

2
Proof

POLICY-ORIENTED METHOD

Health-Related Quality of Life Measurement for Evaluation Research and Policy Analysis

Robert M. Kaplan
*San Diego State University
University of California, San Diego*

James W. Bush
University of California, San Diego

The present article describes a unit of health status, the "Well-Year," which expresses the output of health programs in terms of the number of years and the health-related "quality of life" produced by a treatment or program. Dividing the cost of the program by the number of Well-Years that it produces gives the cost-utility of the program. This cost-utility ratio can be used in a general health policy model to compare the efficiency of different programs or to assess the relative contribution of different programs and providers in the health care system. A comprehensive standardized measure of health status has many advantages for health planning, decision analysis, and program evaluation. An example demonstrates how the relative production of Well-Years by psychologists might be compared to the contribution of other health care services.

The research reported in this paper was supported by Grants 1 ROL HL 25109-1 and 1 K04 HL00809 from the National Institutes of Health (Kaplan), and by Grant 2118 HS00702 from the National Center for Health Services Research (Bust).

Address correspondence to either author, Health Policy Project, M-022 Department of Community Medicine, School of Medicine, University of California, San Diego, La Jolla, CA 92093.

Portions of this paper were based on an invited address to the Measurement and Evaluation Division (Div. 5) of the American Psychological Association, September 1, 1980, Montreal, Canada.

Comments by John Anderson are gratefully acknowledged.

The need for a comprehensive measure of health status suitable for policy analysis has appeared repeatedly in the health services research literature (see Chen, Bush and Zaremba 1975; Fanshel and Bush 1970; Hulka and Cassel 1973; Stewart, Ware, and Davies-Avery 1978 for review). This need has stimulated attempts to develop suitable measures, despite the widely discussed difficulties (Jette 1980; Keeler and Kane 1981; Sullivan 1966; Torrance 1976).

For most problems in medical research, it is possible to measure effectiveness using a single indicator, such as diastolic blood pressure or a laboratory test. These approaches are not suitable, however, for comparing the relative output of different interventions for different disease groups in different populations.

Further, such disease specific measures are of little value for assessing the consequences or side effects of the treatment (Jette 1980; Mosteller 1981). A treatment for hypertension, for example, may cause gastric irritation, nausea, and bed disability. A measure focusing only on blood pressure may miss the overall impact of the treatment upon function and symptoms. Overall assessment and comparison between groups requires a more comprehensive measure of health status that makes the relative importance of each component explicit (Fanshel and Bush 1970).

Many different paths have been used as general health outcome measures. Most of the available measures, however, are incapable of combining mortality and morbidity into the same unit, or of combining specific morbidity measures with each other (Sullivan 1966). As Mosteller noted in his presidential address to the American Association for the Advancement of Science (1981), death rates are too crude to measure the efficacy of surgery because many of its benefits are aimed at improved life quality.

The other extreme from mortality alone is the breakdown of morbidity into multiple categories ("dimensions") that are difficult to comprehend and impossible to rationally compare with one another. A truly comprehensive health status measure must rationally combine mortality with the quality of life.

A major approach to the problem of health program comparisons is "human capital" assessment, which assigns dollar values to people's lives according to their expected lifetime earnings (Mushkin 1962). The general health policy model was developed in the late 1960's to avoid the discriminatory biases of such assessments (Fanshel and Bush 1970). This was the earliest investigation of the methodological foundations of "cost-effectiveness" using preference measures in health (Weinstein and Stason 1977).

Although the activities of different health care providers and programs are diverse, they all share the common goal of improving health status. The new model, integrating substantive utility theory that is common to

economics, psychology, sociology, statistics, decision theory, medicine, public health, and operations research, proposed the concept of differences in the lifetime expected utility for evaluating health services.

Working with a series of colleagues since that early effort, Bush has led and coordinated the continued conceptual and empirical investigation of all aspects of the model and its applications (Bush, Fanshel and Chen 1972; Bush, Chen and Patrick 1973; Bush, Kaplan and Berry, in press; Chen, Bush and Patrick 1975; Chen and Bush 1976; Kaplan, Bush and Berry 1976, 1979; Patrick, Bush and Chen 1973a, b).

This effort, in which persons with extensive training in psychological research have made major contributions, is now known as the "health index approach" to policy analysis. Strange as it may seem to well-trained psychologists, it is not widely recognized among economists and decision theorists that standardization of important elements of the model is not only possible and desirable, but actually necessary for reliability and comparability between analyses and analysts (Culyer 1981; Weinstein 1980; Williams 1981).

The evolution of the terminology is worth noting. The output units of the general model were described in early publications as Quality Adjusted Life Years, derived from differences in the Quality Adjusted Life Expectancy (Bush et al 1972, 1973; Bush et al 1973a). Klarman et al (1968) had used quality-adjusted life years in an earlier study as an ad hoc method to find the minimum cost management of end-stage renal disease (1968). They did not suggest, however, that there was any compelling conceptual basis for accepting the measure, that it could be extended to maximize health across disease and program categories generally, or that preferences could actually be measured or standardized. In short, they did not propose a general approach to health policy analysis.

Nevertheless, the "quality of life" terminology was soon abandoned for the general health policy model because it has surplus meaning. In common speech and in social indicators research, that term includes all the circumstances of living, such as housing, work, recreation, environment, etc. (Campbell, Converse and Rodgers 1976; Dalkey, Rourke, Lewis and Snyder 1972; Environmental Protection Administration 1973; Hill, Chaples, Downey, Singell, Solzman and Schwartz 1973; Wingo and Evans 1978). It did not seem desirable to "medicalize" such a general term.

