

Probabilistic Sensitivity Analysis Using Monte Carlo Simulation A Practical Approach

*Peter Doubilet, M.D., Ph.D.,
Colin B. Begg, Ph.D.,
Milton C. Weinstein, Ph.D.,
Peter Braun, M.D.,
and Barbara J. McNeil, M.D., Ph.D.*

The data for medical decision analyses are often unreliable. Traditional sensitivity analysis — varying one or more probability or utility estimates from baseline values to see if the optimal strategy changes — is cumbersome if more than two values are allowed to vary concurrently. This paper describes a practical method for probabilistic sensitivity analysis, in which uncertainties in all values are considered simultaneously. The uncertainty in each probability and utility is assumed to possess a probability distribution. For ease of application we have used a parametric model that permits each distribution to be specified by two values: the baseline estimate and a bound (upper or lower) of the 95 percent confidence interval. Following multiple simulations of the decision tree in which each probability and utility is randomly assigned a value within its distribution, the following results are recorded: (a) the mean and standard deviation of the expected utility of each strategy; (b) the frequency with which each strategy is optimal; (c) the frequency with which each strategy “buys” or “costs” a specified amount of utility relative to the remaining strategies. As illustrated by an application to a previously published decision analysis, this technique is easy to use and can be a valuable addition to the armamentarium of the decision analyst. (Med Decis Making 5:157-177, 1985)

This work was supported in part by grants from the National Library of Medicine (1-K04-LM00082), the National Cancer Institute (CA-31247), the National Center for Health Services Research (HS-03314), and the Hartford Foundation.

From the Department of Radiology, Harvard Medical School and Brigham and Women's Hospital; the Division of Biostatistics, Dana-Farber Cancer Institute; and the Institute for Health Research, Harvard School of Public Health. Please address requests for reprints to Dr. Doubilet, Department of Radiology, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA.

The data for medical decision analyses are often uncertain or unreliable. Two of the main routes to obtaining such information – studies reported in the medical literature and experts' subjective estimates – are frequently imprecise. Medical studies may be poorly designed and are often based on small numbers of patients. Furthermore, the patients in the study may differ from the individual or group to whom the decision analysis is being applied. Subjective estimates may be afflicted by a variety of biases [1].

