

Methodological issues in evaluating measures of health as outcomes for COPD

P.W. Jones*, R.M. Kaplan[#]

Methodological issues in evaluating measures of health as outcomes for COPD.
P.W. Jones, R.M. Kaplan. ©ERS Journals Ltd 2003.

ABSTRACT: The selection of an appropriate outcome measure depends on the aspect of the disease being addressed and the purpose of the study being conducted. The most fundamental property of any outcome is its ability to measure the biological variable under question. Other properties such as discriminative and evaluative abilities are also important.

Specific outcomes measure a single biological variable, such as forced expiratory volume in one second or depression. The specificity of such measures is attractive but requires precise definition of what is being measured and why. Other, summative, outcomes are used to quantify the overall effect of a number of different biological processes.

The simplest summative measures are global questions such as "How would you rate your health overall?" Others are complex with many items. If designed and used correctly, these questionnaires can provide an estimate of the overall impact of disease or response to therapy and an index of whether that response was clinically worthwhile.

Standardisation of measurements is important to permit comparisons between patients and studies, which makes the measurement of an individual's "quality of life" difficult. The term "health-status measurement" may be better when referring to the use of standardised questionnaires. Utility-based measures help address concerns regarding clinical *versus* statistical improvement and place outcomes for chronic obstructive pulmonary disease treatment trials in the context of all healthcare treatments.

Eur Respir J 2003; 21: Suppl. 41, 13s-18s.

*St George's Hospital Medical School, London, UK. [#]Dept of Family and Preventive Medicine, University of California, San Diego, CA, USA.

Correspondence: P.W. Jones, St George's Hospital Medical School, London SW17 0RE, UK.
Fax: 44 2087255955
E-mail: p.jones@sghms.ac.uk

Keywords: Chronic obstructive pulmonary disease
health status
measurement standardisation
outcome measures
utility-based measures

Received: August 21 2002
Accepted after revision: February 20 2003

Chronic disease usually has three types of effects, which in chronic obstructive pulmonary disease (COPD) would be defined as follows: primary effects in lungs, which may be structural or mechanical; secondary effects in other organs, such as muscles and circulation; and tertiary effects, which involve an interaction between patients and their environment.

From the patient's perspective, health is related to better functioning, symptom relief and longer life [1-3]. However, life duration and quality of life by themselves are not the only important outcomes; all effects need to be taken into account when evaluating treatment. Effects on pulmonary function or secondary effects on organs are important because they may reduce quality of life or shorten life expectancy [4, 5]. If pulmonary function had no effect on these outcomes, it would be of little concern [1].

Against what criteria should measures be judged? Is it appropriate to evaluate outcome measures against forced expiratory volume in one second (FEV₁), maximum oxygen consumption and diffusing capacity? A substantial number of studies in the literature show that the correlations between physiological outcomes and measures of health-related quality of life (health status) are modest, but much of the variance in the latter is not explained by physiological variables [6, 7].

Nevertheless, several lines of evidence suggest that health-status measures are important. For example, it is a significant, prospective, predictor of mortality for patients with advanced lung disease [5]. Furthermore, improved quality of life is what

patients want to achieve with their medical treatment. When seeking care, patients want relief from shortness of breath, the ability to function in the community and the capacity to perform activities of daily living [1]. A treatment that alters a physiological parameter, such as FEV₁, but does nothing for quality of life, may not be successful.

A wide variety of measures suitable for assessing outcomes are available [8-12]. The selection of which measure to use will depend on the aspect of the disease that is being addressed and the purpose of the study [13-15]. Measurement of outcome for patients with COPD may be different than for patients with other chronic diseases and may involve more significant challenges. To illustrate this point, consider comparing outcomes for patients treated for COPD with those of patients treated for osteoarthritis of the hip or for cataracts. Patients with osteoarthritis of the hip are often treated with total joint replacement surgery. Many studies using both generic and disease-targeted measures have demonstrated clinical improvement [16-20]. A significant number of studies have evaluated patients undergoing cataract extraction with lens replacement and show substantial changes with disease-targeted measures, but only modest changes using generic measures [21, 22]. In both osteoarthritis of the hip and cataract disease, surgical interventions produce substantial treatment benefit.

