

Measuring health status in chronic airflow limitation

G. Guyatt

Measuring health status in chronic airflow limitation. G. Guyatt.

ABSTRACT: A health status instrument for use in clinical trials must be valid (measuring what it is supposed to measure) and responsive (able to detect clinically important change). Approaches to measuring health status in clinical trials include using a battery of instruments, a general instrument which provides a profile of the patient's health, an instrument that generates a health utility, or an instrument that focuses on the problems associated with a particular disease. Disease-specific instruments have been used in clinical trials in chronic airflow limitation (CAL). The Oxygen Cost Diagram is simple and easy to administer, but responsiveness and validity are unproven. The Transition Dyspnoea Index is valid and responsive, but is difficult to use in trials in which multiple measurements are desired. The Chronic Respiratory Disease Questionnaire has proved valid and responsive in controlled trials in CAL patients. Health status measures should be included in all clinical trials in CAL.

Eur Respir J., 1988, 1, 560-564.

Departments of Medicine and Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario.

Correspondence: Dr. G. Guyatt, Dept of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre, Room 3H7, 1200 Main Street West, Hamilton, Ontario, Canada, L8N 3Z5.

Keywords: Quality of life; health status; obstructive lung disease; validity; responsiveness.

Received February 15, 1988; accepted February 19, 1988.

This work was supported in part by the Ontario Ministry of Health. Dr. Guyatt is a Career Scientist of the Ontario Ministry of Health.

Measuring health status in CAL

During the last decade, the importance of measuring subjective aspects of health status in patients with chronic airflow limitation (CAL) has become increasingly recognized. The term "quality of life" has appeared as a label for the measurement of physical and emotional (as opposed to physiological) function [1]. Quality of life is influenced by many factors other than health (*e.g.* income, job satisfaction, social opportunities); clinicians are interested in "health-related quality of life". In the present discussion the term "health status" will be used to refer to the wide variety of subjective experiences (including symptoms, physical function, and emotional function) which are related to health.

This report will be organized into a number of sections, firstly, a discussion of the necessary attributes of a health status measurement instrument, secondly, a review of the approaches available and finally, a more detailed consideration of approaches specific to chronic airflow limitation, including our work in the area.

Necessary attributes of a health status measurement instrument

There are two essential attributes necessary for a health status instrument for use in clinical trials: validity and responsiveness. An instrument is valid if it

measures what it is supposed to measure [2, 3]. Since there is no gold standard measure for health status in CAL, the validity of an index or instrument is established by comparing it with other questionnaires examining similar attributes, with the results of physiological tests, and with ratings made by clinicians and relatives [4]. Many simple questionnaires used in clinical trials rely on face validity: intuitively, the questions appear to relate to aspects of health status. Unfortunately, it is difficult to interpret the results of such *ad hoc* instruments. For example, questionnaires asking patients if their function improved after a rehabilitation programme may be measuring satisfaction with the programme, rather than health status. The validity of a questionnaire must be established before it can be applied as a meaningful outcome measure in clinical trials.

The second important property is responsiveness. Responsiveness (or sensitivity to change) refers to the instrument's ability to detect clinically important change, which is determined by two properties [5]. Firstly, to be responsive a questionnaire must yield more or less the same scores when subjects are stable, *i.e.* it should be reproducible. Secondly, it must register changes in score when subjects' health status improves or deteriorates; this property can be called changeability. If an instrument's responsiveness is unproved, and a controlled trial in which the instrument is used is negative, there remain two interpretations. Firstly, treatment doesn't work; secondly, the instru-

ment is not responsive. For example, ALEXANDER *et al.* [6] used an *ad hoc* measure of symptoms in a controlled trial of theophylline in CAL, and failed to detect a difference between periods on and off the drug. Subsequent studies [7, 8] have shown that theophylline does improve health status in CAL, and the negative result of ALEXANDER *et al.* was probably due to the unresponsiveness of their measure.

Approaches to measuring health status in clinical trials

Multiple attributes, multiple measures

There are many aspects of peoples' health status which one may want to measure. The World Health Organization has identified physical, emotional, and social aspects of health [9]. Within the dimension of physical function, one can identify components *e.g.* mobility, physical activity, self care, and role performance [10]. In the multiple attributes, multiple measures approach, one uses a different instrument to evaluate each aspect of health status separately.

This approach has not been used extensively in patients with CAL. It is comprehensive, but its limitations include the need to find a valid, responsive instrument for every attribute one wishes to measure, and the possibility that only some of the instruments chosen will show differences between the treatments under investigation. Unless one of the instruments has been designated as the primary measure of outcome before the trial started, different results in different measures may make interpretation difficult.

