Primer on Medical Decision Analysis:

Part 3-Estimating Probabilities and Utilities

GARY NAGLIE, MD, MURRAY D. KRAHN, MD, MSc, DAVID NAIMARK, MD, DONALD A. REDELMEIER, MD, MS(HSR), ALLAN S. DETSKY, MD, PhD

This paper describes how to estimate probabilities and outcome values for decision trees. Probabilities are usually derived from published studies, but occasionally are derived from existing databases, primary data collection, or expert judgment. Outcome values represent quantitative estimates of the desirability of the outcome states, and are often expressed as utility values between 0 and 1. Utility values for different health states can be derived from the published literature, from direct measurement in appropriate subjects, or from expert opinion. Methods for assigning utilities to complex outcome states are described, and the concept of quality-adjusted life years is introduced. Key *words:* decision analysis; expected value; utility; sensitivity analysis; decision trees; probability. (Med Decis Making 1997;17:136-141)

Probabilities and outcome values are two of the basic elements of a decision analysis. A probability is a quantitative estimate of the likelihood that a given outcome depicted in the tree will occur. An outcome value is a quantitative expression of the desirability of such an outcome. The validity of a decision analysis depends on the accuracy of these numerical estimates. This paper reviews some practical approaches for estimating probabilities and outcome values.

Estimating Probabilities

The goal of estimating probabilities for a decision tree is to find the most accurate estimate for the probability of each event in the model. The best estimate for each probability value is called the "baseline" estimate. The analysis that uses the best estimates of the probabilities is called a "baseline" analysis. Since there is usually some uncertainty about the best estimate for each probability, the range of reasonable estimates should be specified. This range may reflect the variety of estimates from different studies or may be based on the 95% confidence interval from a single study. The less confidence you have in the numerical estimate of a probability value, the wider the range should be. The range of values for each probability can be used in a sensitivity analysis to assess how different numerical estimates can affect the overall result of the decision analysis (see Part 4 of this series).[1]

In order to estimate probabilities, the best available information should be sought.[2-4] You should start with a systematic search of the literature, which generally involves the following steps: a computerized literature search, a search of personal files and the files of content experts, and a review of reference lists from retrieved articles.[2,5,6] Once published studies have been identified, the next step is to evaluate the validity of their results by applying critical appraisal criteria.[7] When the quality of a study is poor, you cannot have much confidence in any probability estimate derived from it. Even when high-quality published studies exist, the results of the studies may not apply to your model if the study population or the treatment intervention differs from that in the model. Additionally, if the study assesses the treatment under optimal circumstances of adherence and follow-up, the results may overestimate the effectiveness that you may expect in your population.'

After completing a systematic search of the literature, you will usually have several relevant published papers. If a single study stands out as being exemplary in methodologic quality and relevance to your analysis, use its results for your probability es-

Received November 27, 1995, from the University of Toronto Programme in Clinical Epidemiology and Health Care Research (The Toronto Hospital and The Sunnybrook Health Science Centre Units) and the Departments of Health Administration, Medicine, and Clinical Biochemistry, Toronto, Ontario, Canada. Revision accepted for publication December 12, 1996. Drs. Naglie and Krahn are partially supported by Arthur Bond Fellowships from the Physicians' Services Incorporated Foundation. Drs. Detsky and Redelmeier are partially supported by Career Awards from the National Health Research and Development Programme and the Ontario Ministry of Health, respectively.

Address correspondence and reprint requests to Dr. Detsky: EN G-246, General Division, The Toronto Hospital, 200 Elizabeth Street, Toronto, ON M5G 2C4, Canada. e-mail: (detsky@utstat.toronto.edu).

timates. If several relevant studies exist, eliminate studies that are of poor methodologic quality, and then use an average of the results from the remaining studies to estimate the probability values. You may be lucky enough to find a published meta-analysis, which averages the results of several studies, taking into account factors such as study size and study quality.',' Once you have obtained your probability estimates, a useful way to display them is in a table that contains the baseline estimate for each probability, the range of values considered reasonable, and the reference sources used (see table 1).[2]

