Eur Respir J, 1995, 8, 888–898
DOI: 10.1183/09031936.95.08060888

European Respiratory Journal
ISSN 0903 - 1936

Comparison of performance of four instruments in evaluating the effects of salmeterol on asthma quality of life

M.P.M.H Rutten-van Mölken*, F. Custers**, E.K.A Van Doorslaer+, C.C.M. Jansen*, L. Heurman**, F.P.V Maesen**, J.J. Smeets**, A.M. Bommer++, J.A.M. Raaijmakers++

Comparison of performance of four instruments in evaluating the effects of salmeterol on asthma quality of life. M.P.M.H. Rutten-van Mölken, F. Custers, E.K.A. Van Doorslaer, C.C.M. Jansen, L. Heurman, F.P.V. Maesen, J.J. Smeets, A.M. Bommer, J.A.M. Raaijmakers. ©ERS Journals Ltd 1995.

ABSTRACT: Quality of life measures are increasingly used as important efficacy endpoints in studies of drugs for asthma. The purpose of this study was to assess both the sensitivity to change and the construct validity of four different quality of life instruments in patients with asthma.

In a double-blind, parallel group study, 120 moderate asthma patients, aged between 18–70 yrs, received either inhaled salmeterol 50 μg *b.i.d.* or inhaled salbutamol 400 μg *b.i.d.* In addition to respiratory outcomes, quality of life was measured at a 6 weeks follow-up using: 1) Asthma Quality of Life Questionnaire (AQLQ); 2) Living With Asthma Questionnaire (LWAQ); 3) Sickness Impact Profile (SIP); 4) Rating Scale (RS); and Standard Gamble (SG) utilities.

Salmeterol led to significant improvements over salbutamol on virtually all clinical outcomes. Although all the quality of life instruments showed the same trend in favour of salmeterol, only the AQLQ and RS utilities showed significantly greater improvement on salmeterol than on salbutamol. Except for the AQLQ, the correlation between change in lung function and change in quality of life was generally low. Whereas, the AQLQ correlated well with the patient's overall assessment of efficacy (r=0.64), the LWAQ, SIP and utilities failed to show such a correlation. The AQLQ showed the best correlation with symptom scores. The cross-sectional correlation between the AQLQ and the LWAQ was 0.73, whereas the longitudinal correlation was only 0.29. The SG generally showed poor correlation with other measures, including the RS.

In conclusion, patients given salmeterol showed a greater improvement in quality of life compared to patients given salbutamol. Of the disease-specific questionnaires the Asthma Quality of Life Questionnaire was found to be more responsive to change than the Living With Asthma Questionnaire and showed greater validity. Of the generic instruments, the rating scale utilities were most responsive. The Standard Gamble showed poor correlation with other measures.

Eur Respir J., 1995, 8, 888–898.

*Dept of Health Economics, University of Limburg, Maastricht, The Netherlands. **Dept of Respiratory Medicine, De Wever Hospital, Heerlen, The Netherlands. +Institute for Medical Technology Assessment, Erasmus University Rotterdam, Rotterdam, The Netherlands. ++Glaxo B.V., Zeist, The Netherlands.

Correspondence:
M.P.M.H. Rutten-van Mölken
University of Limburg
Dept of Health Economics
P.O. Box 616
6200 MD Maastricht
The Netherlands

Keywords: Asthma
quality of life
reliability
salmeterol
sensitivity to change
validity

Received: April 29 1994
Accepted after revision February 25 1995

The study has been financially supported by Glaxo B.V. The Netherlands.

One of the most important goals of asthma management is to improve a patient's everyday functioning and subjective well-being, often referred to as "quality of life". Improvement in quality of life can be achieved through improvement of respiratory function and reduction of symptoms and exacerbations. Whereas the impact of treatment on respiratory functioning can be measured exactly, the assessment of change in quality of life is often based on the physician's intuition about the relevance of lung function changes to a patient's daily functioning. Until recently, the measurement of quality of life by means of questionnaires had not gained widespread acceptance in asthma research. A bibliography of quality of life measures published in 1990 did not include a single asthma-specific quality of life questionnaire [1]. However, quality of life measures are increasingly being proposed as important efficacy endpoints, especially in studies of drugs [2]. Quality of life monitoring is important, because it reflects patients' concerns and perceptions and it can be one of the predictors of compliance. Moreover, the relatively weak correlations between lung function and quality of life found in a number of earlier studies [3–8], indicate that quality of life outcomes focus on rather different aspects of a disease than the physiological outcomes. Together, these outcomes may give a complete picture of a patient's health status.

Recently, a number of asthma-specific quality of life measures have been proposed for use in clinical trials

[9–14]. Assessing the measurement properties of these instruments, such as validity, reliability and responsiveness, is an ongoing process that has to be repeated in different countries, patient populations and treatment settings. Based on the current knowledge of measurement properties, none of the recently developed asthma quality of life questionnaires can be said to be superior in particular circumstances.

In this study, we evaluated the measurement properties of the Dutch versions of four quality of life questionnaires in a clinical trial comparing salmeterol with salbutamol. Two of the questionnaires, the Living with Asthma Questionnaire (LWAQ) and the Asthma Quality of Life Questionnaire (AQLQ), were disease specific, whereas the other two, the Sickness Impact Profile (SIP) and the Patient Utilities, were generic. The advantage of incorporating both types of instruments into one study is that we can make a detailed study of those aspects of quality of life that are characteristic for asthma, as well as a global assessment of overall quality of life. Such an overall assessment enables the comparison of results across different interventions for different diseases

In the present study, we were primarily interested in the responsiveness of the four questionnaires to clinical improvements that were expected to result from salmeterol compared to salbutamol. Salmeterol is a new long-acting β-agonist which, when compared to either placebo or salbutamol, has been shown to increase clinic lung function and both morning and evening peak flow, reduce diurnal variation in peak flow, and daytime and nocturnal asthma symptoms and exacerbations, as well as the need for additional bronchodilator rescue drugs [15–18]. Furthermore, salmeterol has been shown to improve the duration and quality of sleep [19]. This, in turn, may be associated with better daytime functioning and well-being.