Although treatments for COPD may not produce dramatic benefits seen for total joint replacement or cataract replacement, quality of life can improve after some treatments for COPD patients. In particular, several studies have

documented improvements in quality of life following participation in rehabilitation programmes [23]. Quality-of-life measures, such as the 36-item Short Form (SF-36) of the Medical Outcomes Study [24], the Quality of Well-Being Scale (QWB) [25], the St George's Respiratory Questionnaire (SGRQ) [26] and the University of California at San Diego Shortness of Breath Questionnaire [6] are sensitive to relatively minor changes for COPD patients.

Classifying outcomes

In respiratory medicine, the many different kinds of outcomes may be classified into two broad categories: specific and global.

Specific outcomes

Specific outcomes measure a single biological variable such as FEV₁ or depression. Their characteristic attribute is that they address a unidimensional construct (for example, the degree of airway obstruction or a particular mood disturbance).

Use of a specific outcome is attractive because it should be clear what is being measured. That very specificity, however, requires a precise definition of the question being asked. This may be illustrated by the choice of outcome used to assess the effect of long-acting bronchodilators, agents that act by inducing airway smooth muscle relaxation, which cannot be measured *in vivo*. Indeed, it is noteworthy that even at the level of physiological function, outcome parameters other than those produced by the immediate action of the drug must be used. In other words, the outcomes used in practice are often surrogate measures of the drug's basic physiological action. The inability to measure airway smooth muscle relaxation directly may be important only occasionally, because it is the consequence of that process that is clinically relevant. Furthermore, the process of interest does not occur in isolation, but is taking place in the context of other primary and secondary effects of the disease and may be modified by them.

Measurable outcomes of airway smooth muscle relaxation caused by long-acting bronchodilators include FEV₁, forced inspiratory flow, inspiratory capacity, slow vital capacity and end-expiratory lung volume. Some of these measurements are the direct result of changes in the airway wall, but others are influenced by lung volumes, which may themselves be improved through a reduction in volume of trapped gas. Other relevant physiological variables that are more difficult to collect are dynamic end-expiratory and end-inspiratory lung volumes, yet these may be more closely associated with breathlessness during exercise than spirometric measurements obtained at rest [27, 28]. Some benefits of reduced bronchomotor tone, such as those that occur during sleep or an acute exacerbation, may be timing- or state-specific and only vaguely related to spirometric measurements made during the day in a laboratory and in a stable state.

The selection of a particular specific outcome should depend on the study's purpose, clinical efficacy or a mechanistic explanation of drug effects. If the study is not directed primarily at elucidating mechanisms, a specific outcome should be chosen because it may provide pathophysiological confirmation that the therapy produced clinical benefit through its postulated mechanism of action. Unfortunately, a specific mechanistic outcome is more often chosen because of ease of measurement than for sound scientific reasons.

Global outcomes

Global or summative outcomes are used to quantify the overall effect of a number of biological steps [29], but they may not be recognisable immediately as being such, because some of them measure factors that appear to be unidimensional. One example is exercise performance because this physiological outcome is determined by cardiac, pulmonary, circulatory, and peripheral muscle function, taken together with the sensations of breathlessness and fatigue [6, 7, 30]. Even the FEV₁ is a summative measure in COPD (as opposed to asthma) because it reflects both disease in the airway wall and the loss of alveolar attachments caused by emphysema. This summative property of the FEV₁ is employed in practice because it is used to define the severity of COPD regardless of underlying pathophysiology [9, 31].

Health status is more readily recognisable as a summative measure. In theory, it is easy to conceptualise health as being a single construct, but in practice, such measurements address a range of different aspects of disturbance to health and well-being. Some questionnaires, such as the generic SF-36 [32], do not even provide a single summative scale of overall health impairment, and present their results as a profile of scores or as physical or mental summary scores. Conversely, other generic instruments, such as the Sickness Impact Profile [33] and the QWB [34], do provide a total score, as do some disease-specific questionnaires for COPD, such as the SGRQ [35].

