

The Medical Literature

Users' Guides to the Medical Literature

XII. How to Use Articles About Health-Related Quality of Life

Gordon H. Guyatt, MD, MSc; C. David Naylor, MD, MSc, DPhil; Elizabeth Juniper, MCSP, MSc; Daren K. Heyland, MD; Roman Jaeschke, MD, MSc; Deborah J. Cook, MD, MSc; for the Evidence-Based Medicine Working Group

CLINICAL SCENARIO

You are a physician following a 35-year-old man who has had active Crohn disease for 8 years. The symptoms were severe enough to require resectional surgery 4 years ago, and despite treatment with sulfasalazine and metronidazole, the patient has had active disease requiring oral steroids for the last 2 years. Repeated attempts to decrease the prednisone have failed, and the patient has required doses of greater than 15 mg per day to control symptoms. You are impressed by both the methods and results of a recent article¹ documenting that such patients benefit from oral methotrexate and suggest to the patient that he consider this medication. When you explain some of the risks of methotrexate, particularly potential liver toxicity, the patient is hesitant. How much better, he asks, am I likely to feel while taking this medication?

INTRODUCTION

There are 3 reasons we offer treatment to our patients. We believe our interventions increase longevity, pre-

vent future morbidity, or make patients feel better. The first 2 of these 3 end points are relatively easy to measure. At least in part because of difficulty in measurement, clinicians have for many years been ready to substitute physiological or laboratory tests for the direct measurement of the third. In the last 20 years, however, clinicians have recognized the importance of direct measurement of how people are feeling and how they are able to function in daily activities. Investigators have developed increasingly sophisticated methods of making these measurements.

Since, as clinicians, we are most interested in aspects of life quality directly related to health rather than issues such as finances or the quality of the environment, we frequently refer to measurements of how people are feeling as health-related quality of life (HRQL).² Investigators measure HRQL using questionnaires that typically include questions about how patients are feeling or what they are experiencing associated with response options such as yes or no, 7-point scales, or visual analogue scales. Investigators aggregate responses to these questions into domains or dimensions (such as physical or emotional function) that yield an overall score.

Controversy exists concerning the boundaries of HRQL and the extent to which individual patient's values must be included in its measurement.³⁻⁵ Is it sufficient to know that patients with chronic obstructive lung disease in general value being able to climb stairs without getting short of breath, or does one need to establish that the individual patient values climbing stairs without dyspnea? Further controversy exists about

how the relative values of items and domains need to be established and how these values should be determined. Is it enough to know that both dyspnea and fatigue are important to people with lung disease, or does one need to establish their relative importance? If establishing their relative importance is necessary, which of the many available approaches should one use?

In this article, we take a simple approach. We use HRQL to refer to the health aspects of their lives that people, in general, value, and we are ready to accept patients' statement of what they value without precise determination of ranking of items or domains.

Clinicians often have limited familiarity with methods of measuring how patients feel. At the same time, they are facing articles that recommend administering or withholding treatment on the basis of its impact on patients' well-being. This Users' Guide is designed for clinicians asking the question: Will this treatment make my patient feel better? As in other guides, we will use the framework of the validity of the methods, interpretation of the results, and application of the results to one's patients (Table). In addition, we begin the guide with a commentary on when one should and should not be concerned about HRQL measurement. Our guidelines borrow heavily from our previous work.^{2,5} While this article focuses on using HRQL measures to help with treatment decisions, we hope that it may also improve clinical care by

From the Departments of Clinical Epidemiology and Biostatistics (Drs Guyatt and Cook and Ms Juniper) and Medicine (Drs Guyatt, Heyland, Jaeschke, and Cook), McMaster University, Hamilton, Ontario, and the Institute for Clinical Evaluative Sciences and the Clinical Epidemiology Unit, Sunnybrook Health Science Centre, University of Toronto, Toronto, Ontario (Dr Naylor).

The original list of members (with affiliations) appears in the first article of this series (JAMA. 1993;270:2093-2095). A list of new members appears in the 10th article of the series (JAMA. 1996;275:1435-1439). The following members contributed to this article: Paul Glasziou, MB, PhD; Virginia Moyer, MD, MPH; and Peter Tugwell, MD, MSc.

Reprints: Gordon H. Guyatt, MD, MSc, McMaster University HealthSciences Centre, 1200 Main St W, Room 2C12, Hamilton, Ontario, Canada L8N 3Z5.