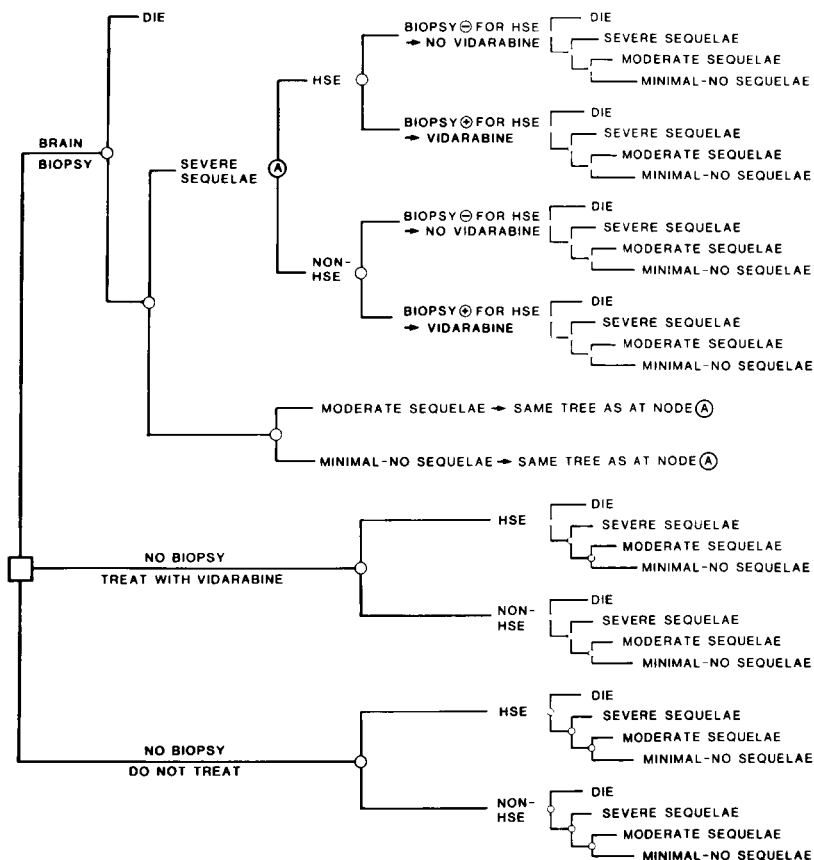


Figure 1. Decision tree for suspected herpes simplex encephalitis (HSE). The decision options are: Do a brain biopsy, followed by therapy with vidarabine only if the biopsy is positive for HSE (*brain biopsy*); Treat with vidarabine without biopsying (*no biopsy; treat with vidarabine*); Do not biopsy and do not treat with vidarabine (*no biopsy; do not treat*). Possible sequelae of brain biopsy and of disease (treated or untreated) are considered separately in the *brain biopsy* subtree. (Adapted from [2])

Consider, for example, the decision tree in Figure 1 (modified from [2]), which examines the decision to biopsy, treat, or not treat a patient with suspected herpes simplex encephalitis (HSE). Sequelae of both brain biopsy and of the disease (treated or untreated) are considered, and several outcomes ranging from death to minimal-to-no sequelae are included. There are few reported studies on which to base estimates of the needed probabilities, and those that are available involve small numbers of patients [3,4]. When experts were polled for subjective probability estimates there was wide variation in the values provided; an extreme example was the range of 0.1 to 0.5 for the probability of death in untreated non-HSE patients [2].

The uncertainties of the probability estimates in this case are further highlighted by the marked discrepancies between a number of the values used by Braun [2] and those used in another analysis [5] of the same clinical problem. Assignment of utility values to outcomes is at least as difficult as probability estimation. There are no objective guidelines for evaluating moderate or severe sequelae relative to death and minimal-to-no sequelae, and subjective estimators might not even agree on the rank order of death and severe sequelae.

There are at least two reasons why the decision analyst must deal with uncertainty in available data. First, physicians are unlikely to be influenced by a decision analysis whose result is contrary to their clinical judgment unless the analyst provides an argument that addresses the uncertainty in the probability and utility estimates in the tree. Second, unless the effect of uncertainty is examined there is no guide as to whether it would be worthwhile to seek better data (e.g., by carrying out clinical trials) for future decisions in similar patients.

The usual tactic for dealing with uncertain data is to carry out a sensitivity analysis by varying one or more of the probability or utility estimates from baseline values and observing the effect on the choice of strategy [6]. If a single strategy has the highest expected utility when estimates are varied within a reasonable range, then that strategy can be recommended with confidence (provided that the decision tree itself is an adequate model of the clinical problem). On the other hand, if the optimal strategy is sensitive to variation of baseline estimates, then it is appropriate to treat the results skeptically and to consider areas in which further data collection may be valuable.

Conventional sensitivity analysis, however, has a number of limitations. It is cumbersome when more than two or three quantities are allowed to vary simultaneously, and the results of a multiple-way analysis cannot readily be presented. Even a three-way sensitivity analysis using families of curves on a graph [7] may be difficult for a mathematically unsophisticated reader to follow. This restriction to no more than three-way analysis is a major drawback for trees with many uncertain estimates. If the tree has, for example, 50 probabilities and utilities, then the results of a three-way analy-

sis are conditional upon the validity of the estimates selected for the remaining 47 quantities. A further limitation of conventional sensitivity analysis is that it does not permit the decision analyst to make a summary statement, often requested by the physician, about the certainty that the strategy selected by the analysis is in fact optimal.

