
Articles

A General Health Policy Model: Update and Applications

Robert M. Kaplan and John P. Anderson

This article describes the development of a General Health Policy Model that can be used for program evaluation, population monitoring, clinical research, and policy analysis. An important component of the model, the Quality of Well-being scale (QWB) 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. The level of wellness at particular points in time is governed by the prognosis (transition rates or probabilities) generated by the underlying disease or injury under different treatment (control) variables. Well-years result from integrating the level of wellness, or health-related quality of life, over the life expectancy. Several issues relevant to the application of the model are discussed. It is suggested that a quality of life measure need not have separate components for social and mental health. Social health has been difficult to define; social support may be a poor criterion for resource allocation; and some evidence suggests that aspects of mental health are captured by the general measure. Although it has been suggested that measures of child health should differ from those used for adults, we argue that a separate conceptualization of child health creates new problems for policy analysis. After offering several applications of the model for the evaluation of prevention programs, we conclude that many of the advantages of general measures have been overlooked and should be given serious consideration in future studies.

This study was supported under Grants K04 HL 00809 and R01 HL 34732 from the Lung Division, National Heart, Lung, and Blood Institute. It was prepared for the Brookings Institution Conference on "Prevention as a Way to Improve Work Capacity," Washington, DC, May 8-9, 1986.

Address correspondence and requests for reprints to Robert M. Kaplan, Ph.D., Professor and Acting Chief, Division of Health Care Sciences, Department of Community and Family Medicine, Mail Code M-022, University of California, San Diego, La Jolla, CA 92093. John P. Anderson, Ph.D., is a Research Specialist in the Department of Community and Family Medicine at the University of California at San Diego.

OVERVIEW

The conceptualization and measurement of health status and quality of life is gaining increasing attention in the health services and biomedical literature (Wenger, Mattson, and Furberg, 1984; Bergner, 1987). Today, quality of life measures are commonly used in evaluating clinical trials, and the Food and Drug Administration is urging quality of life assessments for new products (Walker and Ross, 1988).

In 1970, Fanshel and Bush proposed a comprehensive model for health measurement, policy analysis, and program planning. By 1973, the model was operationalized (Patrick, Bush, and Chen, 1973a and 1973b) and preliminary analyses were performed (Bush, Chen, and Patrick, 1973). Over the next decade, the model underwent many revisions (Chen, Bush, and Patrick, 1975; Kaplan and Bush, 1982). Field testing helped establish the validity (Kaplan, Bush, and Berry, 1976), generalizability (Kaplan, Bush, and Berry, 1978), and reliability of the model (Anderson et al., unpublished), and provoked additional refinements.

Specific problems with the administration and the working of questionnaires were also evaluated (Anderson et al., 1986). Within the last few years there have been several new refinements and a growing number of applications. In addition, some appropriate questions about the value of the model for certain situations have been raised. In this article, we bring together some of the developments and applications of the model, and we respond to some of the concerns that have been raised in the literature.

Before describing the general model, a discussion of some pervasive issues in health measurement is presented. Then the article describes applications of the model in screening, prevention, and tertiary care. Finally, we discuss benefit-cost/utility studies in policy analysis.

HEALTH STATUS MEASUREMENT

In a variety of publications, we have argued that a single index of health status is both feasible and highly desirable (Kaplan and Bush, 1982; Kaplan, Bush, and Berry, 1976; Kaplan, Bush, and Berry, 1978; Kaplan, 1982; Kaplan, Bush, and Berry, 1979; Kaplan and Ernst, 1983; Kaplan, 1985a and 1985b; Kaplan, Atkins, and Timms, 1984; Kaplan, Ries, and Atkins, 1985). The original model was cast in an operations research-engineering framework (Anderson and Moser,

1985; Fanshel and Bush, 1970). Later, the value of the model was described in relation to social indicators in health planning research (Chen, Bush, and Patrick, 1975) and linear programming for policy analysis (Chen and Bush, 1976).

Health status measurement has been characterized by competing traditions. One of the major issues is disease specificity. Some investigators argue that specific measures are required for each disease category. Others, including our group, believe that there are many advantages to a general approach. Among those favoring the general approach to health status measurement, some groups have focused on mortality while others have focused on morbidity. Our approach is to integrate morbidity and mortality into common units of health status. In the following paragraphs, we will elaborate on each of these issues.

GENERAL VERSUS SPECIFIC HEALTH MEASURES

Most health-related quality of life measures are designed for use with any population. However, some investigators feel that it is necessary to develop quality of life measures for specific diseases. For example, the RAND Corporation has produced a series of booklets describing the conceptualization and measurement of "physiologic health." Each booklet describes the problems that arise in conceptualizing and measuring a specific condition—such as anemia, acne, vision impairment, and so on. The rationale underlying the development of these measures is largely clinical. It suggests that medical conditions have very specific outcomes. Diabetic patients are evaluated according to blood glucose, chronic obstructive lung disease patients are evaluated according to pulmonary function, and so on. Clearly there are advantages to the clinician in considering outcomes relative to specific diseases. In addition to general physiological indicators, there are also quality of life measures designed specifically for particular disease groups. For instance, Meenan and colleagues have developed a specific quality of life measure for arthritis patients (1982), and this is only one among many approaches to health status assessment for this particular disease (Liang, Cullen, and Larson, 1982).

Though useful for many purposes, these disease-specific measures have a weakness from the policy point of view: their use precludes the possibility of comparing the benefits of programs that are directed at different populations or groups suffering from different diseases. In addition, many preventive programs affect outcomes that are not system specific. For example, cigarette smoking may increase the probability of coronary heart disease, peripheral artery occlusion, cerebrovascular dis-

ease, and cancer of the larynx, lung, mouth, esophagus, bladder, pancreas, and stomach. It is also the major cause of chronic obstructive pulmonary disease (Fielding, 1985). The impact of smoking upon health is truly overwhelming. But only a general health status measure can provide a comprehensive summary of these heterogeneous health effects. Policy analysis also requires a more general approach to health status assessment.

Another advantage of general measures is their ability to capture side effects and benefits that were not anticipated. For example, anti-arthritis drugs can produce gastric irritation that is sometimes debilitating. A disease-specific measure that focuses on grip strength and range of motion would miss these problems. The general system expresses outcomes as the net benefits minus side effects. Disease-specific measures that miss unexpected side effects can overestimate treatment effects. Interestingly, the few studies that include both general and disease-specific measures have shown that the general measures are equally, if not more, sensitive to disease-specific changes (Deyo, 1988). Exclusion of general measures essentially eliminates the possibility of performing policy analysis or comparison of expenditures in different disease groups. Even if an investigator requires disease-specific measures, there is often no disadvantage to including both general and specific approaches.

MORTALITY

Mortality remains the major outcome measure in most epidemiologic studies and clinical trials. Typically, mortality is expressed in unit of time. In order for mortality data to be meaningful, they must be expressed in the form of a rate, that is, the proportion of deaths from a particular cause occurring within some defined time interval (usually per year). Usually, mortality rates are 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).

There are many advantages to reporting mortality rates. They are "hard" data (despite some misclassification bias, NIH-1979), and the meaning of the outcome is not difficult to comprehend. Despite the many advantages of mortality outcomes, there are also some obvious limitations. Mortality rates consider only the dead and ignore the living. Many important health care services, including prevention, can be expected to have little or no impact upon mortality rates. For example, each year approximately 1.2 million cataract removal procedures

are performed in the United States (Jaffe, 1981). Although the procedure is essentially noncontroversial, cataract removal has little or no impact on mortality and is certainly unrelated to infant mortality. An outcome measure focusing only on mortality would miss the value of the surgery, which proves to have benefits in as many as 95 percent of the cases.

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 working days missed or bed disability days. Most approaches to health status assessment are essentially morbidity indicators. For example, the Sickness Impact Profile (Bergner et al., 1981) represents the effect of disease or disability upon a variety of categories of behavioral function. The RAND Health Insurance Experiment measures include 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, there is accrual of mortality cases. 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 (as opposed to no support or treatment) for randomly assigned groups of very ill, elderly nursing home residents. Let us suppose that the program maintained them all at a very low level of function throughout the year, while in the comparison group, the sickest 10 percent died. Looking just at the living in the follow-up, one finds the comparison group to be healthier, since the sickest had been removed by mortality. By this standard, the program of no support or treatment might be put forth as the better alternative. With a measure that combined morbidity and mortality, however, the story would be very different, with mortality effects dragging overall health of the comparison group to a very low level. We will return to the importance of mortality later.

