

Original article

Can the standard gamble and rating scale be used to measure quality of life in rhinoconjunctivitis? Comparison with the RQLQ and SF-36

Background: With interest in health economics growing, it is important to know whether utilities may be used to measure health-related quality of life in patients with rhinoconjunctivitis. The objective was to compare the validity and measurement properties of disease-specific versions of the standard gamble and rating scale with those of the Rhinoconjunctivitis Quality of Life Questionnaire (RQLQ) and the Short-Form 36 (SF-36).

Methods: One hundred adults with symptomatic rhinoconjunctivitis participated in a 5 week observational study, completing the standard gamble, rating scale, RQLQ and SF-36 at baseline and after 1 and 5 weeks. Symptom diaries were completed for 1 week before each follow-up visit.

Results: Reliability was highest for the RQLQ (intraclass correlation coefficient = 0.97), followed by the rating scale (0.75), the SF-36 physical (0.75), the SF-36 mental (0.74) and the standard gamble (0.12). The responsiveness index was highest for the RQLQ (0.76), followed by the rating scale (0.56) and the SF-36 mental (0.28). Both cross-sectional and longitudinal validity were strongest for the RQLQ and the rating scale.

Conclusions Both the rating scale and the RQLQ have strong evaluative and discriminative properties. The SF-36 has acceptable discriminative properties but its evaluative properties are poor. All measurement properties for the standard gamble are inadequate. Poor correlation between the standard gamble and the rating scale indicates that utilities cannot be derived from rating scale data.

**E. F. Juniper,¹ A. K. Thompson,²
J. N. Roberts³**

¹Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton Ontario, Canada;

²Aventis Pharmaceuticals, Bridgewater, New Jersey, USA; ³University of Western Ontario, London, Ontario, Canada

Key words: conjunctivitis; health status; measurement; quality of life; questionnaires; rhinitis; rhinoconjunctivitis; utilities.

Professor Elizabeth Juniper MCSP, MSc, 20 Marcuse Fields, Bosham, West Sussex, PO18 8NA, UK

Accepted for publication 2 October 2001

Health-related quality of life (HRQL) has been defined as 'the functional effect of an illness and its consequent therapy on a patient, as perceived by the patient' (1). There is no 'best' instrument for measuring HRQL in patients with rhinoconjunctivitis because each approach has its strengths and weaknesses. Generic health profiles, such as the SF-36 (2), have been designed to measure the burden of illness experienced by patients with a wide range of medical conditions and can be used to compare the functional impairments caused by different diseases. However, generic questionnaires often lack sufficient depth to capture the small but important changes that may occur in a particular disease either spontaneously or as the result of an intervention (3). This lack of responsiveness in generic questionnaires has led to the development of disease-specific questionnaires, such as the RQLQ (4), which focus on the functional impairments associated with a single illness.

The third type of HRQL instrument is the utilities which estimate the value that patients and society place on various health states and provide a single number between 0 (death) and 1 (perfect health). This approach is founded in modern utility theory, a normative rational model of decision under uncertainty (5). Utilities are used by health economists for cost/utility analyses and to derive Quality Adjusted Life Years (QALYs) (6) which are used by funding agencies to guide the allocation of health care resources.

The standard gamble is a true utility and measures the value that patients place on their own health state. The technique was developed by von Neumann and Morgenstern (5) and adapted for clinical use by Torrance (6). The rating scale (feeling thermometer) (6) measures patients' perception of their own health state but it can not be considered a true utility because it does not meet the utility theory requirement of 'decision under uncertainty'. Nevertheless, it is being used with increasing frequency in clinical studies because it is much easier than the standard gamble both to administer and to understand.

In recent studies, we modified the original generic versions of the standard gamble and the rating scale to be disease-specific for adults and children with asthma (7, 8). In the present study, we have modified both the standard gamble and the rating scale to be specific for adults with rhinoconjunctivitis and compared their measurement properties with those of the RQLQ and the SF-36.

Recognizing the weakness of the measurement properties of the standard gamble in some medical conditions, Torrance has suggested that there is a power curve relationship between standard gamble and rating scale so that utilities may be interpolated from rating scale data (9). In this study, we have examined the relationship between the standard gamble and the rating scale to determine whether utilities can be interpolated in this way in rhinoconjunctivitis.