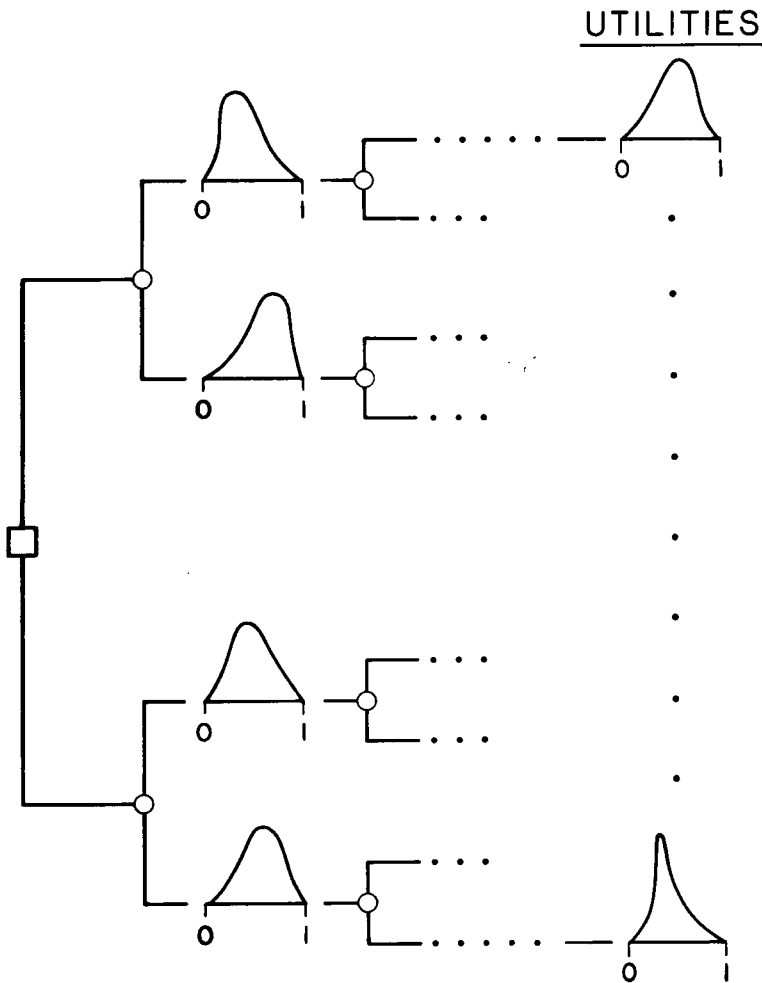


Figure 2. Basis for probabilistic sensitivity analysis. All probabilities and utilities are variable quantities and have associated distribution functions (the density function is illustrated here). The range of possible values is shown as 0-1, but may differ in other examples.

An alternative, probabilistic approach to sensitivity analysis, without these limitations, has been described by Pass and Goldstein [8]. We will summarize their technique, which is based on Monte Carlo simulation, and then present a method by which probabilistic sensitivity analysis can be a practical, easily applied tool for the decision analyst.

Methods

PROBABILISTIC SENSITIVITY ANALYSIS

This section summarizes the approach of Pass and Goldstein. Each probability and utility in the decision tree, instead of taking on a single value, is assumed to be a variable quantity with a range of possible values and to have an associated distribution function (Figure 2). The expected utility of each decision option, computed by "folding back" the decision tree, is a sum of products of these probabilities and utilities, and hence is itself a variable quantity whose distribution function depends on those of the individual probabilities and utilities.

Precise calculation of the distributions of expected utilities, and of the confidence with which an optimal strategy can be selected, is impossible unless the decision tree is very simple (and usually unrealistic). Instead, a Monte Carlo approach can be used. Each probability and utility is randomly assigned a value from its distribution, and the expected utility of each option is computed. This process is repeated a large number of times, and the following results are recorded:

- (a) The mean and standard deviation, over all runs, of the expected utility values for each strategy.
- (b) The frequency with which each strategy is optimal. This provides a measure of the confidence with which an optimal strategy can be selected on the basis of the decision tree and data at hand. If no single strategy is optimal in a large fraction of the runs, this indicates that too much uncertainty exists to rely confidently on the results of the analysis. Conversely, if one strategy has the highest expected utility in a large majority (e.g., 95%) of the runs, then — barring systematic errors in the model or the data — that strategy can confidently be selected as the optimal one.
- (c) The frequency with which each strategy "buys" or "costs" a specified amount of utility. A strategy "buys" a given amount of utility if its expected utility is at least this amount greater than that of all other strategies. It "costs" a given amount of utility if its expected utility is at least this amount less than that of any other strategy. These frequencies can be of value when the magnitude of the difference in expected utility is important (e.g., if there is an amount below which differences are felt to be clinically unimportant).

A PRACTICAL APPROACH TO PROBABILISTIC SENSITIVITY ANALYSIS

Specifying Distribution Functions. Specifying distribution functions about all probabilities and utilities can be a time-consuming task. We therefore sought to make probabilistic sensitivity analysis a practical tool by devising a method to specify distributions easily and quickly. In addition to ease of application, our method had to satisfy a second condition: If the user has already carried out a decision analysis with a set of initial, or baseline, values for each probability and utility, and then wants to perform a sensitivity analysis centered on these values, the mean of each distribution function must equal the baseline value. This condition assures that the mean value of the expected utility of each strategy, over all runs of a Monte Carlo simulation, will converge to the baseline expected utility [9].

