MEDICAL CARE
March 1989, Vol. 27, No. 3, Supplement

The Quality of Well-Being Scale

Applications in AIDS, Cystic Fibrosis, and Arthritis

ROBERT M. KAPLAN, PHD,* JOHN P. ANDERSON, PHD,* ALBERT W. WU, MD,*
WM. CHRISTOPHER MATHEWS, MD,* FRANKLIN KOZIN, MD,†
AND DAVID ORENSTEIN, MD‡

The Quality of Well-being (QWB) Scale combines preference-weighted measures of symptoms and functioning to provide a numerical point in-time expression of well-being that ranges from zero (0) for death to 1.0 for asymptomatic optimum functioning. The QWB includes three scales of function: mobility, physical activity, and social activity. Each step of these scales is associated with preference weights. Preference adjustments for symptoms are also included. This paper describes how this general system was used to evaluate outcomes in three different clinical conditions: acquired immune deficiency syndrome (AIDS), cystic fibrosis, and arthritis. In one study, the QWB was administered to 31 patients participating in evaluation of azidothymidine (AZT) treatment for AIDS. The QWB system demonstrated substantial benefits of AZT treatment in comparison to placebo. In a second study, the QWB and a series of pulmonary function measures were administered to 44 patients with cystic fibrosis. The QWB was demonstrated to be significantly correlated with measures of pulmonary function, including FEV_1 and maximal midexpiratory flow rate (MMEFR). In addition, there were significant associations between the QWB and measures of exercise tolerance. In the third study, the QWB and an arthritis-specific measure were administered to 83 arthritis patients before and after their treatment. The QWB was at least as capable of detecting clinical change in this population as was the disease-specific measure. For all three conditions, the QWB considered side effects and benefits of treatment in a common unit. Clinical trial data are cited to suggest that the QWB is a valuable outcome measure in arthritis treatment evaluation. We conclude that the QWB has substantial validity as a general health outcome measure and that the system can be used with different populations. Key words: quality of well-being; health status assessment; AIDS; cystic fibrosis; arthritis. (Med Care 1989; 27:S27–S43)

* From the University of California, San Diego, California.

† From the Scripps Clinic and Research Foundation, La Jolla, California.

‡ From the University of Pittsburgh, Pittsburgh, Pennsylvania.

Supported in part by the Multipurpose Arthritis Center Grant AR 33489 from the National Institutes of Health.

Address correspondence to: Robert M. Kaplan, Professor and Acting Chief, Division of Health Care Sciences M-022, University of California, San Diego, La Jolla, CA 92093.

Within the last few years, quality of life data have been used increasingly to evaluate the cost-utility and cost-effectiveness of health care programs. Such analyses require the evaluation of very different types of health care interventions using the same outcome unit. Currently, the utility of general health status measures in clinical studies is an open issue. This report presents new data from studies that have used the Quality of Well-being (QWB) Scale in three different clinical populations. It highlights

some of the strengths and weaknesses of general health outcome measures and considers the issue of general versus disease-specific measures within clinical populations.

Measurement of Health Status

The conceptualization and measurement of health status has interested scholars for many decades. Following the Eisenhower administration, a President's Commission on National Goals identified health status measurement as an important objective. Shortly after, John Kenneth Galbraith, in *The Affluent Society,* described the need to measure the effect of the health care system on "quality of life." Recent years have seen many attempts to define and measure health status.[1-3] Before considering any specific approach, it is worth noting that traditional indicators of "health" have well-identified problems.

Mortality

Mortality remains the major outcome measure in most epidemiologic studies and clinical trials. Typically, mortality is expressed in a unit of time. For mortality data to be meaningful, they must be expressed as a rate, that is the proportion of deaths from a particular cause occurring in some defined time interval (usually a year). Mortality rates are often age-adjusted. Case fatality rates express the proportion of persons who died of a particular disease divided by the total number with the disease (including those who die and those who live). Reporting mortality rates has its advantages. They are "hard" data (despite some misclassification bias[4]) and the meaning of the outcome is not difficult to comprehend. Despite their many advantages, mortality outcomes have some obvious limitations. Mortality rates consider the dead and ignore the living. Many important treatments or programs might have little or no impact on mortality rates and important illnesses, such as arthritis, have relatively little relationship to mortality.

Morbidity

The most common approach to health status assessment is to measure morbidity in terms of function or role performance. For example, morbidity estimates often include work days missed or bed disability days. Many different approaches to health status assessment using morbidity indicators have been introduced. These include, for example, the Sickness Impact Profile (SIP)[5] which represents the effect of disease or disability upon a variety of categories of behavioral function, and the RAND health status measures, which have separate categories for the effects of disease or health states upon physical function, social function, and mental function. These measures do not integrate morbidity and mortality, although as each birth cohort ages, mortality cases accrue.

Death is a health outcome, and it is important that this outcome not be excluded from any expression of health status. For example, suppose we were evaluating the effect of Program A, integrated support and treatment, against that of Program B, no support or treatment, for randomly assigned groups of very ill, elderly, nursing home residents. Let us suppose that Program A maintained patients at a low level of function throughout the year, but that in the comparison group (Program B), the sickest 10% died. Looking just at the living in the follow-up, one finds Program B patients to be healthier, because the sickest had been removed by death. By this standard, the program of no supportive treatment might be put forth as the better alternative. With a measure that combined morbidity and mortality the outcome would be very different, because mortality effects would reduce the overall health of Program B to a low level.

The Value Dimension

Scholars have debated the components of "health" for many centuries.[6] Sullivan, in a review of literature from various fields noted that most concepts of morbidity involved three types of evidence: clinical, subjective, and behavioral.[7] Clinical outcomes include clinical judgment, physical findings, laboratory tests, or results of invasive procedures. Clinical evidence is valuable if, and only if, it is clearly related to well-defined behavioral health outcomes. For example, significant abnormalities in certain blood proteins are of concern only if these deviations correlate with dysfunction and early mortality. The burden of proof is on the scientist to demonstrate these associations.

Subjective evidence includes symptoms and complaints that are also very important in health care. Although symptoms are a major correlate of health care utilization, not all symptoms should be given equal weight, because the number of symptoms does not necessarily depict the severity of health status. For example, an adult with an acute 24-hour flu may have an enormous number of symptoms. Although these can include nausea, headache, cough, sneezing, aches and pains, vomiting, and diarrhea, it is not clear that this condition is more severe than the single symptom of a very severe headache.

