide optimal care for patients with chronic obstructive pulmonary disease physicians need to understand if their patients benefit from an intervention. The objective of this study was to assess agreement between patients and physicians on health-related quality of life (HRQL) changes in response to respiratory rehabilitation and to explore sources for disagreement.

Methods: Sixty-one patients rated their health states on a validated preference-based instrument, the feeling thermometer (FT). In an analogous manner, the eight treating physicians rated the patients' health states on the FT. Patients and physicians were blinded to each other's ratings. We calculated intraclass correlation coefficients (ICC) to assess agreement between patients and physicians and used HRQL instruments and the 6-min walking test to assess the evaluative properties of the FT.

Results: We found moderate agreement at baseline (ICC 0.40, $P = 0.018$) and follow-up (ICC 0.49, $P = 0.008$) but large disagreement about change scores (ICC 0.02, $P = 0.46$). Patients' FT change scores correlated well with change scores of the Chronic Respiratory Questionnaire, SF-36 and the Borg scale for dyspnoea whereas physicians' FT change scores correlated significantly with the change score of the 6-

Abbreviations: HRQL, Health-related quality of life; COPD, Chronic obstructive pulmonary disease; FT, Feeling thermometer; CRQ, Chronic Respiratory Questionnaire

*Corresponding author. Tel.: +41-1-255-87-09.
E-mail address: milo.puhan@evimed.ch (M.A. Puhan).

0954-6111/$ - see front matter
doi:10.1016/j.rmed.2004.04.011

min walking test ($r = 0.33$). Physicians' ratings showed an inconsistent pattern for correlations with HRQL measures.

Conclusions: There is large disagreement between patients and physicians on HRQL changes in response to respiratory rehabilitation. Investigators should assess whether the introduction of HRQL instruments into clinical practice raises the awareness of physicians towards HRQL and improves agreement with their patients.

Introduction

In patients suffering from chronic diseases, ascertainment of treatment effects must include individual patient's symptoms and health-related quality of life (HRQL).[1,2] Physicians should have a vivid understanding of patients' symptoms and HRQL to adequately evaluate the benefits from interventions such as respiratory rehabilitation. Agreement on HRQL changes between patients and physicians indicates that physicians are able to understand the patients' perspective. This will help physicians to inform future patients about the effects of respiratory rehabilitation on patient-important outcomes.

The aim of respiratory rehabilitation for patients with chronic obstructive pulmonary disease (COPD) is to improve HRQL and exercise capacity.[1,2] There is strong evidence that respiratory rehabilitation leads to clinically relevant improvements of HRQL.[3] Physicians should act on this evidence, but they also need to evaluate the benefits and harms of respiratory rehabilitation in their individual patients. However, there is little evidence about physicians' assessment of health status and HRQL changes in individual patients with COPD and their agreement with patients.[1]

Agreement between the patients' and physicians' estimates of HRQL is an ongoing focus of interest. Ratings scales and multidimensional HRQL instruments have been used frequently to assess patients' and physicians' estimations of general health status,[4,5] disease-specific HRQL[6,7] or pain.[8,9] In most of these studies, agreement between the patients' and physicians' perspective was moderate (intraclass correlation coefficients for most studies between 0.5 and 0.6). Investigators[10] suggested that some disagreement may result from different measurement properties of the HRQL instruments in terms of validity and reliability when used by patients and physicians.

Investigators use preference based instruments to measure HRQL and to obtain utility estimates,[11–13] including the feeling thermometer (FT), a visual analogue scale presented in the form of a thermometer with 100 marked intervals.[14] When completing this instrument, patients choose the score on the thermometer that represents the value they place on their health state. The FT works well as an evaluative instrument to assess HRQL. Two recent randomised trials,[15,16] the EuroQol[17] and other studies[18–21] have shown the good responsiveness and strong longitudinal validity of the FT.

We assessed how patients and their treating physicians evaluated HRQL changes in response to respiratory rehabilitation using the FT and we explored the measurement properties of the FT when used by patients and physicians as a potential determinant influencing the level of agreement.