Users' Guides to the Medical Literature section editor: Drummond Rennie, MD, Deputy Editor (West), JAMA.

Are the results valid?

Primary guides

Have the investigators measured aspects of patients' lives that patients consider important?

Did the HRQL instruments work in the way they are supposed to?

Secondary guides

Are there important aspects of HRQL that have been omitted?

If there were trade-offs between quality and quantity of life, or an economic evaluation, have the investigators used the right measures?

What were the results?

What was the magnitude of effect on HRQL?

Will the results help me in caring for my patients?

Will the information from the study help me inform my patients?

Did the study design simulate clinical practice?

emphasizing aspects of patients' experience, including functional, emotional, and social limitations, which clinicians sometimes neglect.

DO YOU NEED TO WORRY ABOUT HRQL?

In the early days of clinical trials, few if any treatment studies included measurements of HRQL, and no one worried much. When should you be concerned if investigators have not paid adequate attention to how patients feel?

In general, delaying mortality is sufficient reason to administer a treatment. Some years ago, investigators showed that around-the-clock oxygen therapy for patients with severe chronic airflow limitation improved mortality.⁶ The fact that HRQL data weren't reported in the original article turns out not to be an important omission. Since the intervention prolongs life, our enthusiasm for continuous oxygen administration is not blunted by a subsequent report suggesting that more intensive oxygen therapy had little or no impact on HRQL.⁷ Similarly, while feeling better is important to patients with heart failure, when interventions either extend⁸ or shorten⁹ life span, we usually do not need an HRQL assessment to inform our clinical decisions.

There are exceptions to this rule. While many of our life-prolonging treatments have a negligible impact on or actually improve HRQL, this is not always the case. If treatment leads to a deterioration in HRQL, patients may be concerned that small gains in life span come at too high a cost. Interventions that highlight this concern include chemotherapy for cancer and human immunodeficiency virus disease. In the extreme, life may be prolonged, but patients' families may wonder if, for example, their fate is a persistent vegetative state, they are not better off dead.

A patient's own preferences expressed through an advance directive may support this view.

When the goal of treatment is to improve how people are feeling (rather than to prolong their lives) and physiological correlates of patients' experience are lacking, HRQL measurement is imperative. For example, we would pay little attention to studies of antidepressants that failed to measure patients' mood, or trials of antimigraine medication that failed to measure pain.

The difficult decisions occur when the relation between physiologic or laboratory measures and HRQL outcomes is uncertain. Practitioners have relied on substitute end points not because they weren't interested in making patients feel better, but because they assumed a strong link between physiologic measurements and patients' well-being. A recent trial in patients with symptomatic postmenopausal osteoporosis examined the effect of sodium fluoride on bone density and vertebral fractures.¹⁰ The investigators believed that increased bone mass and fewer vertebral fractures would lead to decreased pain and increased function. Does their failure to measure the effect of treatment on areas of unequivocal importance to patients, including pain, physical function, and household and leisure activities,¹¹ affect the clinical message of the results? Similarly, investigators measuring the effects of antianginal medication have often been satisfied with increased duration of exercise on the treadmill without direct measurement of decreased symptoms or increase in activity in day-to-day life. Are we ready to prescribe medication on the basis of increased laboratory exercise capacity?

Bone density, vertebral fractures, and exercise capacity, or similar measures such as joint count, ejection fraction, or pulmonary function, are surrogate end points for what we really want to measure: the effect of treatment on our patients' lives. Whether these surrogate measures are adequate depends on how confident we are of the link with how people feel. When this issue has been investigated empirically, the relation between physiologic and clinical measures and patients' symptoms is usually modest and often highly variable.²⁻¹⁷ Though these findings lead us to recommend caution in assuming that improvement in physiologic or clinical function will result in patients feeling better, each clinician (and, when appropriate, the patient) must decide on her own threshold.

Referring to the opening scenario, investigators reported the results of a randomized trial of methotrexate in 141 patients with chronically active Crohn

disease despite at least 3 months of prednisone therapy.¹ Patients who received methotrexate were twice as likely to be in clinical remission following 16 weeks of treatment than those who received placebo (39.4% vs 19.1%, $P=.02$), and actively treated patients received less prednisone and showed less disease activity. Is additional information regarding HRQL necessary to interpret the results of this study? As depicted in the scenario, the decision to give methotrexate depends on weighing the benefits and risks, and the patient's question about how much better he is likely to feel with medication may well be relevant to his decision. Without information about the effect of the medication on HRQL, therefore, neither the clinician nor the patient can make a fully informed choice.