As an example of how to derive probability estimates from the literature, let us once again consider the giant cell arteritis (GCA) decision tree shown in Part 2 of this series.' The key probability estimates for this tree are: the probability of an adverse outcome from GCA, the sensitivity and specificity of a temporal artery biopsy in diagnosing GCA, the effectiveness of prednisone in reducing the risk of an adverse outcome from GCA, and the probability of a serious side effect from prednisone.[10]

The major adverse outcome from GCA is permanent blindness. A systematic review of the literature for studies to estimate the probability of permanent blindness and the effectiveness of prednisone in preventing this complication revealed several observational studies, but no randomized trial. Three studies were found that identified cases of GCA with normal vision at the time of diagnosis and assessed the development of blindness in patients treated with prednisone and in patients given no treatment (i.e., historical controls).[10] We calculated the average for the three studies, and obtained probabilities of blindness of 0.120 without prednisone and 0.013 with prednisone. The baseline estimate of the effectiveness of prednisone in preventing this complication was then calculated by using the formula outlined in Part 2 of the series':

$$\text{HO.120} - 0.0131 \div 0.1201 = 0.89$$

The lowest and highest estimates from the three studies were then used to establish the plausible range for effectiveness. Similar techniques were used to find baseline estimates and ranges for the sensitivity and specificity of temporal artery biopsy, and for the probability of a major complication of prednisone use (see table l)[10]

In some circumstances, you may create a decision tree and discover that there are only one or two very small, poor-quality published studies, or no published studies, on which to base your probability estimates. In such situations, you will need to use alternative sources of information such as expert judgment, existing databases, and primary data collection? We recommend that you begin with expert judgment and/or existing databases to make initial probability estimates. Since these estimates are subject to bias,[11-13] a wide range of possible values should be considered in a sensitivity analysis. If the results of your decision model prove to be sensitive to a probability value derived in this way, the answer to the decision problem will remain uncertain until further information is derived from primary data collection.

Table 1 • Example of Probability Table

Variable	Probability*	
	Baseline	Range
Probability of major complication of giant cell arteritis	0.12	0.05-0.40
Temporal artery biopsy		
Sensitivity	0.80	0.58-0.97
Specificity	1 .oo	0.90-l .oo
Effectiveness of prednisone	0.89	0.69-l .00
Probability of major complication of prednisone use	0.19	0.05-0.40

*Baseline probabilities are the averages of estimates from published studies: ranges are based on the highest and lowest estimates from published studies. The specific references for the probability estimates can be found in Buchbinder and Detsky.[10]

Estimating Outcome Values

The final component involved in constructing a decision model is to assign a quantitative value to the outcome at the end of each branch of the tree. Outcome values can be expressed in several ways: life years, quality-adjusted life years (QALYs), cases of disease or complications prevented, or utilities, The simplest type of decision model is that that has only two possible outcomes (e.g., alive or dead, disease or no disease, complication or no complication). In such circumstances, a common convention is to assign the value 1 to the better outcome and the value 0 to the worse outcome.[3] When this convention is applied, the outcome value for each treatment option will represent the overall probability that the better outcome will occur if this treatment option is chosen. Assigning outcome values is usually more complicated because most decision problems have more than two possible outcomes.

A "utility" is a measure of a decision maker's relative preference for an outcome, and is expressed as a single value between 0 and l.[14] Utilities for outcomes are usually assessed relative to two extremes, referred to as "anchor states." The commonly used anchor states are "death," assigned a value of 0, and "full health," assigned a value of 1. Utility measures provide summary scores that aggregate the positive

Table 2 • Outcomes for Decision Options in the Management of Giant Cell Arteritis (GCA)

