

Fundamental Methodologies In Identifying Values In The Evaluation Of Health Programs

M. Jane Fulton
School of Health Administration
University of Ottawa

RÉSUMÉ

L'évaluation de programmes de soins médicaux débute par la détermination de la valeur attachée par le patient aux divers résultats possibles d'un traitement. A partir d'une analyse de documents publiés en psychologie et en économie, l'article qui suit décrit plusieurs stratégies visant à mesurer ces valeurs ou préférences. Il explore également l'application de techniques de classement, de graduation de préférences, d'arbitrage du temps et d'incorporation du risque dans la mesure de l'utilité d'un traitement.

ABSTRACT

The evaluation of programs that deliver medical care begins with a determination of the value a patient places on the outcomes of treatment. This paper describes several strategies drawn from the fields of psychology and economics that identify these values or preferences and attempt to measure them. Techniques for ranking, preference scaling, time-trade-off, and incorporating risk into a measure of utility of a treatment are explored.

Introduction

Evaluating health care programs is a process of dealing with uncertainty and incomplete information whose objective is to improve decision making about future resource allocation. In order to achieve these objectives, a number of methods of processing available information have been borrowed from disciplines such as economics and adapted to the health care environment.

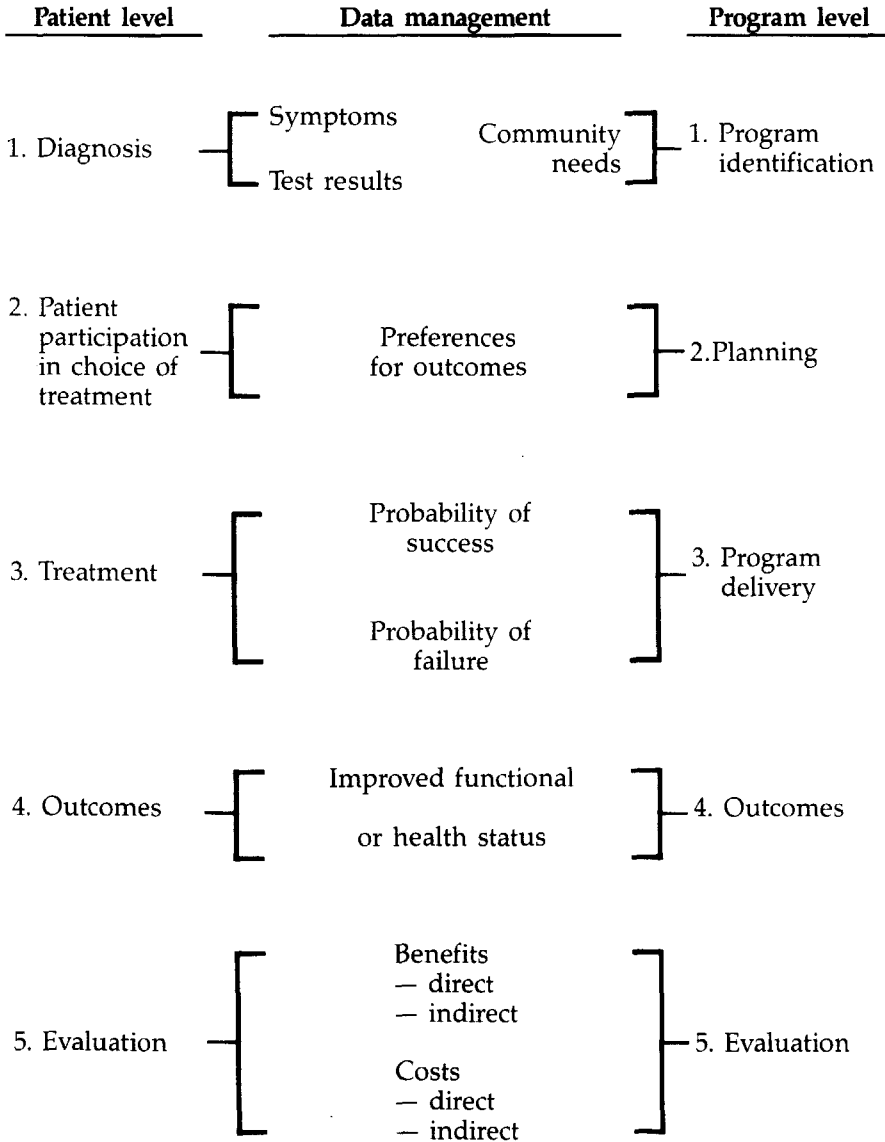
There is no shortage of articles on cost-benefit and cost-effectiveness methods of evaluation, data analysis is a burgeoning field, and program planning commonly incorporates an element of evaluation. An aspect of evaluation that may have been overlooked is the contribution to decision making that can be made by an improved understanding of the preferences of those people whom health programs aim to serve. For without an understanding of the value to patients that may be attributed to the delivery of health care and to its consumption, there is no *real* measure of the *benefit* of a program. We can compute any set of data, and if it is not a meaningful set, managerial decisions will likely range from less than optimal to useless.