All of these clinical effects of salmeterol may be expected to improve quality of life. Thus far, few studies have addressed this question. ULLMAN *et al.* [15] found a nonsignificant trend towards a better sense of well-being during salmeterol as compared to salbutamol treatment. A study by PALMER and HYLAND [20] showed that the addition of salmeterol instead of placebo to a patient's standard therapy improved the quality of life, as measured with the LWAQ. However, more than 50% of the patients were excluded from this study because they did not complete valid responses to the questionnaire. JUNIPER *et al.* [21] showed that scores on the AQLQ were significantly better on salmeterol than on a placebo or salbutamol.

In addition to the ability of the four quality of life questionnaires to detect clinically important differences, we also looked into their longitudinal and cross-sectional construct validity [22]. Both types of validity were addressed, because the questionnaires can be used to measure within-patient changes as well as to measure differences between patients with various degrees of disease severity. Assessing these psychometric properties may aid future selection of measures, assist in prioritizing outcomes, and permit more accurate sample size calculations.

Methods

Patients

Subjects were out-patients with stable asthma, aged 18–70 yrs, who consented to participate in the study. On entering the study, they were required to have a forced expiratory volume in one second (FEV_1) of 50–70% of predicted normal values, not having used bronchodilators for at least 6 h previously. The FEV_1 had to increase by at least 15% after inhalation of 400 μg salbutamol. Patients were excluded if they had suffered a serious concomitant disease, an upper or lower respiratory tract infection during the 6 weeks prior to the study, used a booster course of oral steroids during the 6 weeks prior to the study, were likely to change their concurrent medication for diseases other than asthma, or had a poor understanding of Dutch. The use of theophyllines, other β-agonists apart from study medication, β-blockers, methylxanthines, or a booster course of corticosteroids was not allowed. All other maintenance drugs were continued at constant dosage throughout the study.

Study design

The study had a double-blind, randomized, parallel group design. After a 2 week run-in period, 120 patients eligible for treatment were randomized to receive 6 weeks of treatment with either salmeterol 50 μg *b.i.d.* or salbutamol 400 μg *b.i.d.* Both drugs were administered as dry powder *via* the diskhaler device. Rescue salbutamol in a diskhaler device was supplied for symptomatic relief. Patients were seen at the end of the run-in period, after 3 weeks and 6 weeks of treatment, and 2 weeks after cessation of treatment. Quality of life measurements were conducted at the end of the run-in period and after 6 weeks of treatment. The study was approved by the Medical Ethics Committee of the Wever Hospital.

Clinical outcome measures

FEV_1 and forced vital capacity (FVC) were measured at each clinic visit. They were obtained from the best of three measurements using a Sensor Medics Pulmograph. Prior to these clinic measurements the patients withheld inhaled bronchodilators for 12 h. Reversibility was calculated as the change in FEV_1 after inhaling 400 μg salbutamol divided by the prebronchodilator FEV_1 and expressed as a percentage. Peak expiratory flow rate (PEFR) measurements were obtained using the mini-Wright PEFR meter each day on awakening and in the evening before inhaling the study medication. Patients kept a diary to record the number of daytime and night-time rescue salbutamol blisters (each blister contained 400 μg salbutamol), and to record daytime and night-time asthma symptoms. The daytime symptom score documented the frequency and severity of symptoms on a scale from 0 (no symptoms during the day) to 5 (symptoms so severe that you

could not go to work or perform normal daily activities). The night-time symptom score ranged from 0 (no symptoms during the night) to 4 (symptoms so severe that you did not sleep at all).

An overall assessment of efficacy was obtained at each clinic visit by asking the patients how their asthma/breathing had been since the last visit, using a five point response scale ranging from much improved to much worse.

Quality of life instruments

All four quality of life measures used in this study originate from the English language and were translated into Dutch. The advantage of translating existing measures over developing new measures is that it provides a basis for comparison of investigations of quality of life across different countries and cultures [23]. To a large extent, the translations were performed according to recently published methods for good quality translations [23, 24].

Asthma Quality of Life Questionnaire. The AQLQ was developed by JUNIPER and co-workers [14]. The questionnaire is interviewer-administered and designed for use in clinical trials involving adult asthma patients. It contains 32 questions in four domains: activity limitation (11 items); symptoms (12 items); emotional function (5 items); and environmental stimuli (4 items). The AQLQ is made up of questions which ask patients to think about the last 2 weeks, and quantify, either in terms of frequency, duration or amount, a number of asthma-related problems. At follow-up, patients are shown the answers they have given at baseline. The first five questions in the activity limitation domain are "patient-specific". This means that at baseline each patient is asked to state which five of his or her regular activities are most troublesome because of asthma. At follow-up, the patient is again asked how severe the limitation is in doing these same five activities. Answers to each of the 32 items can be scored on a 7 point scale, ranging from 1, which indicates maximal impairment, to 7, which indicates no impairment. The results are expressed as the mean score for each domain as well as for overall quality of life. The AQLQ was found to be valid, reproducible, and responsive to change in time [8, 21, 25, 26].

The translation procedure of the AQLQ started with forward translations by two independent translators, who were aware of the objectives of the instruments and the concepts involved. When consensus had been reached between the two translators and the researchers about the forward translation, a backtranslation was performed by a bilingual professional translator not involved in health care research. To check the face validity, the questionnaires were completed by a number of patients who were encouraged to comment on how they interpreted the questions. If needed, alterations were made in line with the suggestions.