Several factors need to be considered. First, we must determine the degree to which the symptoms limit functioning. One individual with the following five symptoms—an itchy eye, a runny nose, coughing, fatigue, and headache—may still feel well enough to work and to perform all usual activities. Another person with the single symptom of a severe headache may be limited to bed and not be able to move around. Would we want to call the person with five symptoms less well? Another dimension is the duration of the symptoms. A year in pain is certainly worse than a day in pain. The final, and perhaps the most often neglected, factor is the value or preference associated with different types of dysfunction.

Biomedical investigators often avoid reference to values or preferences because these constructs are considered not "scientific." However, the value dimension in health status is inescapable. Fishburn defined value as the quantification of the concept of worth, importance, or desirability.[8] Ultimately, our judgments of the value of health states, and whether one level of functioning is "better" than another level of functioning, depend on subjective evaluations. If we advise individuals to change their diet to avoid heart disease, we inherently assume that the reduced probability of heart disease later in life is valued more than the immediate but enduring mild displeasure of dietary change. The phrase "quality of life" necessarily presumes a qualitative judgment.

Behavioral Dysfunction

When Sullivan[7] reviewed the literature on health measurement more than 20 years ago, he emphasized the importance of behavioral outcomes. Bolstered by the accomplishments of behavioral scientists, a convincing argument was developed suggesting that behavioral indicators such as absenteeism, bed-disability days, and institutional confinement will be the most important consequences of disease and disability. Performance of activities at different ages could be compared to societal standards for these behaviors. Restrictions in usual activity were seen as *prima facie* evidence of deviation from well-being. Many other investigators have focused on point-in-time measures of dysfunction as measures of health,[3,8,9] and these clearly are crucial in our quantification of health.

It is important not to neglect what will

happen in the future. The spectrum of medical care ranges from public health, preventive medicine, and environmental control through diagnosis to therapeutic intervention, convalescence, and rehabilitation. Many programs affect the probability of occurrence of future dysfunction, rather than altering current functional status. In many aspects of preventive care, for example, the benefit of the treatment cannot be seen until many years after the intervention. A supportive family that instills proper health habits in its children, for example, may also promote better health in the future, yet the benefit may not be realized for years. The concept of health must consider not only the ability to function now but also the probability of future changes in function. A person who is very functional and asymptomatic today may harbor a disease with a poor prognosis. Thus, many individuals are at high risk of dying from heart disease even though they are perfectly functional today. Should we call them "healthy?" The term "severity of illness" should take into consideration both dysfunction and prognosis.

Many medical treatments may cause near-term dysfunction to prevent future dysfunction. For example, coronary artery bypass surgery causes severe dysfunction for a short time, yet the surgery is presumed to enhance function or decrease mortality at a later time. Patients may be incapacitated after myocardial infarction and restricted to coronary care units. Yet the treatment is designed to help them achieve better future outcomes. Papanicalau (Pap) smears and hysterectomies are performed to decrease the probability of future cancer deaths. Much of health care involves looking into the future to enhance outcomes over the life span. Therefore, it is essential to divide health into current and future components. We prefer the term "prognosis" to describe the probability of transition among health states over the course of time.[10]

Health-related Quality of Life

The objectives of health care are twofold. First, health care and health policy should increase life expectancy. Second, the health care system should improve the quality of life during the years that people are alive. Consider various measures in health care in light of these two objectives. Traditional biomedical indicators and diagnoses are important to us because they may be related to mortality or to quality of life. We prefer the term health-related quality of life to refer to the impact of health conditions on function. Thus, health-related quality of life may be independent of quality of life relevant to work setting, housing, air pollution, or similar factors.[11]

Numerous quality-of-life measurement systems have evolved during the last 20 years. These systems are based primarily on two different conceptual approaches. The first approach grows out of the tradition of health status measurement. In the late 1960s and early 1970s, the National Center for Health Services Research funded several major projects to develop general measures of health status. Those projects resulted in the Sickness Impact Profile (SIP),[5] the Quality of Well-being Scale,[12] and the General Health Rating Index. The latter measure, originally developed at Southern Illinois University, was adapted by the RAND Corporation under grants from the office of the Assistant Secretary for Planning and Evaluation (ASPE; of the Department of Health and Human Services) and has become known as the RAND Health Status Measure.[9] These efforts usually involved extensive multidisciplinary collaboration between behavioral scientists and physicians and, perhaps not surprisingly, most are focused on the impact of disease and disability on function and observable behaviors, such as performance of social role, ability to get around the community, and physical func-

tioning. Some systems include separate components for the measurement of social and mental health. All were guided by the World Health Organization's (WHO) definition of health status: "Health is a complete state of physical, mental, and social well-being and not merely absence of disease."[13]

The second conceptual approach is based on quality of life as something independent of health status. Some investigators now use traditional psychologic measures and call them "quality-of-life" outcomes. For instance, Follick et al.[14] suggest that quality of life represents psychologic status in addition to symptoms and mortality. In fact, most investigators believe that symptoms and mortality do represent quality of life.[15] Croog et al.[16] used a wide variety of outcome measures and collectively referred to them as "quality of life." These measures included the patients' subjective evaluation of well-being, physical symptoms, sexual function, work performance and satisfaction, emotional status, cognitive function, social participation, and life satisfaction. Other investigators, including Hunt and colleagues[17] regard quality of life as subjective appraisals of life satisfaction. In summary, a wide variety of different dimensions has been described as "quality of life." Although agreement is lacking on which dimensions should be considered the standard for assessing quality of life in research studies, recurrent themes in the methodologic literature can assist in the evaluation of existing instruments.

Unidimensional Versus Multidimensional Constructs

Although all experts agree that quality of life is a multidimensional construct, they debate whether outcome measures must necessarily represent this multidimensional structure. Quality-of-life assessment can take essentially one of two major approaches: a psychometric approach and a decision–theory approach. The psychometric or profile approach attempts to provide separate measures for the many different dimensions of quality of life. Perhaps the best known example of the psychometric tradition is the SIP, which is a 136-item measure that yields 12 different scores. The scores are displayed as a profile similar to a Minnesota Multiphasic Personality Inventory (MMPI).

The decision–theory approach attempts to weight the different dimensions of health to gain a single unitary expression of health status. Supporters of this approach argue that psychometric approaches fail to consider that different health problems are not of equal concern: 100 runny noses are not the same as 100 severe abdominal pains.[15] Not uncommonly, experimental trials using the psychometric approach will find that some aspects of quality of life improve whereas others get worse. For example, a medication might reduce high blood pressure but also be associated with headaches and impotence. The decision–theory approach attempts to place an overall value on health status by weighting the different dimensions and combining them into an aggregate quality score on the grounds that the quality notion is the subjective evaluation of observable or objective health states. It thus aims to provide an overall summary measure of quality of life that integrates subjective function states, preferences for these states, morbidity, and mortality.