This paper explores a cross-section of methods designed to elicit the preferences of patients in a health care setting. The discussion is structured to first, define utility as a measure of value and, second, to describe the most vigorously tested research methods whose aim is to measure values.

Understanding Values and Choices

There are many factors relevant to a patient's choice of treatment for disease. These include, but are not limited to: pain, morbidity, length of hospital

Figure 1
Decision Making in Health Care



stay, and health outcome. Choices based on real income and net worth measures have been avoided because they have very little influence on a patient's access to health care in Canada.

Klein (1983) has stated that people think about only a subset of all relevant attributes when the choices are broad or varied. For each patient, only a few attributes are relevant across the range of treatment options. Many attributes will be evaluated by the physician acting as an agent or advisor to the patient, while a number of attributes will be too difficult for some patients to utilize because of limits on ability to process information (Klein, 1983).

Some problems may arise due to the choice of outcome attributes. Values close together may demonstrate errors of proportional difference of greater magnitude than values near the extreme points. For example, patients may readily distinguish between two weeks in hospital with pain and two weeks in good health, but they may have difficulty distinguishing between two weeks in hospital recovering from surgery and two weeks as an invalid at home with severe pain. So, even if there is a real value difference for the patient between these two states, measurement of this difference may not be possible.

From Values to a Measure of Utility

Utility reflects the desirability of an outcome or the value of one outcome as compared to another (Keeny and Raiffa, 1976; Merz, 1983). Its numerical value is attached to the "worth of a consequence" or the outcome of a decision when compared to an alternative outcome (Lindley, 1976, p. 101). A utility scale is a description of how preferences change as various factors change in either the patient's health or in the environment. For example, if a patient with cancer of the lung has the lung removed surgically and is otherwise capable of living a normal life, the worth of living with only one lung has a measurable utility. This number is usually expressed as a proportion of the value of living in perfect health.

For patients with chronic disease, the treatment utility is made up of a number of components, both positive and negative. Positive components (or outcomes) include: 'cure' of the disease, relief from symptoms such as pain, elimination or reduction of dollar costs of hospital treatment or drug therapy, and establishing a firm diagnosis (Berwick and Weinstein, 1985). Negative outcomes include: Physical or emotional disability, family disruption, lengthy hospitalization, side effects from drugs, continued or more severe pain, continuing costs, and death (Kassirer, 1976, p. 155). Stated preferences for a treatment outcome are indices not only of the subjective value of the treatment to the patient but also its value to the patient's family.

Measurement Strategies

Simple measurement strategies can be followed in identifying values. As Beach and Barnes (1983) found, patients tried to describe their preferences regardless of how they were asked to communicate. The methods with the

best acknowledged success include: the standard gamble, scenarios, ranking, time-trade-off techniques, and ordering.

Standard Gamble, Scenarios and Ranking

The 'standard gamble' method of eliciting personal preferences in a situation of uncertainty has been attributed to Von Neumann and Morgenstern (1944). In this method, subjects are asked to state an equivalent outcome to a 50-50 gamble described in a scenario. They may also be asked to choose between a 50-50 gamble and an intermediate outcome that occurs with certainty. Such an intermediate outcome is called a 'certainty equivalent'.

Vertinsky and Wong (1974) have evaluated methods of preference measurement using scenarios to present different alternatives to subjects. The alternatives described the probability of a positive outcome coupled the probability of a negative outcome. These gambles offered equal chances at either a small or large number of bed rest days due to treatment outcomes for an illness. Subjects were asked to state how many days they would rest in bed for sure rather than take a gamble offering either a long or a short hospital stay.

Each subject's utility was calculated from stated utilities of the alternative bed rest days. The following formula was used:

$$u(x) = 0.50 u(a) + 0.50 u(b)$$

The small number of days of bed rest was set to zero and the large number (up to 15) was set to one. In this way, each choice of a certainty equivalent generated a data point used to create a utility measure. Subjects in the Vertinsky and Wong (1975) study found the standard gamble method to be 'easier to handle'.