Living with Asthma Questionnaire. The LWAQ was developed by HYLAND *et al.* [10]. This questionnaire was translated according to a similar procedure as the AQLQ, which was extended by the use of several backtranslations. The LWAQ consists of a list of 68 statements covering 11 domains: social/leisure, sport, holidays, sleep, work/other activities, colds, mobility, effects on others, medication usage, sex, and dysphoric states and attitudes. Factor analysis has shown that the LWAQ also covers two distinct constructs, which are labelled "health knowledge" and "health appraisal" [27]. The first construct, which contains 49 items, refers to the patient's knowledge of functional limitation imposed by illness. The second construct, which covers 19 items, refers to the patient's evaluation of how much distress those limitations cause. The LWAQ does not ask patients to consider a well-defined time period. To compensate for the tendency of patients to agree with a statement irrespective of its content, the LWAQ contains both positive and negative statements. Patients are asked to indicate for each statement whether it is untrue, slightly true, very true, or not applicable. For negative statements, not applicable and untrue are scored 0, slightly true is scored 1, and true is scored 2. For positive items, not applicable and very true are scored 0, slightly true is scored 1, and untrue is scored 2. Results can be expressed in an overall score, scores for the 11 domain subscales and scores for the two construct subscales. Scores are based on a summation of the patient's responses to all statements divided by the number of valid statements (thus, compensating for statements that are "not applicable" or missing) [20]. Thus, the overall score and the subscores can range from 0 to 2. The lower the score, the better the quality of life. The LWAQ has been shown to be reliable and valid [9, 10], and there is some indication that it may be responsive to change [27].

Sickness Impact Profile. In contrast to the two disease-specific questionnaires described above, the SIP is a comprehensive generic quality of life measure. It was developed by BERGNER *et al.* [28], and has been shown to be reliable and valid. A validated Dutch version of the SIP is available [29]. The SIP measures health-related dysfunction in 12 domains: ambulation, mobility, body care and movement, social interaction, emotional behaviour, alertness behaviour, communication, work, eating, sleep and rest, household management, and recreation and pastimes. It asks patients to consider their situation on the day they complete the questionnaire. The SIP provides summary scores for physical, psychosocial and overall behavioural dysfunction, as well as separate scores for the 12 domains. Scores, which are calculated by addition of predetermined weights, range from 0 to l00, and express the percentage of impairment. The larger the score the greater the impairment. In this study, we applied the original SIP item weights. The SIP has been used in many studies, including studies of patients with asthma [6, 25, 26, 30, 31].

Patient utilities. The primary objective of the instruments outlined above is to describe the (change in) quality of life. Utility measurement goes one step further by explicitly valuing the quality of life. A utility is a single

number on a scale ranging from perfect health, which has a utility of 1, to death, which has a utility of 0. This number represents the value or preference weight assigned to a particular health state. We have elicited utilities from patients by means of a slightly adapted Dutch version of the McMaster Utility Measurement Questionnaire (MUMQ), [32, 33]. This instrument was translated following exactly the same procedure as followed for the AQLQ. The back-translation of the MUMQ has been published [34]. The patients are first asked to describe their own health state over the past 2 weeks by indicating the level of dysfunction on six domains, *i.e.* physical state and mobility, self-care, emotions, leisure activities, pain or other complaints, and side-effects of treatment. The actual valuation of their own health state is preformed by means of rating scale and standard gamble techniques [35]. The rating scale looks like a thermometer with "perfect health" equal to 100 at the top and "death" equal to 0 at the bottom. Patients are asked to place their own health state on that scale, in such a way that it indicates how much better or worse their health state is relative to death and perfect health. In the standard gamble, patients are offered a choice between the certainty of living in their current health state for the rest of their life, or taking a gamble in which they have chance p of returning to perfect health for the rest of their life and chance l-p of dying immediately. Probability p is varied until patients are indifferent between the gamble and living in their current health state. At this indifference point, the utility equals p. For example, if a patient is indifferent between continuing to live in his current health state and a gamble offering 70% chance of gaining perfect health and 30% chance of dying, the utility or preference weight the patient assigns to his own health state on the 0 to 1 scale is 0.7.

Analysis

Analyses of differences in baseline characteristics between the two treatment groups were based on t-tests for continuous data, and Mann-Whitney U-tests for ordinal data. Differences in proportions were analyzed using χ^2 tests. A p-value equal to or less than 0.05 was considered statistically significant. Diary card data were summarized per patient by calculating means and standard deviations for the 2 week run-in period and the 6 week treatment period.

Responsiveness can be operationalized in three different ways. Firstly, the ability of the questionnaires to detect within-subject changes was compared by means of paired t-tests. Secondly, we compared the ability of the questionnaires to detect differences in change between the group receiving salmeterol and the group receiving salbutamol. This was done by subjecting the change scores to analysis of covariance (ANCOVA) with the treatment group used as a factor and the baseline scores used as covariates. ANCOVA was also used to assess differences in measures of lung function. Further analyses were performed to investigate whether the effect of treatment was different between patients using corticosteroids and patients not using corticosteroids. Thirdly, we calculated the effect sizes in the salmeterol and salbutamol groups. Effect size was calculated as the mean change from baseline to follow-up in the salmeterol group divided by the pooled within-subject standard deviation of change from both the salmeterol and salbutamol groups [22, 36].

Cross-sectional construct validity was measured by correlating quality of life scores at each visit with lung function parameters and other quality of life scores. Longitudinal construct validity was assessed by calculating correlations between within-subject changes in quality of life scores and within-subject changes in lung function parameters, symptom scores and other quality of life scores. The strength of the correlation was reported using Spearman's rank correlation coefficients. The correlation between change in quality of life and patient's overall assessment of efficacy expressed on a five point scale was also seen as an indicator of longitudinal construct validity. The scores from the overall assessment were reversed, so that a higher score means a better efficacy. The patient's overall assessment of efficacy was seen as a reference measure of whether or not a patient had changed. If the change in quality of life scores reflects this overall assessment of change (*i.e.* a strong correlation), this is seen as evidence of longitudinal validity.

Results

Patients

After randomization, one of the 120 patients had to be excluded because his understanding of the Dutch language was too poor to complete the questionnaires. Eight of the remaining 119 patients did not complete the 6 week follow-up measurement. Three of these patients (one salmeterol and two salbutamol) dropped out because of pulmonary problems. The pulmonary problems of the two patients receiving salbutamol were so serious that they required hospitalization. Two patients (one salmeterol and one salbutamol) dropped out because of other illnesses, and two patients (one salmeterol and one salbutamol) dropped out due to family circumstances (*e.g.* death of a close relative). One patient receiving salmeterol was no longer willing to participate. Another four patients (two salmeterol and two salbutamol) were not included in the analysis because of protocol violations (they were given antibiotics to treat respiratory infections). Thus, the analyses were performed on 107 patients, 53 receiving salmeterol and 54 receiving salbutamol.