We achieved these goals — simplicity and internal consistency — by assuming that each distribution can be approximated by a parametric distribution. While there is no type of distribution that can be considered correct a priori, we have found the logistic-normal distribution (Figure 3) to be a convenient and mathematically tractable model for this purpose. That is, we assume that the logit transform — $\log(X/1-X)$ — of each probability and utility is normally distributed. This transformation has been suggested before for modeling probabilities [10], and its statistical properties have been studied [11]. (If the range of possible values is other than 0–1, say A – B , then $\log(X/1-X)$ is replaced by $\log((X-A)/(B-X))$. Unless otherwise stated, we assume a range of 0–1; all equations, including those in Appendix A, can easily be modified to apply to the A – B range.)

The benefits of assuming that each quantity X (probability or utility) has a logistic-normal distribution are twofold. First, the normal distribution associated with $\text{logit}(X)$ is fully determined by two values: its mean μ and standard deviation s . Second, random selection of a value r (between 0 and 1) from the distribution of X is easily accomplished by first randomly choosing a value z (between $-\infty$ and $+\infty$) from the associated normal distribution [12] and then taking r to be the value whose logit is z . That is,

$$r = \frac{e^z}{1 + e^z}.$$

This would complete the description of our method if the mean μ and standard deviation s of the logit transform of a probability or utility were directly obtainable. In fact, they are not. We show instead how μ and s can be derived from two values of the probability or utility X : its mean, or baseline value, and the upper or lower bound of its 95 percent confidence range.¹ Estimates of the mean and bound can be based on subjective estimates or on objective data. In the former case, an expert source specifies the two values instead of a baseline value alone, as in a conventional decision analysis. As an example of the latter, suppose a probability of 0.8 is estimated on the basis of an event occurring in 80 out of 100 experimental

subjects. Using standard statistical techniques for a normal approximation to the binomial distribution [13], the 95 percent confidence range is approximately

$$0.8 \pm 1.96 \sqrt{\frac{(0.2)(0.8)}{100}} \text{ or } 0.80 \pm 0.08.$$

Therefore, our baseline value is 0.80, with a lower bound of 0.72 and an upper bound of 0.88.

The derivation of the mean μ and standard deviation s of $\text{logit}(X)$ from the mean m and a bound b of X would appear, at first sight, to be a straightforward undertaking. In particular, one might assume that μ is equal to

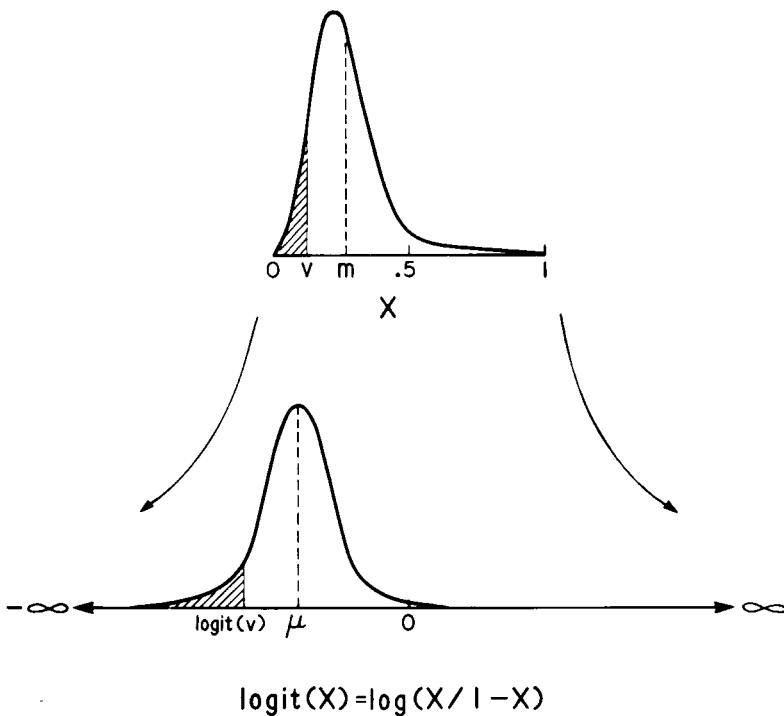


Figure 3. A logistic-normal random variable and its associated normal distribution. A random variable X , whose density function is shown in the upper half of the figure, is logistic-normal if there is a normal distribution (bottom half of the figure) such that, for any value v between 0 and 1, the areas under the curves below v and $\text{logit}(v)$ (shaded regions) are equal. m is the mean value of X , μ of the normal distribution representing $\text{logit}(X)$. Because X has a skewed distribution, μ is not equal to $\text{logit}(m)$.