Well-Years

Our approach is to express the benefits of medical care, behavioral intervention, or preventive programs in terms of well-years. Others have chosen to describe the same outcome as Quality Adjusted Life Years (QALYs) (Weinstein and Stason, 1976). Well-years integrate mortality and morbidity to express health status in terms of equivalents

of well-years of life. If a male cigarette smoker died of heart disease at age 50 and we would have expected him to live to age 75, it might be concluded that the disease caused him to lose 25 life-years. If 100 male cigarette smokers died at age 50 (and also had life expectancies of 75 years), we might conclude that 2,500 (100 men x 25 years) life-years had been lost.

Yet death is not the only outcome of concern in heart disease. Many adults suffer myocardial infarctions, leaving them somewhat disabled over a longer period of time. Although they are still alive, the quality of their lives has diminished. Our model permits all degrees of disability to be compared to one another. A disease that reduces the quality of life by one-half will take away .5 well-years over the course of one year. If it affects two people, it will take away 1.0 well-year (equal to $2 \times .5$) over a one-year period. A medical treatment that improves the quality of life by .2 for each of five individuals will result in a production of one well-year if the benefit is maintained over a one-year period. Using this system, it is possible to express the benefits of various programs by showing how many equivalents of well-years they produce (Kaplan and Bush, 1982; Anderson and Moser, 1985). Yet not all programs have equivalent costs. In periods of scarce resources, it is necessary to find the most efficient use of limited funds. Our approach provides a framework within which to make policy decisions that require selection between competing alternatives. Preventive services may in this way compete with traditional medical services for the scarce health care dollar—and we believe preventive services can be competitive in such analyses. Performing such comparisons requires the use of a general health decision model. In the next section, the general model of health status assessment and benefit-cost/utility analysis will be presented.

THE GENERAL MODEL

BUILDING A HEALTH DECISION MODEL

The Health Decision Model grew out of a substantive body of theory in economics, psychology, medicine, and public health. These theoretical linkages have been presented in several previous papers (Bush, Chen, and Patrick, 1973; Bush, Chen, and Patrick, 1975; Fanshel and Bush, 1970). Building a health decision model requires at least five distinct steps.

Step 1: Defining a Function Status Classification

During the early phases of the Health Index project, a set of mutually exclusive and collectively exhaustive levels of functioning were defined. After an extensive, specialty-by-specialty review of medical reference works, we listed all of the ways that disease and injuries can affect behavior and role performance. Without considering etiology, it was possible to match a finite number of conditions to items appearing on standard health surveys, such as the Health Interview Survey (National Center for Health Statistics), the Survey of the Disabled (Social Security Administration), and several rehabilitation scales and ongoing community surveys. These items fit conceptually into three scales representing related but distinct aspects of daily functioning: Mobility, Physical Activity, and Social Activity. The Mobility and Physical Activity scales have three levels, while Social Activity has five distinct levels. Table 1 shows the steps from the three scales. Several investigators have used this function status classification (or a modified version of it) as an outcome measure for health program evaluation (Reynolds, Rushing, and Miles, 1974; Stewart et al., 1978). However, the development of a truly comprehensive health status indicator requires several more steps.

Step 2: Classifying Symptoms and Problems

There are many reasons why a person may not be functioning at the optimum level. Subjective complaints are an important component of a general health measure because they relate dysfunction to a specific problem. Thus, in addition to Function Level classifications, an exhaustive list of symptoms and problems has been generated. Included in the list are 21 complexes of symptoms and problems representing all of the possible symptomatic complaints that might inhibit function. These symptoms and problems are shown in Table 2.

Step 3: Using Preference Weights to Integrate the Quality of Well-being Scale

We now have described the three scales of function and 21 Symptom/Problem complexes. With these, all we can do is compare populations in terms of frequencies of each scale step (and, if necessary, Symptom/Problem complex). Although comparisons of frequencies are common in health services research, our system offers a strategy for integrating the frequencies into a single comprehensive expression. If our intent is to say which of these distributions is "better off" and which "worse,"

Table 1: Quality of Well-being General Health Policy Model, Elements and Calculating Formulas (Function Scales, with Step Definitions and Calculating Weights)

Step No.	Step Definition	Weight
<i>Mobility Scale (MOB)</i>		
5	No limitations for health reasons	-.000
4	Did not drive a car, health related; did not ride in a car as usual for age (younger than 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	-.062
2	In hospital, health related	-.090
<i>Physical Activity Scale (PAC)</i>		
4	No limitations for health reasons	-.000
3	In wheelchair, moved or 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	-.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	-.077
<i>Social Activity Scale (SAC)</i>		
5	No limitations for health reasons	-.000
4	Limited in other (e.g., recreational) role activity, health related	-.061
3	Limited in major (primary) role activity, health related	-.061
2	Performed no major role activity, health related, but did perform self-care activities	-.061
1	Performed no major role activity, health related, and did not perform or had more help than usual in performance of one or more self-care activities, health related	-.106

Calculating Formulas

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

$$W = 1 + (CPX_{wt}) + (MOB_{wt}) + (PAC_{wt}) + (SAC_{wt})$$

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

CPX-11	Cough, wheezing, or shortness of breath, with or without fever, chills, or aching all over	-.257
MOB-5	No limitations	-.000

Continued

Table 1: Continued

Quality of Well-being Element	Step Definition	Weight
PAC-1	In bed, chair, or couch for most or all of the day, health related	-.077
SAC-2	Performed no major role activity, health related, but did perform self-care	-.061
$W = 1 + (-.257) + (-.000) + (.007) + (-.061) = .605$		
<p>Formula 2. Well-years (WY) as an output measure:</p>		
$WY = \{ \text{No. of Persons} \times (\text{CPXwt} + \text{MOBwt} + \text{PACwt} + \text{SACwt}) \times \text{Time}$		

simple frequency distributions may not be able to help much. For example, is a group with 80 people able to travel but limited in their mobility and with 5 restricted to their homes worse off than a group in which 85 can travel freely but 10 are restricted to home? Obviously, comparing frequency distributions is a complex undertaking. Further, the example involves frequencies for only one scale. How can one make decisions when there are three scales and Symptom/Problem complexes to consider?

Another step is necessary to integrate the three scales and the Symptom/Problem complexes in a manner that will allow a single numerical expression to represent each combination of steps on the scales and Symptom/Problem complexes. The empirical means of accomplishing this is measured *preferences* for the health states. These might be regarded as "quality" judgments. As we noted earlier, the Health Decision Model includes the impact of health conditions upon the quality of life. This requires that the desirability of health situations be evaluated on a continuum from death to completely well. An evaluation such as this is a matter of utility or preference and, thus, function level-symptom/problem combinations are scaled to represent precise degrees of relative importance.

Human judgment studies are needed to determine weights for the different states. We have asked random samples of citizens from the community to evaluate the relative desirability of a good number of health conditions. Random sample surveys were conducted in the San Diego community during two consecutive years. The probability sample included 866 respondents ethnically representative of the population. When necessary, interviews were conducted in Spanish. From a

Table 2: List—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)	-.727
2	Loss of consciousness such as seizure (fits), fainting, or coma (out cold or knocked out)	-.407
3	Burn over large areas of face, body, arms, or legs	-.387
4	Pain, bleeding, itching, or discharge (drainage) from sexual organs—does not include normal menstrual (monthly) bleeding	-.349
5	Trouble learning, remembering, or thinking clearly	-.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	-.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	-.299
8	Pain, burning, bleeding, itching, or other difficulty with rectum, bowel movements, or urination (passing water)	-.292
9	Sick or upset stomach, vomiting or loose bowel movement, with or without fever, chills, or aching all over	-.290
10	General tiredness, weakness, or weight loss	-.259
11	Cough, wheezing, or shortness of breath, <i>with</i> or <i>without</i> fever, chills, or aching all over	-.257
12	Spells of feeling upset, being depressed, or of crying	-.257
13	Headache, or dizziness, or ringing in ears, or spells of feeling hot, or nervous, or shaky	-.244
14	Burning or itching rash on large areas of face, body, arms, or legs	-.240
15	Trouble talking, such as lisp, stuttering, hoarseness, or being unable to speak	-.237
16	Pain or discomfort in one or both eyes (such as burning or itching) or any trouble seeing after correction	-.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	-.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	-.170
19	Taking medication or staying on a prescribed diet for health reasons	-.144
20	Wore eyeglasses or contact lenses	-.101
21	Breathing smog or unpleasant air	-.101
22	No symptoms or problem (not on respondent's card)	-.000
23	Standard symptom/problem	-.257

Source: Kaplan and Anderson (1987).

listing of all possible combinations of the scale (Mobility, Physical Activity, Social Activity, and Symptom/Problem complexes), we drew a stratified random sample of 343 case descriptions (items) and divided them into eight sets of computer-generated booklets. All respondents were assigned randomly to one of the eight booklets, creating eight subgroups of approximately 100 respondents each. In a series of studies, a mathematical model was developed to describe the consumer decision process. The validity of the model was cross-validated with an R^2 of .94 (Kaplan, Bush, and Berry, 1978). These weights, then, described the relative desirability of all of the function states on a scale from zero (for death) to 1.0 (for asymptomatic optimum function). Thus, a state with a weight of .5 was viewed by the members of the community as being about one-half as desirable as optimum function, or about halfway between optimum function and death.