Treat all
No GCA, prednisone treatment
No GCA, prednisone treatment, major prednisone complication
GCA, prednisone treatment
GCA, prednisone treatment, major prednisone complication
GCA, prednisone treatment, major GCA complication
GCA, prednisone treatment, major prednisone complication, major GCA complication
Treat none
No GCA
GCA
GCA, major GCA complication
Biopsy and treat positives
Biopsy positive
No GCA, prednisone treatment, TA* biopsy
No GCA, prednisone treatment, major prednisone complication, TA biopsy
GCA, prednisone treatment, TA biopsy
GCA, prednisone treatment, major prednisone complication, TA biopsy
GCA, prednisone treatment, major GCA complication, TA biopsy
GCA, prednisone treatment, major prednisone complication, major GCA complication, TA biopsy
Biopsy negative
No GCA, TA biopsy
GCA, TA biopsy
GCA, major GCA complication, TA biopsy

***TA** = temporal artery.

and negative aspects of quality of life, and can incorporate attitudes towards risk and length of life.'*

Utilities can be used as the actual outcome values in your decision tree, or they can be used as weights to calculate quality-adjusted life expectancy. A simple, and widely accepted, approach to estimating quality-adjusted life expectancy is to multiply the length of life in a health state by the utility of the health state.[14] For example, if an individual lives 10 years in full health (utility = 1.01 and 10 years with a severe disabling stroke (utility = 0.51, the quality-adjusted survival would be:

$$[(10 \times 1.01 + (10 \times 0.5)] = 15 \text{ QALYs}$$

Utilities can be estimated in many ways: 1) arbitrarily assign values based on your judgment; 2) have a group of experts reach a consensus on the estimates for the utility values; 3) search for relevant, published utility values in the literature; or 4) measure the values directly in appropriate subjects, using reliable and valid methods.** Because of the significant amount of work involved in collecting utility measurements from a group of subjects, we generally recommend beginning with utility estimates from the literature, or from the judgment of experts. Given the inaccuracies associated with these methods, a wide range of possible values should be considered for each utility estimate, allowing for extensive sensitivity analyses.[14] You can then consider directly measuring utilities for those health states that have major impacts on the results of the analysis.

There are several publications that describe the utilities of a wide range of health states,[14-16] and if you are very fortunate, the utility values required for your decision tree may already have been measured. The principles described earlier about using published studies to estimate probability values apply equally to using published studies to estimate utility values. You should search the literature in a systematic fashion, you should assess the validity of the published utility estimates by applying critical appraisal criteria to judge the study quality, and you should ensure that the published utilities are applicable to your decision model.[5,6,17]

For the GCA decision tree, there are several possible outcomes for each decision option (see table 2). The outcomes include various combinations of the following health states: symptoms of GCA, permanent blindness as a major complication of GCA, the negative impact on quality of life associated with taking daily prednisone tablets, a major complication from prednisone treatment, and the negative impact on quality of life associated with undergoing a temporal artery biopsy. Unfortunately, our search of the literature yielded no relevant data on which we could base utility estimates, so we had to derive our own utility estimates.

When outcomes consist of combinations of different health states, the utility of an outcome can be assessed as a whole, or in parts. For example, assessing the utility of the entire combination of having GCA, undergoing a temporal artery biopsy, being on prednisone treatment, and experiencing a major GCA complication and a major complication of prednisone therapy represents a utility assessment of the whole outcome. Alternatively, the utility of this outcome could be assessed by individually assessing the utility of undergoing a temporal artery biopsy, the utility of being on prednisone therapy, the utility of a major GCA complication, and the utility of a major prednisone complication, and then combining these utilities in some way. These two approaches are known as the "holistic" method and the "decomposed" method, respectively." In general, we suggest that if the outcomes of the decision tree are simple and easily ranked from most to least preferred, the holistic approach should be used. If the outcomes consist of combinations of several different health states, as in the GCA tree, or if they are

difficult to rank with respect to utility values, the decomposed method should be used.

For the decomposed approach, we recommend dividing the health states into short-term and long-term states. Short-term states are those that have impacts on quality of life for defined, short periods of time (e.g., days to weeks). Examples include temporary hospitalizations and unpleasant diagnostic procedures. Long-term **states** are those that have enduring impacts on quality of life, such as chronic symptoms from a disease, the negative impact on quality of life related to persistently being on a medication, and major complications from disease or from treatments that have lasting sequelae. Once the health states have been separated into short-term and long-term states, assign them utility values relative to the anchors of full health (utility = 1) and death (utility = 0). As with probability estimates, a useful way to display utility data is in a table that contains the baseline estimates, the range of plausible values, and the reference sources (when published articles are used).