Disease-Specific Versus General Measures

Most health-related quality-of-life measures are designed for use with any population. Some investigators feel it is necessary to develop quality-of-life measures for specific diseases, such as cardiovascular disease. For example, The RAND Corporation Health Insurance Experiment has produced a series of booklets describing the conceptualization and measurement of "physiologic

health." Each booklet describes the problems in conceptualization and measurement of a specific condition, such as coronary heart disease.

The rationale underlying these measures is largely, but not exclusively, clinical and suggests that specific medical conditions have very specific outcomes—an advantage to clinicians. For example, heart patients are evaluated according to ejection fractions, blood gases, etc. In addition to general physiologic indicators, quality-of-life measures are also designed specifically for particular disease groups. This is best represented in the arthritis literature, where several measures have appeared in recent years.[18]

In criticizing disease-specific approaches, many investigators believe that all diseases and disabilities affect overall quality of life. In fact, the purpose of quality-of-life measurement is not to identify clinical information relevant to the disease. Instead, it seeks to determine the impact of the disease on general function. For example, a lower ejection fraction may be associated with shortness of breath, weakness, and increased risk of mortality. Medications used to control cardiovascular diseases might cause headaches, irritability, and general confusion. By focusing too specifically on clinical correlates of disease, it is argued, the general impact is overlooked. Conversely, general quality-of-life measures adequately capture a wide variety of dysfunction associated with cardiovascular diseases. This dysfunction might be in many different systems and recognized in symptoms such as confusion, tiredness, sexual impotence, and depression. These outcomes may not be specific to the disease condition.

The Quality of Well-being Scale

In this report, we apply a general health-related quality-of-life measure in three clinical populations. The clinical populations represent different age groups and clinical characteristics. One group is from the center of the life span (acquired immune deficiency syndrome, or AIDS); one group is young (cystic fibrosis); and one group is older (arthritis). One group is characterized by clinical outcomes that combine mortality and morbidity (AIDS) whereas one group is characterized by high morbidity but low mortality (arthritis).

The QWB Scale combines preference-weighted measures of symptoms and functioning to provide a numerical point-in-time expression of well-being that ranges from zero (0) for death to one (1.0) for asymptomatic optimal functioning, i.e., higher scores represent better health. Table 1 presents the symptom-problem complexes (CPX) and their preference weights. Using these symptoms does not require any assumptions either about the intensity or the duration of the symptoms and problems, or about the underlying pathology. The measure simply indicates the symptom's presence or absence on a given day.

The QWB has three function scales: mobility (MOB), physical activity (PAC), and social activity (SAC). Each step of these scales has its own associated preference weight. These are recorded in Table 2 along with the single-day QWB calculating formula (Formula 1). In the General Health Policy Model (GHPM), the QWB inputs are integrated with terms for the number of people affected and the duration of time affected to produce the output measure, which is known as the "well-year" (Formula 2).

Over the last 15 years, operational components have been defined for the GHPM to aid evaluation and resource allocation for any public issue concerning health. An early conceptual paper demonstrated that well-years are the necessary and final result of applying expected utility (decision) analysis to treatments and policies.[10] The level of wellness at particular points (or over short

TABLE 1. List of Quality of Well-being General Health Policy Model Symptom/Problem Complexes (CPX) with Calculating Weights

CPX No.	CPX Description	Weights
1	Death (not on respondent's card)	−0.727
2	Loss of consciousness such as seizure (fits), fainting, or coma ("out cold" or "knocked out")	−0.407
3	Burn over large areas of face, body, arms or legs	−0.387
4	Pain, bleeding, itching, or discharge (drainage) from sexual organs—does not include normal menstrual bleeding	−0.349
5	Trouble learning, remembering, or thinking clearly	−0.340
6	Any combination of one or more hands, feet, arms, or legs either missing, deformed (crooked), paralyzed (unable to move), or broken—includes wearing artificial limbs or braces	−0.333
7	Pain, stiffness, weakness, numbness, or other discomfort in chest, stomach (including hernia or rupture), side, neck, back, hips, or any joints or hands, feet, arms, or legs	−0.299
8	Pain, burning, bleeding, itching, or other difficulty with rectum, bowel movements, or urination (passing water)	−0.292
9	Sick or upset stomach, vomiting, or loose bowel movement, with or without fever, chills, or aching all over	−0.290
10	General tiredness, weakness, or weight loss	−0.259
11	Cough, wheezing, or shortness of breath with or without fever, chills, or aching all over	−0.257
12	Spells of feeling upset, being depressed, or crying	−0.257
13	Headache, dizziness, ringing in ears, or spells of feeling hot, nervous, or shaky	−0.244
14	Burning or itching rash on large areas of face, body, arms, or legs	−0.240
15	Trouble talking such as lisp, stuttering, hoarseness, or being unable to speak	−0.227
16	Pain or discomfort in one or both eyes (such as burning or itching) or any trouble seeing after correction	−0.230
17	Overweight for age and height or skin defect of face, body, arms, or legs such as scars, pimples, warts, bruises, or changes in color	−0.188
18	Pain in ear, tooth, jaw, throat, lips, tongue: several missing or crooked permanent teeth—includes wearing bridges or false teeth: stuffy, runny nose: or any trouble hearing—includes wearing a hearing aid	−0.170
19	Taking medication or staying on a prescribed diet for health reasons	−0.144
20	Wore eyeglasses or contact lenses	−0.101
21	Breathing smog or unpleasant air	−0.101
22	No symptoms or problems (not on respondent's card)	−0.000
23	Standard symptom/problem	−0.257

From Kaplan RM. Anderson JP. In: Walker S. ed. Quality of life: Assessment and applications. London: MTP Press, 1988.

intervals) is combined with prognoses (transition rates or probabilities) generated by the underlying disease or injury under different treatment conditions. Well-years result from integrating the level of wellness, or health-related quality of life, over the life expectancy.

A major advantage of the GHPM analysis is that it integrates many diverse health system inputs, such as dollar costs, dollar savings, discount rates, morbidity, mortality, consumer preferences for health states, and time dimension to create a standardized program output, i.e., dollars per well-year. This single evaluative expression allows the outputs of different health programs to be compared directly with one another.