ARE THE RESULTS VALID?

Primary Guides

Have the Investigators Measured Aspects of Patients' Lives That Patients Consider Important?—We have described how investigators often substitute end points that make intuitive sense to them for those that patients value. Clinicians can recognize these situations by asking themselves the question: If the end points measured by the investigators were the only thing that changed, would patients be willing to take the treatment? In addition to change in clinical or physiologic variables, patients would require that they feel better or live longer.

How can clinicians be secure that investigators have measured aspects of life that patients value? Investigators may show that the outcomes they have measured are important to patients by asking them directly. For example, in a study examining HRQL in patients with chronic airflow limitation, we used a literature review and interviews with clinicians and patients to identify 123 items reflecting possible ways their illness might affect patients' HRQL.¹⁸ We then asked 100 patients which of the items were problems for them and how important those items were. We found that the most important problem areas for patients were their dyspnea during day-to-day activities and their chronic fatigue. An additional area of difficulty was emotional function, including feeling frustrated and impatient.

If the authors don't present direct evidence that their outcome measures are important to patients, they may cite prior work. For example, a randomized trial of respiratory rehabilitation in patients with chronic lung disease used an HRQL measure based on the responses of patients in the study we've described above and referred to that study.¹⁹ Ideally, the report

will include a summary of the developmental process sufficiently detailed to obviate the need to go back to the prior report.

Alternatively, investigators may describe the content of their measures in detail. An adequate description of the content of a questionnaire allows clinicians to use their experience to decide whether what is being measured is important to patients. For instance, the authors of an article describing a randomized trial of surgery vs watchful waiting for benign prostatic hyperplasia "assessed the degree to which urinary difficulties bothered the patients or interfered with their activities of daily living, sexual function, social activities, and general well-being."²⁰ Few would doubt the importance of these items.

In the study of methotrexate for patients with inflammatory bowel disease (IBD), the patients completed the Inflammatory Bowel Disease Questionnaire (IBDQ), which addresses patients' bowel function, emotional function, systemic symptoms, and social function. Although the authors don't mention this in their article, the 32 items in the IBDQ were chosen because patients with IBD labeled them as the most important in their daily lives.²¹

Did the HRQL Instruments Work in the Way They Are Supposed to?—Measuring how people are feeling is not easy. Investigators must demonstrate that their instruments allow strong inferences about the effect of treatment on HRQL. We will now review how an HRQL measure should perform (we call the way it performs its *measurement properties*) if it is going to be useful.

Signal and Noise.—There are 2 distinct ways in which investigators use HRQL instruments. They may wish to help clinicians distinguish between people who have a better or worse HRQL, or to measure whether people are feeling better or worse over time.²² For instance, suppose a trial of a new drug for patients with heart failure shows that it works best in patients with the New York Heart Association (NYHA) functional classification class IV symptoms. We could use the NYHA class for 2 purposes. One would be to discriminate between patients as to their NYHA class in deciding who to treat. We might also want to determine whether the drug was effective in improving an individual patient's functional status and therefore monitor changes in patients' NYHA functional class.

While for both purposes we require a high ratio of signal to noise, when we are discriminating between people at a single point in time, the signal comes from differences between patients (if ev-

eryone gets the same score, we can't tell who is better off and who is worse off), and the noise comes from variability within subjects (if patients' scores fluctuate wildly, we're not going to be able to say much about their relative well-being).²³ The technical term usually used for the ratio of variability between patients to the total variability is *reliability*.

Instruments used to evaluate change over time must, in contrast, be able to pick up any important changes in the way patients are feeling, even if those changes are small. Thus, the signal comes from the difference in score in patients who have improved or deteriorated, and the noise from the variability in score in patients who have not changed. The term we use for the ability to detect change (the ratio of signal to noise over time) is *responsiveness*.

An unresponsive instrument can result in a false-negative trial in which the intervention improves how patients feel, and yet the instrument fails to detect the improvement. This problem may be particularly salient for questionnaires that have the advantage of covering all relevant areas of HRQL, but the disadvantage of covering each area superficially. A crude instrument such as the NYHA functional classification (with only 4 categories) may work well for stratifying patients, but may not be able to detect small but important improvement with treatment.