Some critics have expressed concern that community rather than specific-population weights are used. The advantage of community weights is that they are general (like the model) and do not bias policy analysis toward any interest group. More important, however, is the consistent failure of empirical studies to show systematic differences among demographic groups (Kaplan, Bush, and Berry, 1978), providers, students, and administrators (Patrick, Bush, and Chen, 1973b), and Americans versus British (Patrick et al., 1985). Relevant to the general versus disease-specific issue, Balaban and colleagues (1986) found that weights provided by rheumatoid arthritis patients are remarkably similar to those we obtained from members of the general population.

Using preference weights, one component of the general model of health is defined. This is the "Quality of Well-being (QWB) scale," which is the point-in-time component of the Health Status Index (Kaplan, Bush, and Berry, 1976; Fanshel and Bush, 1970). The Quality of Well-being score for any individual can be obtained from preferences or "quality" judgments associated with his or her function level, adjusted for symptom or problem.

The example in Table 1 describes a person classified on the three scales of observable function and on a symptom/problem. The table shows the adjustments for each of these components. Using these, a weight of .605 is obtained. By including symptom/problem adjustments, the Health Status Index becomes very sensitive to minor "top-end" variations in health status. The adjustments for particular symptom/problems are shown in Table 2. For example, there are Symptom/Problem complexes for wearing eyeglasses, having a runny nose, or breathing polluted air. These symptom adjustments apply

even if a person is at the top step in the other three scales. For example, a person with a runny nose receives a score of .83 on the Quality of Well-being scale when he or she is at the highest Function Level (that is, the top step on each scale shown in Table 1). Thus, the Index can make fine as well as gross distinctions.

Mathematically, the Quality of Well-being score may be expressed as:

$$W = \frac{1}{N} \sum_{x=1}^L W_x N_x \quad (1)$$

where

W = the symptom-standardized, time-specific Quality of Well-being score.

x indexes the Function levels [$x = 1, \dots, L$].

W_x = the Quality of Well-being (weight, utility, relative desirability, social preference) for each Function level, standardized (adjusted) for all possible Symptom/Problem complexes.

N_x = the number of persons in each Function level.

N = the total number of persons in the group, cohort, or population.

Thus, Quality of Well-being is simply an average of the relative desirability scores assigned to a group of persons for a particular day or a defined interval of time.

Several studies attest to the reliability (Kaplan, Bush, and Berry, 1978; Bush, Kaplan, and Berry, unpublished) and validity (Kaplan, Bush, and Berry, 1976) of the Quality of Well-being scale. For example, convergent evidence for validity is given by significant positive correlations with self-rated health, and negative correlations with age, number of chronic illnesses, symptoms, and physician visits. However, none of these other indicators was able to make the fine discrimination between health states which characterizes the Quality of Well-being scale. These data support the convergent and discriminant validity of the scale (Kaplan, Bush, and Berry, 1976).

Step 4: Estimating Transitions among Health States

The Quality of Well-being scale is the point-in-time component of the model. A comprehensive measure of health status also requires an expression of prognosis or the probability of moving between health

states over time. People who are well now want to remain well. Those who are at suboptimal levels want to become well, or at least not worse. A health decision model must consider both current functioning and the probability of transition to other Function levels over the course of time. When transition is considered and documented in empirical studies, the consideration of a particular diagnosis is no longer needed. We fear diseases because they affect our current functioning or raise the probability of a limitation in our functioning some time in the future. A person at high risk for heart disease may be functioning very well at present, but may have a high probability of transition to a lower level (or death) in the future. Cancer would not be a concern if the disease did not affect current functioning or the probability that functioning would be affected at some future time.

When weights have been properly determined, health status can be expressed precisely as the expected value (product) of the preferences associated with the states of function at a point in time and the probabilities of transition to other states over the remainder of the life expectancy. Quality of Well-being (W) is a static or time-specific measure of function, while the Well-life Expectancy (E) also includes the dynamic or prognostic dimension. The Well-life Expectancy is the product of Quality of Well-being times the expected duration of stay in each Function level over a standard life period. The equation for the Well-life Expectancy is

$$E = \sum W_k Y_k \quad (2)$$

where

E = the symptom-standardized Well-life Expectancy in equivalents of completely well years.

Y = the expected duration of stay in each Function level or case type estimated with an appropriate statistical (preferably stochastic) model.

An example computation of the Well-life Expectancy is shown in Table 3. Suppose that a group of individuals was in a well state for 65.2 years, in a state of non-bed disability for 4.5 years, and in a state of bed disability for 1.9 years before their deaths at the average age of 71.6 calendar years. In order to make adjustments for the diminished quality of life they suffered in the disability states, the duration of stay in each state is multiplied by the preference associated with the state. Thus, the 4.5 years of non-bed disability become 2.7 equivalents of

Table 3: Illustrative Computation of the Well-Life Expectancy

<i>State</i>	<i>k</i>	Y_k	W_k	$W_k Y_k$
Well	A	65.2	1.00	65.2
Non-bed disability	B	4.5	.59	2.7
Bed disability	C	1.9	.34	.6
Current life expectancy	71.6 Life Years			
Well-life expectancy	68.5 Well-Years			

Source: Kaplan and Bush (1982).

well-years when we adjust for the preferences associated with inhabiting that state. Overall, the Well-life Expectancy for this group is 68.5 years. In other words, disability has reduced the quality of their lives by an estimated 3.1 years.

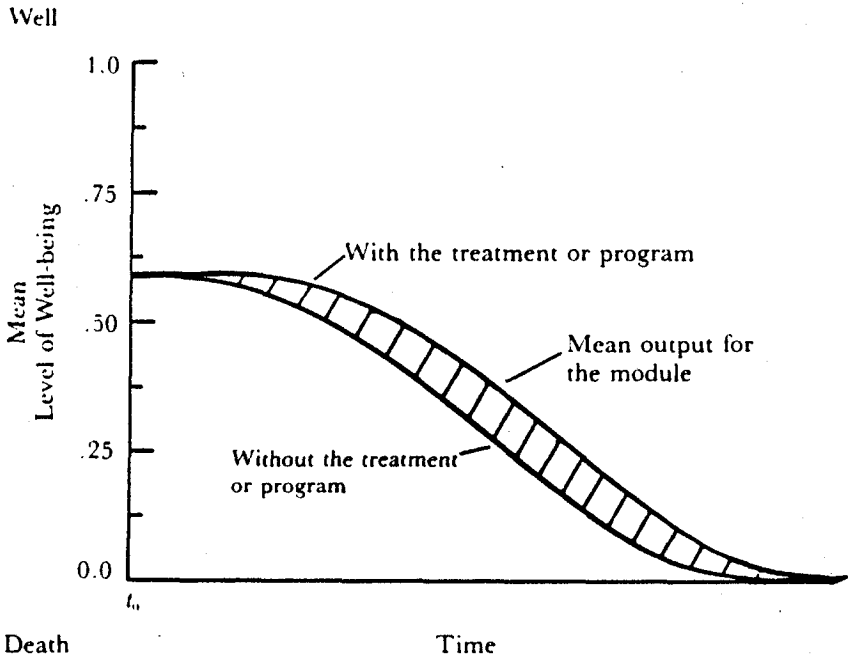
Step 5: Estimating the Benefit-Cost/Utility Ratio

The San Diego Health Index group has shown in a variety of publications how the concept of a well or weighted life expectancy can be used to evaluate the effectiveness of programs and health interventions. The output of a program has been described in a variety of publications as quality-adjusted life years (Bush, Chen, and Patrick, 1973; Bush, Fanshel, and Chen, 1972), well-years, equivalents of well-years, or discounted well-years (Patrick, Bush, and Chen, 1973a, 1973b; Kaplan, Bush, and Berry, 1976). Weinstein (1980, 1983) has popularized the concept, calling the same output Quality-Adjusted Life Years (QALYs), and this has been adopted by the Congressional Office of Technology Assessment (1979). It is worth noting that the Quality-Adjusted Life Years terminology was originally introduced by Bush, Patrick, and Chen (1973), but later was abandoned because it has surplus meaning. The term "wellness" or "well-years" implies a more direct linkage to health conditions. Whatever the term, the San Diego Health Index shows the output of a program in years of life adjusted by the quality of life which has been lost because of disease or disability.