Although the QWB system has been used in a variety of policy analyses and has undergone considerable methodologic development, its application in clinical populations has been somewhat limited. This re-

TABLE 2. Quality of Well-being General Health Policy Model and Sample Calculation

Step No.	Step Definition	Weight
Mobility scale (MOB)		
5	No limitations for health reasons	−0.000
4	Did not drive a car, health related: did not ride in a car as usual for age (15 yr) (health related), *and/or* did not use public transportation (health related), *or* had or would have used more help than usual for age to use public transportation (health related)	−0.062
2	In hospital, health related	−0.090
Physical activity scale (PAC)		
4	No limitations for health reasons	−0.000
3	In wheelchair, moved of controlled movement of wheelchair without help from someone else, *or* had trouble or did not try to lift, stoop, bend over, or use stairs or inclines (health related) *and/or* limped, used a cane, crutches, or walker (health related), *and/or* had any other physical limitation in walking, or did not try to walk as far or as fast as others the same age are able (health related)	−0.060
1	In wheelchair, did not move or control the movement of wheelchair without help from someone else, *or* in bed, chair, or couch for most or all of the day (health related)	−0.077
Social activity scale (SAC)		
5	No limitations for health reasons	−0.000
4	Limited in other (e.g., recreational) role activity (health related)	−0.061
3	Limited in major (primary) role activity (health related)	−0.061
2	Performed no major role activity (health related) but did perform selfcare activities	−0.061
1	Performed no major role (health related) *and* did not perform or had more trouble than usual in performance of one or more self-care activities (health related)	−0.106

Calculating formulas

Formula 1: Point-in-time well-being score for an individual (W):

$$W = 1 + (CPXwt) + (MOBwt) + PACwt + SACwt,$$

where wt is the preference-weighted measure for each factor and CPX is the symptom/problem complex. For example, the W score for a person with the following description profile may be calculated for one day as follows:

Quality of Well-being Element	Step Definition	Weight
CPX-11	Cough, wheezing, or shortness of breath, with or without fever, chills, or aching all over	−0.257
MOB-5	No limitations	−0.000
PAC-1	In bed, chair, or couch for most or all of day (health related)	−0.077
SAC-2	Performed no major role activity (health related) but did perform selfcare.	−0.061

$$W = 1 + -0.257 + -0.000 + -0.007 + -0.061 = 0.605$$

Formula 2: Well years (WY) as an output measure:

$$WY = [\text{No. of persons} \times (CPXwt + MOBwt + PACwt + SACwt)] \times \text{time}$$

From Kaplan RM, Anderson JP. In: Walker S, ed. Quality of life: Assessment and applications. London: MTP Press, 1988.

port presents new data on applications of the QWB system in three clinical populations: AIDS, cystic fibrosis, and arthritis.

Clinical Populations for QWB Applications AIDS

AIDS has commanded great attention as a new infectious disease that is increasing at a logarithmic rate and is uniformly fatal. As of March 1988, a total of 81,433 AIDS cases had been reported from 133 nations worldwide. In the United States, 55,167 cases have been reported.[19] Between 1 and 1.5 million Americans are estimated to be infected with the AIDS virus, now known as the human immunodeficiency virus (HIV).[20]

In addition to opportunistic infections and malignancies that define AIDS, HIV infection may cause a broad range of disease, including persistent lymphadenopathy, thrombocytopenia, immune complex disease, wasting, various constitutional symptoms, and HIV neurologic diseases. The impact of HIV infection on functioning is equally diverse. For example, HIV infection may result in fatigue, arthritis, blindness, memory loss, or paraplegia. Treatments for HIV infection should be designed to prevent early mortality and to reduce morbidity during periods before death.

Because of the major public health threat associated with AIDS, efforts to find new therapeutic approaches to manage this serious condition have been intense. In this section we report preliminary data using the QWB system in an experimental trial evaluating azidothymidine (AZT)* treatment for AIDS patients.

AZT has been available by prescription since September 1987. Its use in treating patients with advanced HIV infection is predicated on the encouraging results of early clinical trials. In the recently completed multicenter phase II AZT trial, 19 of 137 placebo recipients died as compared with 1 of 145 AZT recipients. The incidence of opportunistic infections was also significantly reduced among AZT recipients. Thus, in certain groups of patients, AZT may profoundly lower both mortality and morbidity.[21] However, serious side effects are frequently associated with AZT, including anemia, neutropenia, nausea, myalgia, insomnia, and severe headache. Nearly one third (31%) of patients who received AZT required blood transfusions for anemia.[22]

Local Trial

In the San Diego arm of the multicenter AZT trial, Wu et al. obtained outcome data using the QWB and the Karnofsky Performance Status measure.[23] The participants in the study were 31 patients (27 male, 4 female) with a clinical diagnosis of either AIDS or severe AIDS-related complex (ARC). They were randomly assigned to receive AZT treatment or placebo and were evaluated using the QWB before beginning the trial and at eight follow-up visits over the next 52 weeks. The value of the treatment was estimated using the repeated measures analysis of variance (calculated using a general linear model).

The baseline characteristics of the subjects are given in Table 3. The patients were divided into those with CD4 cell (also known as T_4 lymphocytes or T-helper cells) counts less than or greater than 100×10^9/L. Patients in both groups were comparable at baseline with regard to age, CD4 group, sex, diagnosis (AIDS or ARC), Karnofsky score, and QWB (*t*-test and chi-square, $P > .15$). The mean initial CD4 count was significantly higher in the AZT group ($P < .03$).

For the QWB measure, the repeated measures analysis of variance (ANOVA) showed a significant effect of time ($F\ 8/172 = 9.97$, $P < .0001$). The interaction between group and time of testing ($F\ 8/172 = 4.01$, $P < .0002$), illustrated in Figure 1, is the crucial component in evaluating treatment effectiveness. It suggests that there

* Azidothymidine (AZT) is also known by the generic name zidovudine (ZOV).

TABLE 3. Baseline Characteristics of Patients in San Diego Arm of AZT Trial

Characteristics	AZT	Placebo	Total	*P* value
N	16	15	31	
Male	14	13	27	
Female	2	2	4	
Mean Age	36.9	34.7	35.8	>.54
N AIDS	5	8	13	
N ARC	11	7	18	
CD4 <100	9	8	17	
>100	7	7	14	
Mean CD4	188 ± 198	116 ± 106		<.03
Mean QWB	.6486 ± .0735	.6340 ± .0844		>.60

AIDS: acquired immune deficiency syndrome, ARC: AIDS-related complex, CD4: also known as T_4 lymphocytes or T-helper cells, QWB: Quality of Well-being Scale.

was a differential rate of change between AZT treated and control groups. As Figure 1 demonstrates, QWB scores remained relatively constant over the course of time for the AZT group, while they declined substantially for the placebo group.