The baseline characteristics of the patients who were not included in the analysis did not differ from those who completed the study. Table 1 shows no statistically significant differences between the treatments groups at baseline. Also, no significant differences were found on social indicators, such as marital status, educational level, income and employment, except for the fact that a smaller percentage of the patients in the salmeterol group

Table 1. – Baseline characteristics by treatment group

	Treatment groups		
	Salmeterol n=53	Salbutamol n=54	p=value#
Age yrs*	51 (15)	55 (13)	0.103
Sex M/F	27/26	33/21	0.289
Age of asthma onset yrs*	31 (20)	33 (20)	0.695
FEV1 % pred*	59 (7)	59 (7)	0.958
PEFR prebronchodilator L·s*	4.0 (1.2)	4.2 (1.4)	0.473
PEFR postbronchodilator L·s^{-1}*	5.3 (1.7)	5.1 (1.5)	0.546
Reversibility %+	27 (17, 35)	22 (17, 31)	0.154
Using steroids n			
Inhaled	37	33	0.077
Oral	4	0	
Both	3	4	
Asthma Quality of Life Quest.*	5.55 (1.01)	5.62 (0.84)	0.702
Living with Asthma Quest*	0.78 (0.36)	0.77 (0.31)	0.963
Sickness Impact Profile*	6.99 (6.43)	9.03 (7.56)	0.137
Utilities*			
Rating scale	69.91 (13.95)	67.13 (17.44)	0.366
Standard gamble	0.87 (0.13)	0.87 (0.16)	0.760

*: data are presented as mean and SD in parenthesis. +: median and (quartiles). M: male; F: female; FEV1: forced expiratory volume in one second; PEFR: peak expiratory flow rate; % pred: percentage of predicted value; Quest.: Questionnaire. #: p-value, salmeterol *vs* salbutamol.

had children (64 *vs* 84%; p=0.024). No differences were found between the groups on baseline quality of life scores.

Clinical outcome measures

Table 2 shows that salmeterol led to significant improvements over salbutamol on all respiratory outcomes, except for PEFR in the evening. No significant difference in these improvements was observed between patients taking corticosteroids and those not taking corticosteroids. Thus, an interaction between corticosteroid use and type of bronchodilator was not present. The mean use of supplemental salbutamol during daytime was significantly lower for salmeterol than for salbutamol (0.64 *versus* 1.24 blisters·day^{-1}; p=0.013). The additional salbutamol use during the night was not significantly different between the groups (0.13 *versus* 0.25 blisters·day^{-1}; p=0.076). The mean daytime and night-time symptom scores were very low and not significantly different between the two treatment groups (p=0.08 and p=0.67, respectively). The percentage of days without any symptoms was significantly higher for salmeterol than for salbutamol (63 *versus* 48%; p=0.039), but the percentage of nights without awakenings was not significantly different between the groups (80 *versus* 70%; p=0.108). Asthma-related adverse events and pharmacologically predictable adverse events occurred at the same rate in both treatment groups. After 3 weeks of treatment the patients' overall assessments of efficacy clearly favoured salmeterol (table 3). After 6 weeks of treatment no further changes in the patients' overall assessments were found.

Responsiveness

Table 4 shows that the AQLQ, the LWAQ and the rating scale utilities were highly responsive to the within-patient improvements in the salmeterol group. These measures also showed some within-patient improvement in the salbutamol group. Only the standard gamble utilities showed a within-patient deterioration of quality

Table 2. – Effectiveness of the two treatments in terms of lung function by treatment group

	Mean change in lung function				
	Salmeterol	Salbutamol	Δ(salm - salb)	CI of diff.	95% p-value#
PEFR a.m. L·min^{-1}	+32.9	+8.6	24.3	8.94–39.6	0.002
PEFR p.m. L·min^{-1}	+16.8	+4.3	12.5	-1.2–26.1	0.073
FEV1 % pred	+9.3	+1.7	7.5	3.2–11.9	0.001
FVC L	+0.32	+0.04	0.27	0.10–0.44	0.002
Reversibility %	-16.1	-7.3	-8.8	-12.8–4.8	<0.001

FVC: forced vital capacity; salm: salmeterol; salb: salbutamol; Δ: difference; 95% CI: 95% confidence interval. #: salmeterol *vs* salbutamol.

Table 3. – Patients' overall assessment of efficacy after 3 weeks of treatment

	Salmeterol	Salbutamol
Asthma:		
Much improved	9	3
Improved	22	12
Equal	21	34
Worse	1	4
Much worse	0	0

Mann-Whitney test: p=0.001.

of life in the salbutamol group. This deterioration was not found in any of the other quality of life measures.

Although all the questionnaires revealed the same trend in favour of salmeterol, only the AQLQ and the rating scale utilities showed a statistically significant higher improvement in quality of life on salmeterol than on salbutamol. The overall AQLQ score and the domain scores for activity and symptoms showed a significantly greater improvement on salmeterol. The activity domain of the AQLQ was found to be most responsive, and the emotional domain was the least responsive to change. The overall LWAQ score and its construct subscores did not show a significant difference between the groups. However, the table does not show the finding that 2 of its 11 domains, sport and sex, did show a difference in favour of salmeterol. The mean difference between the groups as regards change in the dimension of sport was 0.27 (95% confidence interval (95% CI) 0.09–0.44; p=0.003) and the mean difference for change in the dimension of sex was 0.24 (95% CI 0.05–0.42; p=0.012). No significant interaction effect between treatment and the use of corticosteroids was found for any of the quality of life outcomes.

Table 5 relates the change in the mean score in the salmeterol group to the within-subject standard deviation of change in both the salmeterol and salbutamol groups. The effect sizes that result from this analysis showed the AQLQ to be most responsive, followed by the LWAQ and rating scale utilities.