RELEVANT ISSUES

By comparing experimental and control groups on a general health status measure, it is possible to estimate the output of a program in terms of the well-years it produces. This is shown as the area between curves representing the two groups in Figure 1. Dividing the cost of the program by the well-years it yields gives the benefit-cost/utility ratio.

Figure 1: Theoretical Comparison of Treated and Untreated Groups (the area between the two curves is the output or benefit of a program in Well-Year Units)



Source: Kaplan and Bush (1982).

There are many attractive elements of general health status measures. We argue that the ultimate purpose of health care and prevention is directed toward two simple objectives. First, investments in health care are aimed at extending the duration of life. Second, health care programs should improve the quality of life while individuals are alive. A comprehensive expression of health status can determine the effects of a program using a unit that simultaneously considers risks and benefits. As Mosteller (1981) has suggested, specific measures of health outcome often ignore the side effects of treatment. A treatment for hypertension, for example, may cause gastric irritation, nausea, and bed disability. Health benefits of treatment can be expressed in well-year units, as can health side effects.

Despite the advantages of a general system, reviews of the literature often fault these methods because they do not include separate scores for social health, mental health, and child health. We believe some of these arguments are misdirected. We discuss some of the issues in the next sections.

SOCIAL HEALTH

For nearly 35 years, physicians, psychologists, sociologists, and epidemiologists have been attempting to include social support and social function in a definition of health status. Despite relentless effort, it has been difficult to find a meaningful definition of social support as a component of health. The term "social health" was included in the World Health Organization definition of health that accompanied its charter document in 1948. It defined health as "a state of complete physical, mental, and social well-being and not merely the absence of infirmity." In identifying the dimensions of health, WHO neglected to provide any operational definitions. Thus, different investigators have taken different approaches in their attempts to capture physical, mental, and social dimensions. Since publication of the WHO statement, many investigators have tried to develop measures to operationalize the three components of health status. With surprising consistency, authors quote the WHO definition and then present their methods for measuring the three components. So prevalent is the notion that health status must include these three components that many reviews now negatively evaluate any measure that does not conform to the WHO definition. For example, Meenan (1982) disapproved of several health measures, because "these approaches fall short of conceptualizing or measuring health in the WHO sense of a physical, psychological, and social state" (p. 785).

With the command of the World Health Organization so plainly set forth, many investigators have struggled to develop their measures of social health. Yet there have been consistent problems. For example, Kane and Kane (1985) devoted a substantial section of their monograph to a description of problems in the quantification of social health. These problems included vague concepts, lack of norms, the interactive nature of variables, difficulty in construction of a continuum, and the subjective nature of social health.

Only Donald, Ware, and colleagues have begun to question the meaning of social health (1980a,b). In one paper, Donald, Ware, et al. (1980b) reviewed 70 studies relevant to social health. They selected

from these 11 studies for more detailed analysis. The great majority of these studies focused on what we now call social support. Yet at least two separate components were being assessed by the many investigators contributing to this literature. One component was "social contacts," or the performance of social role; the other component was "social resources," which is more analogous to the concept of social support. This distinction is very important. Social contacts might include participation in work, attendance at school, and other aspects of functioning. Social resources relate to social life, friendships, and family relationships.

In a series of analyses, it has been demonstrated that social support may be a predictor for health outcomes (Ware and Donald, 1980; Kaplan, 1985c)—yet the direction is not always clear. For example, Heitzmann and Kaplan (1984) have demonstrated that social support may predict positive outcomes for women but negative outcomes for men. Social support is not an outcome that can serve as the target of health care. On the other hand, social functioning is a component of health status. Diseases and disabilities affect social function. Social function is a central component in the concept of quality of life.

Optimizing social health raises issues of social control and public policy. Considering the example of function, there is strong consensus that function is desirable. Thus, it seems reasonable to devote public resources to maximize the level of function and quality of life within a community. Optimized health status might be considered a common goal, as is national defense, a strong educational system, and so forth. Social functioning is an important component of health status and is included in our model (in the Social Activity scale). On the other hand, including social support in the definition of health status would imply that community resources should be used to obtain some defined level of social support. We might expect considerable public disagreement about what the social support objective might be. For example, would we want to develop a public policy that requires people to have friends?

Excluding social support from the definition of health makes policy analysis relatively straightforward. There is little disagreement about what states of health are desirable (Patrick, Bush, and Chen, 1973b; Kaplan, Bush, and Berry, 1978). With this consensus, achieving these desired states becomes the objective of health care. A major issue lies in defining a mix of programs that most efficiently and effectively achieves these objectives; programs that enhance social support might be considered in this mix. We hold that including social support

in the definition of health only confuses the definition of these objectives.

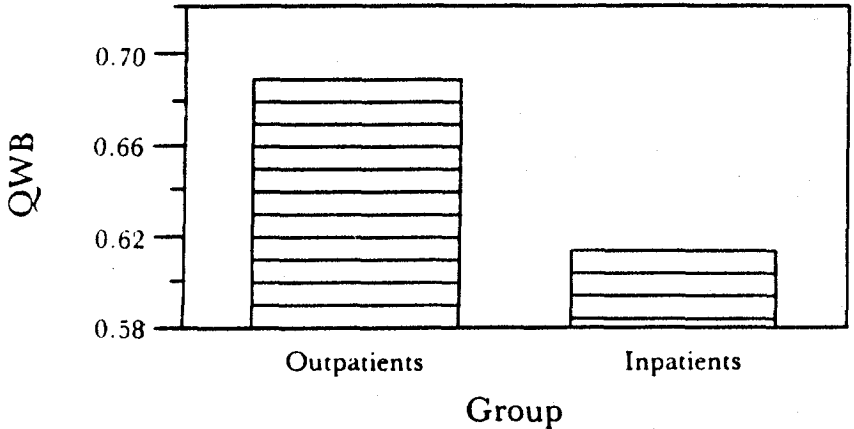
Since social function is a component of health status, it is included as an integral part of our model.

MENTAL HEALTH

The WHO definition of health status included a separate category for mental health. This prompted most investigators to develop separate measures of mental health functioning. Perhaps the best known effort in this area is the work by Ware, Johnston, Davies-Avery, and Brook (1979). These investigators adapted Dupuy's (1974) General Well-being Index which is now known in many circles as the RAND Mental Health Measure. The now classic work by Ware et al. has been used to accentuate the argument that mental health is one component of general health status that is specific and different from physical functioning. In support of this argument, Ware et al. (1984) argue that the correlation between psychological distress and physical functioning is only .25. In addition, they offered comparisons between those with no physical limitations but with differences on items reflective of psychological distress. For this high physically functioning group, those with higher scores on mental distress used three times as many mental health services as those low in distress.

We argue that the WHO definition of health status promotes an artificial dichotomy between mental and physical function. In order to understand this argument, it is important to think about the impact of mental illness, anxiety, or poor social adjustment upon functioning. Mental health affects longevity (Wells, 1985) and quality of life. In other words, the impact of mental health on general health status is expressed through its impact on life expectancy, functioning, and symptoms. Ware et al. ignore the fact that many individuals with perfect physical functioning experience symptoms. For example, an individual experiencing anxiety at work might check a symptom describing anxiety. This anxiety might affect quality of life in a manner similar to a physical symptom such as shortness of breath. Severe anxiety in the form of a phobia presents the disruption of role performance in addition to the experience of symptoms, perhaps limiting an individual to his or her house because of the problem. Many individuals experience depression to the extent that it disrupts their activities of daily living. At the other extreme, anxiety and depression can be so severe that they result in hospitalization. Thus, the impact of the con-

Figure 2: Differences between Psychiatric Outpatients with a Primary Diagnosis of Depression



dition upon functioning is very much the same as the impact of a physical malady.