Methods

Study design

The data for this study were collected in a randomised trial in which we assessed the measurement properties of different administration formats of the German Chronic Respiratory Questionnaire (CRQ).[22,23] In this present study, we compared patients' and physicians' estimations of changes in health status following therapy using the FT.[14] At baseline, we asked patients with COPD to rate their general health status on the FT. After completion of an intense respiratory rehabilitation program we showed patients their baseline score and asked them again to rate their general health status on the FT considering their first score. In an analogous manner, we asked physicians directly involved in the care of the enrolled patients to value the patients' general health status on an identical FT at baseline and follow-up. Physicians and patients were blinded to each other's ratings. We did not intervene in the rehabilitation programme apart from the baseline and follow-up assessments. All local ethics committees approved the study protocol and patients provided informed consent prior to participation in the study.

Participants, physicians and intervention

We recruited consecutive COPD patients in four rehabilitation clinics in Switzerland, Germany and

Austria from August to November 2002 as described previously.[22] We selected respiratory rehabilitation as the intervention because there is strong evidence for its effectiveness to improve HRQL and exercise capacity[3] and it was therefore suitable for our analysis. We included physicians who were directly involved in the rehabilitative care of patients enrolled in this study. The intense respiratory rehabilitation programmes followed the published recommendations and consisted mainly of physical exercise,[2] but included also educational and psychological sessions. The sessions lasted for 2–3 weeks and were continued for another 8 weeks at lower intensity in ambulatory rehabilitation sites.

Outcome assessment

We used the FT as an estimate for general health status for both the patients and physicians. We used defined anchors for the worst (dead = 0) and best (perfect health = 100) health states to facilitate comparisons between individuals and groups.[24] We used an informed version of the FT, i.e. patients and physicians could reflect in their follow-up score the baseline estimate and the improvement since then. The uninformed administration of the FT may cause bias because physicians might recall pretreatment scores more easily than patients or vice versa. This bias would result in a more informed estimation of the after treatment scores in one group. In addition, previous studies showed that informing patient of their pretreatment scores improved measurement properties of HRQL instruments when administered after intervals of approximately 2 weeks.[25]

Additional outcome measures at baseline and follow-up

We used informed versions of the self- and interviewer-administered formats of the CRQ[22,26] to assess COPD-specific HRQL. The 20 questions of the CRQ provide summary scores for four domains that are important to COPD patients (dyspnoea, fatigue, emotional function and mastery). Patients answer each of the 20 questions on a seven-point scale expressing the degree of disability from 1 (maximum impairment) to 7 (no impairment). In addition all patients completed 6-min walking tests, the self-administered uninformed version of the German SF-36[27] and a modified Borg scale in German[28] to assess the intensity of perceived dyspnoea at the end of the 6-min walking test. The Borg scale consisted of a vertical line labeled from 0 to 10 and with verbal descriptors. Zero represented "no dyspnoea at all" and 10 "maximal dyspnoea". Of the HRQL instrument, patients completed the CRQ first, followed by the SF-36 and the FT.

Statistical analysis

We first determined the effectiveness of respiratory rehabilitation in our patients on a group level using established outcomes measures such as the CRQ and the 6-min walking test.[3] We calculated means and standard deviations for baseline, follow-up and change scores of the CRQ and the 6-min walking test and used paired *t*-tests to compare baseline and follow-up scores. We determined the proportion of patients whose change scores were above the minimal clinically importance difference, which is 0.5 for the CRQ[29,30] and around 50 m for the 6-min walking test.[31]

We then calculated means and standard deviations for baseline, follow-up and change scores for the FT of patients and physicians. We constructed Bland Altman plots[32] of the difference between the estimates of each patient–physician pair against the mean of their health status estimates (mean for each patient–physician pair) to visualise if the observed differences depended on the magnitude of the FT estimates (e.g. more disagreement for patients with worse health status). We developed Bland–Altman plots for baseline, follow-up and change scores to assess patient-physician agreement. In addition, we calculated intraclass correlation coefficients (two-way mixed effects model for average measures) for baseline, follow-up and change score agreement. The intraclass correlation coefficient is more appropriate for assessing reliability than the Pearson correlation coefficient because it reflects random as well as systematic differences in test scores.

We were interested in the evaluative properties of the FT. We assessed responsiveness (the ability to detect change in HRQL) using paired *t*-tests to compare baseline and follow-up scores within patients and physicians as well as unpaired *t*-tests to compare change scores between patients and physicians. We also determined the proportion of patients and physicians whose change scores were above the minimal clinically importance difference of the FT of 8.[33] We also assessed longitudinal validity of the FT for patients and physicians, i.e. whether the tendency for improvements or worsening measured by the FT was analogous compared to other outcome measures such as the CRQ, the 6-min walking test, the SF-36 (general health and

vitality index) and the Borg scale. We calculated Pearson correlation coefficients of the FT change scores with change scores of these outcome measures.