Cross-sectional validity

Cross-sectional construct validity correlations are given in table 6. In general, the association between quality of life measures and lung function measures was low. The SIP correlated significantly with all lung function measures except reversibility. The LWAQ and the SIP correlated better with the FVC, and the AQLQ correlated better with reversibility, than the other quality of life measures.

All quality of life questionnaires, except for the standard gamble utilities, correlated significantly with both daytime and night-time symptom scores. This correlation was considerably higher for the AQLQ than for the other questionnaires.

The correlations among quality of life questionnaires were much higher than the correlations between quality of life and lung function. The correlation between the two disease-specific quality of life measures was strong. The SIP correlates well with the LWAQ and rating scale utilities. Standard gamble utilities showed weak correlations with the other quality of life measures.

Table 4. – Responsiveness to changes within subjects and differences in changes between salmeterol (n=53) and salbutamol (n=54)

	Change in Quality of Life				
	Salmeterol	Salbutamol	Diff. (salm - salb)	95% CI of diff.	p-value#
AQLQ	+0.49***	+0.27***	0.22	0.03–0.42	0.022
Activities	+0.57***	+0.22*	0.35	0.14–0.56	0.002
Symptoms	+0.58***	+0.33**	0.25	-0.001–0.49	0.051
Environment	+0.44***	+0.22*	0.22	-0.04–0.48	0.100
Emotions	+0.18*	+0.17*	0.01	-0.19–0.20	0.960
LWAQ	-0.12***	-0.09*	0.03	-0.04–0.09	0.425
Health knowledge	-0.13***	-0.08**	0.05	-0.02–0.12	0.146
Health appraisal	-0.06**	-0.10	0.04	-0.12–0.03	0.290
SIP	-1.59	-0.93	0.66	-1.11–2.43	0.463
Physical	-0.83	-0.20	0.62	-0.73–2.00	0.363
Psychosocial	-2.82**	-1.91*	0.91	-1.35–3.16	0.429
Utilities					
Rating Scale	+7.35***	+4.08**	3.27	0.08–6.46	0.045
Standard Gamble	+0.008	-0.034*	0.042	-0.005–0.09	0.077

Covariance analysis: improvement is indicated by a plus for the AQLQ and the Utilities, and by a minus for the LWAQ and the SIP. AQLQ: Asthma Quality of Life Questionnaire; LWAQ: Living with Ashma Questionnaire; SIP: Sickness Impact Profile; For further abbreviations see legend to table 2. Within-patient changes; *: p≤0.05; **: p≤0.01; ***: p≤0.001. #: p-value, salmeterol *vs* salbutamol.

Table 5. – Effect sizes in salmeterol group by quality of life instrument

Quality of life instrument	Within-subject Δsalm	Pooled within-subject SD of change[#]	Effect size[##]
AQLQ	0.492	0.600	0.820
Activities	0.573	0.666	0.860
Symptoms	0.581	0.803	0.723
Environment	0.438	0.797	0.550
Emotions	0.176	0.583	0.302
LWAQ	0.118	0.170	0.694
Health knowledge	0.130	0.208	0.625
Health appraisal	0.064	0.192	0.333
SIP	1.585	4.956	0.320
Physical	0.827	3.950	0.209
Psychosocial	2.815	6.648	0.423
Utilities			
Rating Scale	7.345	10.381	0.708
Standard Gamble	0.008	0.161	0.050

Δslm: change in salmeterol. [#]: SD Δsalm+salb; [##]: Δsalm/SD Δsalm+salb. For further abbreviations see legend to tables 2 and 4.

Table 6. Cross-sectional validity: Spearman's rank correlation coefficients

	Quality of Life				
	AQLQ	LWAQ	SIP	RS utilities	SG utilities
Lung function					
PEFR a.m.	0.02	-0.13	-0.16*	0.13	0.05
PEFR p.m.	-0.03[#]	-0.09	-0.17*	0.11	0.06
FEV_1 % pred	0.12	-0.15	-0.17*	0.22**	-0.04[#]
FVC L	0.13	-0.26**	-0.29**	0.10	0.06
Reversibility %	-0.28**	0.17*	0.02	-0.09	0.04[#]
Symptoms					
Daytime score	-0.59**	0.42**	+0.39**	-0.26*	-0.17
Night-time score	-0.70**	0.42**	+0.38**	-0.33**	-0.21
Quality of Life					
AQLQ		-0.73**	-0.48**	0.47**	0.19*
LWAQ			0.56**	-0.43**	-0.13
SIP				-0.59**	-0.15
RS utilities					0.17*

RS: Rating Scale; SG: Standard Gamble. For further abbreviations see legend to tables 1, 2 and 4. *: $p \leq 0.01$; **: $p \leq 0.001$; [#]: paradoxical direction.

Longitudinal validity

Longitudinal construct validity correlations are given in table 7. Of all quality of life measures, the change in AQLQ correlates best with the change in lung function. The longitudinal correlations between these two measures are stronger than the cross-sectional correlations. The other quality of life measures show hardly any correlation with lung function.

The longitudinal correlation with both daytime and night-time symptom score is best for the AQLQ, followed by the rating scale utilities. The change in LWAQ correlates with change in daytime symptom score only.

The relatively high correlation between the AQLQ and the patients' overall assessments of efficacy indicates that the patients' perceptions of change were indeed reflected in changes in the quality of life as measured with the AQLQ. The other quality of life measures did not reflect the patients' global assessments of efficacy.

The change in AQLQ shows significant correlation with change in all other quality of life measures. This longitudinal correlation was strongest between the AQLQ and the rating scale utilities. The change in LWAQ correlates with the change AQLQ and the change in SIP, but not with the change in utilities. The correlation between the two utility measures is nonsignificant.