Health-status questionnaires are complex instruments, but other global outcomes, such as asthma severity scores used on diary cards, use much simpler techniques. Typically, patients are asked to rate their overall symptom level using a three- to seven-point category scale (e.g. none, mild, moderate, severe). Similar techniques are used for assessing the overall efficacy of therapy by patients or physicians. Scores of this type are now being used in COPD clinical trials. In contrast to the total scores obtained from complex questionnaires, such outcomes are pure global scores because they are not calculated from responses to multiple discrete items. Their chief disadvantage is that it is never clear how an individual is making a judgment as to the overall level of symptoms, state of health or the effect of therapy.

Global outcomes offer a number of attractive properties. If designed and used correctly, they may provide a measure of the overall impact of disease or response to therapy [2, 36–39]. This may be especially useful when a treatment has multiple beneficial actions. Global outcomes may also be more sensitive to treatment than specific outcomes because they have the potential to aggregate multiple small effects together. Each treatment effect may not be large in itself, but becomes of significant benefit when seen together with other effects.

Global scores may be useful in one other respect. They are high-level outcomes and thus they may be closer to constructs that are relevant to patients and physicians alike. As a result, concepts such as a "worthwhile" improvement in exercise tolerance or reduction in symptoms may be easier to conceptualise than a worthwhile improvement in FEV₁. The latter has little immediate or obvious worth to a patient, unlike exercise performance or reduced breathlessness, each of which has an intrinsic worth. Thus, improvement in FEV₁ may be perceived as worthwhile only because it is associated with an improvement in other measures of clinical outcome.

When using a global outcome measure, it is important to recognise that its role is to summarise and aggregate. It can demonstrate that a change has occurred and provide an assessment of whether that change is clinically significant, but it may not identify the mechanisms. In this respect, it should be used for hypothesis generation rather than hypothesis testing.

Measurement properties

The most fundamental property of any outcome is its ability to measure the biological or behavioural variable under question, which is usually termed validity. The outcome must possess other important properties if it is to be useful, among which are the abilities to detect differences in disease level between patients (discriminative ability) and changes within a patient (evaluative ability). Many physiological measurements have both properties, but it appears that this is not always the case with questionnaires. For example, some questionnaires were designed to have largely discriminative properties, such as the SF-36, Medical Research Council Dyspnoea Scale [40], Baseline Dyspnoea Index [41] and the University of California at San Diego Shortness of Breath Questionnaire [6]. These can define the patient population in terms of the severity level of the variable under question, but may be insensitive to worthwhile changes. Other questionnaires are designed principally to detect change, such as the Chronic Respiratory Questionnaire (CRQ) [42] and Transitional Dyspnoea Index [41]. Such instruments are sensitive to change but possibly at the cost of sacrificing the ability to distinguish between severity levels in different patients. However, certain measures of symptomatic outcome appear to have both discriminative and evaluative properties, such as the SGRQ.

In terms of questionnaires that were designed to be responsive to therapeutic intervention, the utility of the instrument will depend upon three additional properties: reliability, the ability of the instrument to perform in the same manner in different settings with different operators; repeatability, the stability of the measurement when the testing conditions and patient are stable; and sensitivity, the ability to detect changes [8, 43, 44]. This latter property may depend in part upon a trade-off between the other two properties. For example, a measure that is reliable because it has very broad categories of response and has high repeatability may have poor sensitivity due to lack of precision; it cannot detect small changes or discriminate between small differences. The parameter that most clearly defines an outcome's usefulness is its signal-to-noise ratio (*i.e.* the ratio of sensitivity to repeatability) [45]. A highly sensitive outcome will be of practical value only if it is reliable.

Significance of outcome measurements

Outcome measurements may be used to provide a method of assessing the value or worth of a treatment. For this to occur, the results obtained with an outcome have to acquire meaning. Such meaning requires reference points against which the measured outcome can be compared. The possible maximum and minimum values provide such anchor points and additional guidance about the value that may be attached to a particular outcome [46, 47].

It is of fundamental importance to distinguish between statistical significance and clinical significance. The former depends on the size of the study as much as on the size of the effect. Small effects may be rendered statistically significant if the study is sufficiently large. Clinical significance is a much more useful concept but one that is difficult to define and measure [10]. Value judgments are always required at some stage in the establishment of thresholds for a clinically significant effect or minimum clinically important difference. Such judgments are required whether the outcome being validated is a quality-of-life score or a physiological measure.