These results suggest that the QWB can detect strong treatment effects associated with AZT treatment. One advantage of the QWB system is that it allows the expression of program benefits in terms of well-year units and the comparison of treatment alternatives that are very different from one another. In the AZT trial, the placebo-treated group experienced substantial mortality and greater morbidity than the AZT group. Neither measures of mortality alone nor of morbidity alone were as capable of detecting the potent treatment effect of this medication as the QWB. Further, the comprehensive measure takes the side effects of AZT into account and expresses outcome as the net benefits minus adverse effects.

Cystic Fibrosis

Cystic fibrosis (CF) affects approximately 30,000 Americans, has a prevalence of 1 in 2,000 live births, and is the most common inherited cause of early death. When the disease was first described 50 years ago, virtually all patients died before the age of 2 years. Average survival has now increased to 26 years and, although there is no cure, life expectancy for CF patients continues to improve.

Clinical studies of CF typically use pulmonary function tests, chest roentgenograms, and clinical judgment to assess treatment effectiveness. Although pulmonary function testing and exercise evaluations provide objective physiologic measures of disease severity, they may not be sensitive to important aspects of disease progression and treatment effects. During acute exacerbations of pulmonary infection, CF patients are commonly treated with potent intravenous aminoglycoside antibiotics. Although these antibiotics may reduce the infection and improve pulmonary function, they may have undesirable side effects, in some cases causing irreversible deafness or renal dysfunction.

Thus, although measures of outcome in terms of pulmonary function may show improvement in that clinical domain, an overall measure of well-being should consider both the benefits and consequences of the powerful treatments. Furthermore, many of the important consequences of CF may be overlooked with traditional measures and pulmonary function tests. These include upset stomachs, headaches, chest pain, bone and joint pain, coughing, and short-

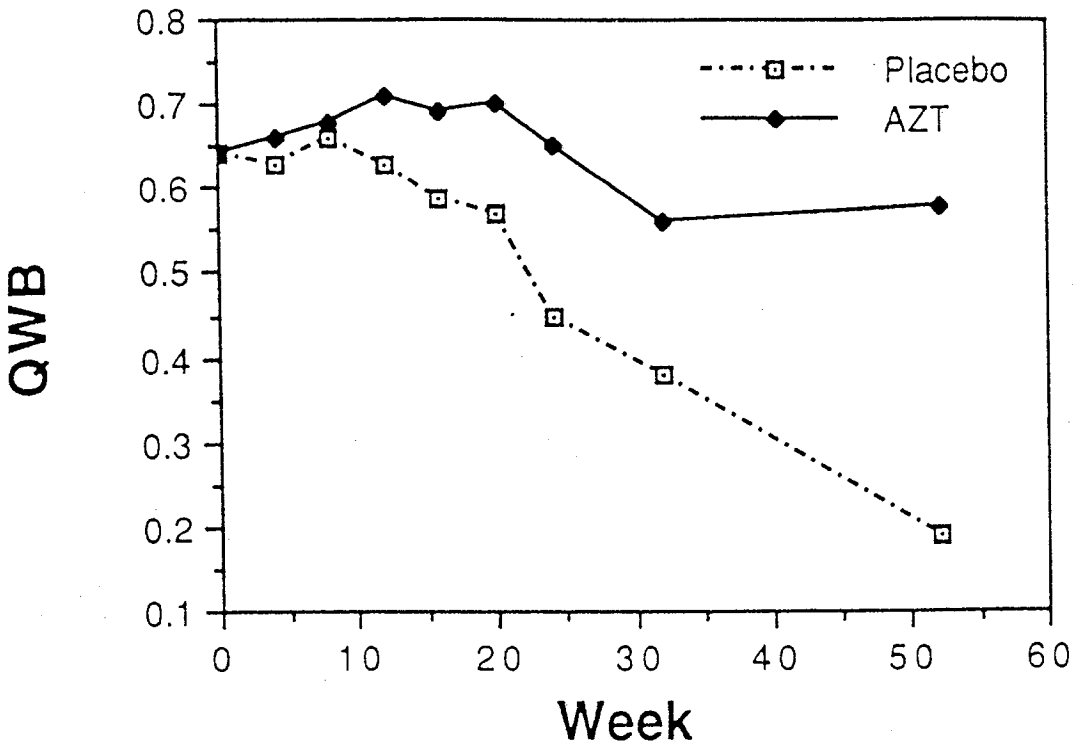

FIG. 1. QWB outcomes of AIDS patients treated with AZT or placebo.

ness of breath. Considering the goals of extending life and improving quality of life, a general health status measure seems quite appropriate for studies of CF patients.

Physicians caring for CF patients have expressed considerable interest in developing quality of life outcome measures for their patients. The goal of any treatment program for patients with CF should be to improve, or at least prevent the deterioration of, the quality of life. Because death is an important outcome in CF, measures that integrate mortality and morbidity are needed. Currently, CF patients are typically rated on the basis of simple clinical judgment scales, but the QWB is now used in several CF centers.

In preparation for a clinical trial evaluating exercise treatments for CF patients, Orenstein and his colleagues completed a QWB validity study for a group of CF patients.[24] They administered the QWB to 44 patients (19 female, 25 male) who ranged in age from 7 to 36 years (mean = 16.5, SD = 6.9 years) and represented a wide diversity of severity of their pulmonary disease. Pulmonary function was measured using standard spirometric methods defined by the guidelines of the Snowbird conference.[25]

QWB scores were plotted against two representative tests of pulmonary function: forced expired volume in one second (FEV_1) and maximal midexpiratory flow rate (MMEFR). The QWB also was administered to 15 patients who performed progressive exercise tests; their scores were plotted against one representative test of exercise tolerance (peak oxygen consumption, VO_2 max). Two-dimensional scatter plots for these relationships showed highly significant associations between the QWB scores

and measures of pulmonary function and exercise tolerance. These correlations are summarized as follows:

Criterion	*Correlation*	*P*
FEV_1	.5518	.0001
MMEFR	.4793	.001
VO_2max	.5778	.01

These pilot data strongly indicate the usefulness of the QWB for CF patients. One of the values of the QWB is that it is not age-specific; indeed, the associations appeared to be equally strong for CF patients across different age groups. As with patients with chronic obstructive pulmonary disease (COPD), the correlation between pulmonary function and QWB score among CF patients is substantial. The QWB was significantly correlated with traditional outcome measures such as pulmonary function and exercise tolerance. However, the correlation was not perfect ($r < 1.0$), indicating that QWB captures aspects of life quality in addition to pulmonary function. Pulmonary function and exercise testing do provide objective measures of disease severity and progression. However, they may not be sensitive to many other important aspects of the disease and within the same level of pulmonary function, patients may show considerable variability in their daily activities. These early data suggest that the QWB has validity as an outcome measure in this specific clinical population. In clinical trials, CF-specific measures may also be used even though the use of a general measure allows the cost-utility for interventions in this patient population to be compared with those for populations that suffer from different clinical conditions.