Finally, in a post hoc analysis we used linear regression models with the difference between patient and physician change scores as the dependent variable to assess whether physician or patient characteristics such as working experience, specialisation, rehabilitation centres (two teaching and one non-teaching hospital), patients' age or disease severity were significant independent predictors for disagreement.

We set the statistical significance level at *P*-values <0.05. We performed all statistical analyses with SPSS for Windows version 10.0 (SPSS Inc, Chicago, IL).

Results

We initially enrolled 80 patients in this study, but we had insufficient data for 9 patient–physician pairs for the following reasons: three patients withdrew for non-specified reasons, two did not complete the rehabilitation (one patient suffered from acute exacerbation and one discontinued the rehabilitation program), three physicians did not provide a FT score for the follow-up assessment and, upon review, one patient did not meet the a priori inclusion criteria. In addition, we excluded all patients ($n = 10$) recruited in Austria from the analysis because the treating physician did not remain blind to the patients' FT scores. Thus, we included data from 61 pairs with complete data sets (61 patients and 8 physicians).

Table 1 presents the patient and physicians characteristics at baseline. The physicians of the four rehabilitation centres differed in their clinical experience. Residents in one Swiss teaching hospital for rehabilitation medicine had a working or training experience of 3 years or less. They attended 1 year in the rehabilitation clinic as part of their 5 years training program in internal medicine or another discipline. In the other hospitals, the physicians had completed their training programs in internal medicine and two of them had subspecialty training in respiratory or rehabilitation medicine.

We observed not only significant differences between baseline and follow-up scores in all CRQ domains and the 6-min walking test (Table 2), but 61–77% of the patients had change scores above the minimal clinically important difference on the CRQ scores and 32% improved their functional exercise capacity clinically meaningfully (>50 m). Thus the patients' health status improved over the course of rehabilitation.

Table 1 Characteristics of participants.

Baseline characteristics	
Physicians	$N = 8$
Age, mean years (SD)	38.1 (7.8)
Women, *N*	4
Status, *N*	
Residents	4
Specialists	4
Internal medicine	4
Respiratory medicine	1
Rehabilitation medicine	2
Years since graduation	
Residents, mean (SD)	2.75 (0.5)
Specialists, mean (SD)	15.74 (7.4)
Working in teaching hospital, No.	6
Patients	$N = 61$
Women, *N* (%)	18 (34.5)
Age, years, mean (SD)	69.0 (7.2)
FEV_1, mean % predicted (SD)	43.8 (14.7)
FEV_1/FVC, mean % predicted (SD)	48.5 (12.3)
6-min walking test, mean meters (SD)	358 (118)
Smoking, mean pack-years (SD)	45.8 (27.6)
Social status, *N* (%)	
Living alone	20 (32.8%)
Working	9 (14.8)
Not working	8 (13.1)
Retired	44 (72.1)
Duration of rehabilitation, mean days (SD)	16.8 (3.7)

At baseline, patients estimated their health states on the FT significantly higher than physicians ($P = 0.015$), while estimations at follow-up did not differ significantly ($P = 0.23$, see Table 3). The FT in both patients and physicians indicated good responsiveness. But the physicians' change scores were on average 9.5 points larger ($P<0.001$) compared to the patients' estimates. Accordingly, a larger proportion of patients had change scores above 8 from the physicians' perspective (93%) compared to the patients' perspective (66%).

The intraclass correlation coefficient was 0.40 (95% CI 0.01–0.63, $P = 0.018$) for baseline agreement, 0.46 (95% CI 0.11–0.68, $P = 0.008$) for follow-up agreement and 0.02 (95% CI −0.39–0.34, $P = 0.46$) for agreement of change scores. The Bland Altman plots in Fig. 1 show the differences between patients and physicians (*y*-axis) and across the spectrum of disease severity (*x*-axis). There was a systematic trend towards higher baseline ratings by patients (data symbols above

Table 2 Effectiveness of respiratory rehabilitation.