When establishing thresholds for clinical significance, it is necessary to prespecify the criteria used for assessing what

magnitude of change in an outcome will be judged clinically significant. These criteria will also require a selection of other outcomes to be used as a reference standard for what constitutes clinical significance [10]. Such references may include patient/physician global judgments, clinically significant changes in another clinical variable and prediction of future events (*e.g.* death, exacerbations and hospital admission).

It may not always be necessary, or even a worthwhile enterprise, to produce thresholds for clinical significance for all outcomes used in COPD. Although it may be possible to produce reliable estimates for a clinically significant threshold for changes in FEV₁ in patients with COPD, is it worthwhile to do so? The outcomes used to establish criteria for a clinically significant improvement in COPD can themselves be measured in a clinical trial, and the degree of association between FEV₁ and these outcomes is only modest, such that a clinical threshold for FEV₁ would merely be a weak surrogate for a clinical outcome's threshold.

Issues surrounding the establishment of thresholds for clinical significance are complex and are reviewed in detail elsewhere [10]. Thresholds for the 6-min walking distance [48], CRQ [49] and SGRQ [50] are available, but it is important to appreciate that these are mean estimates obtained from patient groups. These thresholds are helpful, but they should be used only as indicative values, not as rigid or high-precision boundaries between that which is worthwhile to a patient and that which is not [10].

Measurement standardisation

The purpose of taking measurements is to make comparisons between or within patients in tests for trends or treatment effects. These comparisons can only be valid if all measurements are made in exactly the same way.

Some physiological measures, such as FEV₁, are expressed in agreed-upon standard units and have criteria for the adequacy of a measurement. Such standardisation is the result of years of custom, practice and international agreement. By contrast, psychology has a number of scales for measuring depression, but none is universally accepted and consequently there is no standard measurement unit. That said, one or two depression scales are now widely used in respiratory medicine, an example being the Hospital Anxiety and Depression Scale [51]. However, it will be some time before a particular scale becomes the *de facto* standard.

The concept of a "health-related, quality-of-life measurement" provides a challenge to standardisation [2, 10]. Life is potentially too rich and varied to capture standardised quality-of-life effects in individuals, even for the most socially restricted of COPD patients. For example, the inability to play with grandchildren may be an important factor in the lives of many patients with COPD, but often, reasons unrelated to health may restrict this activity. As a result, an item in a questionnaire would need to be worded along the lines of "If you have suitably aged grandchildren with whom you would wish to play but are unable to do so solely because of breathlessness or fatigue, please check the box." This complex item, with conditional and specific requirements, may have low repeatability and would certainly present the developer of the questionnaire with the problem of how to handle the "not applicable" responses in the scoring system. Furthermore, the presence of items that are "not applicable" reduces the number of items that can be used by some patients and, thereby, the instrument's precision.

In clinical trials, all measurements should be made using an instrument that is appropriate to the task. Each patient must be evaluated using a standardised questionnaire that is

suitable for every individual who is being assessed. In this context, standardisation means that all items in the questionnaire are common (at least potentially) to all patients with the disease. The consequence of this item selection process is that the resulting scores are population-based estimates of health that may not reflect precisely any given individual's actual health impairment. This is in no way different from the use of the FEV₁ expressed as a percentage of age-, sex-, height- and race-matched predicted values for assessing an individual patient's degree of airway obstruction. Such estimates are based upon population norms and not the patient's own premorbid state. By analogy, the items in health-status questionnaires are those that reflect the usual effect of the disease in a population of patients with COPD.

Standardisation should also apply to symptom measurement. Diary cards have been used for many years as an outcome in asthma and they are now being used in COPD. Data from such diaries are used to calculate mean symptom scores and also to calculate derived parameters such as "symptom-free days" in asthma and "bad days" in COPD. Both are potentially valuable measurements, but there is no consensus (except perhaps within a given pharmaceutical company) concerning the wording and number of response categories in the diary. This is important because a recent study has shown that one diary card question, phrased to address the level of asthma symptoms, produced a more severe mean score over 14 days than a similar question in the same diary that addressed the effect of asthma on daily life [52]. Diary cards are also being used to identify exacerbations prospectively [53], but again, there is no consensus concerning the level and duration of change in symptoms (or FEV₁ or peak expiratory flow) that constitutes an exacerbation. In view of the increased appreciation of the importance of these events, agreement on methods of identifying the occurrence of an exacerbation must be sought soon.