Alternatives considered included Function Years (Bush et al 1973); Value-Adjusted Life Years (Chen et al 1975); and since 1976, Well-Years (Chen and Bush 1976; Epstein, Schneiderman, Bush and Zettner 1980; Kaplan et al 1976, 1979). Later, when several investigators at Harvard first became interested in health status measures as criteria for resource allocation, the acronym QALY was coined for the older terminology (Zechhauser and Shepard 1976; Weinstein and Stason 1977). This acronym has recently been

adopted by the Congressional Office of Technology Assessment (1980).

The term "wellness" or "Well-Years" was chosen to imply a more direct linkage to health conditions; i.e., to denote the *health-related* quality of life. It also distinguishes an approach that uses standardized scales and measured preferences as opposed to ad hoc state definitions and arbitrary preference assignments. Regardless of the terminology, the general health policy model expresses the output of health programs in comparable units of life years adjusted for lost "quality" due to disease or disability.

Well-Years

The media often assess the effect of a disaster—a volcano, a tornado, a train wreck—by the number of lives it takes. Many other people, however, may be made partially dysfunctional by such events. To understand the full impact, we need a means of including their distress in our measure.

The total number of life-years lost is another way to think about health effects. For example, if a 45-year-old man lost his life in an accident, and we would have expected him to live to the age of 75, we might say that the incident cost him 30 life-years. Furthermore, if 18 individuals each had 30 years shaved off their life expectancy, then the total impact of the disaster may be thought of as costing 540 life-years ($= 18 \text{ people} \times 30 \text{ years/person}$).

Two categories of persons who remain alive must also be considered: individuals for whom the probability of premature death may have been increased, and those for whom the quality of life may have been diminished. The general health policy model permits various degrees of disability (including death, symptoms, and probabilities of future dysfunction) to be compared to one another.

When the proper steps have been followed, the model quantifies the health output of any treatment in terms of the years of life, adjusted for their diminished quality, it produces or saves. Thus, a "Well-Year" can be defined conceptually as the equivalent of a year of completely well life, or a year of life free of dysfunction, symptoms, and health-related problems.

A disease that reduces the health-related quality of life by one-half, for example, will take away .500 Well-Years over the course of one year. If it affects two people, it will take away 1.0 Well-Year ($= 2 \times .500$). A medical treatment that improves the level of well-being by .100 for each of 10 individuals will produce one Well-Year, if this benefit is maintained over the course of one year.

The effectiveness of health programs and treatments can be compared with each other by the number of Well-Years that they produce. Dividing the cost of a program by the number of Well-Years gives its relative efficiency or "cost-effectiveness."

Table 1 compares several health programs that have been evaluated using

TABLE 1
Comparative Cost/Well-Year of Various Programs*

<i>Program</i>	<i>Estimated Cost-Utility</i>	<i>Reference</i>
PKU screening	\$ 2,900/Well-Year	Bush et al 1973
T4 (Thyroid) Screening	\$ 3,600/Well-Year	Epstein et al 1981
Severe Hypertension Screening (diastolic > 105mm Hg)	\$ 4,850/Well-Year	Weinstein and Stason 1977
Tuberculin Testing	\$ 6,000/Well-Year	Bush et al 1972
Mild Hypertension Screening (diastolic 95-104mm Hg)	\$ 9,800/Well-Year	Weinstein and Stason 1977
Estrogen Therapy for postmenopausal symptoms	\$18,600/Well-Year	Weinstein 1980
Hospital Renal Dialysis	>> \$50,000/Well-Year	

*Inflation has produced discrepancies in the value of the dollar at different points in time. Thus the year of publication should be considered when evaluating the relative cost/Well-Year. The discount rates and preference weights are also not completely consistent. For details, see original sources.

the general health policy model. As the table demonstrates, the Well-Year concept is a powerful tool for comparing the relative efficiency of various programs. To measure Well-Years meaningfully, however, we must understand their derivation from the general framework of decision theory.

Decision Theoretic Basis

Treatment and policy decisions involve many different factors. The general model adopts the widely accepted social and legal precedent of "one person-one vote" and treats days in all lives as of equivalent social value, regardless of each person's economic status or other social attributes. With this egalitarian basis for comparing the lives and preferences of different persons, we can focus directly on the expected change in health status from potential health treatments, programs, and policy alternatives.

Improving health status means that we try to make persons live longer lives of higher "quality." With this simple statement of purpose, decision theory guides us to clearly distinguish three separate concepts or dimensions: (1) the states that a person may occupy at any point in time, (2) the probabilities ("risks") of being in the states at different times and (3) the relative desirability of occupying the states. This conceptual sequence, from the analysis of the purpose of health treatments and programs to the disaggregation of that purpose into its component concepts, is crucial to

understanding the construction of the general health policy model and the derivation of Well-Years.

Instead of "operationalizing" health status by developing a miscellaneous list of attributes, which are then related and reduced by statistical methods such as factor analysis, the utility maximization framework dictates the dimensions to be included, the dimensions to be omitted, and the model for relating the dimensions. The usual decision model must be refined, for example, by recognizing that disease states (e.g., diagnoses, blood chemistries or tissue alterations) do not affect well-being or produce dissatisfaction directly; they must be related (via risk factors, prognoses or transition probabilities) to the symptomatic and dysfunctional attributes that they generate.

It is these attributes that constitute the health-related "quality of life" and that are associated directly with satisfaction, desirability or utility. Derived in this way, the representation of health status (and its changes) has the required mathematical properties, not possessed by other "indexes" and aggregation methods, to express relative importance and to be used in cost-effectiveness and other optimization models.