Table 7. – Longitudinal construct validity: Spearman's rank correlation coefficients between change values

	Change in Quality of Life				
Change in	AQLQ	LWAQ	SIP	RS utilities	SG utilities
Lung function					
PEFR a.m.	-0.36**	-0.04	-0.19	0.19	0.23*
PEFR p.m.	0.35**	-0.20	-0.28*	0.08	0.16
FEV_1 % pred	0.32**	-0.05	-0.11	0.19	0.09
FVC	0.28*	-0.05	-0.12	0.21	0.24*
Reversibility %	-0.21	-0.02#	0.01	-0.12	-0.03
Symptoms					
Daytime score	0.51**	-0.24*	-0.20	0.24*	0.08
Night-time score	0.54**	-0.06	-0.04	0.23*	0.14
Overall assessment of efficacy ~	0.64**	-0.20	-0.29	0.24	0.20
Quality of Life					
AQLQ		-0.29*	-0.31**	0.36**	0.29**
LWAQ			0.31**	-0.11	-0.09
SIP				-0.17	-0.18
RS utilities					0.16

RS: Rating Scale; SG: Standard Gamble. For further abbreviations see legends to table 1, 2 and 4. *: $p \leq 0.01$; **: $p \leq 0.001$; #: paradoxical direction; ~: the patient's overall assessment of efficacy is not a change score.

Discussion

Whether a quality of life questionnaire is suitable for use in clinical trials depends largely on how well it detects treatment-related changes in the impact of a disease on quality of life. Of all the quality of life measures we used in this study, only the AQLQ and the rating scale utilities reflected the greater improvement that was found in virtually all clinical outcome measures on treatment with salmeterol as compared to salbutamol. The LWAQ domain scores for sport and sex also showed a significant difference in favour of salmeterol. However, despite the presence of clear clinical differences, the overall LWAQ and its two construct scores failed to detect a difference in quality of life between the treatment groups.

The AQLQ was found to be the most sensitive to improvements in health status. The greater responsiveness of the AQLQ may result from "informed" administration [37], *i.e.* that patients read the answers they had given at baseline just before answering the questions at follow-up. This "informed answering" is a part of the administration procedure prescribed for this questionnaire, and contrasts with the procedure for the other questionnaires.

The activity domain of the AQLQ was particularly highly responsive. The responsiveness may be increased because 5 of the 8 items in this domain are "patient-specific", which means that patients themselves indicate which five most important activities should form the items. Thus, the activity domain focuses on what really matters to a patient. The most frequently identified individual activities that were impaired because of asthma were housework (*e.g.* vacuuming, scrubbing the floor) which was mentioned by 37% of the patients, walking uphill or upstairs (33%), sleeping (25%), hurrying (25%), exercising (*e.g.* running, football, ballet) (22%) and cycling (20%). The disadvantage of using individualized items may be that comparing the activity scores between patients is somewhat difficult, since they relate to different activities. Moreover, individualization is not a prerequisite for responsivenes: a recent study showed that the nonindividualized St. Georges Respiratory Questionnaire Impact score was able to detect a difference between nedocromil sodium and placebo, even though the clinical improvements due to this drug seemed to be less than the improvements achieved by salmeterol in our trial [31]. The emotional and environmental domains of the AQLQ were less responsive than the activities and symptom domains. The environmental domain examines how environmental stimuli, such as smoke, weather, perfumes and smells, affect the patient's asthma. Probably, this is unlikely to be influenced to any great extent by drugs that primarily aim at bronchodilation. The emotional domain addresses topics such as concern, frustration and fear, which may be highly related to a patient's psychological disposition, and is unlikely to change in 6 weeks.

This explains why the developers of the LWAQ predicted that the health appraisal construct of the LWAQ would be less responsive to change over time caused by drug treatment than the health knowledge construct [38]. Since the presence of problems is more closely related to morbidity, the patient's knowledge of these problems is expected to be more sensitive to changes in morbidity than the patient's emotional appraisal of these problems, which is more closely related to personality. This hypothesis is confirmed in our study. It also suggests that for a quality of life instrument to be highly responsive, it should focus on functional limitation more than on emotional dysfunction.

Although the health knowledge construct is more responsive than the health appraisal construct, it is not as responsive as the activity domain of the AQLQ, even though it covers a greater range of activity restriction. Along with the patient-specific questions of the AQLQ, this may be due to the process through which the items were reduced. The item reduction process in constructing the LWAQ was based on factor analysis, whereas the item reduction process of the AQLQ was based on patient responses. Thus, the latter includes only those activities which were frequently reported and judged important [39].

Of the two utility measurement techniques used in our study, only the rating scale measured a significantly greater improvement on salmeterol than on salbutamol. It was more responsive to change than the standard gamble. Utilities elicited *via* the standard gamble technique were the only outcomes that showed a deterioration in the salbutamol group. All other measures showed a trend indicating an improvement of quality of life in both groups, but a greater improvement in the salmeterol than in the salbutamol group. This deterioration was not a result of outlier values. It was mainly due to the fact that salbutamol patients who stated in the overall assessment that they had not changed showed a significantly larger reduction in mean utility than salmeterol patients who stated they had not changed. We do not have a good explanation for this.

That the observed trend was not significant in a number of quality of life measures does not necessarily imply that there are no real differences in the domains of these questionnaires. Since salmeterol is clinically effective, it could have led to an increased level of activity as a result of which the burden of symptoms is also increased [40]. Furthermore, the extra effects of salmeterol over salbutamol may be camouflaged by the higher use of rescue salbutamol (especially during the day) in patients given salbutamol.

This difference in the use of rescue salbutamol may also explain why there seems to be no improvement in night-time asthma due to salmeterol. Since, the percentage of nights without awakenings did not show a significant difference between the two groups. However, during both the day and the night, patients receiving salbutamol used more supplemental salbutamol than patients receiving salmeterol, although only the daytime use was significantly different between the groups. Furthermore, in both treatment groups the symptom scores were very low at baseline, thus leaving little room for improvement.