In studies that show no difference in change in HRQL when patients receive a treatment vs a control intervention, clinicians should look for evidence that the instruments have been able to detect small or medium-sized effects in previous investigations. In the absence of this evidence, instrument unresponsiveness becomes a plausible reason for the failure to detect differences in HRQL. For example, a randomized trial of a diabetic education program reported no changes in 2 measures of well-being and attributed the result to, among other factors, lack of integration of the program with standard therapy.²⁴ Given that the program improved knowledge and self-care and patients felt less dependent on physicians, another explanation is inadequate responsiveness of the 2 HRQL measures.

In the trial of methotrexate in Crohn disease, concern about responsiveness decreases because the study showed statistically significant differences between treatment and control groups. As it turns out, the IBDQ had detected small to medium-sized differences in previous investigations.^{21,25,26}

Validity.—Validity has to do with whether the instrument is measuring

what it is intended to measure. The absence of a reference or criterion standard for HRQL creates a challenge for anyone hoping to measure how patients are feeling. We can be more confident that an instrument is doing its job if it appears targeted to the right problems (the technical term for this is *face validity*). Empirical evidence that it measures the domains of interest will also help.

To provide such evidence, investigators have borrowed validation strategies from psychologists who have for many years had to decide whether questionnaires assessing intelligence, attitudes, and emotional function were really measuring what is intended. Investigators interested in attitudes may show apparent differences between individuals that really reflect variability in the tendency to provide socially acceptable answers rather than differences in underlying attitudes; investigators may demonstrate apparent effects of rehabilitation on HRQL, but may really be detecting differences in satisfaction with care. In either case, the instrument would be detecting a signal, but it would be the wrong signal.

Establishing validity therefore involves examining the logical relationships that should exist between measures. For example, we would expect that, in general, patients with lower treadmill exercise capacity will have more dyspnea in daily life than those with higher exercise capacity, and we would expect to see substantial correlations between a new measure of emotional function and existing emotional function questionnaires. When we are interested in evaluating change over time, we examine correlations of change scores: patients who deteriorate on their treadmill exercise capacity should, in general, show increases in dyspnea, while those whose exercise capacity improves should experience less dyspnea; a new emotional function measure should show improvement in patients who improve on existing measures of emotional function. The technical term for this process is testing an instrument's *construct validity*.

Clinicians should look for evidence of the validity of HRQL measures used in clinical studies. Reports of randomized trials using HRQL measures seldom review evidence for the validity of the instruments they use, but clinicians can gain some reassurance from statements (backed by citations) that the questionnaires have been previously validated. In the absence of evident face validity or empirical evidence of validity, clinicians are entitled to skepticism about the study's measurement of HRQL.

In the methotrexate in IBD study, the investigators refer to the IBDQ as "previously validated" and provide 2 relevant citations.^{21,25} These articles describe extensive validation of the questionnaire, including correlations of change that document the instrument's usefulness for measuring change over time.

Secondary Guides

Are There Important Aspects of HRQL That Have Been Omitted?—Investigators may have addressed HRQL issues, but have not done so comprehensively. Exhaustive measurement may be more or less important in a particular context. One can think of a hierarchy that begins with symptoms, moves on to the functional consequences of the symptoms, and ends with more complex elements such as emotional function. If, as a clinician, you believe your patient's sole interest is in whether a treatment relieves the primary symptoms and most important functional limitations, you will be satisfied with a limited range of assessment. Recent randomized trials in patients with migraine^{27,28} and postherpetic neuralgia²⁹ restricted themselves primarily to the measurement of pain; studies of patients with rheumatoid arthritis^{30,31} and back pain³² measured pain and physical function, but not emotional or social function.

As a clinician, you can judge whether or not these omissions are important to you or, more importantly, your patients. We would encourage you, however, to bear in mind the broader impact of disease on patients' lives. Disease-specific measures that explore the full range of patients' problems and experience remind us of domains we might otherwise forget. We can trust these measures to be comprehensive if the developers have conducted a detailed survey of patients suffering from the illness or condition.

If you are interested in going beyond the specific illness and comparing the impact of treatments on HRQL across diseases or conditions, you will require a more comprehensive assessment. None of the disease-specific, system- or organ-specific, function-specific (such as instruments that examine sleep or sexual function), or problem-specific (such as pain) measures are adequate for comparisons across conditions. These comparisons require generic measures designed for administration to people with any underlying health problem (or no problem at all) that cover all relevant areas of HRQL.