Having defined the dimensions conceptually, we can develop methods to measure the different components. We first divide the target population into socially and medically similar subgroups (patient types) for separate analysis, and note the numbers in each group. For each patient type, we construct a "decision tree" of the sequence of events that would occur under different treatment and policy alternatives. This diagram includes not only change events determined by forces outside the decision-makers' control (usually the patient's disease), but also treatment choices that must be made at different point in future.

Each computation of the lifetime expected utility, therefore, represents not just one but a stream of decisions (i.e., a policy) over a set of present and future developments in the disease history. In studying tuberculin testing, for example, we must decide for analytical purposes whether later (re)occurrences of active tuberculosis will be managed by hospitalization or not (Bush et al 1972), or, in PKU screening, at what age children will be removed from their special diet (Bush et al 1973).

The well-life expectancy summarizes all available information about the risks, states, and their preferences for all outcomes from a given policy for a defined type of patient (Bush et al 1972; Bush et al 1973; Epstein et al 1981). Although subject to error, this number is an expected value that can be treated as a "certainty equivalent." Since the underlying optimization model is necessarily linear (Chen et al 1976), optimum control (decision or policy) can be achieved over a broad range of conditions by treating this number as though it is known with certainty (Pindyck 1973; Schweppe 1973; Theil 1957).

Formalized and extended over the life expectancy (to avoid bias from analyses involving mortality), the Well-Life Expectancy is nothing more than the logical conclusion of the well-known decision theory paradigm:

$$\text{Expected Utility} = \text{Expectancy (Probability)} \times \text{Value (Utility)}$$

(Edwards 1954; Edwards and Tversky 1961; Luce 1959; Luce and Raiffa 1958; Luce and Suppes 1965; Restle 1961; Tversky 1966, 1967). This general theory has also been adapted to other applications in psychology (e.g., Atkinson 1957).

The differences in the Well-Life Expectancy between the program and the reference or no program case (both expressed in Well-Years) estimates the health output or expected utility of each policy alternative, also expressed in Well-Years (Fig. 1). Thus, Well-Years, the conceptual units or basic building blocks for estimating health program outputs, are derived from differences in the lifetime expected utility. This is contrary to the approach and account given by other authors (Weinstein 1979; Weinstein and Feinberg 1980), for whom the well-life expectancy is a convenient "index"

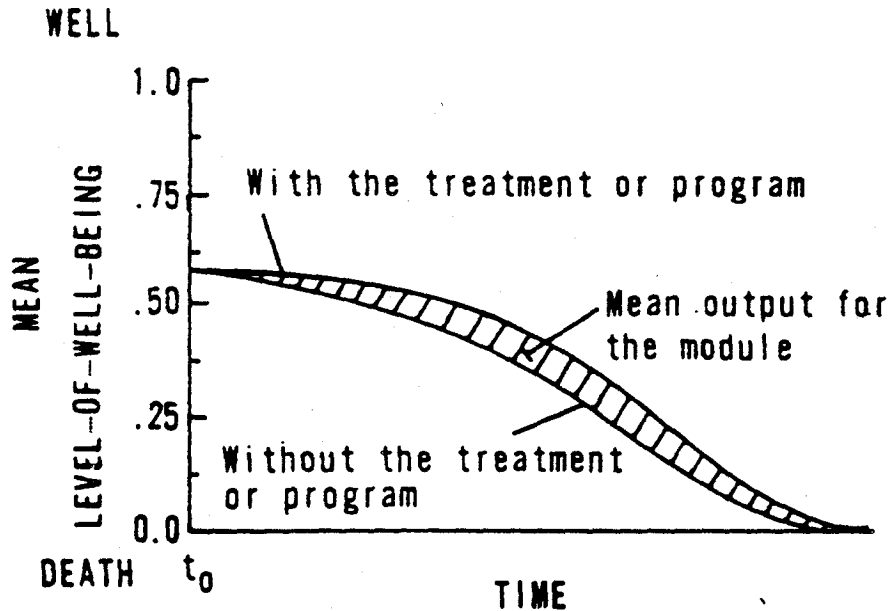


FIG. 1 The area under the upper curve is the Well-Life Expectancy with a treatment or program (in Well-Years). The area under the lower curve is the Well-Life Expectancy without the treatment or program (in Well-Years). The area between the two curves is the mean output of the treatment or program (also in Well-Years) for members of medically similar patient or population subgroups.

for trade-off analysis that "emerges from" a set of states from well to death created for other (unspecified) reasons.

Decision analysis is frequently applied to one-of-a-kind decisions, but in health, similar decisions are repeated frequently and the results of different analyses should be comparable with each other. This gives the opportunity (and the need) for standardization of both state definitions and preferences, to improve the reliability, validity, and comparability of the analyses. In order to make decisions which are comparable to those of other analysts, different decision makers must use the same types of information. This requires a uniform set of health states and a common information about the preferences for these states. Given the same input data, different decision makers can then reach the same conclusions about the relative worth of different programs within a reasonable margin of error. Methods for developing data on all components of the model will be discussed separately.

Well States: Function Level and Symptom/Problem Classification

In a refinement of traditional decision analysis, the general health policy model recognizes that the attributes of function, symptoms and problems exist at every point in time over the patient's life history—not just as final outcomes. Under a reasonable and nonrestrictive set of assumptions, furthermore, a total well-state history can be summarized (Table 4) using a limited set of attributes and a standardized set of associated preferences (Bush et al 1971; Chen et al 1975; Keeney and Raiffa 1976; Koopmans 1972).

During the early phases of the health policy project, a comprehensive set of items from multiple sources was organized into three scales that represent the different "dimensions" or attributes of daily functioning: Mobility, Physical Activity, and Social Activity. Table 2 lists the labels representing the scale steps.