Material and methods

Subjects

One hundred adults (18–75 years) with current symptoms of rhinoconjunctivitis participated in the study. Patients were enrolled from participants in previous research studies and respondents to notices in the local media. Symptoms were required to be sufficiently troublesome to require additional medication and to score 0.5 or more on either the nasal or eye symptom domains of the RQLQ. Patients were not permitted to have nasal or sinus infections requiring treatment with antibiotics or to have any illness, other than rhinoconjunctivitis, that had an impact on health-related quality of life. They were required to understand the study instructions and to be able to complete a daily symptom diary. All patients signed an informed consent. The study was approved by the McMaster University Faculty of Health Sciences Ethics Committee.

Study design

The study was a 5 week observational study with subjects assessed in the clinic at baseline and after 1 and 5 weeks. The standard gamble, rating scale, self-administered version of the RQLQ and the SF36 were completed at each visit. For 1 week before each follow-up visit, patients recorded rhinoconjunctivitis symptoms and medication use in a diary. A global rating of change questionnaire was completed at the end of weeks 1 and 5.

At enrollment, patients were asked to tolerate 1 week without any additional medications so that we could make an estimate of the reliability (reproducibility) of the instruments while the patients were in a stable clinical state. Those who refused were permitted to increase their usual rhinoconjunctivitis medication. After 1 week, all patients were advised to increase their rhinoconjunctivitis medications in the manner recommended by their clinician. Those who did not have a treatment plan were offered fexofenadine (Allegra, Aventis Pharmaceuticals Inc., Bridgewater, NJ, USA) 60 mg when needed up to 120 mg daily. Those who did not like using a non-sedating antihistamine and preferred a nasal steroid were offered fluticasone propionate nasal spray (Flonase, GlaxoSmithKline, Ontario, Canada) two puffs in each nostril up to twice/day. Patients who refused additional rhinoconjunctivitis medication continued on their previously established medications throughout the study.

Outcome measures

Rating scale (6) The rating scale is a preference instrument and

explores how patients feel about their own health state. It looks like a thermometer with clearly defined end points: 0 = least preferred health state (death) and 100 = most preferred health state (perfect health). In this study, patients first read aloud a description of three hypothetical health states which had been derived from the RQLQ:

1) 'You have rhinoconjunctivitis (nose and eye allergies/hayfever) but it bothers you very little. You have mild to moderate nose and eye symptoms but they rarely interfere with your day-to-day activities. Occasionally they prevent you from getting a good night's sleep and you may feel a bit tired during the day. Your symptoms hardly ever make you feel frustrated or irritable.'

2) 'You have rhinoconjunctivitis and it bothers you quite a bit. You have fairly troublesome nose and eye symptoms quite a lot of the time and these sometimes cause you to have a headache. Your symptoms frequently interfere with your day-to-day activities. Often they prevent you from getting a good night's sleep. You sometimes feel tired and have difficulty concentrating. You get frustrated and feel irritable because of your symptoms.'

3) 'You have severe rhinoconjunctivitis and it is very troublesome. Your nose and eye symptoms are extremely bothersome and you often have a bad headache. Your day-to-day activities are very restricted and it is always difficult to sleep at night. You feel very tired during the day and it is very difficult to concentrate. This makes you feel very frustrated and you are very irritable.'

Patients were asked to place three markers on the thermometer to reflect their feelings about each state. We then asked patients to think about how their own rhinoconjunctivitis had been during the previous week and to place a marker on the thermometer to reflect their feelings about this.

Standard Gamble (6) This utility instrument measures the value that patients place on their own health state. Initially, patients are asked to think about a particular health state and then to consider whether they would prefer to remain in that health state for the next 10 years or take a chance with a new (imaginary) treatment. They are told that the new treatment has the ability to return them to perfect health immediately with no side-effects but also has the ability to cause instant death. Initially, the probability of returning to perfect health if they took the new treatment is set at 100% with absolutely no chance of death. Usually, all those who understand the concept choose to take the new treatment rather than stay in their present health state. The probability of returning to perfect health on taking the new treatment is then gradually reduced (and the chance of death increased) until the patient decides to remain in their current health state rather than take a chance on the new treatment. The indifference point represents the value that the patient places on that health state. In this study, patients first completed the standard gamble with each of the three hypothetical rhinoconjunctivitis states used for the rating scale and then it was administered for how their own rhinoconjunctivitis had been during the previous week.