One type of generic measure, health profiles, yields scores for all domains of HRQL (including, for example, mobility,

self-care, and physical, emotional, and social function). There are a number of well-established health profiles, including the Sickness Impact Profile³³ and the short forms of the instruments used in the Medical Outcomes Study^{34,35} that have advantages of simplicity, self-administration, and the ability to put changes in specific functions in the context of overall HRQL. Inevitably, such instruments cover each area superficially. This may limit their responsiveness—indeed, several randomized trials have found that generic instruments were less powerful in detecting treatment effects than specific instruments.^{19,36-40} Ironically, generic instruments may also suffer from not being sufficiently comprehensive: they may completely omit patients' primary symptoms.

Disease-specific measures may comprehensively sample all aspects of HRQL relevant to a specific illness and also be responsive, but they are unlikely to deal with adverse effects. For instance, the IBDQ measures all important disease-specific areas of HRQL, including symptoms directly related to the primary bowel disturbance, systemic symptoms, and emotional and social function. Coincidentally, it measures some methotrexate adverse effects, including nausea and lethargy, because these are also experienced by patients with IBD not taking methotrexate, but not other adverse effects such as rash or mouth ulcers. The investigators could have administered a generic instrument to tap in to non-IBD-related aspects of HRQL, but once again would likely have failed to measure adverse effects in sufficient detail. Adverse effect-specific instruments are limited; the investigators chose a checklist approach and documented the frequency of occurrence of adverse events both severe and not severe enough to warrant discontinuation of treatment.

If There Were Trade-offs Between Quality and Quantity of Life, or an Economic Evaluation, Have the Investigators Used the Right Measures?—While providing information about the broad domains of HRQL and therefore allowing comparisons across conditions, health profiles are ill-suited for health policy decisions that involve integrating costs. Health policy decisions require choices about resource allocation across diseases, conditions, or medical problems, and also involve considerations of cost. These choices require standardized comparisons that allow one to relate the impact of very different treatments (such as drugs, surgery, or rehabilitation programs) on very different conditions (such as chronic lung disease, renal failure, or Parkinson disease). Inevitably, they involve putting a value on health states and may thus re-

quire sophisticated weighting for patient preferences, and necessitate relating health states to anchors of death and full health. Such measures may aid policy-makers in making the right decisions about how public money is allocated.

Measures that provide a single number that summarizes all of HRQL are preference or value weighted, and have the preferences or values anchored to death and full health are called *utility measures*. Typically, utility measures use a scale from 0 (death) to 1.0 (full health) to summarize HRQL. Since they weight the duration of life according to its quality, their output is often called *quality-adjusted life years* (QALYs). Thus, utilities are holistic measures that ask patients to express, in a single value, their strengths of preferences for particular health states.

Boyle and colleagues,⁴¹ in a classic article, used a utility measure to calculate that treating critically ill infants weighing 1000 to 1499 g at birth cost \$3200 per QALY gained, while treating infants with a birth weight of 500 to 999 g cost \$22 400 per QALY gained.⁴¹ Estimates for the cost per QALY for treating patients receiving renal dialysis have ranged from approximately \$30 000 to \$50 000.^{42,43} While different weighting schemes yield different results and may therefore be considered arbitrary, a number of increasingly simple utility measures are now available, have provided interesting results in clinical trials, and may facilitate integrating cost into policy decisions. However, the use, measurement, and interpretation of utility measures remain controversial.⁴⁴ The investigators in the methotrexate trial did not use a health profile or a utility measure, thus limiting use of the data for comparisons across disease states and preventing a formal economic analysis.

What Were the Results?

What Was the Magnitude of Effect on HRQL?—Understanding the results of a trial involving HRQL involves special challenges. Patients with acute back pain who were prescribed bed rest had mean scores on the Oswestry Back-Disability Index, a measure that focuses on disease-specific functional status, 3.9 points worse than control patients.³² Patients with severe rheumatoid arthritis allocated to cyclosporine had a mean disability score 0.28 unit better than control patients.³⁰ Are these differences trivial, small but important, of moderate magnitude, or do they constitute large and extremely important differences between treatments?