In the GCA example, there are five long-term states, which are assumed to persist for the entire time horizon of the analysis: no GCA, GCA symptoms, major GCA complication, prednisone treatment, and major prednisone complication (see table 3). The GCA example has only one short-term state: temporal artery biopsy. We used the consensus of a group of physicians to estimate the utility values. Because prednisone therapy essentially eliminates all the symptoms of GCA, the utility of having GCA symptoms on prednisone therapy was assumed to be equal to the utility of having no GCA symptoms (i.e., utility = 1). However, patients on prednisone therapy are considered to have a negative impact on their quality of life associated with the prednisone treatment itself (i.e., the utility of prednisone treatment).

The next step in the decomposed strategy involves aggregating the separate utilities. There are several ways in which the utilities for decomposed states can be combined to yield an overall utility value for the entire outcome state, including adding the utility values of the different states, multiplying the utility values of the different states, or adding the utility values of some states and multiplying others. Using any of these aggregation methods entails certain assumptions about the independence and interactions of the different dimensions being combined." Ultimately, the only way to establish the accuracy of your combined utility values is to empirically verify your methods, which is a task that is generally beyond the capabilities of the neophyte analyst.

The aggregation scheme that we recommend requires that you convert the utility values of your short-term states into "disutility" values. The "disutility" value of a health state represents the negative impact on quality of life associated with the state. The equation for calculating the disutility value of a health state is very simple:

Disutility value = 1.0 – utility value

Next, you should multiply the utility values of all the long-term states together. Finally, subtract the disutility values fur the short-term states from the product of the utilities of the long-term states. This aggregation scheme will yield a utility value for each outcome state depicted in your decision tree.

For example, consider the "biopsy and treat positives" strategy. The utility for the outcome state "biopsy-proven GCA, on prednisone treatment, with a major prednisone complication and with a major GCA complication" is represented in the decision tree terminal node by the following formula:

[utility of GCA symptoms on prednisone therapy X utility of taking prednisone therapy daily X utility of a major prednisone complication X utility of a major GCA complication1 – [l.0 – utility of undergoing a temporal artery biopsy1 = [1.0 × 0.97 × 0.75 × 0.60] – [1.0 – 0.9953 = 0.432.

Table 4 displays the utility estimates for all the outcome states for the GCA example using the baseline utility values for the decomposed health states, which are show-n in table 3.

Once you have derived your utility estimates for all the outcome states, you should assess the rank order of the utility values to see if the ranking of outcome states meets the minimal requirement of making sense (see table 4). This task is often referred to as an assessment of "face validity," and simply means that you check to make sure that outcomes that are clearly worse than others don't have higher utility estimates. If the utility estimates for your outcome states fail to meet this relatively crude measure of validity, either you have made a mistake in estimating the utilities of the decomposed states or

Table 3 • Utility Estimates for Giant Cell Arteritis (GCA) Decomposed Health States

Health State	Baseline Utility*	Range*
No GCA	1.00	
GCA symptoms	0.85	0.70–0.95
Major GCA complication	0.60	0.20–0.85
Prednisone treatment	0.97	0.90-i .00
Major prednisone complication	0.75	0.60-0.90
temporal artery biopsy	0.995	0.97–1.00

*Baseline utilities and ranges are based on consensus estimates of a group of expert physicians.