Health profiles

Health profiles are single instruments which measure different aspects of health status. They differ from the multiple attributes, multiple measures approach in that they can be aggregated into a single score, or at least a small number of scores. A second difference is that they are general measures designed for use in a wide variety of conditions. One of the most popular health profiles, the Sickness Impact Profile [11], was used in the Nocturnal Oxygen Therapy Trial [12]. Interest in the effect of the intervention on health status was reduced by the mortality benefit found in patients prescribed round the clock oxygen.

Health profiles offer a number of advantages to the clinical investigator. Their reproducibility and validity have been established, often in a variety of populations. They allow determination of the effects of the intervention on different aspects of health status without necessitating the use of multiple instruments (thus saving the time of both the investigator and the patient). Since they are designed for a wide variety of conditions, one can potentially compare the effects on health status of different interventions in different diseases.

Health profiles also have limitations. They may not

focus adequately on the aspects of health status of specific interest to the investigator. For example, there may be few questions relating specifically to the major problems of patients with CAL. Inadequate focus on CAL patients' symptoms is likely to produce an unresponsive instrument which may miss small but still clinically important changes in health status [13, 14].

Utility measurement

Utility measurement refers to any strategy which attempts to quantitate health status as a single number along a continuum from death to full health. Most utility measures give death a value of 0 and full health a value of 1.0. Use of utility measures in clinical trials requires serial measurement of the utility of the patient's health status throughout the study. One utility measure, the Quality of Well-Being Scale [15, 16] has been used in a controlled trial of a compliance enhancing manoeuvre in patients with CAL undergoing rehabilitation. TOEVS, KAPLAN, and ATKINS [17] showed that a programme designed to improve compliance with an exercise programme in CAL could improve health status.

The major advantage of utility measurement is its amenability to cost-utility analysis. In cost-utility analysis the cost of an intervention is related to the number of quality adjusted life years (QUALY) gained through application of the intervention. For example, in the study described above, the cost-utility analysis showed that the cost of the programme was 24,256 \$ for each additional well-year or QUALY gained [17].

Utility measurement also has limitations. Utilities vary depending on how they are obtained, raising questions of the validity of any single measurement [18]. Utilities do not allow the investigator to determine which aspects of health status are responsible for changes in utility. Finally, utilities at least potentially share the disadvantage of health profiles in that they may not be responsive to small, but still clinically important, changes.

Disease specific instruments

A final alternative is to focus on aspects of health status which are specific to the disease, and the treatment, which is being studied [4]. The rationale for this approach lies in the increased responsiveness which may result from including only important aspects of health status which are affected by the underlying disease. The instrument may even focus on problems which are specific to the individual patient [19].

In addition to the likelihood of improved responsiveness, disease specific measures have the advantage of relating closely to areas routinely explored by the physician. The disadvantages of disease specific measures is that they are (deliberately) not comprehensive, and cannot be used to compare across conditions or, at times, even across programmes. Their potential responsive-

ness, and the fact that they make sense to clinicians, has led to their widespread use in trials of patients with CAL.

Use of disease-specific instruments in CAL

At least three instruments designed specifically for patients with CAL have been used as outcomes in clinical trials: the Oxygen Cost Diagram (OCD) [20], the Transition Dyspnoea Index (TDI) [21], and the Chronic Respiratory Disease Questionnaire (CRQ) [19].

The Oxygen Cost Diagram (OCD). Patients are asked to indicate the extent of their dyspnoea by making a mark on a 10 cm line on which activities are written at intervals which correspond to the metabolic equivalents required to carry them out [20]. The OCD has been shown to have moderate correlations with the 12 min walk test, and with spirometry, in CAL patients. EATON *et al.* [22] conducted a double-blind cross-over trial of placebo, low and high-dose theophylline in CAL patients. These investigators found that there were significant differences between active and placebo periods in forced expiratory volume in one second (FEV₁) and vital capacity (VC) but only trends in the OCD, the 12 min walk, and progressive cycle ergometry. The fact that these trends did not reach statistical significance may have been due to the sample size, which was only fourteen patients. In summary, there is some evidence for the validity of the OCD, but the extent to which it is responsive is not yet clear.