Utility-based measures

In addition to performing economic analyses [54], utility-based measures resolve some of the problems in measuring COPD outcomes. For example, they help address concerns about clinical *versus* statistical improvement [42]. Furthermore, these utility-based measures place outcomes for trials of COPD treatments in the context of all treatments in healthcare [15, 55].

Utility-based, health-outcome measures place levels of wellness on a continuum ranging from death (0.0) to perfect health (1.0). These measures represent a significant refinement over traditional survival analysis that considers each individual in a binary (alive or dead) fashion. Utility weights can be used to represent levels of wellness along this continuum and are often applied to "quality adjust" survival time [29]. Utility-based measures put clinical effect size into context by showing how observed differences map on the continuum between optimum function and death. For example, COPD patients participating in rehabilitation programmes improve by ~0.04 units. This is four hundredths of the distance between death and perfect health. The numbers can be used to estimate quality-adjusted life years, as well as the duration of the benefit [34].

The three most commonly used methods are the EuroQol (EQ)-5D [56], the Health Utilities Index (HUI) [57], and the QWB [25]. The EQ-5D has been created by a collaborative group from Western Europe known as the EuroQol group [56]. Its method has been validated in postal surveys in England, Sweden and the Netherlands. More recent versions

of the EQ-5D are now used in a substantial number of clinical and population studies [58, 59].

The HUI, which was developed in Canada by FEENY *et al.* [60], uses a multiattribute model to map preference for the 972,000 possible states onto the 0.0–1.0 continuum. The HUI has been used in many population and clinical studies.

The QWB integrates several components into a single score [25, 61, 62]. Patients are classified according to objective levels of functioning, represented by the scales of mobility, physical activity and social activity. Once observable, behavioural levels of functioning have been classified, each individual is placed on the 0.0–1.0 scale of wellness, a continuum between optimum function and death.

Each of these three methods is well validated and can be used in outcome studies for patients with chronic disease. Most importantly, the methods are required if the investigator intends to perform cost-utility studies [54].

Conclusions

The methods for measuring health-related quality of life and population health have been tested and progress has been made in outcome measures for patients with COPD. These measures are useful for population monitoring, clinical trials and resource allocation.

The choice of an outcome should reflect a study's purpose. Studies of basic physiological mechanisms or pharmacological efficacy should use specific outcomes that assess biological variables as close to the site of action or process as possible. In studies where the result of interest is the product of multiple effects or mechanisms, the chosen outcome should be at the point of convergence or the end of a sequence of effects. The outcome should be clinically relevant and provide a measure of overall efficacy and an estimate of clinical value. At present, clinical outcomes, such as breathlessness, exercise capacity and health status, may provide the closest approaches to this ideal.

Several excellent quality-of-life measures are designed specifically for evaluating outcomes in chronic obstructive pulmonary disease patients. Beyond these measures, other, more generic methods are available for estimating outcomes in clinical trials. Utility-based measures offer the extra advantage of contributing to economic analysis, and these methods should be given careful consideration for clinical studies of patients with chronic obstructive pulmonary disease.

References

1. Kaplan RM. Behavior as the central outcome in health care. *Am Psychol* 1990; 45: 1211–1220.
2. Kaplan RM. The Ziggy theorem: toward an outcomes-focused health psychology. *Health Psychol* 1994; 13: 451–460.
3. Kaplan RM. Measuring health outcomes for resource allocation. In: Glueckauf E, Robert G, Frank E, eds. *Psychological Practice in a Changing Health Care System: Issues and New Directions*. New York, Springer Publishing Co., Inc., 1996; pp. 101–133.
4. Carone M, Ambrosino N, Bertolotti G, *et al.* Quality of life evaluation and survival study: a 3-yr prospective multinational study on patients with chronic respiratory failure. *Monaldi Arch Chest Dis* 2001; 56: 17–22.
5. Squier HC, Ries AL, Kaplan RM, *et al.* Quality of well-being predicts survival in lung transplantation candidates. *Am J Respir Crit Care Med* 1995; 152: 2032–2036.
6. Eakin EG, Resnikoff PM, Prewitt LM, Ries AL, Kaplan RM.

- Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest* 1998; 113: 619-624.
7. Eakin EG, Sassidamborn DE, Ries AL, Kaplan RM. Reliability and validity of dyspnea measures in patients with obstructive lung disease. *Int J Behav Med* 1995; 2: 118-134.
8. Guyatt GH, King DR, Feeny DH, Stubbing D, Goldstein RS. Generic and specific measurement of health-related quality of life in a clinical trial of respiratory rehabilitation. *J Clin Epidemiol* 1999; 52: 187-192.
9. Jones PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax* 2001; 56: 880-887.
10. Jones PW. Interpreting thresholds for a clinically significant change in health status in asthma and COPD. *Eur Respir J* 2002; 19: 398-404.
11. Kaplan RM, Feeny D, Revicki DA. Methods for assessing relative importance in preference based outcome measures. *Qual Life Res* 1993; 2: 467-475.
12. Orenstein DM, Kaplan RM. Measuring the quality of well-being in cystic fibrosis and lung transplantation - the importance of the area under the curve. *Chest* 1991; 100: 1016-1018.
13. Russell LB. Improving the panel's recommendations. *Med Decis Making* 1999; 19: 374-383.
14. Russell LB. The methodologic partnership of effectiveness reviews and cost-effectiveness analysis. *Am J Prev Med* 2001; 20: Suppl. 3, 10-12.
15. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the panel on cost-effectiveness in health and medicine. *JAMA* 1996; 276: 1253-1258.
16. Soderman P, Malchau H, Herberts P. Outcome after total hip arthroplasty: Part I. General health evaluation in relation to definition of failure in the Swedish National Total Hip Arthroplasty register. *Acta Orthop Scand* 2000; 71: 354-359.
17. O'Connell T, Browne C, Corcoran R, Howell F. Quality of life following total hip replacement. *Ir Med J* 2000; 93: 108-110.
18. March LM, Cross MJ, Lapsley H, et al. Outcomes after hip or knee replacement surgery for osteoarthritis. A prospective cohort study comparing patients' quality of life before and after surgery with age-related population norms. *Med J Aus* 1999; 171: 235-238.
19. Mangione CM, Goldman L, Orav EJ, et al. Health-related quality of life after elective surgery: measurement of longitudinal changes. *J Gen Intern Med* 1997; 12: 686-697.
20. Boardman DL, Dorey F, Thomas BJ, Lieberman JR. The accuracy of assessing total hip arthroplasty outcomes: a prospective correlation study of walking ability and 2 validated measurement devices. *J Arthroplasty* 2000; 15: 200-204.
21. tenHove M, Siegel K, Groll D, Hopman W. Evaluating the quality of life in patients undergoing surgery to restore bilateral sight. *Ann Meet Int Soc Technol Assess Health Care* 1998; 14: 83.
22. Hove MT, Seigel K, Groll D, Hopman W, MacKenzie T. The measurement of quality of life in patients with unilateral visual impairment. *Abst Book Assoc Health Services Res* 1998; 15: 187.
23. Ries AL, Kaplan RM, Limberg TM, Prewitt LM. Effects of pulmonary rehabilitation on physiologic and psychosocial outcomes in patients with chronic obstructive pulmonary disease. *Ann Intern Med* 1995; 122: 823-832.
24. Ware JE Jr, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *J Clin Epidemiol* 1998; 51: 903-912.
25. Kaplan RM, Sieber WJ, Ganiats TG. The Quality of Well-Being Scale: Comparison of the interviewer-administered version with a self-administered questionnaire. *Psychol Health* 1997; 12: 783-791.
26. Jones PW. Issues concerning health-related quality of life in COPD. *Chest* 1995; 107: Suppl. 5, 187S-193S.
27. Belman MJ, Botnick WC, Shin JW. Inhaled bronchodilators reduce dynamic hyperinflation during exercise in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1996; 153: 967-975.
28. O'Donnell DE, Webb KA. Exertional breathlessness in patients with chronic airflow limitation. The role of lung hyperinflation. *Am Rev Respir Dis* 1993; 148: 1351-1357.
29. Kaplan RM. Using quality of life information to set priorities in health policy. *Soc Indic Res* 1994; 33: 121-163.
30. Eakin EG, Kaplan RM, Ries AL, Sassidamborn DE. Patients' self-reports of dyspnea - an important and independent outcome in chronic obstructive pulmonary disease. *Ann Behav Med* 1996; 18: 87-90.
31. Carone M, Donner CF, Jones PW. Health status measurement: an increasingly important outcome evaluation in COPD patients. *Monaldi Arch Chest Dis* 2001; 56: 297-298.
32. Stewart AL, Hays RD, Ware JE Jr. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care* 1988; 26: 724-735.
33. 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.
34. Kaplan RM, Atkins CJ, Timms R. Validity of a quality of well-being scale as an outcome measure in chronic obstructive pulmonary disease. *J Chronic Dis* 1984; 37: 85-95.
35. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992; 145: 1321-1327.
36. Kaplan RM. An outcomes-based model for directing decisions in women's health care. *Clin Obstet Gynecol* 1994; 37: 192-206.
37. Kaplan RM. Two pathways to prevention. *Am Psychol* 2000; 55: 382-396.
38. Kaplan RM, Ganiats TG, Sieber WJ, Anderson JP. The quality of well-being scale: critical similarities and differences with SF-36. *Int J Qual Health Care* 1998; 10: 509-520.
39. Kaplan RM, Mehta R. Outcome measurement in kidney disease. *Blood Purif* 1994; 12: 20-29.
40. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax* 1999; 54: 581-586.
41. 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.
42. 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.
43. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol* 1989; 42: 403-408.
44. Guyatt GH, Tugwell PX, Feeny DH, Drummond MF, Haynes RB. The role of before-after studies of therapeutic impact in the evaluation of diagnostic technologies. *J Chronic Dis* 1986; 39: 295-304.
45. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis* 1986; 39: 897-906.
46. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics* 2000; 18: 419-423.
47. Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol* 1996; 49: 1215-1219.
48. Redelmeier DA, Bayoumi AM, Goldstein RS, Guyatt GH. Interpreting small differences in functional status: the Six Minute Walk test in chronic lung disease patients. *Am J Respir Crit Care Med* 1997; 155: 1278-1282.
49. Jaeschke R, Singer J, Guyatt GH. Measurement of health