	Baseline	Follow-up	Change scores (range) (CI, *P*-value)	Proportion of patients ⩾ MID
CRQ dyspnoea domain	3.78 (1.12)	4.68 (1.10)	0.90 (range −0.8 to 4.8) (0.63; 1.12; $P<0.001$)*	61%
CRQ fatigue domain	3.99 (1.22)	5.01 (0.99)	1.02 (range −1.8 to 3.8) (0.83; 1.35; $P<0.001$)*	77%
CRQ emotional function domain	4.26 (1.15)	5.32 (0.90)	1.06 (range −1.3 to 3.9) (0.83; 1.30; $P<0.001$)*	70%
CRQ mastery domain	4.36 (1.37)	5.39 (0.97)	1.03 (range −0.5–4.3) (0.75; 1.30; $P<0.001$)*	70%
6-min walking test (m)	358.1 (115.2)	393.6 (120.5)	35.4 (range −132 to 140) (22.1–48.8; $P<0.001$)*	32%

Values are mean (standard deviation).
*Paired *t*-tests (95% confidence intervals; *P*-value).

Table 3 Baseline, follow-up and change scores of the patients' and physicians' estimation of health status on the FT.

	Baseline	Follow-up	Change scores	Proportion of patients ⩾ MID
Patient FT	50.3 (14.0)*	63.0 (15.7)*	12.7 (range −14 to 40) (10.0; 15.3; $P<0.001$)‡	66%
Physician FT	44.0 (14.4)*	66.2 (13.4)*	22.2 (range 0–55) (19.2; 25.1; $P<0.001$)‡	93%
Difference between patients and physicians	6.3 (1.2; 11.4 $P=0.015$)†	−3.2 (−8.4; 2.1 $P=0.23$)†	−9.5 (−5.6; −13.4; $P<0.001$)†	

*Values are mean (standard deviation).
†Unpaired *t*-tests (95% confidence intervals; *P*-value).
‡Paired *t*-tests (95% confidence intervals; *P*-value).

zero in upper plot) and higher change scores by physicians (data symbols below zero in lower plot).

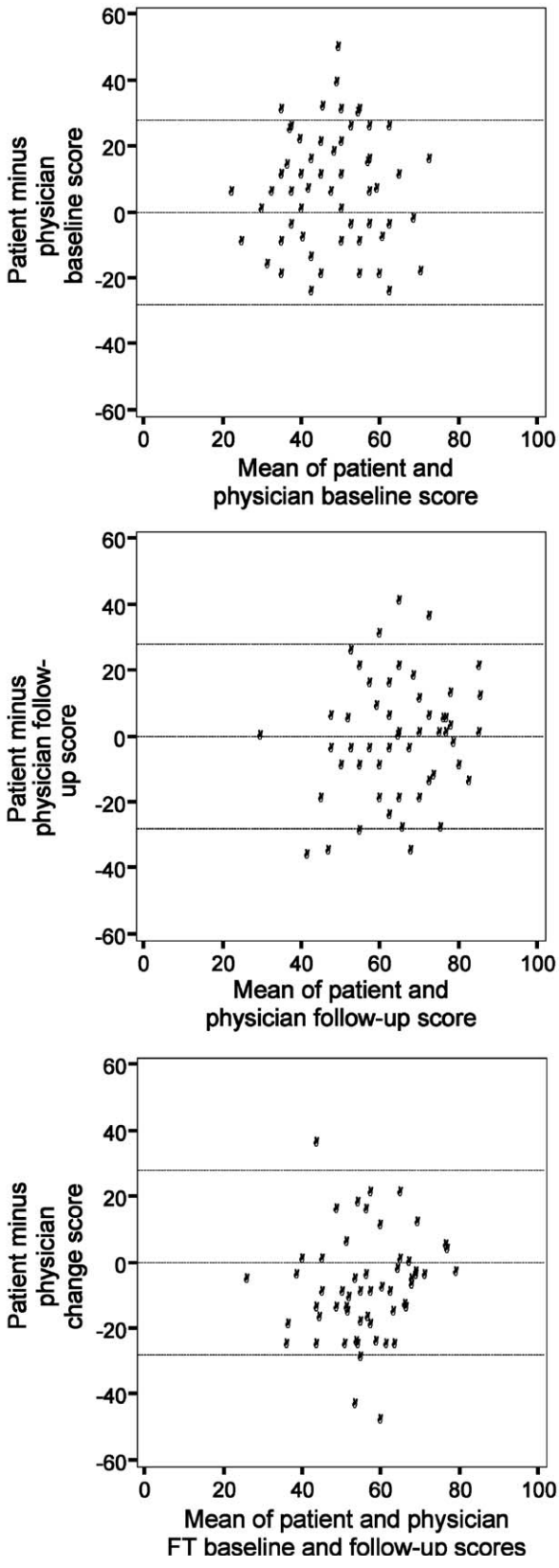
We show the longitudinal validity of the FT in Table 4. The patient FT change scores showed moderate to good correlations with all other HRQL measures indicating that the FT is indeed a global estimate for HRQL and symptoms (dyspnoea). The patient FT change scores did not reflect changes of exercise capacity as measured by the 6-min walking test. In contrast, we observed for the physician FT ratings the highest correlation with changes of exercise capacity ($r=0.36$). Correlations with HRQL measures were inconsistent and only significant with the CRQ emotional function domain ($r=0.33$).