Figure 2 compares Quality of Well-being (QWB) and depression scores for psychiatric inpatients and outpatients. Both groups were suffering from affective disorders. As the figure shows, the QWB is sensitive to differences between populations whose problem is primarily "mental" rather than physical.

As in physical health, mental health conditions must be examined in relation to changes over the course of time. For example, depression may last three days, three weeks, or one year. The total impact needs to be expressed as a function of its duration. More importantly, mental health status may affect differential transition among functional states over the course of time. The term "positive health" typically is used to describe some aspect of life style or mental outlook that is associated with better future health. This is specifically represented in our model as lower probabilities of transition to poor health over the course of time. For example, today an individual with refined skills for coping with stress may be no different from an individual without such coping skills. However, given certain epidemiologic linkages, the former may have a higher probability of better functioning at future points in time.

Much of the confusion about mental health has been generated by a very refined technology for assessing mental states. Often, detailed

questionnaire methods have been factor-analyzed to describe different dimensions of mental health. Nevertheless, these very different levels of functioning may ultimately impact the general well-being. This may be analogous to the many available measures of blood chemistry. For example, indicators of kidney function (creatinine, BUN, and others) may be identified as separate factors. Yet the importance of these measures is their relationship to longevity and to function at particular points in time. The fact that measures of blood chemistry form many factors is of interest and importance to clinicians; to the policy analyst, it may have less meaning.

Another reason for not separating mental and physical function is that mental function may affect physical function and vice versa. The growing literature on psychoneuroimmunology (Biondi and Pancheri, 1985) clearly demonstrates the intertwining nature of these areas. In addition, experiments have demonstrated that general health status can be improved in medical patients even when physical functioning is unaffected (Atkins et al., 1984). The most important point is that all providers in health care are attempting to improve quality of life and extend the duration of life. It is valuable to allow mental health providers and physical health providers to compare the benefits of their services using a common unit.

Finally, it is important to emphasize that the Physical Health Index — and other aggregate indexes used by Ware and colleagues — is not the same thing as our General Health Policy Model. Although it is based on measurement of the same functions, it differs in several significant ways. First, the RAND measure excludes the tabulation of symptoms. Many "mental health" problems are overlooked because their expression is primarily in symptomatology. Second, there is no provision for combining over the measures to give a single numeric expression of well-being, as with the QWB. Third, the RAND measure does not consider transition among levels of function over the course of time. Indeed, then, it would overlook any relationship between positive outlook and transition among states. The separate measurement of mental health remains a major issue in the conceptualization of general health status (Bergner et al., 1976). Although our position is against the norm, we believe mental health can be conceptualized as a portion of general health status and that there is considerable disadvantage in attempting the separate measurement and specification of mental function. We do understand that some investigators are interested in specific subcomponents of mental health, such as cognitive functioning. In these cases, more detailed measures might

be considered as additions to—but not replacements for—the general measures.

CHILD HEALTH

In an important review of health status measurement, Bergner (1985) stated that “health status indexes for children are about where adult health status measures were 18 years ago” (p. 702). The thrust of Bergner’s argument was that a conceptualization of child health “*distinct from health in adults, is needed*” (p. 702). Bergner argued that a child health measure must be equally applicable across developmental levels and that the measure should be appropriate for the study of congenital abnormalities, family planning services, postnatal care, and health insurance.

Since the inception of the original Health Index Project, we have included measures appropriate for children. Our large 1974–1975 general population survey included a supplementary probability sample of 368 children. Questionnaires have been devised to ask about appropriate social activities for children with focus on infants and school-age children. In addition, one of the first major demonstration projects used this system to evaluate the benefits of phenylketonuria (PKU) screening.

Evaluation of the health status of children is like evaluation of the health status of adults. First, we must consider mortality. Second, we must consider any effects upon current functioning and, third, we must consider probabilities of transition to other levels of function over the course of time. Assessment of current functioning involves an evaluation of symptoms.

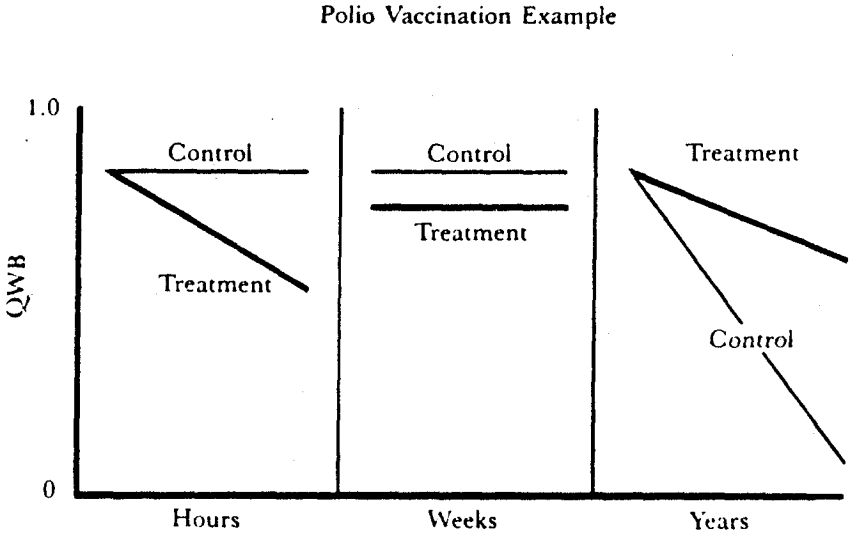
Children, like adults, experience the impact of some diseases and disabilities upon function. Yet the categories in which children are affected by health conditions should be the same as those for adults. That is why Eisen and colleagues (1980) chose to evaluate the physical health of children in the same way they did for adults. The exact content of the items changes somewhat to focus on role-appropriate activities for the younger group. Yet a child with nausea reports nausea just as an adult does. If the nausea is severe, it will limit the child’s functioning and performance of social role (attending school or playing with the usual vigor). If an adult has severe nausea, it will also interfere with performance of the social role. It is worth noting that in our community surveys, over 40 percent of the children had one-day scores less than 1.0. Children do have symptoms that are demonstratively captured by the system. Many people believe that illnesses of childhood

are less serious than adult illnesses because children rebound well from various ailments. This is reflected in the prognosis dimension of the system. If children are afflicted for a fewer number of days, then they have a higher probability of transition to better levels of functioning for the future. Pediatricians have complained that traditional methods of health status assessment are inadequate for proper assessment of their contributions (Pantell and Lewis, 1986). Because pediatricians often focus on preventive measures and education, these efforts cannot be adequately evaluated by concurrent assessment of health status.

A comprehensive view of health status might show the preventive contribution of pediatric care. For example, vaccinating children against measles, mumps, and rubella (MMR) has little beneficial effect on current health status. Indeed, it may have a detrimental effect since some children will develop aches, pains, and minor fevers from the vaccine. Measures that consider only the day of the inoculation visit will show this problem. Yet no pediatrician provides these inoculations to affect concurrent health status; the value is in the prevention of diseases that occur later in life (see example in Figure 3). The left-hand portion of Figure 3 shows the hypothetical effect of polio vaccine if the measurements are taken within hours of treatment. Notice that health status for the treatment condition is *poorer* than in the control condition. This is because of reactions to the vaccine. If the analysis is done over weeks (center portion) there will be very little difference between treatment and control. Only when we consider the entire life cycle (right-hand portion) will the obvious effect of the treatment be apparent.¹ It is naive, then, to evaluate the benefit of this service without taking the life-span perspective possible under the system. Berwick, Cretin, and Keeler (1980) have attempted to evaluate dietary interventions for the primary prevention of heart disease. Using a years-of-life-added approach, they estimated the cost to produce a life-year.

Recently, we have been collaborating with several pediatricians. In one study (with David Ornstein of the University of Pittsburgh), we found that the QWB provides considerable information about children with cystic fibrosis. For children with compromised lung function ($FEV_{1,} < 80$ percent of predicted), the association between pulmonary function and QWB was substantial ($R = .96$). Other studies are currently evaluating the validity of the QWB system in children with asthma and with diabetes, and among well children with episodic acute illnesses. In the next section, we demonstrate how the General Health Policy Model has been effectively used to evaluate PKU screening.