Arthritis

Arthritis is the major cause of activity limitation in the United States. We consider arthritis in this report for two reasons. First, arthritis has relatively little impact upon mortality in contrast to cystic fibrosis and AIDS. Second, many published arguments favor disease-specific measures of arthritis over general outcome measures. We summarize the results of a published clinical trial and present new evidence that the general QWB performs as well as disease-specific measures in capturing clinical changes.

Clinical outcomes in studies of rheumatology have been difficult to evaluate.[26] Clinical measures often include joint tenderness, grip strength, and joint circumference. Some studies show that often the reliability of these measures is poor.[27] Fries[28] questioned the relevance and reliability of various traditional outcome measures, including laboratory measures of erythrocyte sedimentation rate (ESR), latex fixation titer, and hemoglobin, since rheumatoid arthritis (RA) patients may develop serologic abnormalities that do not precisely coincide with joint inflammation.[29] In addition, Fries suggested that traditional clinical measures such as grip strength, walking time, and patient global assessment are merely surrogates for true outcome in arthritis, which, he argued, are disability, physical discomfort, and financial loss. He asserted that pain and functional outcomes are most meaningful to the patients. Patients strive for extended life expectancy and for improved function during the years they are alive. Laboratory findings may predict this dysfunction and are important only for that reason.

Arthritis-specific measures

Most health-related quality-of-life measures developed before 1980 were designed for use with any disease. Recently, some investigators have promoted using disease-specific quality-of-life measures, usually by combining a general measure with disease-specific measures. Nowhere is there a better example of this interplay than in research on rheumatoid arthritis. Investigators studying new treatments for arthritis have developed a series of quality-of-life measures that

apply only to patients with arthritis. This field was reviewed in a special issue of the *Journal of Rheumatology* in 1982. Most of the new scales have been adapted from ADL[30] or other functional status measures,[31] but they typically include a series of items that specifically measure the impact of arthritis on daily functioning.

Despite the wide variety of arthritis-specific measures now available, it is unclear whether they have advantages over traditional nondisease-specific approaches. There are several difficulties. First, arthritis-specific measures may overestimate the impact of innovative arthritis treatments, because they are not sensitive to many of the side effects of arthritis treatments on nonmusculoskeletal systems. For example, oral gold therapy may cause gastrointestinal disturbances that may be severe on occasion, and anti-inflammatory and immunosuppressive drugs can cause a variety of side effects. Because measures that focus only on arthritis symptoms and dysfunction may neglect some of the consequences of treatment, global measures are better able to capture the net side effects as well as benefits.

To compare a general with a specific measure, we administered both the specific Arthritis Impact Measurement Scale (AIMS) and the general QWB to 83 adults who were being treated for various musculoskeletal disorders, including osteoarthritis, rheumatoid arthritis, and back pain. Each patient was evaluated twice—at the initiation of treatment and again 2 to 4 weeks later. The AIMS is a multidimensional health index designed at the Multipurpose Arthritis Center (MAC) at Boston University; it is intended to measure physical health and social and psychologic well-being specifically in patients with arthritis. It consists of 66 questions divided into nine scales with four to seven items in each scale. These scales are combined into three health status components: physical function, psychologic status, and pain level. The scaling properties, reliability, and validity of the AIMS have been studied extensively,[32,33] and it is widely regarded as a representative arthritis-specific measure.

TABLE 4. Point in Time Correlations Between QWB and AIMS

Variable	QWB	AIMS-Pys	AIMS-Pain	AIMS-Psych
QWB	1.00	−.57	−.40	−.17
AIMS-Pys		1.00	.46	.33
AIMS-Pain			1.00	.34
AIMS-Psych				1.00

QWB is Quality of Well-being Scale; AIMS is Arthritis Impact Measurement Scale (AIMS-Pys is the physical dimension, and Psych is the psychologic dimension).

The matrix of correlations between various pairs of measures is given in Table 4. Higher QWB scores would be expected to be associated with lower AIMS scores, because high QWB scores represent good function whereas the other high scores represent poor function. The QWB shows substantial negative correlations with the other measures (Table 4). The largest correlation (−.57) was between the QWB and the physical function component of the AIMS. Dynamic correlations based on changes between introductory and follow-up interviews are shown in Table 5. These

TABLE 5. Dynamic time Correlations Between QWB and AIMS

Variable	QWB	AIMS-Pys	AIMS-Pain	AIMS-Psych
QWB	1.00	−.28	−.28	−.32
AIMS-Pys		1.00	.20	.34
AIMS-Pain			1.00	.39
AIMS-Psych				1.00

Note: See Table 4 for definitions.