Rhinoconjunctivitis Quality of Life Questionnaire (4) This self-administered, 28-item questionnaire has been validated to measure the functional impact of rhinoconjunctivitis in seven domains (activity limitation, sleep, non nose/eye symptoms, practical problems, nose symptoms, eye symptoms and emotional function). Patients score their experiences during the previous week on a 7-point scale (0=no impairment, 6 = severe impairment).

The Medical Outcomes Study Short Form 36 (SF-36) (2) This 36-item, self-administered generic health profile questionnaire provides summary scores in two domains, physical health and mental health. Good internal consistency and cross-sectional validity have been demonstrated for patients with rhinitis (10).

Global Rating of Change Questionnaire (10) The global rating of change questionnaire used in this study asked patients about change in their rhinoconjunctivitis quality of life since their previous clinic visit. The questionnaire has a 15 point scale (+7 = very much better, 0 = no change and -7 = very much worse).

Clinical measures of rhinoconjunctivitis For 1 week before each follow-up clinic visit, patients completed a diary of their symptoms and medication requirements. Each morning and each evening they scored the severity (0 = nil, 1 = mild, 2 = moderate, 3 = severe) and frequency (0 = nil, 1 = a few short episodes, 2 = frequent episodes, 3 = continuous) of: sneezing; stuffy nose; runny nose; itchy nose; and eye symptoms. Results are expressed as mean scores for nose and eye symptoms (0 = asymptomatic, 3 = very symptomatic). In addition, each evening patients recorded the amount of rhinoconjunctivitis medication used during the previous 24 h.

Statistical analysis

The objective was to compare the measurement properties (discriminative: reliability and cross-sectional construct validity; and evaluative: responsiveness and longitudinal construct validity) (11) of the four instruments.

The estimate of reliability required defining a group of patients whose rhinoconjunctivitis did not change between consecutive clinic visits (0–1 and 1–5 weeks). Patients were included in the reliability analysis if they scored -1, 0 or +1 on the global rating of change questionnaire completed at weeks 1 and 5. If a patient's clinical state was stable during both time periods (0–1 and 1–5 weeks), a single point was blindly selected using a random number generator. Test-retest reliability has been estimated as the within-subject standard deviation and related to the overall standard deviation as an intraclass correlation coefficient (ICC). This statistic also provides evidence of the instrument's ability to discriminate between patients of different levels of impairment.

Responsiveness was determined in three ways. First, in the patients who experienced a change in their rhinoconjunctivitis between weeks 1 and 5, we examined whether the instruments could detect change using a paired *t*-test. Patients who either improved or deteriorated by a score of 2 or greater on the global rating completed at week 5 were included. Second, we determined whether the instruments could detect a difference between these patients who changed during this time period and those patients who remained stable using an unpaired *t*-test. Third, we calculated a responsiveness index ($\Delta/\text{sd}\Delta$) for each instrument (12). Difference in responsiveness indices between instruments was tested using a paired *t*-test.

Both cross-sectional and longitudinal construct validity were first evaluated by correlating the four instruments with each other. Our hypothesis was that the validity of the RQLQ, as both an evaluative and discriminative instrument, has already been established (3) and that if the SF-36, rating scale and standard gamble measure health-related quality of life in rhinoconjunctivitis, these instruments should correlate moderately to strongly with the RQLQ. In addition, we examined the correlations between the four instruments and other measures of clinical status (diary symptoms).

Results

All 100 patients completed the study. There were 47 males and 53 females and their mean age was 39 years ($\text{sd} = 11.95$). The mean overall RQLQ score at enrollment was 2.58 ($\text{sd} = 1.16$).

Measurement properties

Reliability Forty-four patients remained stable bet-

Table 1. Reliability ($n = 44$)

Instrument	Within subject SD	Between subject SD	ICC
RQLQ	0.22	1.24	0.97
SF-36 Physical	4.05	7.00	0.75
SF-36 Mental	4.24	7.11	0.74
Rating scale	9.96	17.21	0.75
Standard gamble	0.10	0.04	0.12

ween clinic visits and contributed data to the reliability analysis. The within-subject standard deviations and intraclass correlation coefficients for each of the instruments are shown in Table 1. Highest reliability was shown by the RQLQ (ICC = 0.97), followed by the SF-36 Physical (ICC = 0.75) and the rating scale (ICC = 0.75).