Figure 3: The Effectiveness of a Hypothetical Program when the Evaluation Period is Conceptualized as Hours, Weeks, or Years

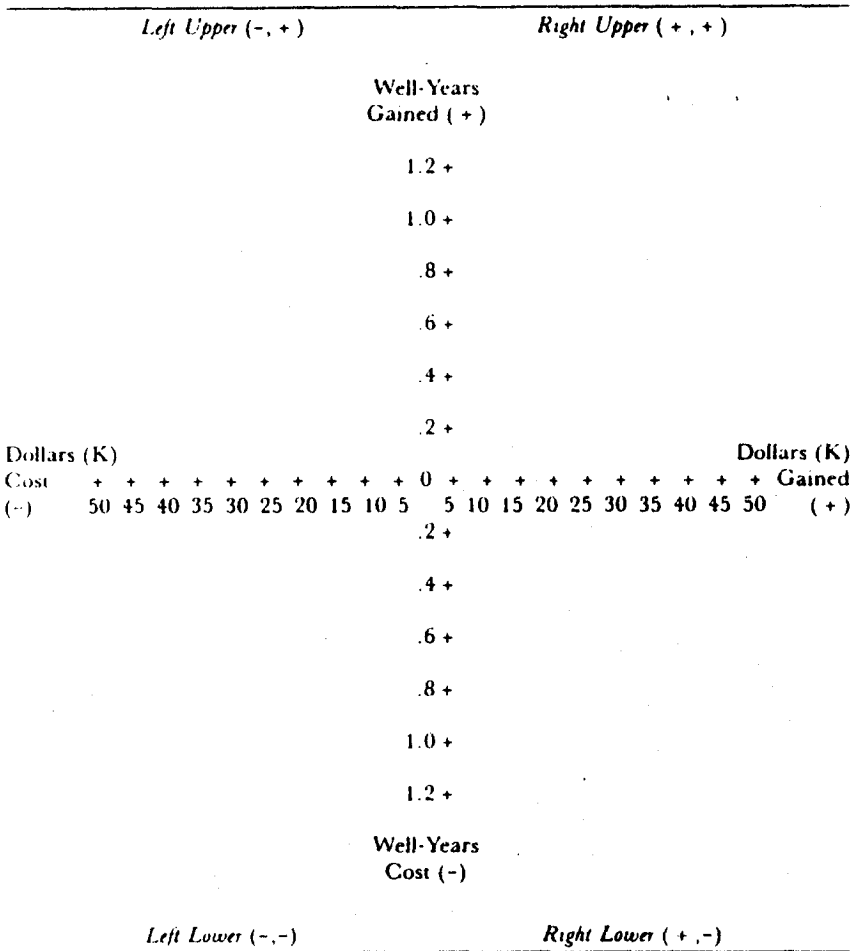


POLICY SPACE

Various approaches to cost/benefit and cost/utility analysis occasionally produce different results. The output for cost/benefit analysis is in monetary terms—a program that produces cost savings. Cost/utility analysis focuses on the cost to produce a well-year of life. Recently, Anderson et al. (1986) integrated the concepts of well-years and net dollars returned within a common framework that they called benefit-cost/utility analysis ($b-c/u$). This was accomplished by creating a two-dimensional policy space as illustrated in Figure 4. The x -axis in the figure represents net dollars returned per person. Returns are defined as benefits minus costs in dollar units. The y -axis displays well-years lost or gained through a particular treatment program, clinical intervention, or policy change.

The right half of the plane would be used to represent programs in which benefits exceed costs, while the left half would display situations in which costs exceed benefits. The upper half of the figure displays

Figure 4: Coordinates for Health Policy Space to Compare Efficiency of Alternative Health Interventions



outcomes that have positive health effects in terms of well-years. Those in the bottom half of the figure would be used to represent negative health outcomes in the well-year units.

The two-dimensional space yields four quadrants. One quadrant, the lower left, represents unsuitable alternatives. In these cases, dollars are spent and negative health consequences occur. Administration of a

uniformly toxic treatment might be represented by this quadrant. The upper right quadrant represents the most attractive alternatives. Here, well-year health benefits are gained, and there are also economic benefits. Prevention of early heart disease might be an example. The upper left quadrant shows well-year gains, but with more significant costs associated with these improvements. Transplantation surgery for the elderly might be described by this quadrant. Here, significant health benefits accrue, but the recipients may not return to the productive economic sector.

The lower right quadrant represents another level of economic trade-off. Here, society may be willing to sacrifice some health benefits in exchange for cost savings. Anderson and colleagues suggested that these trade-offs may be common in studies involving nuclear power, pollution control, occupational, environmental, and consumer product safety, highway speed limits, and so on.

This policy space conceptually identifies areas where standard *b-c/u* trade-offs of dollars for well-years work (upper left and lower right), and areas where they do not work (upper right and lower left). It also provides a basis for understanding the differences between *cost/utility* and *benefit/cost* trade-offs. In the upper right, both dollars and well-years are being produced, rather than one being lost and the other being gained, and in the lower left, both well-years and dollars are being lost.

Furthermore, the policy space shows that programs analyzed in the upper right are *de facto* superior to programs analyzed in the upper left and lower right, and programs analyzed in these areas are similarly superior to those in the lower left. Thus, simply by considering in which area of policy space a program is located provides important information.

APPLICATIONS OF THE MODEL IN PREVENTION

A common misconception is that health status measures and models are designed for studies of disability and have no value for studies of preventive alternatives in health care. Indeed, the system can be used to evaluate any alternative which affects current or future health outcomes. In this section, we review several applications of the General Health Policy Model—in screening, and in primary, secondary, and tertiary prevention. Most of these programs represent options in the upper left quadrant of the policy space.

SCREENING

The General Health Policy Model has been used to evaluate several screening programs. For example, the system has been used to evaluate the value of following up abnormal screening serum panel results. Screening panels can provide early information that may guide the physician toward a preventive medical intervention. However, "abnormal" results could also be statistical artifacts that lead to unnecessary workup, anxiety, and other consequences of labeling. Epstein and colleagues (1981) investigated the value of serum thyroxine (T4) determination, because early detection of hypo- or hyperthyroidism might lead to significant preventive efforts. Yet those with true disease are only a small portion of those with abnormal values. Among 3,603 patients screened, Epstein and colleagues found 91 patients with elevated T4 and 20 patients with abnormally low values. Although the follow-up testing and evaluation is expensive (\$11,136), true disease was found in 2 of the 91 hyperthyroidism screens and 6 of the 20 hypothyroidism screens. Despite the high costs of case finding, screening was shown to be cost-effective because of the potential health benefits attributable to early intervention. Ultimately, the cost per well-year was estimated at \$3,600. In contrast, Amberg et al. (1982) used the same methods to evaluate the value of serum alkaline phosphatase in the screening panel. Among 1,027 patients, nearly 30 percent (304) showed elevated alkaline phosphatase levels. Amberg and associates chose to follow 118 unexpected deviations in more detail. They found that only one of these had an undiagnosed disease—viral hepatitis. The system was used to estimate the health effects of the interventions. It was shown that few health benefits were attributable to screening and that the cost/utility exceeded \$85,400 per well-year (in 1982 dollars). Thus, some components of the screening panel produced benefits at a low cost while others produced benefits at a significantly higher cost.

In one recent application of the model, Anderson and Moser (1985) evaluated the benefit-cost/utility of a screening and treatment program for intestinal parasites among Indochinese refugees to the United States. Using the system, they were able to place persons hospitalized due to liver abscess or plugged hollow viscus into the defined levels of mobility, physical activity, and social activity. In addition, weights for the symptom/problem of abdominal pain were used. Their analysis also took into consideration death rates for untreated infection.

Anderson and Moser used analysis of stool samples from 2,978 Indochinese refugees. Considering that various parasites could be

treated, the screening resulted in between .056 and 1.838 well-years in various examples. In these analyses, it was shown that some programs actually save money (\$8,778), while others are extremely expensive (\$1,238,329) on their way to production of a well-year of life.