These examples show that the interpretability of most HRQL measures is not self-evident. There are a number of

methods available for understanding the magnitude of HRQL effects. Investigators may relate changes in HRQL questionnaire score to well-known functional measures (such as the NYHA functional classification), to clinical diagnosis (such as the change in score needed to move people in or out of the diagnostic category of depression), or to the impact of major life events.⁴⁵ They may relate changes in HRQL score to patients' global ratings of the magnitude of change they have experienced,⁴⁶ or to the extent they rate themselves as better or worse than other patients.⁴⁷ Whatever the strategy, if investigators don't provide an indication of how to interpret changes in HRQL score, the findings are of limited use to clinicians.

Even if we did know that 3.9 points on the Oswestry Back-Disability Index or 0.28 unit on a rheumatoid arthritis disability index signified, for instance, small but important changes, mean differences between groups may be difficult to interpret. Clinicians may find the proportion of patients who achieved small, medium, and large gains due to treatment more informative.

The investigators who conducted the trial of methotrexate for Crohn disease do not help clinicians interpret the magnitude of difference in HRQL. The mean difference in IBDQ score between treatment and control groups at 16 weeks was 0.59. Other investigations suggest that differences of approximately 0.5 may represent small but important changes, while large improvements correspond to a difference in score of greater than 1.0.⁴⁶⁻⁴⁹ Thus, the mean difference between treated and control patients in the methotrexate study likely falls into the category of small but important change in HRQL.

Will the Results Help Me in Caring for My Patients?

Will the Information From the Study Help Me Inform My Patients?—People with the same chronic disease often vary markedly in the problems they experience. Even if the problems are the same, the magnitude of the impact of those problems in their lives may differ. Assessment of HRQL will only help in the care of an individual patient if that patient's problems are similar to those of patients in the trial.

Knowing whether HRQL results of a study are relevant for your patients means understanding their experience of illness. Even the most common problems of a chronic disease don't affect all those afflicted. For instance, 92% of patients with IBD complain of frequent bowel movements, and 82% complain of abdominal cramps.⁵⁰ With respect to

emotional function, 78% feel frustrated and 76% feel depressed. The patients who experienced these difficulties vary in the extent to which they felt the problems were important. Thinking back to the scenario, before answering the question about how the treatment would affect the patient's life, the clinician would have to find out the problems the patient was currently experiencing, the importance he attached to those problems, and the value he might attach to having the problems ameliorated.

Reflecting further on the process of communicating with patients, HRQL instruments that focus on specific aspects of patients' experience may be more useful than global measures. Patients with chronic lung disease may find it more informative to know that their compatriots offered a treatment became less dyspneic and fatigued in daily activity, rather than simply that they judged their HRQL as improved. HRQL measures will be most useful when the results facilitate their practical use by you and your patients.

Did the Study Design Simulate Clinical Practice?—Treatments affect HRQL both by reducing disease symptoms and consequences and by creating new problems. Adverse effects may make the cure worse than the disease. Clinicians conducting clinical trials are usually blind to treatment allocation and try to maintain patients on the study medication as long as possible. Patients may therefore soldier on in the face of considerable adverse effects, and this may be reflected in their HRQL.

This is not how we conduct our clinical practice. If patients experience significant adverse effects, we discontinue the medication, particularly if there is a suitable alternative. Thus, the design of the clinical trial may create an artificial situation with misleading estimates of the impact of treatment on HRQL. This issue is of particular concern for treatments such as antihypertensive drugs in which much of the impairment in HRQL may be due not to the medical condition, but to the treatment.

The trial of methotrexate in Crohn disease simulated clinical practice well. If the patient is experiencing problems similar to those of the trial patients, and if those problems are important to him, he is likely to achieve comparable benefit to patients enrolled in the trial.

CONCLUSION

We encourage clinicians to consider the impact of their treatments on patients' HRQL, and to look for information regarding this impact in clinical trials. Responsive, valid, and interpretable instruments measuring experiences of

importance to most patients should increasingly help guide our clinical decisions.

We acknowledge a useful review of the manuscript by Brian Feagan, MD, who reassured us we were on the right track with our scenario. We offer special thanks to Deborah Maddock who has provided outstanding administrative support and coordination for the activities of the Evidence-Based Medicine Working Group.