McNeil and Pauker (1979) used a method similar to scenarios in a study of patient preferences for treatment for carcinoma of the bronchus and lung. The scenario was separated from the probability of outcomes statement and a diagram was used to assist patients in visualizing the probabilities.

McNeil et.al. (1978) used a 50/50 gamble strategy to develop utility scales for patients exposed to surgical risk. Their method involved five basic steps. First, patients were asked to choose how many years they would like to live in good health: 10 or 25 years. Second, these patients were asked to think of a 50/50 gamble in terms of a coin toss. If heads appeared they would get the long term survival, if tails appeared they would not survive treatment. Third, these patients were offered a choice between taking the gamble or stating a fixed period of life equivalent to that gamble. (The lower this certainty equivalent, or guaranteed survival, the more averse a patient was to risk).

Fourth, McNeil et.al. (1979) took the certainty equivalent (for example, 5 years) and asked the gamble again, using the new equivalent in place of death (written as 0 years of survival). The question was then a choice between a 50/50 gamble of 5 years before death and the choice of 10 or 25 years from step 1 and certainty equivalent selected by the patient. Finally, step four was repeated using the 50/50 gamble between the certainty equivalent and death or a choice of a guaranteed survival. Patients who were averse

to the risk of surgical death, chose less than 3 years of survival rather than take the 50-50 gamble between death and 10 years of good health.

An adaptation of the McNeil et al. (1978) method was used by Fulton (1986) in a study of patients with heart disease. To measure the value of reducing treatment risk for each patient, subjects made five consecutive choices. First, each patient decided whether to accept a hypothetical treatment for heart disease that offered a 99% change of a normal life expectancy with no symptoms of disease. Almost all agreed. Second, patients decided whether to accept a 100% chance of living in perfect health for only one year. Almost all refused.

Then patients made three choices of accepting a treatment with a 25%, a 50% or a 75% risk of death or stating the least number of years of life they would accept rather than take the risk. Patient values for risk avoidance were identified by plotting the certainty equivalents (years accepted for sure).

Ranking and Preference Scaling

Beach and Barnes (1983) expanded on the work of Stillwell et al. (1982) to demonstrate several approximate measures that were useful in eliciting subject preferences. These included a simple ranking method, a seven point preference scale rating, and voting. The study determined that "the simple seven point rating scales yield an ordering that correlates .9 with the ordering derived from the ranking method and .89 with the ordering derived from the points method" (Beach and Barnes, 1983, p. 423). The authors recommend against the voting method because it had the lowest correlation with other methods studied. Isen and Patrick (1983) also report the successful use of scaling to demonstrate preferences in an hypothetical risk situation.

Pliskin et al. (1985) have stated that a patient need not attempt an assessment of choices in order to create a two attribute utility function. Instead, the patient may be presented with pairs of choices and asked to rank the pairs. Ranking was feasible in the Pliskin (1985) study because the outcome measure (the number of radiographs performed per year) was also a decision variable that had influence over the probability of the second attribute (probability of a tooth lesion).

Time-Trade-Off

Time trade-off has been considered to be empirically equivalent to the Von Neumann and Morgenstern standard gamble by Torrance et al. (1982). In the Torrance study, patients were asked to determine a point of indifference between a lifetime (70 years) with a chronic illness and a shorter but healthy life. The interviewer supplemented the question with visual aids. Since no lotteries were used, the researchers could not evaluate the patient attitudes toward risk, instead they developed a measure of the population mean of health states preferred to death. The authors recommended that questions asking patients to scale attributes of health states offer the subjects one very bad and one very good outcome. Statistical analysis of the ordered preferences is simplified by this strategy.

An example of the time-trade-off strategy has been reported by Read et al.

(1984), who found that subjects in a study of physician responses to choices of health outcomes stated that 15 years of life with severe angina was worth less than 10 years of life with moderate angina.

Ordering

Torrance et.al. (1982) have described a method they used to develop a cardinal ordering of the utility of various health states. Subjects were asked to place arrows corresponding to health states on a 'feeling thermometer'. The spacing between the arrows was the source of data on the relative difference between the value of various health states. Perfect health was listed as 100 and death was listed as zero.

Using measures of value

Using a utility measure allows the combination of many attributes of health into a common denominator, such a 'quality of life years'. This measure discounts years of sickness or disability so that more disabled years of life are then equivalent to one year of good health.

Several variables that could influence values have been explored, including age and time. Although patient preferences based on age of the patient have not been explored in detail, some evidence has been reported on macro-choice differences that appear in older patients. Cassileth et.al. (1980) found that younger patients wanted to participate in decisions, while many older patients wanted to abdicate and let doctors make decisions.