One of the most important applications of the model has been an analysis of PKU screening conducted for the State of New York. Between 1965 and 1970, PKU screening in New York successfully detected approximately 22 cases each year. In late 1960s dollars, the costs were close to \$850,000 per year. Bush, Chen, and Patrick (1973) used a national panel of experts to estimate the functional states that would be entered by patients with and without early detection of PKU. Although PKU affects a very small number of children, the impact of the disease is substantial. With the majority of their life expectancy in front of them, losses in well-years accumulate. According to the data obtained from the judges, PKU screening produced 289 well-years for each year of its operation. In 1973 dollars, the cost to produce a well-year with the program was \$2,896.

The PKU study is an example of the benefits of pediatric screening and treatment. The true benefits of the treatment are in avoiding death or profound retardation throughout the life cycle. If different measures of function were used for childhood, adolescence, adulthood, and later life, there would essentially be no way to aggregate the total impact of this disease or its treatment.

TERTIARY PREVENTION

Most of the analyses described above used secondary data for the cost/utility calculations. Typically, these analyses depend on a variety of assumptions that are difficult to verify. In this section, we use data obtained from a prospective study that involved actual observations of function in treated and control subjects.

The study involved an exercise rehabilitation approach for patients with chronic obstructive pulmonary disease (COPD). COPD is the second most rapidly increasing of the top ten causes of death (behind only AIDS) and is the second leading cause of permanent disability in older adults. COPD is responsible for approximately 34 days of restricted activity per 100 persons per year. Total costs of COPD have been estimated to be \$27 billion per year (Lenfant, 1982). Health Interview Survey data suggest that COPD is responsible for over 2 million hospital stays each year and may result in the loss of 1,140,000 years of potential life (Kaplan, Ries, and Atkins, 1985).

Since COPD is generally recognized at an advanced stage, medi-

cal and surgical interventions have limited effectiveness. No behavioral, surgical, or medical intervention can repair damaged lung tissue and restore pulmonary function (Unger, Moser, and Hansen, 1980). The objective of intervention is improved function and quality of life. Several studies have demonstrated that functioning can improve in the absence of changes in pulmonary function (Atkins et al., 1984; Petty and Cherniak, 1981). Capturing these benefits requires a technology for measuring functioning and quality of life.

Currently, one of the most important treatment modalities for COPD patients is a multidisciplinary therapeutic rehabilitation program (Lertzman and Cherniak, 1976). Such a program typically includes a supervised exercise component. Although the data on exercise and rehabilitation programs for COPD appear promising, we still do not have evidence from randomized, controlled studies of their efficacy. We recently reported the results of an experimental trial designed to evaluate behavioral intervention programs for COPD patients (Atkins et al., 1984). COPD patients were exercise-tested on a treadmill and given an exercise prescription. Then they were randomly assigned to treatment or control groups. The experimental or treatment groups were designed to improve compliance with an exercise prescription using behavioral technology. Control groups received either no treatment or attention not designed to affect the outcome. In addition to the Quality of Well-being scale for the General Health Policy Model, the outcome measures included measures of exercise tolerance, compliance, and pulmonary function.

Analysis of the data suggested that those in treatment conditions complied more with the advice to walk than those in the control groups. These changes were reflected in changes in exercise tolerance measured one month after the treatment. However, there were no significant changes in spirometric (pulmonary function) measures. Several analyses were performed using the General Health Policy Model (Kaplan, Atkins, and Timms, 1984; Toevs, Kaplan, and Atkins, 1984). Over the course of 18 months, the experimental and control groups showed significant differences on the quality of life measure. These differences were used to calculate well-years and to perform cost/utility studies. The analyses suggested that behavioral and rehabilitation programs for COPD patients produced a well-year of life at approximately \$23,000. This is comparable to other widely advocated health care programs.

The results of the COPD trial are of interest for several reasons. First, it is significant that there were no noticeable changes in traditional measures of pulmonary function. In other words, we have no

evidence that the programs improved the disease state. Yet the intervention programs may have been successful because they trained patients to do more, given their condition. In the words of Robert Lewis Stevenson, himself a victim of chronic lung disease, "Life is not a matter of holding good cards, but of playing a poor hand well."

A second interesting aspect of the analysis is that the participants in the COPD study were all older adults. As Russell (1986) has noted, there has been concern that benefit-cost/utility analysis discriminates against the elderly. As this analysis suggests, behavioral intervention programs for the sick elderly can be shown to produce benefits at a cost comparable to widely advocated programs. One of the reasons for this is that these programs produce benefits in the near term. Thus, they are relatively insensitive to discounting assumption (Toevs, Kaplan, and Atkins, 1984). Finally, the results are noteworthy because the intervention was not provided by traditional health care providers. All providers within the health care system attempt to produce a common unit of benefit. Yet, third-party insurers do not reimburse all providers for their appropriate services when such services yield the same desired benefit at a competitive cost. Had the COPD project been analyzed in strict cost/benefit terms, it might have been in the upper left quadrant of the policy space. Most of the benefits accrue to the elderly who are beyond retirement age. Thus, there are health benefits but not necessarily monetary benefits.

SUMMARY

In this article, we have argued that the General Health Policy Model can be used to quantify the health effects of medical, preventive, and health policy interventions. The model integrated point-in-time estimates of function, transition among functional levels over time, utilities for health states, and mortality outcomes. Although it is responsive to the general definition of health status proposed by the World Health Organization, the model integrates the different components of health within a single unit. It also can account for problematic areas in health measurement, such as child health.

Examples from prevention and tertiary care suggest the general applicability of the model. The major advantage of the system is that it permits direct comparisons between interventions and programs with different specific objectives. For example, investments in preventive programs can be compared directly with those for traditional medical services.

Several of the arguments against general health status and quality of life measures were considered in this article. We found that separate measures of social, mental, physical, and child health have some important disadvantages. In particular, they make impossible comparisons of programs designed to affect different populations, such as children, the mentally ill, or the elderly. We argue that many of the advantages of general measures have been overlooked and that these approaches should be given serious consideration in future studies.

NOTE

1. The authors thank Leighton Read, M.D., for suggesting this example.

REFERENCES

- Amberg, J. M., et al. The abnormal outpatient chemistry panel serum alkaline phosphatase: Analysis of physical response, outcome, cost, and health effectiveness. *The Journal of Chronic Diseases* 35(1):81-88, 1982.
- Anderson, J. P., and R. J. Moser. Parasite screening and treatment among Indochinese refugees: Cost-benefit/utility and the general health policy model. *Journal of the American Medical Association* 253(6):2229-35, 1985.
- Anderson, J. P., et al. Interday Reliability of Function Assessment for a Health Status Measure: The Quality of Well-Being Scale Instrument. University of California at San Diego. Unpublished.
- Anderson, J. P., et al. Policy space areas and properties of benefit-cost/utility analysis. *Journal of the American Medical Association* 255(6):794-95, 1986.
- Atkins, C. J., et al. Behavioral programs for exercise compliance in chronic obstructive pulmonary disease. *Journal of Consulting and Clinical Psychology* 52(4):591-603, 1984.
- Balaban, D. J., et al. Weights for scoring the quality of well-being instrument among rheumatoid arthritics. *Medical Care* 24(11):973-80, 1986.
- Bergner, M. Development, Testing, and Use of the Sickness Impact Profile. In S. Walker (ed.). *Quality of Life: Assessment and Application*. London: MTP Press, 1987.
- _____. Measurement of health status. *Medical Care* 23(5):696-704, 1985.
- Bergner, M., et al. The sickness impact profile: Development and final revision of a health status measure. *Medical Care* 19(8):787-806, 1981.
- Bergner, M., et al. The sickness impact profile: Validation of a health status measure. *Medical Care* 14(1):57, 1976.
- Berwick, D. M., S. Cretin, and E. B. Keeler. *Cholesterol, Children, and Heart Disease: An Analysis of Alternatives*. New York: Oxford University Press, 1980.
- Biondi, M., and P. Pancheri. Stress, Personality, Immunity, and Cancer: A Challenge for Psychosomatic Medicine. In R. M. Kaplan and M. H.