Combinations of steps from the three scales are referred to as Function Levels; detailed definitions have been published elsewhere (Patrick et al 1973a, b; Chen et al 1975). Several investigators have used this classification (or modified versions of it) as an outcome measure for health program evaluation (Meenan et al 1981; Reynolds, Rushing and Miles 1974; Stewart et al 1978).

Classification of Function Levels alone is insufficient as a criterion for evaluation and resource allocation, however, since over 80% of ambulatory patients are not dysfunctional. Furthermore, preferences for states in the same level of function differ depending on which symptom or problem is causing the deviation from state of complete wellness.

All policy analyses and outcome evaluations should include the impact of

TABLE 2
Dimensions and Steps for Function Levels in the Quality
of Well-Being Scale

<i>Mobility</i>	<i>Physical Activity</i>	<i>Social Activity</i>
Drove car and used bus or train without help (5)	Walked without physical problems (4)	Did work, school, or housework and other activities (5)
Did not drive, or had help to use bus or train (4)	Walked with physical limitations (3)	Did work, school, or housework but other activities limited (4)
In house (3)	Moved own wheelchair without help (2)	
In hospital (2)	In bed or chair (1)	Limited in amount or kind of work, school, or housework (3)
In special care unit (1)		Performed self-care but not work, school or housework (2)
		Had help with self-care (1)

the treatment on relevant symptoms and problems, and not just their impact on function. So a comprehensive list of symptom/problem complexes has been added to the Function Level attributes to represent almost all the symptomatic complaints that might inhibit function (See examples, Table 3).

TABLE 3
Examples of Symptom/Problem Complexes and Linear
Adjustments for Level-of-Well-Being Scores

<i>Symptom/Problem Complex</i>	<i>Adjustment</i>
19. Pain, stiffness, numbness, or discomfort of neck, hands, feet, arms, legs, or several joints.	- .034
20. One hand or arm missing, deformed (crooked), paralyzed (unable to move), or broken (includes wearing artificial limbs or braces).	- .061
27. Burn over large areas of face, body, arms or legs.	- .110

The function level scales and the symptom/problem complexes, which describe the health experience of a person on a particular day, are the attributes that define the states of wellness or "well-states"—a term that is more comprehensive than "function states" because it includes symptoms and problems in addition to function. Using steps from the scales in Table (1), an example of such a state might be:

In house (3)

Walked with physical limitations (3)

Performed self-care, but not work, school, or housework (2)

Pain, stiffness, numbness, or discomfort of neck, hands, feet, arms, legs or several joints (19)

This standardization of the state descriptions is one aspect of the "health index approach" to policy analysis, an approach that dramatically simplifies the representation of the complexity of the disease treatment and outcome process.

The standardized case descriptions serve as the basis for the preference studies so the preferences can be meaningfully applied to actual persons. Furthermore, the same state definitions must be translated into accurate, reliable questionnaires to determine the transitions between the different states and the outcomes of treatments (Bush et al 1971; Chen and Bush 1975; Berry and Bush 1978). In studies by different groups, our instruments have now been used to classify over 50,000 person days with a classification accuracy that exceeds 96% (Anderson et al, in press; Bush 1981; Bush et al, in press). As simple as it sounds, the need for such measurements has not yet been recognized by many policy analysts.

Preferences in the Quality of Well-Being Scale

The impact of health conditions upon the quality of life is a matter of preference, value, or utility. Although a value element in definitions of health has long been recognized (Parsons 1951), Fanshel and Bush (1970) were the first to separate the dimensions and propose that preferences could be measured and incorporated into health status and outcome measurement in a systematic way.

Human judgement studies are necessary to determine preferences for the different states. For scaling purposes we can arbitrarily anchor the scales at 0.0 for death and 1.0 for completely well. These anchors do not limit the preference ratings when it is desirable to have ratings above 1.0 ("positive health") or below 0.0 ("worse than death"). For policy purposes, very precise public preferences for the states can be measured in household interview surveys. In several studies, random samples of citizens from the com-

munity evaluated the relative desirability of over 400 case descriptions or well-state profiles.

The measurement methods used were conjoint analysis (Green and Srinivasan 1978; Luce and Tukey 1964; Tversky 1967) and functional measurement (Anderson 1974; Luce 1981), based on category ratings of multi-dimensional stimuli. Using these methods, a model of the Preference structure has been developed that assigns weights to each function level scale step and symptom/problem complex to provide overall scores for all possible states of wellness with a high degree of accuracy ($R = .95$). This model has been cross-validated on a totally new set of case descriptions with an R^2 of .94 (Kaplan et al 1978).

These preferences differ little if at all between social groups and remain stable over time (Bush et al, in press). They remain invariant from one analysis to another, insuring comparability across decision situations, across analysts, and across different disease programs and treatment outcomes. Furthermore, the sensitivity of outcome measurements and estimates to variations in the preference scores can be tested very efficiently, because of the standardization. Together, the state definitions and the social preferences define the Quality of Well-Being scale (formerly the Index of Well-Being), the time specific component of the general health policy model (Fanshel and Bush 1970; Kaplan et al 1976).

The Quality of Well-Being score for different individuals can be obtained from preferences associated with their Function Levels and an adjustment for the most undesirable symptom or problem. The preference for the Function Level described in the previous section has been measured as .582 (Kaplan et al 1976), and the adjustment for the symptom or problem as -.034 (see Complex 19 in Table 3). Therefore, the Quality of Well-Being score associated with this well-state is .548 ($= .582 - .034$).