- status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10: 407-415
50. Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med* 1991; 85: Suppl. B, 25-33.
51. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; 67: 361-370.
52. Barley EA, Jones PW. A comparison of global questions versus health status questionnaires as measures of the severity and impact of asthma. *Eur Respir J* 1999; 14: 591-596.
53. Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998; 157: 1418-1422.
54. Gold MR. Cost-Effectiveness in Health and Medicine. New York, Oxford University Press, 1996.
55. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on cost-effectiveness in health and medicine. *JAMA* 1996; 276: 1172-1177.
56. Kind P. The performance characteristics of EQ-5D, a measure of health related quality of life for use in technology assessment. *Ann Meet Int Soc Technol Assess Health Care* 1997; 13: 81.
57. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. Health Utilities Index. *Pharmacoeconomics* 1995; 7: 490-502.
58. Gudex C, Dolan P, Kind P, Williams A. Health state valuations from the general public using the visual analogue scale. *Qual Life Res* 1996; 5: 521-531.
59. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Br J Rheumatol* 1997; 36: 551-559.
60. Feeny D, Furlong W, Mulhern RK, Barr RD, Hudson M. A framework document for assessing health-related quality of life among children with cancer. *Int J Cancer Suppl* 1999; 12: 2-9.
61. Kaplan RM, Anderson JP, Patterson TL, et al. Validity of the Quality of Well-Being Scale for persons with human immunodeficiency virus infection. HNRC Group. HIV Neurobehavioral Research Center. *Psychosom Med* 1995; 57: 138-147.
62. Kaplan RM, Feeny D, Revicki DA. Methods for assessing relative importance in preference based outcome measures. In: Joyce CR, O'Boyle CA, McGee HM, eds. *Individual Quality of Life: Approaches to Conceptualisation and Assessment*. The Netherlands, Hardwood Academic Publishers, 1999; pp. 135-149.