- Criqui (eds.). *Behavioral Epidemiology and Disease Prevention*. New York: Plenum, 1985, pp. 271-98.
- Bush, J. W., M. M. Chen, and D. L. Patrick. Cost-Effectiveness Using a Health Status Index: Analysis of the New York State PKU Screening Program. In R. Berg (ed.). *Health Status Index*. Chicago, IL: Hospital Research and Educational Trust, 1973, pp. 172-208.
- Bush, J. W., S. Fanshel, and M. M. Chen. Analysis of a tuberculin testing program using a health status index. *Social-Economic Planning Sciences* 6(1):49-69, 1972.
- Bush, J. W., R. M. Kaplan, and C. C. Berry. Sources of Variability in a Health Policy Model. Unpublished.
- Chen, M. M., and J. W. Bush. Maximizing health system output with political and administrative constraints using mathematical programming. *Inquiry* 13(2):215-27, 1976.
- Chen, M. M., J. W. Bush, and D. L. Patrick. Social indicators for health planning and policy analysis. *Policy Sciences* 6(1):71-89, 1975.
- Deyo, R. A. Measuring "Quality of Life" in Patients with Rheumatoid Arthritis. In S. Walker and R. Rosser (eds.). *Quality of Life: Assessment and Application*. London: MTP Press, 1988, pp. 205-22.
- Donald, C. A., et al. Conceptualization and Measurement of Health for Adults in the Health Insurance Study. In *Social Health (R-198714-HEW)*, Volume IV. Santa Monica, CA: RAND Corporation, 1980.
- Dupey, H. Utility of the National Center for Health Statistics General Well-Being Schedule in the Assessment of Self-Representation of Subjective Well-Being and Distress. National Conference on Education in Alcohol, Drug Abuse, and Mental Health Programs. Washington, DC, 1974.
- Eisen, M., et al. Conceptualization and Measurement of Health for Children in the Health Insurance Study. In *Social Health*. Santa Monica, CA: RAND Corporation, 1980.
- Epstein, K. A., et al. The "abnormal" screening serum thyroxin (T4): Analysis of physician response, outcome, cost and health effectiveness. *The Journal of Chronic Diseases* 34(5):175-90, 1981.
- Fanshel, S., and J. W. Bush. A health status index and its application to health-services outcomes. *Operations Research* 18(6):1021-66, 1970.
- Fielding, J. E. Smoking, health effects and control. *New England Journal of Medicine* 313(8):491-98, 1985.
- Heitzmann, C., and R. M. Kaplan. Interaction between sex and social support in the control of type II diabetes mellitus. *Journal of Consulting and Clinical Psychology* 52(6):1087-89, 1984.
- Jaffe, N. *Cataract Surgery and its Complications*. 3rd Edition. St. Louis, MO: Mosby, 1981.
- Kane, R. A., and R. L. Kane. *Assessing the Elderly*. Boston, MA: D. C. Heath and Company, 1985.
- Kaplan, R. M. Human Preference Measurement for Health Decisions and the Evaluation of Long-Term Care. In R. L. Kane and R. A. Kane (eds.). *Values and Long-Term Care*. Lexington, MA: Lexington Books, 1982, pp. 157-88.
- . Quality-of-Life Measurement. In P. Karoly (ed.). *Measurement Strategies in Health Psychology*. New York: Wiley-Interscience, 1985b, pp. 115-46.

- _____. Quantification of Health Outcomes for Policy Studies in Behavioral Epidemiology. In R. M. Kaplan and M. H. Criqui (eds.). *Behavioral Epidemiology and Disease Prevention*. New York: Plenum Press, 1985a, pp. 31-54.
- _____. Social Support and Social Health: Is It Time To Rethink the WHO Definition of Health? In I. G. Sarason and B. R. Sarason (eds.). *Social Support: Theory, Research, and Applications*. The Hague: Martinus Nijhoff, 1985c, pp. 95-112.
- Kaplan, R. M., and J. W. Bush. Health-related quality of life measurement for evaluation research and policy analysis. *Health Psychology* 1(1):61-80, 1982.
- Kaplan, R. M., and J. A. Ernst. Do category scales produce biased preference weights for a health index? *Medical Care* 21(2):193-207, 1983.
- Kaplan, R. M., C. J. Atkins, and R. M. Timms. Validity of a quality of well-being scale as an outcome measure in chronic obstructive pulmonary disease. *Journal of Chronic Disease* 37(1):85-95, 1984.
- Kaplan, R. M., J. W. Bush, and C. C. Berry. Health status: Types of validity for an index of well-being. *Health Services Research* 11(4):478-507, 1976.
- _____. Health status index: Category rating versus magnitude estimation for measuring levels of well being. *Medical Care* 17(5):501-23, 1979.
- _____. The Reliability, Stability, and Generalizability of a Health Status Index. American Statistical Association, Proceedings of the Social Statistics Section, Washington, DC, 1978, pp. 704-709.
- Kaplan, R. M., A. Ries, and C. J. Atkins. Behavioral management of chronic obstructive pulmonary disease. *Annals of Internal Medicine*, 7(1):5-10, 1985.
- Lenfant, C. Government and community. *American Review of Respiratory Disease* 126(15):753-57, 1982.
- Lertzman, M. M., and R. M. Cherniak. Rehabilitation of patients with chronic obstructive pulmonary disease. *American Review of Respiratory Disease* 114(6):1145-65, 1976.
- Liang, M. H., K. Cullen, and M. Larson. In search of a more perfect mouse trap (health status or a quality of life instrument). *Journal of Rheumatology* 9(5):775-79, 1982.
- Meenan, R. F. AIMS approach to health status measurement: Conceptual background and measurement properties. *Journal of Rheumatology* 9(5):785-88, 1982.
- Mosteller, R. Innovation and evaluation. *Science* 211(4523):881-86, 1981.
- Pantell, R. H., and C. C. Lewis. What Is Medical Care Doing for Children? Issues in Assessment. Paper presented at the Conference on Advances in Health Assessment, Palm Springs, CA, February 1986.
- Patrick, D. L., J. W. Bush, and M. M. Chen. Methods for measuring levels of well-being for a health status index. *Health Services Research* 8(2):228-45, 1973b.
- _____. Toward an operational definition of health. *Journal of Health and Social Behavior* 14(1):6-23, 1973a.
- Patrick, D. L., et al. A cross-cultural comparison of health status values. *American Journal of Public Health* 75(12):1402-1407, 1985.
- Petty, T. L., and R. M. Cherniak. Comprehensive care of CPOD. *Clinical Notes on Respiratory Diseases* 20(1):3-12, 1981.

- Reynolds, W. J., W. A. Rushing, and D. L. Miles. The validation of a function status index. *Journal of Health and Social Behavior* 15(4):271-83, 1974.
- Russell, L. *Is Prevention Better Than Cure?* Washington, DC: Brookings, 1986.
- Stewart, A. L., et al. *Conceptualization and Measurement of Health for Adults: Vol. 2, Physical Health in Terms of Functioning*. Santa Monica, CA: RAND Corporation, 1978.
- Toevs, C. T., R. M. Kaplan, and C. J. Atkins. The costs and effects of behavioral programs in chronic obstructive pulmonary disease. *Medical Care* 22(12):1088-1100, 1984.
- Unger, K., K. Moser, and P. Hansen. Selection of an exercise program for patients with chronic obstructive pulmonary disease. *Heart and Lungs* 9(1):68-76, 1980.
- U.S. Congress, Office of Technology Assessment. *A Review of Selected Federal Vaccine and Immunization Policies: Bases on Case Studies of Pneumococcal Vaccine*. Washington DC: Government Printing Office, 1979.
- Walker, S., and R. Rosser. *Quality of Life: Assessment and Application*. London: MTP Press, 1988.
- Ware, J. E., Jr., and C. A. Donald. *Social Well-Being: Its Meaning and Measurement*. Santa Monica, CA: RAND Corporation, 1980.
- Ware, J. E., Jr., et al. Conceptualization and Measurement of Health for Adults. In *The Health Insurance Study, Vol. III, Mental Health*. Santa Monica, CA: RAND Corporation, 1979.
- Ware, J. E., Jr., et al. Health status and the use of outpatient mental health services. *American Psychologist* 30(10):1090-1100, 1984.
- Weinstein, M. C., and H. Feinberg. *Clinical Decision Analysis*. Philadelphia, PA: W. B. Saunders, 1980.
- Weinstein, M. C., and W. B. Stason. *Cost-effectiveness of Coronary Artery Bypass Surgery*. Cambridge, MA: Harvard University Center for Analysis of Health Practice, 1983.
- _____. *Hypertension: A Policy Perspective*. Cambridge, MA: Harvard University Press, 1976.
- Wells, K. B. *Depression as a Tracer Condition for the National Study of Medical Care Outcomes*. Santa Monica, CA: RAND Corporation, 1985.
- Wenger, N. K., M. E. Mattson, and C. D. Furberg. *Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies*. New York: LeJacq Publishing, 1984.