Using the symptom/problem adjustments, the scale is sensitive to variations within "high-level wellness." There are, for example, symptom/problem complexes for wearing eyeglasses, having a nasal discharge, or breathing polluted air. The adjustments apply even when a person is completely functional on the other three scales. For example, persons with a "runny nose" receive a score of .837 on the Quality of Well-Being scale when they are at the highest Function Level (see Kaplan et al 1976).

Several studies attest to the reliability (Kaplan et al 1978; Bush et al, in press) and validity (Kaplan et al 1976) of the Quality of Well-Being Scale. Convergent evidence for validity is given, for example, by high positive correlations with ratings of actual persons in the different states, and substantial negative correlations with age, number of chronic conditions, total number of symptoms, and utilization of health services.

None of these other characteristics, however, were able to make such fine distinction between levels of wellness in different persons or populations.

These data strongly support the convergent and discriminant validity of the Quality of Well-Being Scale (Kaplan et al 1976).

Still more importantly, the ratings for the well-state profiles correspond exactly to the interpersonal trade-offs that citizens wish to see implemented in health policies. This property, which is essential to the validity of all approaches to health policy analysis, has not previously been tested or demonstrated for any other preference measurement technique in health decision research (Patrick et al 1973b). These studies also provide strong convergent evidence for the validity of the preference scores.

State Transitions and the Well-Life Expectancy

Another component of the general health policy model considers transitions among states over time. The fact that different individuals are in the same state for different reasons is reflected in different expected transitions (prognoses) to the other states over the course of time. The medical characteristics of the person, including the disease or injury causing the dysfunction, determine the "health hazards" or "risk factors" both for arriving in a particular state and for departing from it, for better or for worse.

Consider two different persons in the state described earlier: one who was in this condition because of participation in a marathon race, and another because of arthritis. The marathon runner, although sore from the ordeal, is expected to be off and running again within a few days. The arthritis sufferer may, however, continue at a low level of function.

A comprehensive health status measure must include not only the current state—it must include the expected transitions to other states of wellness over the course of time. 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 limited at some future time. In terms of decision analysis, the present evaluation of these future events is captured in the lifetime expected utility or the Well-Life Expectancy. Persons with different risk factors or health hazard status have a lower well-life expectancy.

Another requirement for a health policy model is that it consider risk aversion. That is most easily implemented in a general policy model via a discount rate, an inverse process to the interest rate which can be applied to life years and health status. The discount rate simply and systematically represents the net social preference for health program outputs that are sooner and more nearly certain than later and more uncertain.

Decision analysis should consider the entire life expectancy, because health policies have long and short consequences. If the analysis considers only a limited time interval (say, five years) it ignores outcomes which occur

after the study period. A treatment which prevents early death will continue to produce life years through the remainder of the life expectancy. If the decision tree is truncated or pruned at an earlier time, benefits of averting death are attenuated. This can bias the analysis, usually against the treatment. Therefore, great care must be exercised in interpreting the results of short-term follow-up studies.

The Well-Life Expectancy is the product of Quality of Well-Being score times the expected duration of stay in each function level over a standard life period (Table 4). The expected duration of stay in each state is determined by the transition rates (Bush et al 1971; Chen et al 1975). Suppose that a group of individuals was in a completely well state, on the average, 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 death at the average age of 71.6 life years.

TABLE 4
Illustrative Computation of the Well-Life Expectancy

State	k	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: Chen, M., Bush, J. W., & Patrick D. L., Social indicators for health planning and policy analysis. *Policy Sciences*, 1975, 6, 71-89.

To adjust for the diminished well-being that they suffered in the disability states, the duration of stay in each state is multiplied by the preference measured for each state. Thus, the 4.5 years on non-bed disability becomes 2.7 Well-Years. Overall, the Well-Life Expectancy for this group is 68.5 years—a reduction of approximately 3.1 years (Chen et al 1975).

Methods to estimate the transitions among the different disease categories and states of wellness are a major problem in health outcome measurement, but detailed discussion is beyond the scope of this article (Bush et al 1971; Berry and Bush 1978). Nevertheless, disaggregating the health outcomes into the probabilities of being in particular diagnostic categories and particular states of wellness, and then applying standardized measured preferences, markedly decreases the possibilities for error and arbitrariness in the computation of Well-Years (Bush et al 1972; Bush et al 1973; Willems et al 1980; Epstein et al 1981).

The ultimate resolution of the estimation problem is to routinely incorporate standardized state definitions into randomized, prospective, and

other follow-up studies of all types (Kaplan and Atkins 1981). Such studies would relate diagnoses, disease forms, and other characteristics of the patient's condition directly to changes in the states of wellness, so policy analyses can be performed using empirical data.

Relation to Costs: The Cost/Utility Ratio

The Health Policy Project has shown in a series of publications how the Cost/Well-Year can be used to evaluate the relative efficiency of programs and health interventions. "Cost-effectiveness" is a term frequently used to refer to measures deliberately chosen to avoid problems of valuation, so the term "cost-utility" is more appropriate for output assessments based on measured preferences and expected utilities (Torrance, 1976). The "costs" associated with health programs should include not only production costs (the labor and material inputs to the treatment), but also future averted or incurred direct (health system) and indirect (economic) costs.

A major extension of the general health policy model permits a complete integration of standard economic cost-benefit analysis with costs/Well-Year for lives and health. That model consists of production costs (in dollars) minus direct and indirect economic benefits (averted future costs in dollars) divided by the expected utility of the treatment or program (in Well-Years).

The method is totally general. It rationally and comparably includes health considerations in analyses of "non-health" policies, such as coal vs. nuclear energy, or transportation policy. The health effects of an overpass to prevent accidents at a proposed intersection can be evaluated, for example, in dollars per Well-Year. In this way, all types of expenditures or regulations to improve health status can be evaluated in comparable terms.