It is known that generic quality of life measures, such as the SIP are usually less responsive to small changes than disease-specific measures. Although the SIP showed the overall trend for greater improvement in salmeterol than in salbutamol, it failed to identify a significant difference between patients receiving salmeterol and patients receiving salbutamol. The latter is similar to the result of a recent study in mild to moderate asthma patients, in which nedocromil sodium led to significant improvements in asthma severity, night-time asthma, daytime bronchodilator use and the St George's Respiratory Questionnaire Impacts score, but not in the SIP score [31]. Nevertheless, the use of generic quality of life measures has the advantage that it allows comparisons across diseases and patient populations, which is of interest especially when the results of a study are presented to policy makers. Inclusion of generic quality of life measures in addition to disease-specific measures is useful in providing a broader range of domains in which changes may be detected. A new drug, such as salmeterol, may have unanticipated effects on quality of life that are not detected by disease-specific questionnaires. However, the SIP did not detect any such effects. Our patients' SIP scores were skewed towards the less impaired end of the scale, thus leaving little room for improvement. The same is true for the standard gamble, the mean score of which was already 0.87 at baseline. Patients with the best possible score can still have substantial quality of life impairment. But improvements in these patients simply cannot be detected by the instrument. Such a ceiling effect [41] decreases the responsiveness of an instrument.

Assessing the statistical significance of changes in quality of life measures is relatively easy, but a meaningful interpretation of the magnitude of the significant changes that have been found is difficult. This is because we have not yet had the opportunity to gain wide experience with these instruments, and are still unfamiliar with their units of change. By relating changes in AQLQ scores to patients' overall assessments of change, JUNIPER and co-workers [42] demonstrated that a within-subject change of 0.5 both for the overall and the AQLQ domain scores can be seen as the minimal important difference. This is the smallest difference which patients perceive as beneficial. Thus, in our study, the change in overall AQLQ score and in the activities and symptoms domain scores within the salmeterol group can be judged clinically relevant. No such studies of the minimal important differences have been carried out for the other instruments.

Not only patients given salmeterol, but also patients given salbutamol show some improvement in respiratory function and quality of life after 6 weeks of treatment. These quality of life improvements were large enough to be detected by the AQLQ, the LWAQ, and the rating scale utilities. Thus, salbutamol appears to have a positive effect on quality of life as well. To some extent this effect may result from patients receiving increased attention in a clinical trial situation compared to routine clinical practice. Furthermore, the knowledge that the effects of treatment will be carefully measured may have led to an increased compliance of patients with the drug therapy.

In the literature, evidence of responsiveness is seen both as a distinct property of a quality of life measure and as an indication of a measure's validity [43]. In the absence of a gold standard for quality of life validity can be assessed by examining the correlations between the questionnaires and measures of respiratory function or symptoms, as these are assumed to influence physical functioning and perhaps other aspects of quality of life. This kind of validity is referred to as construct validity. With respect to the longitudinal correlation with lung function and symptom scores, the AQLQ again performed best. The AQLQ showed relatively high correlation both with daytime and night-time symptom scores.

The changes in LWAQ, SIP and utilities hardly showed any correlation with changes in lung function parameters. These three instruments also failed to correlate with the patients' overall assessments of efficacy. A number of correlations had paradoxical directions. However, these correlations were all nonsignificant.

Construct validity can also be assessed by comparing the correlations among various quality of life measures. The correlations among quality of life measures were much higher than the correlations between quality of life measures and respiratory function. For example, the cross-sectional correlation between the LWAQ and the AQLQ was 0.73. This might be expected given the large degree of overlap in the content of the items which comprise the scales. However, longitudinally, the correlation between the LWAQ and the AQLQ was only 0.29. The change in AQLQ correlates better with the change in rating scale utilities than with the change in LWAQ.

Construct validity evaluation showed that rating scale utilities correlated better with symptom scores than standard gamble utilities. It also showed that rating scale utilities correlated better with other quality of life measures than standard gamble utilities. The correlation between the two types of utilities was weak. Either the standard gamble measures largely different aspects of health - in which case our validity concept is not valid - or the standard gamble has indeed lower construct validity. It may be that it chiefly measures patients' risk behaviour, since patients are asked to value their health states under risk conditions, whereas with the rating scale health states are valued under certainty.

In summary, the results of this study indicate that salmeterol tends to have a more favourable effect on quality of life than salbutamol. The AQLQ and the rating scale utilities were found to be sufficiently responsive to detect the clinically important changes due to the use of salmeterol, whereas the LWAQ, the SIP and the standard gamble utilities were not. Of the two disease-specific instruments, the AQLQ was found to be more valid than the LWAQ. Of the generic instruments, the validity of the standard gamble was worst.

Acknowledgement: The authors would like to thank M. Hyland and E. Juniper for discussion of the results.