Responsiveness Eighty-three patients experienced changes of 2 or greater on the global rating of change completed at week 5 and thus contributed data to the unstable category in the responsiveness analysis. Seventeen patients scored less than 2 on the global rating and were placed in the stable category. The RQLQ and rating scale showed the greatest responsiveness to within-patient changes in the unstable group ($P < 0.001$) followed by the SF-36 mental and standard gamble ($P < 0.05$) (Table 2). The SF-36 physical failed to detect these changes ($P > 0.05$). Only the RQLQ and rating scales were able to detect the differences between the stable and unstable groups ($P = 0.003$ and $P = 0.03$, respectively). The responsiveness index was highest for the RQLQ and statistically different from the other instruments ($P < 0.01$) (Table 3).

Validity Correlation with the RQLQ was highest for the rating scale ($r = 0.60$) followed by the SF-36 mental ($r = 0.45$) (Table 4). Both cross-sectional and longitudinal correlations between the instruments and diary data showed the highest correlations for the RQLQ, followed by the rating scale (Tables 5, 6). Although the SF-36 physical showed modest cross sectional correlation with diary data ($r = 0.24$), longitudinal correlations were poor ($r = 0.14$).

Table 2. Responsiveness: change in score between consecutive visits

Instrument	Stable group mean $\pm$ sd ($n = 17$)	Unstable group mean $\pm$ sd ($n = 83$)	Difference <i>P</i> -value	Responsiveness index
RQLQ	0.00 $\pm$ 0.36	1.13 $\pm$ 1.50**	0.003	0.76
SF-36 Physical	0.04 $\pm$ 4.17	0.18 $\pm$ 6.51	0.93	0.03
SF-36 Mental	2.10 $\pm$ 4.11	2.38 $\pm$ 8.43*	0.89	0.28
Rating scale	0.24 $\pm$ 16.34	12.69 $\pm$ 22.58**	0.03	0.56
Standard gamble	0.006 $\pm$ 0.03	0.03 $\pm$ 0.13*	0.43	0.25

* $P < 0.05$

** $P < 0.001$

Table 3. Statistical comparison of responsiveness indices (*P*-values)

Instrument	RQLQ (0.76)	Rating scale (0.56)	SF-36 Mental (0.28)	Standard gamble (0.25)
Rating scale (0.56)	<0.001			
SF-36 Mental (0.28)	<0.001	0.059		
Standard gamble (0.25)	<0.001	0.014	0.813	
SF-36 Physical (0.03)	<0.001	<0.001	0.163	0.025

Discussion

The measurement properties of the RQLQ are very similar to those demonstrated previously (4, 13) and confirm that the instrument is reliable and responsive and is valid for use in both clinical trials and epidemiological surveys.

The generic SF-36 health profile is used extensively in a wide range of medical conditions and has been shown to be a good instrument for measuring burden of illness and for detecting problems not usually associated with a particular disease. In this study we have shown that the SF-36 has acceptable discriminative properties (reliability and cross-sectional validity), thus providing evidence that in cross-sectional evaluations and surveys the instrument can detect differences between patients of different levels of impairment. These results are in keeping with those of a previous study in which the instrument was shown to have good internal consistency and acceptable cross-sectional validity (14). However, the evaluative properties (responsiveness and longitudinal validity) of the SF-36 in this study were poor. This indicates that the instrument is relatively insensitive to changes that occur within patients over time either as a result of treatment or natural fluctuations in the disease. Although patients in the unstable group experienced clinically important changes in their RQLQ scores (mean = 1.13, $P=0.003$) (minimal important difference = 0.5), the SF-36 failed to detect these changes and also failed to detect the difference between the stable and unstable groups. There is no doubt that if a treatment effect is large and there is a big sample size, the SF-36 will detect this effect (15) but there is the risk in clinical trials that the SF-36 may miss small but clinically important treatment effects.