Given a comparable health output unit, standard marginal economic analysis applies; that is, if a dollar cost/Well-Year is considered socially efficient for one program, then programs with similar cost/utility ratios are also justifiable. Although no definitive rules determine when the efficiency of a program is sufficient to justify its adoption, the following guidelines emerge from several previous analyses:

<i>Cost per Well-Year</i>	<i>Policy Implication</i>
Less than \$20,000 per Well-Year	Cost effective by current standards
\$20,000 to \$100,000 per Well-Year	Possibly controversial, but justifiable by many current examples
Greater than \$100,000 per Well-Year	Questionable in comparison with other health care expenditures

The guidelines, which will be refined as other analyses are completed, suggest to policy makers and to the public the relative efficiency of new programs in comparison with the spectrum of previous analyses.

The costs of most programs analyzed to date fall below \$20,000 per Well-Year, or well within a range that is cost-effective by existing practices and policies. The appropriateness of this \$20,000/Well-Year figure is justified by many current expenditures for tertiary medical care, and also by analyses of the economic value of human labor and consumption (human capital). The standard is further justified by the amounts that most persons are willing to pay for themselves, their families, or others, for one more year of well life.

It is more difficult to say that a program or treatment is *not* justifiable from a cost standpoint, even if it exceeds the upper extreme of about \$100,000 per Well-Year. This extreme is presently not well defined, and more evidence needs to be compiled. Fortunately, such a cutoff point is not relevant to the consideration of most policy analyses done to date. For such analyses to have comparable results, however, they should be done with careful accounting of all costs and health effects.

Application of Well-Years in Policy Analysis

A variety of health programs has been analyzed using the general health policy model, and their relative efficiency is becoming established. Hypertension screening programs have been estimated to produce a Well-Year for about \$10,000 (Stason and Weinstein 1977). Hospital renal dialysis, known to be an effective treatment because of its life-prolonging capacity, costs more than \$50,000 to produce a Well-Year.

The New York State PKU screening program (which finds only about 22 cases per year at a cost of nearly a million dollars) is still very efficient; i.e., effective relative to costs, because it yields a Well-Year for about \$2,900 in 1970 currency (Bush et al 1973). The Congressional Office of Technology Assessment (OTA) recently used the general model, with Function Levels and preference weights from our 1973 PKU study, to analyze a national pneumococcal vaccine program (1979). From existing clinical data, they estimated that the vaccine prevents approximately 60% of all pneumococcal pneumonias, that about 5% of the vaccine recipients react with swelling and fever, and that more severe reactions, such as temporary paralysis from the Gullian-Barre Syndrome, occur in one case per 100,000.

The average cost per Well-Year expected from the vaccine across all age groups is \$4,800 (1979 dollars). For young children, who rarely die of pneumonia, the costs were as high as \$77,200/Well-Year, while the costs for the aged, the most frequent victims, are less than \$1,000 per Well-Year.

The general health policy model considers both positive and negative health effects by mapping them on to the same scale. Reducing the yield of

Well-Years by subtracting the Gullian-Barre disabled (at the same rate as from the swine flu program), or adding to the program costs by the increase in insurance premiums, increases the overall cost/utility ratio by a modest \$100/Well-Year. For the elderly, it had no measurable effect. The Well-Years produced by avoiding pneumonias greatly outweigh the setbacks caused by severe side effects with very low probabilities. This illustrates how the general health policy model, with a common unit of health output, can simplify a complex situation and make the appropriate decision become obvious. In early 1980, for the first time in history, Congress authorized reimbursement for a preventive procedure under Medicare (P.L. 96-611).

Psychologists, Evaluation Research and Health Services Reimbursement

One of the major issues facing professional psychologists is the struggle for reimbursement under various health insurance plans. As the American public becomes increasingly sensitive to increases in health care costs, proposed new health expenditures will be more carefully evaluated.

If psychologists are to gain a larger place in the delivery of health care, they must demonstrate that their services are cost-effective. Attempts to persuade policy makers that the skills of psychologists are worthwhile (Kiesler, Cummings, & Vanden Bos 1979) are often not convincing because psychological and medical services are evaluated, if at all, using different outcome measures.

Using Well-Years, cognitive and behavioral interventions can be evaluated and related to costs in the same way as medical interventions.

In a randomized trial now under way, for example, Kaplan and Atkins (1981) are evaluating inducements to exercise for patients with chronic obstructive pulmonary disease. At the end of three months, the treated and control groups display a statistically significant difference of .110 on the Quality of Well-Being Scale.

If this difference persists, preliminary analysis of the costs of the treatment suggest that the cost/utility ratio will be considerably less than \$5,000 per Well-Year. Most of the other treatment modalities practiced by psychologists can be evaluated in the same way, at least as far as the final effects of treatment on function, symptoms, and well-being are concerned.

Conclusions

In this brief article we have only been able to introduce a few aspects of the general health policy model. Despite the difficult problems associated with its development, it has many practical uses: (1) Most importantly, the general health policy model is necessary for cost/utility analysis and

resource allocation. In addition, different components of the model are useful (1) to measure the effectiveness of medical interventions; (2) to assess the quality of health care; (3) to assess the health care needs of different populations; (4) to improve clinical decision making; and (5) to help understand causes of variations in health (Ware, Brook, Davies and Lohr 1981).

Health psychologists and other methodologists interested in evaluation may be interested in a variety of issues relevant to the continued development and utilization of these measures. Some of these issues include the value of using aggregate scales versus separate indicators, the validity of general measures for dysfunction caused by mental symptoms, the statistical power of the scales for detecting differences with small samples, the role of discounting, better data on costs, appropriate methods for preference measurement, improved questionnaire techniques, the application to disease specific groups, methods of estimating transition probabilities, ethical issues with strict efficiency measures (distributional fairness), and many others. With continued development, we expect more widespread use of the general health policy model and Well-Years in the future.