Experience of physicians ($B=4.8$, $P=0.63$), specialisation ($B=-7.7$, $P=0.74$), rehabilitation centres ($B=-1.6$, $P=0.75$), disease severity ($B=-0.2$, $P=0.16$) and patients' age ($B=-0.5$, $P=0.09$) were not significant predictors for disagreement between patient and physician change scores.

Discussion

In this analysis, we studied agreement between patients and treating physicians on therapy related HRQL gains. While agreement for health status at baseline and follow-up was moderate, there was large disagreement on change of health status. The FT showed excellent responsiveness when used by patients and physicians but only the patients' ratings on the FT reflected HRQL changes well. The patients' FT changes scores correlated consistently with established HRQL measures such as the CRQ and the SF-36, while the physicians'

FT showed a different and inconsistent pattern of correlations with the change scores of other outcome measures.



The strength of our study includes the prospective study design that allowed blinding the patients and physicians to each other's ratings. In addition we could not only explore patient and physician factors that may explain some of the disagreement but we also had outcome measures to assess the evaluative properties of the FT.

The limitations of the study include the short time between baseline and follow-up and one could criticise that patients may not perceive the full effect from respiratory rehabilitation despite gains in function, until later.[34] However, we did observe treatment effects similar to those of other studies about respiratory rehabilitation indicating that patients achieved large improvement in HRQL.[35,36] In addition, our study resembles clinical practice in which the clinician has to assess patients after the intense respiratory rehabilitation and before referring them to subsequent less intense programs. If physicians caring for patients with COPD cannot adequately evaluate the effect from respiratory rehabilitation, the decision for referral is insecure. Our study included 61 patient–physician pairs and a larger number of pairs would have led to more precise estimates of agreement. However, agreement for baseline and follow-up scores, i.e. single points of time, were similar compared to previous studies that assessed agreement.[10]

There are several reasons for disagreement between physician and patient ratings. We eliminated bias by keeping patients and physicians blind to each other's HRQL measures. Therefore, therapeutic actions were taken without discussing these estimates and these conditions mimic current clinical practice.[37] Our data indicate that physicians focused on objective outcomes such as the 6-min walking test[2] and included little patient important outcomes in their rating of patients' HRQL. In addition, physicians are generally more familiar with outcome measures that reflect physiological processes. The correlation of the emotional function domain of the CRQ with the physicians' FT scores indicates that physicians consider emotional function in their evaluation of HRQL. For the patients, we confirmed the consis-

Figure 1 The three Bland Altman plots show the difference between patients' and physicians' baseline, follow-up and change scores on the FT against the mean of their baseline and follow-up scores. The middle line represents full agreement between patients and physicians (difference = zero) and the upper and lower line two standard deviations of the FT baseline scores. There is no clear trend that disease severity influenced differences between patients and physicians (smaller differences towards higher FT scores).

Table 4 Correlation* between the FT change scores and the other outcome measures.

	Patient FT	Physician FT
Dyspnoea CRQ	0.43† (0.31–0.55)	0.04 (−0.09–0.17)
Fatigue CRQ	0.42† (0.30–0.54)	0.17 (0.04–0.30)
Emotional function CRQ	0.43† (0.31–0.55)	0.33‡ (0.21–0.45)
Mastery CRQ	0.51† (0.40–0.62)	0.16 (0.03–0.29)
SF-36 general health perception	0.30‡ (0.17–0.43)	0.04 (−0.12–0.14)
SF-36 vitality index	0.27‡ (0.14–0.40)	−0.05 (−0.19–0.09)
Borg dyspnoea scale	−0.31‡ (−0.19 to −0.43)	0.08 (−0.05–0.21)
6-min walking test	0.03 (−0.10–0.16)	0.36‡ (0.24–0.48)