Table 4 • Rank Ordering of Giant Cell Arteritis (GCA) Outcome Values

Outcome State	Utility*	QALYs†
No GCA, no prednisone, no TA‡ biopsy	1.000	13.600
No GCA, no prednisone, TA biopsy	0.995	13.595
No GCA, prednisone	0.970	13.192
GCA, prednisone§	0.970	13.192
No GCA, prednisone, TA biopsy	0.965	13.187
GCA, prednisone, TA biopsy	0.965	13.187
GCA, no prednisone	0.650	11.560
GCA, no prednisone, TA biopsy	0.845	11.556
No GCA, prednisone, major prednisone complication	0.726	9.694
GCA, prednisone, major prednisone complication	0.726	9.894
No GCA, prednisone, major prednisone complication, TA biopsy	0.723	9.890
GCA, prednisone, major prednisone complication, TA biopsy	0.723	9.890
GCA, prednisone, major GCA complication	0.562	7.915
GCA, prednisone, major GCA complication, TA biopsy	0.577	7.912
GCA, no predinsone, major GCA complication	0.510	6.936
GCA, no prednisone, major GCA complication, TA biopsy	0.505	6.933
GCA, prednisone, major prednisone and GCA complications	0.437	5.936
GCA, prednisone, major prednisone and GCA complications, TA biopsy	0.432	5.934

*Utilities are calculated by multiplying the baseline utilities of the long-term states and then, when applicable, subtracting the disutility (i.e., 1 − utility) value for temporal artery biopsy.

†QALYs = quality-adjusted life years are calculated by, when applicable, subtracting the time period of negative impact of temporal artery biopsy from the life expectancy and then multiplying the difference by the product of the baseline utilities of the long-term states.

‡TA = temporal artery.

§The utility of No GCA, prednisone equals that of GCA, prednisone, since we assume that prednisone completely eliminates GCA symptoms.

this method of aggregating utilities is not appropriate for the given decision tree. Even if face validity is achieved, caution is required, since the aggregation method is arbitrary and may misrepresent the complexity of interactions between health states.

As an alternative approach, you could express the outcome values for the GCA example in terms of quality-adjusted life expectancy (QALE). To simplify this example, we assume that GCA and its treatment have no direct effect on life expectancy (LE), so that LE estimates can be derived directly from published life tables for the general population. Other sources describe how to adjust LE for the presence of one or more diseases that have impacts on LE.[4,19,20]

For the purpose of estimating QALE, we recommend that you represent the negative impacts on quality of life of short-term states by assigning values in units of time roughly equivalent to the periods of time that the states have negative impacts on the individual. For example, a consensus group of experts estimated that temporal artery biopsy has a negative impact on patients for two days, or 0.005 years. The time periods of negative impacts on quality of life associated with short-term states are then subtracted from the LE. The implicit, conservative assumption associated with this method is that the quality of life is zero during the period of time experienced in the short-term state.

Once the short-term states have been dealt with, you should aggregate the utilities of the long-term states by multiplying them together. The product of the utilities of the long-term states should then be multiplied by the difference of the LE and the time periods of negative impacts on quality of life associated with short-term states. This will give you the overall QALE for each outcome state.

For example, the QALE for the outcome state "biopsy-proven GCA, on prednisone treatment, with a major prednisone complication and with a major GCA complication" in a cohort of 70-year-olds, with a LE of 13.6 years,[21] is represented by the following formula:

> [utility of GCA symptoms on prednisone therapy X utility of taking prednisone therapy daily X utility of a major prednisone complication X utility of a major GCA complication] X [LE − time period of negative impact from temporal artery biopsy1 = [1.0 X 0.97 X 0.75 X 0.601 X [13.6 years − 0.005 years] = 5.934 QALYs.

The most ambitious approach to estimating utility values for your decision tree is direct measurement, and this approach is often reserved for utility variables that have major impacts on the results of the analysis. Measuring utility values involves the following steps: developing health-state descriptions, choosing the subjects, and choosing the method of measurement.[14,17] A detailed explanation of how to develop health-state descriptions and measure utilities is beyond the scope of this primer; we refer you to several reviews for more information.[14,17,18,22-24]