REFERENCE

- Anderson, J. P., Bush, J. W., & Berry, C. C. Performance, capacity and self-classification in health outcome and quality of life studies. *American Journal of Public Health*, in press.
- Anderson, N. H. Information integration theory: a brief survey. In D. H. Krantz, R. G. Atkinson, R. D. Luce and P. Suppes (Eds.), *Contemporary Developments in Mathematical Psychology* (Vol. 2). New York: Freeman, 1974.
- Atkinson, J. W. Motivational determinants of risk-taking behavior. *Psychological Review*, 1957, 64, 359-372.
- Berry, C. C., & Bush, J. W. Estimating prognoses for a dynamic health index, the weighted life expectancy, using the multiple logistic with survey and mortality data. In Proceedings of the *American Statistical Association, Social Statistics Section*, 1978.
- Bush, J. W. *Quality of Well-Being Scale: Function Status Profile and Symptom/Problem Complex Questionnaire*. San Diego: Health Policy Project, University of California.
- Bush, J. W., Chen, M., & Patrick, D. L. Cost-effectiveness using a health status index: analysis of the New York State PKU screening program. In R. Berg, (Ed.), *Health Status Indexes*. Chicago: Hospital Research and Educational Trust, 1973.
- Bush, J. W., Chen, M., & Zaremba, J. Estimating health program outcomes using a Markov equilibrium analysis of disease development. *American Journal of Public Health*, 1971, 61, 2362-2375.
- Bush, J. W., Fanshel, S., & Chen, M. Analysis of a tuberculin testing program using a health status index. *Social-Economic Planning Sciences*, 1972, 6, 49-68.
- Bush, J. W., Kaplan, R. M., & Berry, C. C. A standardized quality of well-being scale for cost-utility and policy analysis in health: reliability and generalizability. *Medical Care*, in press.
- Campbell, A., Converse, P. E., & Rodgers, W. L. *The Quality of American Life: Perceptions, Evaluations, and Satisfaction*. New York, N.Y.: Russell-Sage Foundation, 1976.

- Chen, M., & Bush, J. W. Maximizing health system output with political & administrative constraints using mathematical programming. *Inquiry*, 1976, 13, 215-227.
- Chen, M., Bush, J. W., & Patrick, D. L. Social indicators for health planning and policy analysis. *Policy Sciences*, 1975, 6, 71-89.
- Chen, M., Bush, J. W., & Zaremba, J. Effectiveness Measures, Chapter 12 in L. J. Shuman, R. D. Speas, & J. P. Young (Eds.), *Operations Research in Health Care—A Critical Analysis*. Baltimore: The Johns Hopkins University Press, 1975.
- Culyer, A. J. Health, economics, and health economics. In J. Van der Gaag & M. Perلمان (Eds.), *Health, Economics, and Health Economics*. New York: North Holland, 1981.
- Dalkey, N. C., Rourke, D. L., Lewis, R., & Snyder, D. *Studies in the Quality of Life*. Lexington, MA: Lexington Books, D. C. Heath & Co., 1972.
- Edwards, W. The theory of decision-making. *Psychological Bulletin*, 1954, 51, 380-417.
- Edwards, W., & Tversky, A. *Decision Making*. Baltimore: Penguin, 1967.
- Environmental Protection Administration, Office of Research and Monitoring. *The Quality of Life Concept: A Potential New Tool for Decision-Makers*. Washington, D.C., U.S. Government Printing Office, 1973.
- Epstein, K. A., Schneiderman, L. J., Bush, J. W., & Zettner, A. The abnormal screening serum thyroxine (T4): analysis of physician response, outcome, cost and health effectiveness. *Journal of Chronic Disease*, 1981, 134, 175-190.
- Fanshel, S., & Bush, M. W. A health status index and its applications to health-services outcomes. *Operations Research*, 1970, 18, 1026-1066.
- Green, E. P., & Srinivasan, V. Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, 1978, 5, 103-123.
- Hill, A. D., Chaples, E. A., Downey, M. T., Singell, L. D., Solzman, D. M., & Swartz, G. M. *The Quality of Life in America: Pollution, Poverty, Power & Fear*. Chicago: Holt, Rinehart, and Winston, Inc., 1973.
- Hulka, B. S., & Cassel, J. C. The AAFP-UNC study of the organization, utilization, and assessment of primary medical care. *American Journal of Public Health*, 1973, 63, 494-501.
- Jette, A. M. Health status indicators: their utility in chronic disease evaluation research. *Journal of Chronic Disease*, 1980, 33, 567-579.
- Kaplan, R. M., & Atkins, C. J. Effects of exercise upon health status and life qualities among patients with chronic lung disease. Presented at the *American Psychological Association Meeting*. Los Angeles, August, 1981.
- Kaplan, R. M., Bush, J. W., & Berry, C. C. Health status: types of validity for an index of well-being. *Health Services Research*, 1976, 11, 478-507.
- Kaplan, R. M., Bush, J. W., & Berry, C. C. Health status index: category rating versus magnitude estimation for measuring levels of well-being. *Medical Care*, 1979, 17, 501-523.
- Kaplan, R. M., Bush, J. W., & Berry, C. C. The reliability, stability and generalizability of a health status index. *American Statistical Association, Proceedings of the Social Statistics Section*, 1978, 704-709.
- Keeler, E., & Kane, R. What is special about long-term care? in R. L. Kane & R. S. Kane (Eds.), *Value Preferences and Long-Term Care*. Lexington, MA: D. C. Heath & Co., 1981.
- Keeney, R. L., & Raiffa, H. *Decisions with Multiple Objectives: Preferences & Value Trade-offs*. New York: John Wiley & Sons, 1976.
- Kiesler, R. A., Cummings, N. A., & Van den Box, G. R. *Psychology and National Health Insurance*. Washington, D.C.: APA, 1979.
- Klarman, H. E., Francis, J. O., Rosenthal, G. D. Cost-effectiveness analysis applied to the treatment of chronic renal disease. *Medical Care*, 1968, 6, 48-54.
- Koopmans, T. C. Representation of preference orderings over time. In C. B. McGuire, R. Radmar (Eds.), *Decision & Organization*. Amsterdam: North Holland Publishing Co., 1972.