The Transition Dyspnoea Index (TDI). A health worker judges whether patients with CAL have had changes in their functional impairment, the magnitude of tasks that evoke dyspnoea, and the associated magnitude of effort. While a basic structure is provided, the details of how the interview is to be conducted is left to the health worker. In one investigation, the TDI showed excellent agreement between observers, but bore little correlation with changes in spirometry or changes in the 12 min walk test distance [21]. However, in a randomized cross-over trial of theophylline in twelve patients with CAL the TDI distinguished between theophylline and placebo periods, whereas spirometry did not [8]. We found the TDI highly responsive to improvement following a respiratory rehabilitation programme [19]. Thus, the TDI appears responsive but its validity remains to be confirmed.

The TDI has other disadvantages. The results depend on interviewers' questioning and their interpretation of patients' responses. In addition, administration of the questionnaire, and interpretation, are difficult if serial measurements of function are required. Finally, both the OCD and the TDI are concerned exclusively with dyspnoea and do not attempt to measure other aspects of subjective health status.

The Chronic Respiratory Disease Questionnaire (CRQ). Because of limitations in existing instruments, we de-

veloped a new disease-specific measure of health status for CAL. The approach we used to develop the questionnaire followed a plan which can be applied to any disease-specific health status measure [4]. The development and testing of the questionnaire is described in detail in another publication [19]. We began by constructing a list of items likely to be important to patients with CAL. Our Item Selection Questionnaire contained 123 items, of which 62 dealt primarily with physical function, and 32 with emotional function. The other items dealt with areas in which there is an overlap between physical and emotional function, such as dyspnoea-induced fear and panic, with problems in concentration and reasoning, and with problems of inconvenience, such as the need to avoid smoky rooms. The item selection questionnaire was administered to one hundred patients chosen at random from among those seen in the previous year at a regional respiratory referral centre which provides secondary and tertiary care. Patients were included if their FEV₁ was consistently less than 70% of predicted.

Patients were initially asked to volunteer physical and emotional problems they experienced as a result of their lung disease. They were then asked specifically if the 123 items represented ways in which their lives were effected by their breathing problem. Patients rated the importance of each affirmatively answered item on a five point Likert scale (extremely important, quite important...not very important).

The results of these interviews are presented in detail in another publication [23]. In summary, we found that the items chosen most frequently and rated most important by the subjects fell into four dimensions: shortness of breath, fatigue, emotional function, and mastery, or a feeling of control over the disease. To increase reproducibility, we stipulated that each dimension should have at least four items. The fatigue, emotional function, and mastery dimensions were constructed by choosing the relevant items which obtained the highest product of frequency and importance on the Item Selection Questionnaire. For the emotional function dimension we added three questions concerning positive affect, including feeling relaxed and happy.

For the dyspnoea dimension, we took a different approach. Reasoning that items associated with dyspnoea would vary widely depending on the patient's sex, range of activities, and level of disability, we individualized the questions. Patients were asked to list activities associated with shortness of breath which they perform frequently and which are important in their daily lives. Twenty-three activities were offered as probes to aid recall. Patients were then asked to choose the five most important activities from among those listed. These items constitutes the dyspnoea dimension for that patient for the duration of the study.

The next task was to decide how to present the items to the patients. One important issue in selecting response options for an evaluative instrument (one designed to measure change over time) is ensuring item responsiveness. Responsiveness was not crucial for the Item Selection Questionnaire, so a five point

Likert scale was considered adequate. However, if the CRQ is to be responsive, each item must be responsive. Although there are no data allowing a rational choice of the number of response options (and therefore the choice is somewhat arbitrary) we judged that a seven-point Likert scale would allow patients to make distinctions as finely as they would wish.

In the dyspnoea section of the questionnaire, patients are asked about how short of breath they have felt during various activities in the previous fortnight. The response options are as follows:

1. Extremely short of breath
2. Very short of breath
3. Quite a bit short of breath
4. Moderate shortness of breath
5. Some shortness of breath
6. A little shortness of breath
7. Not at all short of breath.

It may be noted that we could have used a Borg scale, which is a popular way of measuring dyspnoea, as a way of presenting response options. However, the Borg scale would not have been appropriate for other sorts of questions. For example, we ask patients about how much of the time they have felt frustrated or impatient. The response options in this case are as follows:

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. Hardly any of the time
7. None of the time.

We felt that use of seven point Likert scales was the best way of standardizing response options across all questions.

The CRQ which emerged from this process contained twenty items which take a maximum of 30 min, and usually between 10 and 25 min to administer. To test the reproducibility of the CRQ, a single interviewer administered the CRQ to 25 patients with stable CAL six times at two week intervals. The coefficient of variation (the within-person standard deviation divided by the mean) was 6% for the dyspnoea dimension, 9% for both fatigue and emotional function, and 12% for mastery. These results compare favourably with most functional status and physiological measures [19].