References

1. Feagan BG, Rochon J, Fedorak RN, et al. Methotrexate for the treatment of Crohn's disease: the North American Crohn's Study Group Investigators. *N Engl J Med*. 1995;332:292-297.
2. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life: basic sciences review. *Ann Intern Med*. 1993;70:225-230.
3. Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. *JAMA*. 1994;272:619-626.
4. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life: a conceptual model of patient outcomes. *JAMA*. 1995;273:59-65.
5. Guyatt GH, Cook DJ. Health status, quality of life, and the individual. *JAMA*. 1994;272:630-631.
6. Nocturnal Oxygen Therapy Trial Group. Continuous or nocturnal oxygen therapy to hypoxemic chronic obstructive lung disease. *Ann Intern Med*. 1980;93:391-398.
7. Heaton RK, Grant I, McSweeney AJ, Adams KM, Petty TL. Psychologic effects of continuous and nocturnal oxygen therapy in hypoxemic chronic obstructive pulmonary disease. *Arch Intern Med*. 1983;143:1941-1947.
8. Mulrow CD, Mulrow JP, Linn WD, Aguilar C, Ramirez G. Relative efficacy of vasodilator therapy in chronic congestive heart failure: implications of randomized trials. *JAMA*. 1988;259:3422-3426.
9. Packer M, Carver JR, Rodeheffer RJ, et al. Effect of oral milrinone on mortality in severe chronic heart failure: the PROMISE Study Research Group. *N Engl J Med*. 1991;325:1468-1475.
10. Pak CY, Sakhaee K, Piziali V, et al. Slow-release sodium fluoride in the management of postmenopausal osteoporosis. *Ann Intern Med*. 1994;120:625-632.
11. Cook DJ, Guyatt GH, Adachi JD, et al. Quality of life issues in women with vertebral fractures due to osteoporosis. *Arthritis Rheum*. 1993;36:750-756.
12. Guyatt GH, Thompson PJ, Berman LB, et al. How should we measure function in patients with chronic heart and lung disease? *J Chronic Dis*. 1985;38:517-524.
13. 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.
14. Rector TS, Kubo SH, Cohn JN. Patients' self-assessment of their congestive heart failure, II: content, reliability and validity of a new measure—the Minnesota Living With Heart Failure Questionnaire. *Heart Failure*. 1987;3:198-209.
15. Juniper EF, Guyatt GH, Ferrie PJ, Griffith LE. Measuring quality of life in asthma. *Am Rev Respir Dis*. 1993;147:468-479.
16. Wiklund I, Comerford MB, Dimenas E. The relationship between exercise tolerance and quality of life in angina pectoris. *Clin Cardiol*. 1991;14:204-208.
17. Osteoporosis Quality of Life Research Group. Measuring quality of life in women with osteoporosis. *J Bone Miner Res*. In press.
18. 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.
19. Goldstein RS, Gort EH, Guyatt GH, Stubbings D, Avendano MA. Prospective randomized controlled trial of respiratory rehabilitation. *Lancet*. 1994;344:1394-1397.
20. Wasson JH, Reda DJ, Bruskevitz RC, Elinson J, Keller AM, Henderson WG. A comparison of trans-