- Luce, R. D. *Individual Choice Behavior*. New York: John Wiley & Sons, 1959.
- Luce, R. D. Axioms for the averaging and adding representations of functional measurement. *Mathematical Social Sciences*, 1981, 1, 139-144.
- Luce, R. D. & Raiffa, H. *Games and Decisions*. New York: John Wiley & Sons, 1958.
- Luce, R. D., & Suppes, P. Preference, utility, subjective probability. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. III). New York: John Wiley & Sons, 1965.
- Luce, R. D., & Tukey, J. W. Simultaneous conjoint measurement: a new type of fundamental measurement. *Journal of Mathematical Psychology*, 1961, 1, 1-27.
- Meenan, R. J., Gertman, P. M., & Mason, J. H. Measuring health status in arthritis: the arthritis impact measurement scales. *Arthritis & Rheumatism*, 1980, 23, 145-152.
- Mosteller, F. Taking science out of social science. *Science*, 1981, 212, 291.
- Mushkin, S. Health as an investment. *Journal of Political Economy*, 1962, 70, 129.
- Office of Technology Assessment, United States Congress. *A Review of Selected Federal Vaccine and Immunization Policies: Based on Case Studies of Pneumococcal Vaccine*. Washington, D.C.: U.S. Government Printing Office, 1979.
- Parsons, T. *The Social System*. New York: The Free Press, 1951.
- Patrick, D. L., Bush, J. W., & Chen, M. Toward an operational definition of health. *Journal of Health & Social Behavior*, 1973a, 14, 6-23.
- Patrick, D. L., Bush, J. W., & Chen, M. Methods for measuring levels of well-being for a health status index. *Health Services Research*, 1973b, 8, 228-245.
- Pindyck, R. S. *Optimal Planning for Economic Stabilization*. Amsterdam: North Holland, 1973.
- Restle, F. *Psychology of Judgement and Choice*. New York: John Wiley & Sons, 1961.
- Reynolds, W. J., Rushing, W. A., & Miles, D. L. The validation of a function status index. *Journal of Health and Social Behavior*, 1974, 15, 271.
- Riddiough, M. A., & Willems, J. S. Federal policies affecting vaccine research and production. *Science*, 1980, 209, 563-566.
- Schweppe, F. C. *Uncertain Dynamic Systems*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1973.
- Stason, W. B., & Weinstein, M. C. Allocation of resources to manage hypertension. *New England Journal of Medicine*, 1977, 296, 732-739.
- Steward, A. L., Ware, J. E., Brook, R. H., & Davies-Avery, A. *Conceptualization and Measurement of Health for Adults: Vol. 2 Physical Health in Terms of Functioning*. Santa Monica: Rand Corp., 1978.
- Sullivan, D. F. *Conceptual Problems in Developing an Index of Health*. Washington, D.C.: National Center for Health Statistics, 1966.
- Theil, H. A note on certainty equivalence in dynamic planning. *Econometrica*, 1957, 25, 346-349.
- Torrance, G. W. Toward a utility theory foundation for health status index models. *Health Services Research*, 1976, 11, 439-469.
- Tversky, A. Utility theory and additivity analysis of risky choice. *Journal of Experimental Psychology*, 1966, 71, 680-683.
- Tversky, A. Additivity, Utility, and subjective probability. *Journal of Mathematical Psychology*, 1967, 4, 175-201.
- Tversky, A. A general theory of polynomial conjoint measurement. *Journal of Mathematical Psychology*, 1967, 4, 1-20.
- Weinstein, M. C. Economic evaluation of medical procedures and technologies: progress, problems and prospects. In J. L. Wagner (Ed.), *Medical Technology*, U.S. Dept. of HEW Pub. (PHS) 79-3254. Washington, D.C.: U.S. Government Printing Office, 1979.
- Weinstein, M. C. Estrogen use in post-menopausal women—costs, risks & benefits. *New*

- England Journal of Medicine*, 1980, 303, 308-316.
- Weinstein, M. C. & Fineberg, H. *Clinical Decision Analysis*. Philadelphia: Saunders Co., 1980.
- Weinstein, M. C. & Stason, W. B. Foundations of cost-effectiveness analysis for health and medical practices. *New England Journal of Medicine*, 1977, 296, 716.
- Willems, J. S., Sanders, C. R., Riddiough, M. A. & Bell, J. C. Cost-effectiveness of vaccination against pneumococcal pneumonia. *New England Journal of Medicine*, 1980, 302, 553-559.
- Williams, A. Welfare economics & health status measurement. In J. Van der Gaag & M. Perman (Eds.), *Health, Economics and Health Economics*. New York: North Holland, 1981.
- Wingo, L. & Evans, A. *Public Economics and the Quality of Life*. Baltimore: Johns Hopkins University Press, 1978.
- Zechhauser, R. & Shepard, D. Where now for saving lives? *Law and Contemporary Problems*, 1976, 40, 5-45.