Although the properties of the disease-specific rating scale tended to be weaker than those of the RQLQ, they are still very acceptable and indicate that

Table 4. Pearson correlation coefficients between instruments at the end of week 1 ($n=100$)

Instrument	SF-36 Physical	SF-36 Mental	Rating Scale	Standard Gamble
RQLQ	0.35	0.45	0.60	0.13
SF-36 Physical		0.07	0.29	0.20
SF-36 Mental			0.26	0.04
Rating scale				0.42

Table 5. Cross-sectional construct validity (Pearson correlation coefficients at the end of week 1)*

Instrument	RQLQ	SF-36 Physical	SF-36 Mental	Rating scale	Standard gamble
Nasal diary	0.69	0.24	0.18	0.32	0.09
Eye diary	0.61	0.18	0.20	0.39	0.08

* Expected correlations based on data from previous studies: $r=0.40-0.65$ (4, 20)

this instrument is valid for use in rhinoconjunctivitis. In addition, the correlation between the RQLQ and the rating scale ($r=0.60$) indicates that the two instruments are probably measuring the same construct (concept). Although the rating scale is not a utility and cannot be used for cost/utility analyses or for calculating QALYs, this study raises the question as to whether it can be used as an alternative to the RQLQ for measuring rhinoconjunctivitis specific quality of life. The rating scale yields a single number between 0 (death) and 100 (perfect health). It is also quick and easy to administer, though probably not as quick as the Mini Rhinoconjunctivitis Quality of Life Questionnaire (14). However, the RQLQ has stronger measurement properties, is self-administered (the rating scale requires a trained administrator), allows the different domains of impairment to be examined and enables identification of an individual patient's problems.

The standard gamble showed poor measurement properties and therefore cannot be considered a valid instrument for measuring quality of life in patients with rhinoconjunctivitis either in clinical trials or in cross-sectional surveys. Although the change in score in the unstable group just reached statistical significance (Table 2), it was uncorrelated with the diary data (Table 6) suggesting that the standard gamble may have detected something unrelated to rhinoconjunctivitis or the change may have been significant by chance alone. Since the standard gamble asks patients about their willingness to take a risk of dying in order to have complete recovery and most patients with seasonal allergic rhinoconjunctivitis will have a spontaneous and assured recovery at the end of the pollen season, it is not surprising that the standard gamble does not perform well in this condition.

It has been suggested that in some medical conditions, there is a curvilinear relationship between the standard gamble and the rating scale, thus allowing the interpolation of utilities from rating scale data (9). In this study, data were very poorly distributed with 82% of patients scoring at the two

Table 6. Longitudinal Construct Validity (Pearson correlation coefficients)*

Instrument	ΔRQLQ	ΔSF-36 Physical	ΔSF-36 Mental	ΔRating scale	ΔStandard gamble
ΔNasal diary	0.65	0.14	0.13	0.47	0.09
ΔEye diary	0.67	0.05	0.22	0.52	0.27

* Expected correlations based on data from previous studies: $r=0.40-0.65$ (4, 20)

highest levels of the standard gamble. This very poor distribution prohibits power curve fitting and indicates that utilities should not be interpolated from rating scale data in rhinoconjunctivitis.

With the increased interest in health economics, this lack of validity of the standard gamble in rhinoconjunctivitis is an important concern. The only alternative method for deriving utilities is from generic health profile questionnaires like the Multi-attribute Health Utilities Index (HUI) (16) and the EuroQol (17), which yield societal utilities by applying an algorithm (derived usually using the standard gamble or the time trade-off) to the health profile score. If one considers that the SF-36 is a reasonable representative of generic health profiles, the poor evaluative and modest discriminative measurement properties of the SF-36 shown in this study suggest that weighted health profiles, such as the HUI (16) and EuroQol (17), may not be a valid method for determining utilities in rhinoconjunctivitis. This hypothesis is supported by the poor performance of the HUI in children with asthma (7). Therefore, for pharmacoeconomic evaluations in rhinoconjunctivitis clinical trials, it may be more appropriate to forget about utilities and instead derive comparative costs from the number-needed-to-treat (this is the number of patients that need to be treated with an intervention for one patient to have a clinically important improvement in their quality of life over and above

that which would have experienced if the control treatment had been taken) (18). Such an approach is already being investigated for clinical trials in asthma.

There is currently a trend for governments and other resource allocation agencies to use utilities to determine which medical conditions and interventions should receive funding. This approach may be appropriate for chronic and life-threatening conditions but the results of this study suggest that it is invalid for rhinoconjunctivitis which causes considerable distress to patients (14) but which is not life-threatening.

In conclusion, the results of this study show that both the RQLQ and the disease-specific rating scale have strong discriminative and evaluative properties for measuring rhinoconjunctivitis-specific quality of life. The SF-36 has acceptable discriminative properties but its evaluative properties are poor. All the measurement properties of the disease-specific standard gamble are inadequate. The weakness of the relationship between the standard gamble and the rating scale and the poor distribution of standard gamble data indicate that utilities should not be interpolated from rating scale data in rhinoconjunctivitis.