To test the responsiveness of the instrument we conducted two separate studies. We began by administering the CRQ to thirteen patients with chronic lung disease whose respiratory physicians predicted improvement with institution or modification of treatment. The CRQ was administered at the time of consultation and at a follow up visit 2-6 weeks later. Despite only small improvements in spirometry, scores on each of the four CRQ dimensions were substantially improved.

In the second responsiveness study we administered the CRQ to 28 patients with CAL entering our

multidisciplinary in-patient respiratory rehabilitation programme. Questionnaire administration was repeated two weeks after discharge. Substantial improvement in scores occurred on all four dimensions. In this study, we found the CRQ more responsive to change in dyspnoea than either the OCD, the Rand dyspnoea questionnaire (a modified version of the British MRC dyspnoea questionnaire constructed by investigators at the Rand Corporation) [24], or a general measure of health status, the Rand physical and emotional function questionnaires [25]. Responsiveness of the CRQ and the TDI were comparable.

Finally, as part of the same study, we examined correlations between changes in CRQ and changes in other relevant measures, such as spirometry and the 6 min walk test. In general, moderate correlations were found, supporting the validity of the CRQ.

We have used the CRQ in a controlled trial of inhaled salbutamol and oral theophylline CAL patients (reported in detail in another publication [7]). Patients underwent four treatment periods, each of two weeks duration, during which they received the following combinations: placebo-placebo, placebo-salbutamol, placebo-theophylline, and salbutamol-theophylline. Outcomes included twice daily recordings of peak flow rates, spirometry, the distance patients could walk in 6 min, and the CRQ. Clinically important and statistically significant differences between the four periods were noted on both physiological measures and on the CRQ. For the group as a whole, improvement with inhaled salbutamol and oral theophylline was comparable, and additional benefit was gained from a combination of the two drugs. We were thus able to show that both inhaled salbutamol and oral theophylline can improve not only airflow obstruction but also subjective health status in patients with CAL.

Measuring other aspects of health status in CAL

Patients with CAL have problems specific to their illness which are not included in the CRQ. These include cough and difficulty clearing their sputum. These items were not included in the CRQ because, examining the results of the Item Selection Questionnaire, they were not sufficiently frequent or important. They are nevertheless important problems in a subgroup of patients with CAL.

A number of well-established questionnaires (such as the American Thoracic Society Questionnaire) are available for determining if patients experience symptoms of cough and sputum production. However, these questionnaires are not designed for (and are thus not suitable for) measuring change over time in response to an intervention. There appear to be no validated or responsive measures of cough, sputum, or difficulty producing sputum, available. We therefore used an abbreviated version of our approach to construction of disease-specific measures to construct a questionnaire for a controlled trial of ambroxol in patients with chronic bronchitis [26]. Since there was no differ-

ence in a wide variety of measures between patients receiving and not receiving the drug, no comments concerning the responsiveness of the questionnaire can be made. Whilst data regarding validity are also limited, the fact that all related measures showed similar results suggest that the questionnaire did measure patients' difficulty in expectorating sputum.

Conclusions

Valid and responsive instruments for measuring important aspects of health status in patients with CAL are available. All clinical trials of treatment in CAL patients should include measures of physical and emotional function.