References

1. Spilker B, Molinek F, Johnston K, *et al.* Quality of life bibliography and indexes. *Med Care* 1989; 27 (Suppl.), S77.
2. Smith ND. Quality of life studies from the perspective of an FDA reviewing statistician. *Drug Inform J* 1993; 27: 617–623.
3. McSweeny AJ, Heaton, RK, Grant I, Cugell D, Solliday N, Timms R. Chronic obstructive pulmonary disease: socioemotional adjustment and life quality. *Chest* 1980; 77: 309–311.
4. Guyatt GH, Berman LB, Townsend M, Pugsley S, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987; 42: 773–778.
5. Williams SJ, Bury MR. Impairment, disability and handicap in chronic respiratory illness. *Soc Sci Med* 1989; 29: 609–616.
6. Schrier AC, Dekker F, Kaptein AA, Dijkman JH. Quality of life in elderly patients with chronic nonspecific lung disease seen in family practice. *Chest* 1990; 89: 894–899.
7. Van Schayck CP, Rutten-van Mölken MPMH, Van Doorslaer EKA, Folgering H, Van Weel C. Two-year bronchodilator treatment in patients with mild airflow obstruction: contradictory effects on lung function and quality of life. *Chest* 1992; 102: 1384–1391.
8. Malo J-L, Boulet L-P, Dewitte J-D, *et al.* Quality of life of subjects with occupational asthma. *J Allergy Clin Immunol* 1993; 91: 1121–1127.
9. Hyland ME. The Living with Asthma Questionnaire. *Respir Med* 1991; 85 (Suppl.): 13–16.
10. Hyland ME, Finnis S, Irvine SH. A scale for assessing quality of life in adult asthma sufferers. *J Psychosom Res* 1991; 35: 99–110.
11. Marks GB, Dunn SM, Woolcock AJ. A scale for the measurement of quality of life in adults with asthma. *J Clin Epidemiol* 1992; 45: 461–472.
12. Creer TL, Wigal JK, Kotses H, McConnaughy K, Winder JA. A life activities questionnaire for adult asthma. *J Asthma* 1992; 29: 393–399.
13. 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 J Respir Dis* 1992; 145: 1321–1327.
14. Juniper EF, Guyatt GH, Epstein RS, Ferrie PJ, Jaeschke R, Hiller TK. Evaluation of impairment of health-related quality of life in asthma: development of a questionnaire for use in clinical trials. *Thorax* 1992; 47: 76–83.
15. Ullman A, Hedner J, Svedmyr N. Inhaled salmeterol and salbutamol in asthmatic patients: an evaluation of asthma symptoms and the possible development of tachyphylaxis. *Am Rev Respir Dis* 1990; 142: 571–575.
16. Britton MG, Earnshaw JS, Palmer JBD, on behalf of a European Study Group. A twelve months comparison of salmeterol with salbutamol in asthmatic patients. *Eur Respir J* 1992; 5: 1062–1067.
17. Lundback B, Rawlinson DW, Palmer JBD, on behalf of a European study group. Twelve month comparison of salmeterol and salbutamol as dry powder formulations in asthmatic patients. *Thorax* 1993; 48: 148–153.
18. Pearlman DS, Chervinsky P, LaForce G, *et al.* A comparison of salmeterol with albuterol in the treatment of mild-to-moderate asthma. *N Engl J Med* 1992; 327: 1420–1425.
19. Fitzpatrick MF, Mackay T, Driver H, Douglas NJ. Salmeterol in nocturnal asthma: a double-blind, placebo-controlled trial of a long-acting inhaled β_2-agonist. *Br Med J* 1990; 301: 1365–1368.
20. Palmer JBD, Hyland ME. Salmeterol in clinical practice: comparator and safety studies, quality of life studies. *Eur Respir J* 1991; 4: 301–303.
21. Juniper EF, Johnston P, Borkhoff C, Haukioja A. Effect of salmeterol on asthma quality of life. *Am Rev Respir Dis* 1993; A60 (Abstract).
22. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 1992; 45: 1341–1345.
23. Guillemin F, Bomhardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993; 46: 1417–1432.

24. European Group for Health Measurement and Quality of Life Assessment: Hunt SM, Alonso J, Bucquet D, Niero M, Wiklung I, McKenna S. Cross-cultural adaptation of health measures. *Health Policy* 1991; 19: 33–44.
25. Juniper EF, Guyatt GH, Ferrie PJ, Griffith E. Measuring quality of life in asthma. *Am Rev Respir Dis* 1993; 147: 832–838.
26. Rowe BH, Oxman AD. Performance of an asthma quality of life questionnaire in an outpatient setting. *Am Rev Respir Dis* 1993; 148: 675–681.
27. Hyland ME, Thompson P, Bellesis M, Kenyon CAP. Quality of life consequences of morbidity reduction in asthma: health-knowledge *versus* health-appraisal. (Submitted).
28. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981; 19: 787–805.
29. Jacobs HM, Luttik A, Touw-Often FWMM, De Melker RA. "Sickness Impact Profile". Resultaten van een valideringsonderzoek van de Nederlandse versie. *Ned Tjdschr Geneesk* 1990; 134: 1950–1954.
30. Marks GB, Dunn SM, Woolcock AJ. An evaluation of an asthma quality of life questionnaire as a measure of change in adults with asthma. *J Clin Epidemiol* 1993; 40: 1103–1111.
31. Jones PW and the Nedocromil Sodium Quality of Life Study Group. Quality of life, symptoms and pulmonary function in asthma: long-term treatment with nedocromil sodium examined in a controlled multicentre trial. *Eur Respir J* 1994; 7: 55–62.
32. Bennett K, Torrance GW, Tugwell P. Methodologic challenges in the development of utility measures of health-related quality of life in rheumatoid arthritis. *Controlled Clin Trials* 1991; 12 (Suppl.): 118–128.
33. Bakker CH, Rutten-van Mölken M, Van Doorslaer E, Bennett K, Van der Linden Sj. Health related utility measurement in rheumatology: an introduction. *Patient Ed Counseling* 1993; 20: 145–152.
34. Bakker C, Rutten M, Van Doorslaer E, Bennett K, Van der Linden Sj. Feasibility of utility assessment by rating scale and standard gamble in patients with ankylosing spondylitis or fibromyalgia. *J Rheumatol* 1994; 21: 269–274.
35. Torrance GW. Measurement of health-state utilities for economic appraisal: a review. *J Health Econ* 1986; 5: 1–30.
36. Deyo RA, Diehr P, Patrick D. Reproducibility and responsiveness of health status measures: statistics and strategies for evaluation. *Controlled Clin Trials* 1991; 12 (Suppl.): 142–158.
37. Guyatt GH, Berman LB, Townsend M, Taylor DW. Should study subjects see their previous responses? *J Chronic Dis* 1985; 38: 1003–1007.
38. Hyland ME. The validity of health assessments: resolving some recent differences. *J Clin Epidemiol* 1993; 46: 1019–1023.
39. Hyland ME. Quality-of-life assessment in respiratory disease. *Pharmacoeconomics* 1992; 2: 43–53.
40. Hughes MD, Van Essen-Zandvliet EEM, Kerstjens H. Issues in the design of long-term clinical trials in asthma. *Drug Inform J* 1993; 27: 793–803.
41. Guyatt GH, Feeny DH, Patrick DL. Measuring health related quality of life. *Ann Intern Med* l993; 118, 622–629.
42. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol* 1994; 47: 81–87.
43. Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Quality Life Res* 1992; 1: 73–75.