*Pearson correlation coefficient (95% confidence interval).
†Significant at <0.01.
‡Significant at <0.05.

tent correlation of the FT with HRQL measures as observed in earlier studies[15–17,38,39] and therefore the good longitudinal validity of the FT. Thus, the different measurement properties had an impact on agreement as hypothesised earlier by Sneeuw et al.[10]

The goals of respiratory rehabilitation are to reduce symptoms and to enhance quality of life. Established outcome measures are patient-important and, therefore, the patient's perspective must be reflected in the physicians' evaluation.[40] However, our data suggest that this is not the case. A recent study has shown that the use of HRQL instruments in daily practice heightened the physicians' awareness of their patients' HRQL.[41] HRQL issues were explicitly discussed more frequently and both patients and physicians believed that the use of HRQL instruments facilitates communications in daily practice. Routine discussion of HRQL in clinical practice may calibrate physicians' perspective of HRQL gains in response to therapy and may lead to realistic evaluations and more adequate recommendations to patients. Thus future research should focus on interventions that increase the physicians' awareness and understanding of the perspective of COPD patients'.

In conclusion, we found large disagreements between patients and their physicians when estimating HRQL gains in response to therapy. A likely explanation is that patients indeed reflect HRQL in their FT ratings whereas physicians include different information when they are asked to evaluate their patients' HRQL gains. Investigators should explore if the use of disease-specific HRQL instruments and subsequent discussion of the results between patients and physicians can raise the physicians awareness of HRQL, reduce disagreement between patients and physicians and whether greater agreement leads to different behaviour for recommendations of respiratory rehabilitation for future patients.

References

1. Pulmonary rehabilitation-1999. American Thoracic Society. *Am J Respir Crit Care Med* 1999; **159**(5 Pt 1):1666–82.
2. Pulmonary rehabilitation. *Thorax* 2001; **56**(11):827–34.
3. Lacasse Y, Brosseau L, Milne S, et al. Pulmonary rehabilitation for chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2002:CD003793.
4. Coucill W, Bryan S, Bentham P, Buckley A, Laight A. EQ-5D in patients with dementia: an investigation of inter-rater agreement. *Med Care* 2001;**39**(8):760–71.
5. Dorman PJ, Waddell F, Slattery J, Dennis M, Sandercock P. Are proxy assessments of health status after stroke with the EuroQol Questionnaire feasible, accurate, and unbiased? *Stroke* 1997;**28**(10):1883–7.
6. Patil R, Cotler S, Banaad-Omiotek G, et al. Physicians' preference values for hepatitis C health states and antiviral therapy: a survey. *BMC Gastroenterol* 2001;**1**(1):6.
7. Cotler SJ, Patil R, McNutt RA, et al. Patients' values for health states associated with hepatitis c and physicians' estimates of those values. *Am J Gastroenterol* 2001;**96**(9): 2730–6.
8. Marquie L, Raufaste E, Lauque D, Marine C, Ecoiffier M, Sorum P. Pain rating by patients and physicians: evidence of systematic pain miscalibration. *Pain* 2003;**102**(3):289–96.
9. Sutherland JE, Wesley RM, Cole PM, Nesvacil LJ, Daly ML, Gepner GJ. Differences and similarities between patient and physician perceptions of patient pain. *Fam Med* 1988;**20**(5): 343–6.
10. Sneeuw KC, Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. *J Clin Epidemiol* 2002;**55**(11):1130–43.
11. Wennberg JE. Outcomes research, cost containment, and the fear of health care rationing. *N Engl J Med* 1990;**323**(17):1202–4.
12. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care* 1989;**5**(4):559–75.
13. Feeny DH, Torrance GW. Incorporating utility-based quality-of-life assessment measures in clinical trials. Two examples. *Med Care* 1989;**27**(3 Suppl):S190–204.