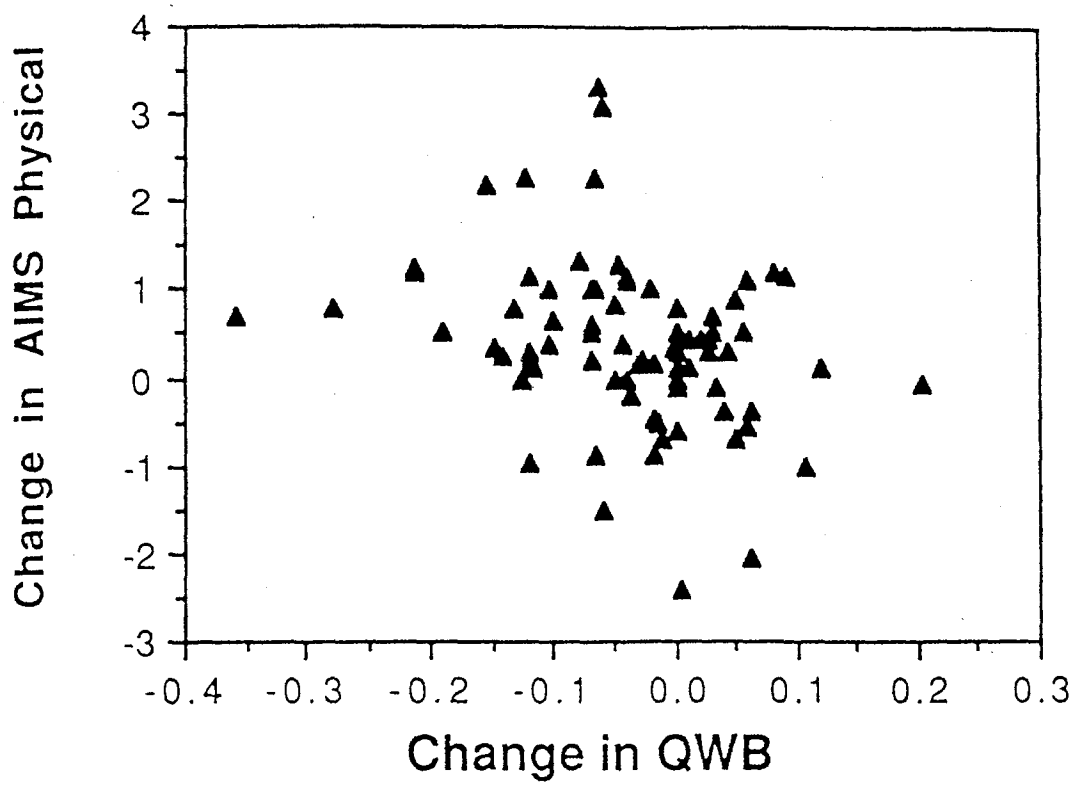

FIG. 2. Scatterplot of dynamic correlations between QWB and AIMS physical.

correlations evaluate whether clinical changes, assessed by different measures, go in the same direction. Changes in the QWB are significantly associated with changes in the physical and the psychologic components of the AIMS scale.

The data in Figure 2 illustrate changes in the physical function component of the AIMS plotted against changes in the QWB; they suggest that the two measures may be tapping something different. The AIMS physical function change varies considerably among those patients who show no change on the QWB. Conversely, the QWB varies greatly for patients who show no change on the AIMS. Changes in the QWB were well explained by changes in the AIMS factors (F 3/75 = 5.06, $P < .003$) with each AIMS score contributing about equal weight. Table 6 shows that there was a significant improvement on all measures. However, the effect size for the QWB was exceeded only by that for the psychologic component for the AIMS.

In summary, we find substantial evidence that a general quality-of-life measure can tap many of the same outcomes that are captured in disease-specific measures of arthritis.

TABLE 6. Pretreatment versus Posttreatment Changes in QWB and AIMS Measures

	Time			
Variable	Pre	Post	t	p
QWB	.609	.647	−3.74	.001
AIMS-Pys	2.67	2.29	3.56	.001
AIMS-Pain	5.06	4.41	2.90	.005
AIMS-Psych	3.09	2.51	4.51	.001

See Table 4 for definitions.

Clinical Trial Applications. Clinical trials of rheumatoid arthritis treatments have considered a wide variety of endpoints. Traditionally clinical outcomes, such as degree of synovitis, are assessed typically by measuring joint swelling or tenderness. The degree of inflammation may be evaluated by noting the duration of morning stiffness or the extent of laboratory abnormalities. Functional measures, such as grip strength and time to walk 50 feet, also may be measured. Speakers at an international conference on outcome measurement in arthritis, however, suggested that comprehensive assessment of quality-of-life outcomes was highly desirable.[34]

Such assessments were obtained in a recent multicenter trial studying 300 patients who were randomly assigned to therapy with oral gold (auranofin) or a placebo. Among a wide variety of traditional and nontraditional measures used in that study, the QWB was found to be quite sensitive to clinical change. The group that received placebo demonstrated essentially no change in the QWB scores, whereas the group receiving auranofin showed a mean improvement of 0.023; this difference was highly statistically significant $P > .005$. Auranofin does not reach pharmacologically effective levels for about 2 months, and QWB scores for the treatment and placebo groups begin to diverge at about this time. Considering the many measures used in the trial, the percentage of variance accounted for with the QWB measure was among the more significant. Traditional clinical measures, such as the 50-foot walk and duration of morning stiffness, were not statistically significant, although they did favor the auranofin group. In addition, simple self-ratings by both patients and physicians failed to detect the significant effect. However, a significant network of associations emerged suggesting that the QWB was associated with similar measures of general function.[35]

What is the clinical meaning of a difference of .023 on the QWB? A difference of .023 translates into 2.3 well-years for each 100 patients who maintain the difference for 1 year. The entire continuum from death to optimal health is represented on a 0 to 1.0 scale in the QWB so that 0.023 represents a change of 2.3%. The differences observed in the auranofin trial are important especially when compared with changes produced through other medical treatments. We believe that a major advantage of the QWB will emerge as newer trials use it to compare treatment effectiveness. For example, it may be possible to say that treatment X produces 10 well-years per 100 patients four times "better" than gold) or that treatment Y produces 1 well-year (i.e., not as effective as gold).

Discussion

General quality-of-life measures have disadvantages and advantages for clinical outcome studies. One unavoidable disadvantage is that they may miss some important improvement that is specific for a particular population. Some observers also have argued that general measures are less sensitive to clinical change. Our data do not support this assertion.

The general outcome measures appear to capture clinical change, as well as, if not better than, disease-specific measures—a distinct advantage. This observation appears to be counter intuitive, but careful inspection of the QWB questions provides a reasonable explanation. For example, an arthritis patient who is unable to button his shirt will need help with self-care. This patient may not see himself or herself as requiring assistance but may be willing to admit to difficulty with the task. The subtle differences are recognized better by the QWB than by disease-specific measures. The general measure may perform more precisely because the questions have been systematically studied and revised on the basis of thousands of administrations with heterogeneous groups of patients and nonpatients. Our current results appear to con-

trast with those of Liang and colleagues[36] who suggested that the QWB is less efficient than the AIMS in detecting improvement in pain level and function in patients with arthritis. However, the same authors found that the QWB was more efficient than several other specific measures in detecting functional improvement on other global measures. Unfortunately, the Liang results are difficult to interpret because the reported mean QWB score in this study was 33.3, a value that is theoretically impossible to obtain.

A major advantage of the general approaches is that they allow the expression of program benefits in terms of well-year units. Using these common units, one can compare treatment interventions that are very different from one another. For example, consider the impact of therapies for AIDS and arthritis. The treatment of rheumatoid arthritis using auranofin produces a net difference of about .023 QWB units per patient. The mean difference between AZT and placebo-treated AIDS patients was .47 QWB units at the final follow-up assessment. To date, the AZT effect is the largest one observed in studies of medical, surgical, or behavioral treatments. The treatment of AIDS with AZT is also considerably more costly than the treatment of arthritis patients with auranofin. Yet the duration of the benefit for arthritis patients may be longer and, over years, the discrepancy between these two cases may diminish. At this point, there are too few data to make further direct comparisons.