- urethral surgery with watchful waiting for moderate symptoms of benign prostatic hyperplasia: the Veterans Affairs Cooperative Study Group on Transurethral Resection of the Prostate. *N Engl J Med.* 1995;332:75-79.
21. Guyatt GH, Mitchell A, Irvine EJ, et al. A new measure of health status for clinical trials in inflammatory bowel disease. *Gastroenterology.* 1989;96:804-810.
22. Kirshner B, Guyatt GH. A methodologic framework for assessing health indices. *J Chronic Dis.* 1985;38:27-36.
23. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol.* 1992;45:1341-1345.
24. De Weerd I, Visser AP, Kok GI, de Weerd O, Van der Veen EA. Randomized controlled multicentre evaluation of an education programme for insulin-treated diabetic patients: effects on metabolic control, quality of life, and costs of therapy. *Diabet Med.* 1991;8:338-345.
25. Irvine EJ, Feagan B, Rochon J, et al. Quality of life: a valid and reliable measure of therapeutic efficacy in the treatment of inflammatory bowel disease. *Gastroenterology.* 1994;106:287-296.
26. Greenberg GR, Feagan BG, Martin F, et al. Oral budesonide for active Crohn's disease: Canadian Inflammatory Bowel Disease Study Group. *N Engl J Med.* 1994;331:836-841.
27. Salonen R, Ashford E, Dahlof C, et al. Intranasal sumatriptan for the acute treatment of migraine. *J Neurol.* 1994;241:463-469.
28. Mathew NT, Saper JR, Silberstein SD, et al. Migraine prophylaxis with divalproex. *Arch Neurol.* 1995;52:281-286.
29. Trying S, Barbarash RA, Nahlik JE, et al. Famiciclovir for the treatment of acute herpes zoster: effects on acute disease and postherpetic neuralgia. *Ann Intern Med.* 1995;123:89-96.
30. Tugwell P, Pincus T, Yocum D, et al. Combination therapy with cyclosporin and methotrexate in severe rheumatoid arthritis: the Methotrexate-Cyclosporine Combination Study Group. *N Engl J Med.* 1995;333:137-141.
31. Kirwan JR. The effect of glucocorticoids on joint destruction in rheumatoid arthritis: the Arthritis and Rheumatism Council Low-Dose Glucocorticoid Study Group. *N Engl J Med.* 1995;333:142-146.
32. Malmivaara A, Hakkinen U, Aro T, et al. The treatment of acute low back pain: bed rest, exercises, or ordinary activity. *N Engl J Med.* 1995;332:351-355.
33. Bergner M, Bobbit 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. Tarlov AR, Ware JE Jr, Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study: an application of methods for monitoring the results of medical care. *JAMA.* 1989;262:925-930.
35. Ware JE, Kosinski M, Bayliss MS, et al. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results of the Medical Outcomes Study. *Med Care.* 1995;33:AS264-AS279.
36. Tandon PK, Stander H, Schwarz RP Jr. Analysis of quality of life data from a randomized, placebo controlled heart-failure trial. *J Clin Epidemiol.* 1989;42:955-962.
37. Smith D, Baker G, Davies G, Dewey M, Chadwick DW. Outcomes of add-on treatment with lamotrigine in partial epilepsy. *Epilepsia.* 1993;34:312-322.
38. Chang SW, Fine R, Siegel D, Chesney M, Black D, Hulley SB. The impact of diuretic therapy on reported sexual function. *Arch Intern Med.* 1991;151:2402-2408.
39. Tugwell P, Bombardier C, Buchanan WW, et al. Methotrexate in rheumatoid arthritis: impact on quality of life assessed by traditional standard-item and individualized patient preference health status questionnaires. *Arch Intern Med.* 1990;150:59-62.
40. Laupacis A, Wong C, Churchill D. The use of generic and specific quality-of-life measures in hemodialysis patients treated with erythropoietin. *Control Clin Trials.* 1991;12:168S-179S.
41. Boyle MH, Torrance GW, Sinclair JC, Horwood SJ. Economic evaluation of neonatal intensive care of very-low-birth-weight infants. *N Engl J Med.* 1983;308:1330-1337.
42. Hornberger JC, Garber AM, Chernen ME. Is high-flux dialysis cost-effective? *Int J Technol Assess Health Care.* 1993;9:85-96.
43. Hornberger JC. The hemodialysis prescription and cost effectiveness. *J Am Soc Nephrol.* 1993;4:1021-1027.
44. Naylor CD. Cost-effectiveness analysis: are the outputs worth the inputs? *ACP J Club.* 1996;124:a12-a14.
45. Testa MA, Anderson RB, Nackley JF, Hollenberg NK, for the Quality-of-Life Hypertension Study Group. Quality of life and antihypertensive therapy in men: a comparison of captopril with enalapril. *N Engl J Med.* 1993;328:907-913.
46. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol.* 1994;47:81-87.
47. Redelmeier DA, Goldstein RS, Guyatt GH. Assessing the minimal important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol.* 1996;49:1215-1219.
48. Jaeschke R, Guyatt G, Keller J, Singer J. Measurement of health status: ascertaining the meaning of a change in quality-of-life questionnaire score. *Control Clin Trials.* 1989;10:407-415.
49. Juniper EF, Guyatt GH, Feeny DH, Ferrie PJ, Griffith LE, Townsend M. Measuring quality of life in children with asthma. *Qual Life Res.* 1996;5:35-46.
50. Mitchell A, Guyatt G, Singer J, et al. Quality of life in patients with inflammatory bowel disease. *J Clin Gastroenterol.* 1988;10:306-310.