14. Bennet KJ. Measuring health state preferences, utilities: rating scale, time trade-off, and standard gamble techniques. *Quality of life and pharmacoeconomics in clinical trials*. Philadelphia: Lippincott-Raven; 1996. p. 259.
15. 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.
16. Schünemann HJ, Armstrong D, Degli'Innocenti A, et al. A randomized multi-center trial to evaluate simple utility elicitation techniques in patients with gastro esophageal reflux disease. *Med Care* 2004, in press.
17. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Br J Rheumatol* 1997;**36**(5):551–9.
18. Alarcon GS, McGwin Jr G, Brooks K, et al. Systemic lupus erythematosus in three ethnic groups. XI. Sources of discrepancy in perception of disease activity: a comparison of physician and patient visual analog scale scores. *Arthritis Rheum* 2002;**47**(4):408–13.
19. Rider LG, Feldman BM, Perez MD, et al. Development of validated disease activity and damage indices for the juvenile idiopathic inflammatory myopathies: I. Physician, parent, and patient global assessments. Juvenile Dermatomyositis Disease Activity Collaborative Study Group. *Arthritis Rheum* 1997;**40**(11):1976–83.
20. Fries JF, Ramey DR. "Arthritis specific" global health analog scales assess "generic" health related quality-of-life in patients with rheumatoid arthritis. *J Rheumatol* 1997;**24**(9):1697–702.
21. Mathias SD, Colwell HH, Miller DP, Moreland LW, Buatti M, Wanke L. Health-related quality of life and functional status of patients with rheumatoid arthritis randomly assigned to receive etanercept or placebo. *Clin Ther* 2000;**22**(1): 128–39.
22. Puhan MA, Behnke M, Laschke M, et al. Self-administration and standardisation of the chronic respiratory questionnaire: a randomised trial in three German-speaking countries. *Respir Med* 2004;**98**:342–50.
23. Puhan MA, Behnke M, Frey M, et al. Self-administration and interviewer-administration of the German Chronic Respiratory Questionnaire: instrument development and assessment of validity and reliability in two randomised studies. *Health Qual Life Outcomes* 2003;**2**:1.
24. Torrance GW, Feeny D, Furlong W. Visual analog scales: do they have a role in the measurement of preferences for health states? *Med Decis Making* 2001;**21**(4):329–34.
25. Guyatt GH, Townsend M, Keller JL, Singer J. Should study subjects see their previous responses: data from a randomized control trial. *J Clin Epidemiol* 1989;**42**(9):913–20.
26. 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.
27. Bullinger M. German translation and psychometric testing of the SF-36 Health Survey: preliminary results from the IQOLA project. International quality of life assessment. *Soc Sci Med* 1995;**41**(10):1359–66.
28. Kirsten DK, Taube C, Lehnigk B, Jorres RA, Magnussen H. Exercise training improves recovery in patients with COPD after an acute exacerbation. *Respir Med* 1998;**92**(10):1191–8.
29. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;**10**(4):407–15.
30. Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol* 1996;**49**(11):1215–9.
31. Redelmeier DA, Bayoumi AM, Goldstein RS, Guyatt GH. Interpreting small differences in functional status: the six minute walk test in chronic lung disease patients. *Am J Respir Crit Care Med* 1997;**155**(4):1278–82.
32. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**(8476):307–10.
33. Schunemann HJ, Griffith L, Jaeschke R, Goldstein R, Stubbing D, Guyatt GH. Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. *J Clin Epidemiol* 2003;**56**(12):1170–6.
34. 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**(2):143–5.
35. Bendstrup KE, Ingemann JJ, Holm S, Bengtsson B. Outpatient rehabilitation improves activities of daily living, quality of life and exercise tolerance in chronic obstructive pulmonary disease. *Eur Respir J* 1997;**10**(12):2801–6.
36. Griffiths TL, Burr ML, Campbell IA, et al. Results at 1 year of outpatient multidisciplinary pulmonary rehabilitation: a randomised controlled trial. *Lancet* 2000;**355**(9201):362–8.
37. Puhan MA, Koller M, Brandli O, Steurer J. Pulmonary rehabilitation of COPD in Switzerland—an assessment of current status. *Schweiz Rundsch Med Prax* 2003;**92**(4):111–6.
38. Juniper EF, Guyatt GH, Feeny DH, Griffith LE, Ferrie PJ. Minimum skills required by children to complete health-related quality of life instruments for asthma: comparison of measurement properties. *Eur Respir J* 1997;**10**(10): 2285–94.
39. Juniper EF, Thompson AK, Roberts JN. Can the standard gamble and rating scale be used to measure quality of life in rhinoconjunctivitis? Comparison with the RQLQ and SF-36. *Allergy* 2002;**57**(3):201–6.
40. Guyatt GH, Devereaux PJ, Montori V., Schunemann HJ, Bhandari M. Putting the patient first: in our practice, and in our use of language. *ACP J Club* 2004; (140): A11.
41. 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.