References

1. Krahn MD, Naglie G, Naimark D, Redelmeier DA, Detsky AS. Primer on medical decision analysis: Part 4-analyzing the model and interpreting the results. Med Decis Making. 1997; 17:142-51.
2. Petitti DB. Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine. New York: Oxford University Press, 1994.
3. Weinstein MC, Fineberg I-IV. Clinical Decision Analysis. Philadelphia, PA: W. B. Saunders, 1980.
4. Sox H, Blatt MA, Higgins MC, Marton KI. Medical Decision Making. London, England: Butterworth & Co., 1988.
5. Richardson WS, Detsky AS, and the Evidence-Based Medicine Working Group. Users' guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of

the study valid? JAMA. 1995;273:1292-5.
6. Richardson WS, Detsky AS, and the Evidence-Based Medicine Working Group. Users' guides to the medical literature. VII. How to use a clinical decision analysis. B. Results and applicability. JAMA. 1995;273:1610-6.
7. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology: A Basic Science for Clinical Medicine. 2nd ed. Boston, MA: Little, Brown and Company, 1991.
8. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Ann Intern Med. 1987;107:224-33.
9. Detsky AS, Naglie G, Krahn MD, Redelmeier DA, Naimark D. Primer on medical decision analysis: Part 2-building a tree. Med Decis Making. 1997;17:126-35.
10. Buchbinder R, Detsky AS. Management of suspected giant cell arteritis: a decision analysis. J Rheumatol. 1992;19: 1220-8.
11. Abrams HB, Detsky AS, Roos LL, Wajda A. Is there a role for surgery in the acute management of infective endocarditis? A decision analysis and medical claims database approach. Med Decis Making. 1988;8:165-74.
12. Poses RM, Anthony M. Availability, wishful thinking, and physicians' diagnostic judgements for patients with suspected bacteremia. Med Decis Making. 1991;11:159-68.
13. Bobbio M, Detrano R, Shandling AH, et al. Clinical assessment of the probability of coronary artery disease judgmental bias from personal knowledge. Med Decis Making. 1992; 12:197-203.
14. Torrance GW. Measurement of health state utilities for economic appraisal: a review. J Health Econ. 1986;5:1-30.
15. Fryback DG, Dasbach EJ,. Klein R, et al. The Beaver Dam health outcomes study: initial catalog of health-state quality factors. Med Decis Making. 1993;13:89-102.
16. Kaplan RM, Anderson JP. A general health policy model: update and applications. Health Serv Res. 1988;23:203-35.
17. Redelmeier DA, Detsky AS. A clinician's guide to utility measurement. Primary Care. 1995;22:271-8.
18. Froberg DG, Kane RL. Methodology for measuring health-state preferences-I. measurement strategies. J Clin Epidemiol. 1989;42:345-54.
19. Beck JR, Kassirer JP, Pauker SG. A convenient approximation of life expectancy (the "DEALE"). I. Validation of the method. Am J Med. 1982;73:883-8.
20. Beck JR, Pauker SG, Gottlieb JE, Klein K, Kassirer JP. A convenient approximation of life expectancy (the "DEALE"). II. Use in medical decision-making. Am J Med. 1982;73:889-97.
21. National Center for Health Statistics. Vital Statistics of the United States, 1988 Vol II, Mortality, Part A, Section 6. Washington, DC: Public Health Service, 1991.
22. Froberg DG, Kane RL. Methodology for measuring health-state preferences-II: scaling methods. J Clin Epidemiol. 1989;42:459-71.
23. Froberg DG, Kane RL. Methodology for measuring health-state preferences-III: populatioin and context issues. J Clin Epidemiol. 1989;42:585-92.
24. Froberg DG, Kane RL. Methodology for measuring health-state preferences-IV: progress and a research agenda. J Clin Epidemiol. 1989;42:675-85.

Glossary

Baseline analysis: An analysis that uses the best estimate for each variable in the model.

Holistic method: A method to derive the utility of the outcome of a branch in the decision tree. The utility of the outcome is assessed as a whole, even if the outcome consists of a combination of different health states.

Decomposed method: A method to derive the utility of the outcome of a branch in the decision tree, when the outcome consists of a combination of different health states. The utility of each health state is assessed independently, and then these utilities are combined into a single value.

Disutility: The disutility of a health state represents the negative impact on quality of life associated with the state. The disutility value is calculated by the equation "1.0 − utility value."