Statement

This manuscript has been prepared in accordance with the recommendations for nomenclature made by the EAACI (19).

References

- SCHIPPER H, CLINCH J, POWELL V. Definitions and conceptual issues. In: SPILKER, B, , ed. *Quality of Life and Pharmacoeconomics in Clinical Trials*. Lippincott-Raven Publishers, Philadelphia, 1996, 11–23.
- STEWART AL, HAYS R, WARE JE. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care* 1988;**26**:724–732.
- RUTTEN-VAN MOLKEN MPMH, CLUSTERS F, VAN DOORSLAER EKA et al. Comparison of performance of four instruments in evaluating the effects of salmeterol on asthma quality of life. *Eur Respir J* 1995;**8**:888–898.
- JUNIPER EF, GUYATT GH. Development and testing of a new measure of health status for clinical trials in rhinoconjunctivitis. *Clin Exper Allergy* 1991;**21**:77–83.
- VON NEUMANN J, MORGENSTERN O. *Theory of Games and Economic Behavior*. Wiley, New York, 1953.
- TORRANCE GW. Measurement of health state utilities for economic appraisal. *J Health Econom* 1986;**5**:1–30.
- JUNIPER EF, GUYATT GH, FEENY DH, GRIFFITH LE, FERRIE PJ. Minimum skills required by children to complete health-related quality of life instruments. comparison of instruments for measuring asthma-specific quality of life. *Eur Respir J* 1997;**10**:2285–2294.
- JUNIPER EF, NORMAN GR, COX FM, ROBERTS JN. Comparison of the standard gamble, rating scale, AQLQ and SF-36 for measuring quality of life in asthma. *Eur Respir J* 2001;**18**:38–44.
- TORRANCE GW. Utility approach to measuring health-related quality of life. *J Chron Dis* 1987;**40**:593–600.
- JUNIPER EF, GUYATT GH, GRIFFITH LE, FERRIE PJ. Interpretation of rhinoconjunctivitis quality of life questionnaire data. *J Allergy Clin Immunol* 1996;**98**:843–845.
- GUYATT GH, KIRSHNER B, JAESCHKE R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 1992;**45**:1341–1345.
- GUYATT GH, WALTER S, NORMAN G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis* 1987;**40**:171–178.
- JUNIPER EF, THOMPSON AK, FERRIE PJ, ROBERTS JN. Development and validation of the Mini Rhinoconjunctivitis Quality of Life Questionnaire. *Clin Exper Allergy* 2000;**30**:132–140.
- BOUSQUET J, BULLINGER M, FAYOL C, MARQUIS P, VALENTIN B, BURTIN B. Assessment of quality of life in patients with perennial rhinitis with the French version of the SF-36 health status questionnaire. *J Allergy Clin Immunol* 1994;**94**:182–188.

15. BOUSQUET J, DUCHATEAU J, PIGNAT JC, et al. Improvement of quality of life by treatment with citirizine in patients with perennial allergic rhinitis as determined by a French version of the SF-36 questionnaire. *J Allergy Clin Immunol* 1996;**98**:309–316.
16. FEENY D, FURLONG W, BARR RD, TORRANCE GW, ROSENBAUM P, WEITZMAN S. A comprehensive multi-attribute system for classifying the health status of survivors of childhood cancer. *J Clin Oncol* 1992;**10**:923–928.
17. THE EUROQOL GROUP. A new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208.
18. GUYATT GH, JUNIPER EF, WALTER SD, GRIFFITH LE, GOLDSTEIN RS. Interpreting treatment effects in randomised trials. *Br Med J* 1998;**316**:690–693.
19. JOHANSSON SGO, HOURIHANE JO'B, BOUSQUET J et al. A revised nomenclature for allergy. An EAACI position statement from the EAACI nomenclature task force. *Allergy* 2001;**56**:813–824.
20. DE GRAFF-INT VELD T, KOENDERS IM, GARRELDs IM, GERTH VAN WIJK R. The relationships between nasal hyperreactivity, quality of life, and nasal symptoms in patients with perennial allergic rhinitis. *J Allergy Clin Immunol* 1996;**98**(3):508–513.