References

- Bergner M. – Measurement of health status. *Medical Care*, 1988, 5, 23, 696–704.
- Ware JE. – Standards for validating health measures: definition and content. *J Chron Dis*, 1987, 40, 473–480.
- Kirshner B, Guyatt GH. – A methodologic framework for assessing health indices. *J Chron Dis*, 1985, 38, 27–36.
- Guyatt GH, Bombardier C, Tugwell PX. – Measuring disease-specific quality of life in clinical trials. *Can Med Assoc J*, 1986, 134, 889–895.
- Guyatt GH, Walter S, Norman G. – Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis*, 1987, 40, 171–178.
- Alexander MR, Dull WL, Kasik JE. – Treatment of chronic obstructive pulmonary disease with orally administered theophylline. A double-blind controlled study. *J Am Med Assoc*, 1980, 244, 2286–2290.
- Guyatt GH, Townsend M, Pugsley SO, et al. – Bronchodilators in chronic airflow limitation, effects on airway function, exercise capacity and quality of life. *Am Rev Respir Dis*, 1987, 135, 1069–1074.
- Mahler DA, Matthay RA, Snyder PE, et al. – Sustained-release theophylline reduces dyspnea in nonreversible obstructive airway disease. *Am Rev Respir Dis*, 1985, 131, 22–25.
- World Health Organization: The constitution of the World Health Organization. *WHO Chron*, 1947, 1–29.
- Torrance GW. – Measurement of health state utilities for economic appraisal. *J Health Economics*, 1986, 5, 1–30.
- Bergner M, Bobbitt RA, Carter WB, Gilson BS. – The sickness impact profile: development and final revision of a health status measure. *Medical Care*, 1981, 19, 787–805.
- McSweeney AJ, Grant I, Heaton RK, Adams KM, Timms RM. – Life quality of patients with chronic obstructive pulmonary disease. *Arch Intern Med*, 1982, 142, 473–478.
- MacKenzie CR, Charlson ME, Digioia D, Kelley K. – Can the sickness impact profile measure change? An example of scale assessment. *J Chron Dis*, 1986, 39, 429–438.
- Deyo RA, Centor RM. – Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chron Dis*, 1986, 39, 897–906.
- Kaplan RM. – Quality of life measurement. *In: Measurement strategies in health psychology*. 1985, pp. 115–146.
- Kaplan RM, Bush JW. – Health-related quality of life measurement for evaluation research and policy analysis. *Health Psychology*, 1982, 1, 61–80.
- Toeys CD, Kaplan RM, Atkins CJ. – The costs and effects of behavioral programs in chronic obstructive pulmonary disease. *Medical Care*, 1984, 22, 1088–1100.
- Sutherland HJ, Dunn V, Boyd NF. – Measurement of values for states of health with linear analog scales. *Med Decis Making*, 1983, 3, 477–487.
- Guyatt GH, Berman LB, Townsend M, Pugsley SO, Chambers LW. – A measure of quality of life for clinical trials in chronic lung disease. *Thorax*, 1987, 42, 773–778.
- McGavin CR, Artvinli M, Naoe H. – Dyspnea, disability and distance walked: comparison of exercise performance in respiratory disease. *Br Med J*, 1978, 2, 241–243.
- Mahler DA, Weinberg DH, Wells CK, Feinstein AR. – The measurement of dyspnea. Contents, interobserver agreement, and physiologic correlates of two new clinical indexes. *Chest*, 1984, 85, 751–758.
- Eaton ML, MacDonald FM, Church TR, Niewoehner DE. – Effects of theophylline on breathlessness and exercise tolerance in patients with chronic airflow obstruction.
- Guyatt GH, Townsend M, Berman LB, Pugsley SO. – Quality of life in patients with chronic airflow limitation. *Br J Dis Chest*, 1987, 81, 45–54.
- Rosenthal N, Lohr KN, Rubenstein RS, et al. – *In: Conceptualization and measurement of physiological health for adults. V. Congestive heart failure*. Rand Corporation, Santa Monica, California, 1981.
- Ware JE, Brook RH, Davies-Avery A, et al. – *In: Conceptualization and measurement of health for adults in the health insurance study. Volume 1. Model of health and methodology*. Rand Corporation, Santa Monica, CA, 1980.
- Guyatt GH, Townsend M, Kazim F, Newhouse MT. – A controlled trial of ambroxol in chronic bronchitis. *Chest*, 1987, 92, 618–620.

RÉSUMÉ: Un instrument pour l'appréciation de l'état de santé dans le cadre d'essais cliniques, doit être valable (mesurer ce qu'il est supposé mesurer) et sensible (apte à détecter des modifications cliniquement importantes). L'approche de la mesure de l'état de santé dans les essais cliniques comporte l'utilisation d'une batterie d'instruments, un instrument général qui indique le profil de santé du patient, un instrument qui met en évidence l'utilité pour la santé, ou un instrument qui se concentre sur les problèmes associés à une maladie bien définie. Des instruments spécifiques à la maladie ont été utilisés dans les essais cliniques pour des limitations chroniques du débit aérien. Le diagramme du coût en oxygène est simple et facile à utiliser, mais sa validité et sa sensibilité restent non prouvées. L'index transitionnel de dyspnée est valable et sensible, mais il est difficile à utiliser dans les essais dans lesquels des mesures multiples sont souhaitées. Le questionnaire pour affections respiratoires chroniques a prouvé sa validité et sa sensibilité dans des essais cliniques chez des patients atteints de limitation chronique du débit aérien. Les mesures de l'état de santé devraient être incluses dans tous les essais cliniques se rapportant aux limitations chroniques du débit aérien.