Curley et.al. (1984) studied 306 outpatients and spouses and found that 33% of older patients wanted to defer decisions to doctors. The study also explored the effect of ambiguity and found that 21% of the subjects refused ambiguous treatment while only 3% refused risky treatment. Pendelton and House (1984) also confirm that older patients with diabetes are less interested in being involved in personal care than younger patients.

Indications are that once patient preferences are identified they demonstrate stability over time. Torrance et.al. (1982) found that in repeated measures of health state preferences on the same individual, choices remained consistent. This was demonstrated by high correlation coefficients (0.86 to 0.94).

Conclusion: Using Values in Evaluation

Decision making and resource allocation in health care depend on program evaluation that begins with an assessment of the values of those who consume health services. These values are most easily identified as preferences for specific treatment outcomes. This paper has briefly described successful methods of this important step in program or treatment evaluation.

References

Beach, L.R. and V. Barnes (1983), *Approximate Measurement in a Multiattribute Utility Context, Organizational Behaviour and Human Performance*, Vol. 32, pp. 417-424.

- Berwick, D.M. and M.C. Weinstein (1985), What do Patients Value? Willingness to Pay for Ultrasound in Normal Pregnancy, *Medical Care*, Vol. 23, No. 7, pp. 881-893.
- Cassileth, B.R., R.V. Zupkis, K. Sutton-Smith and V. March (1980), Information and Participation Preferences Among Cancer Patients, *Annals of Internal Medicine*, Vol. 92, No. 6, pp. 832-836.
- Curley, M.A., S.E. Eraker and J.F. Yates (1984), An Investigation of Patients' Reactions to Therapeutic Uncertainty, *Medical Decision Making*, Vol. 4, No. 4, pp. 501-511.
- Fulton, M.J. (1986), Patient Preferences in the Decision to Treat Coronary Artery Disease, Doctoral dissertation, the University of British Columbia.
- Isen, A.M. and R. Patrick (1983), The Effect of Positive Feelings on Risk Taking: When the Chips are Down, *Organizational Behaviour and Human Performance*, Vol. 31, pp. 194-202.
- Kassirer, J.P. (1976), The Principles of Clinical Decision Making: An Introduction to Decision Analysis, *Yale Journal of Biology and Medicine*, Vol. 9, pp. 149-164.
- Keeney, R.L. and H. Raiffa (1976) Decisions with Multiple Objectives: Preferences and Value Tradeoffs, (New York, Wiley).
- Klein, N.M. (1983), Utility and Decision Strategies: A Second Look at the Rational Decision Maker, *Organizational Behaviour and Human Performance*, Vol. 31, pp. 1-25.
- Lindley, D.V. (1976), Costs and Utilities, in Decision Making and Medical Care, F.T. Dombal and F. Gremy, eds (Armsterdam, North Holland).
- Lusted, L.B. (1968), Introduction to Medical Decision Making (Springfield, Ill., Charles C. Thomas).
- McNeil, B.J. and S.G. Pauker (1979), The Patients' Role in Assessing the Value of Diagnostic Tests, *Radiology*, Vol. 132, No. 3, pp. 605-610.
- McNeil, B.J., R. Weischselbaum and S.G. Pauker (1978), Fallacy of the Five-Year Survival in Lung Cancer, *New England Journal of Medicine*, Vol. 306, pp. 1259-1262.
- Pendelton, L. and W.C. House (1984), Preferences for Treatment Approaches in Medical Care, College Students Versus Diabetic Outpatients, *Medical Care*, Vol. 22, No. 7, pp. 644-646.
- Pliskin, J.S., M. Schwartz, H.G. Grondahn, and J. Boff (1985), Incorporating Individual Patient Preferences in Scheduling Bitewing Radiographs, *Methods of Information in Medicine*, Vol. 24, pp. 213-217.
- Read, J.L., R.J. Quinn, D.M. Berwick, H.V. Fineberg, and M.C. Weinstein (1984), Preferences for Health Outcomes, Comparison of Assessment Methods, *Medical Decision Making*, Vol. 4, No. 3, pp. 315-329.
- Torrance, G.W., M.H. Boyle, and S.P. Horwood (1982), Application of Multi-Attribute Utility Theory to Measure Social Preferences for Health States, *Operations Research*, Vol. 30, No. 6, pp. 1043-1069.

Vertinsky, I. and E. Wong (1975), Preferences and the Construction of Indifference Maps: A Comparative Empirical Evaluation of Two Measurement Methodologies, *Sociology, Economics and Planning Sciences*, Vol. 9, pp. 15-24.

Von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behaviour* (Princeton, Princeton University Press).