This article describes some of our continuing work toward the development of a General Health Policy Model (GHPM). The GHPM can be used for program evaluation, population monitoring, clinical research, and policy analysis. The QWB is an important component of the model, in that it combines preference-weighted measures of symptoms and functioning to provide a numerical point-in-time expression of well-being ranging from 0 (for death) to 1.0 (for asymptomatic optimum functioning). Level of wellness at particular points in time are governed by the prognosis (transition rates or probabilities) generated by the underlying disease or injury. Well-years result from integrating the level of wellness, or health-related quality of life over the life expectancy.

The QWB system has been criticized because it has seen fewer clinical applications than have other health-related quality-of-life measures. In this article, we suggest that the QWB has substantial potential for clinical research. In comparison to other measures, it may be more capable of generating data that can be used in policy analysis, particularly when comparisons across very different treatment options are analyzed. However, because the QWB system is less capable of pin pointing the specific aspects of function that are affected by diseases and their treatments, the GHPM may be better suited for comparisons between treatments or conditions that have different specific objectives. It allows tradeoffs between risks and benefits that are typically expressed in different units. Other measures such as the SIP may be better suited for helping clinicians identify which specific aspects of function are affected by the disease or treatment.

Considerably more research on the validity, reliability, and generalizability of the QWB system will be required. In particular, the preference weighting system needs to be restandardized on a larger sample of respondents. Another problem that deserves greater attention is identification of the appropriate discount rate. Finally, we encourage the development of general health-related quality-of-life data in resources such as the National Health Interview Survey and the National Health and Nutrition Examination Survey.

Editor's note: For discussions related to these points, see the articles by Mulley, by Lipscomb, and by Erickson et al. in this issue.

References

1. Walker S, Rosser R, eds. Quality of life: assessment and applications. London: WTP Press, 1988.

2. Wenger NK, Mattson ME, Furberg CD, Elinson J. Assessment of quality of life in clinical trials of cardiovascular therapies. New York: LaJacq Publishing, 1984.

3. Bergner M. Measurement of health status. Med Care 1985;23:696.

4. National Institutes of Health Epidemiology of Respiratory Diseases Task Force, Washington, D.C., U.S. Government Printing Office, NIH Publication, 1979;81:2019.

5. 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.

6. Fishburn P. Decision and value theory. New York: John Wiley and Sons, 1964.

7. Sullivan DF. Conceptual problems in developing an index of health. Office of Health Statistics, National Center for Health Statistics, Monograph Series II, No. 17, 1966.

8. Katz ST, Downs H, Cash H, Grotz R. Progress and development of an index of EDL. Gerontologist 1970;10:20.

9. Stewart AL, Ware JE, Brook RH, Davies-Avery A. Conceptualization and measurement of health for adults, vol II: physical health in terms of functioning. Santa Monica, CA: The RAND Corporation, 1978.

10. Fanshel S, Bush JW. A health-status index and its applications to health-services outcomes. Operations Research, 1970;18:1021.

11. Rice RM. Organizational work and the overall quality of life. In: Oscamp S, ed. Applied social psychology annual: applications in organizational settings, vol 5. Beverly Hills, CA: Sage 1984:155.

12. Kaplan RM, Bush JW. Health-related quality of life measurement for evaluation research and policy analysis. Health Psychol 1982;1:61.

13. World Health Organization. Constitution of the World Health Organization. Geneva: WHO Basic Documents, 1948.

14. Follick MJ, Gorkin L, Smith T, et al. Quality of life post-myocardial infarction: the effects of a transtelephonic coronary intervention system. Health Psychol 1988;7:169.

15. Bush JW. Relative preferences versus relative frequencies in health-related quality of life evaluations. In: Wenger NK, Mattson ME, Furberg CD, Elinson J, eds. Assessment of quality of life in clinical trials of cardiovascular therapies. New York: LaJacq Publishing 1984:118.

16. Croog SH, Levine S, Testa MA, et al. The effects of anti-hypertensive therapy on quality of life. N Engl J Med 1986;314:1657.

17. Hunt SM, McEwen J. The development of a subjective health indicator. Sociol Health Illness 1983;2:231.

18. Liang NH, Cullen K, Larson M. In search of a more perfect mouse trap (health status or a quality of life instrument). J Rheumatol 1982;9:775.

19. Goldsmith MF. AIDS around the world: analyzing complex patterns. JAMA 1988;259:1917.

20. CDC. Human immunodeficiency virus infection in the United States: a review of current knowledge. MMWR 1987;369(suppl 6):1.

21. Fischl MA, Richman DD, Grieco MH, et al. The efficiency of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex: a double-blind, placebo-controlled trial. N Engl J Med 1987;317:185.

22. Richman DD, Fischl MA, Grieco MH, et al. The toxicity of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex: a double-blind, placebo-controlled trial. N Engl J Med 1987;317:192.

23. Wu AW, Mathews WC, Brysk LT, et al. Quality of life measures in AIDS clinical trials. Presented at the IV International AIDS Conference, Stockholm, June 1988.

24. Orenstein DM, Nixon PA, Ross EA, Kaplan RM. Quality of well-being in cystic fibrosis. Chest (in press).

25. American Thoracic Society. Snowbird Workshop on Standardization of Spirometry, 1979.

26. Deyo, RA. Measuring the quality of life of patients with rheumatoid arthritis in Walker SR, Rosser RM, eds. Quality of life: assessment and application. London: MTP Press, 1988:205.

27. Buchanan WW. Assessment of joint tenderness, grip strength, digital joint circumference and morning stiffness in rheumatoid arthritis. J Rheumatol 1982;9:763.

28. Fries JR. Toward an understanding of patient outcome measurement. Arthritis Rheum 1983;26:697.

29. McCarty DJ. Clinical assessment of arthritis. In: McCarty DJ, ed. Arthritis and allied conditions, 9th ed, Philadelphia: Lea and Febiger, 1986:131.

30. Katz ST, Ford AD, Moskowitz RW, Jackson BA, et al. Studies of illness in the aged: The index of ADL. JAMA 1983;185:914.

31. Kaplan RM, Bush JW, Berry CC. Health status: types of validity for an index of well-being. Health Serv Res 1976;11:478.

32. Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis: the Arthritis Impact Measurement Scales. Arthritis Rheum 1980;23:146.

33. Meenan RF, Yelan EH, Nebbit M, et al. The impact of chronic disease: a sociomedical profile of rheumatoid arthritis. Arthritis Rheum 1981;24:544.

34. Bombardier C, Tugwell P, Sinclair A, et al. Preferences for endpoint measures in clinical trial results of structured workshops. J Rheumatol 1982;9:798.

35. Bombardier C, Ware J, Russell IJ, et al. Auranofin therapy and quality of life for patients with rheumatoid arthritis: results of a multi-center trial. Am J Med 1986;81:565.

36. Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis Rheum 1988;28:542.