$\text{logit}(m)$. Unfortunately, this is not the case, because of skewness in the distribution in the logit scale (Figure 3). Instead, μ and s are computed from m and b using the following formulas (derived in Appendix A):

1. If $m \neq 0.5, 0.025$, and 0.975 , then

$$\mu = \frac{B - E \cdot \sqrt{B^2 - M^2 + M^2 E^2}}{1 - E^2},$$

where $M = \text{logit}(m)$, $B = \text{logit}(b)$, $E = 1.96 / \Phi^{-1}(m)$ (Φ^{-1} represents the inverse² of the normal distribution function Φ).

2. If $m = 0.5$, then $\mu = 0$.

3. If $m = 0.025$ or 0.975 , then $\mu = (M^2 + B^2) / 2B$.

Once μ has been calculated from these formulas, s is computed by

$$s = |\mu - B| / 1.96.$$

(Note that if the quantity inside the radical in the initial formula for m is negative, then m and b cannot be the mean and bound of a logistic-normal distribution. This can occur only if $m > 0.975$ and b is substantially less than 1 (e.g., $m = 0.98$, $b = 0.70$), or if $m < 0.025$ and b is well above 0. Furthermore, as shown in Appendix A, the formulas do not apply [and hence m and b cannot be fitted to a logistic-normal distribution] whenever $m \geq 0.975$ and $b < 0.5$, or $m \leq 0.025$ and $b > 0.5$. When m is greater than 0.025 and less than 0.975, there is always an appropriate logistic-normal distribution.)

If b was specified to be the 95 percent lower bound, then these formulas assure that 2.5 percent of the logistic-normal distribution falls below b . The value above which 2.5 percent of the distribution lies was not explicitly chosen, and so is referred to as the *implicit upper bound*. As shown in Appendix A,

$$\text{Implicit upper bound} = \frac{e^{(\mu + 1.96s)}}{1 + e^{(\mu + 1.96s)}}.$$

If b had been chosen to be the 95 percent upper bound, then the implicit lower bound would be given by an equation similar to this, but with $\mu + 1.96s$ replaced by $\mu - 1.96s$. The term *implicit other bound* will be used to refer to the implicit upper or lower bound.

We illustrate these formulas using the earlier example of a probability estimated to be 0.80 ± 0.08 . Using the mean value $m = 0.80$ and the lower bound $b = 0.72$ in the above formulas, we obtain $\mu = 1.40$ and $s = 0.23$. That is, the uncertainty in this probability is represented by a logistic-normal distribution whose associated normal distribution has mean 1.40 and standard deviation 0.23. The implicit upper bound is 0.86. (If we had chosen to use the upper bound of 0.88, we would have obtained the slightly different parameters $\mu = 1.41$ and $s = 0.30$. The implicit lower bound would have been 0.69.)

Chance Nodes with More than Two Branches. The foregoing discussion of distributions about probabilities applies to two-way chance nodes. The distribution about either of the two probabilities at such a node is derived as described above; the second probability, being one minus the first, is fully determined by the first. It is immaterial which probability is considered first, because if X is a logistic-normal random variable with values between 0 and 1, then $1 - X$ is also logistic-normal (since $\text{logit}(1 - X) = -\text{logit}(X)$).

For a decision tree that contains multiple-way chance nodes, one of two approaches can be taken. The tree can be replaced by an equivalent one that has only two-way nodes. A three-way node can be replaced by a pair of two-way nodes using the method illustrated in Figure 4; this can be extended to nodes with more than three branches. In some cases, however, this process yields a tree that is unwieldy, or one for which it is difficult to estimate mean and boundary values for the resulting chance nodes. If so, a modification of the technique described for two-way nodes can be used. This modification is presented in Appendix B.

Derived Probabilities and Utilities. Some of the probabilities or utilities in a decision tree may be derived from, or best viewed as arising from, a set of more fundamental values. For example, since death in a patient with nonherpes encephalitis treated (unnecessarily) with vidarabine (see Figure 1) can occur as a consequence either of the disease or of drug toxicity, the probability of death in such a patient can be viewed as a derived value that

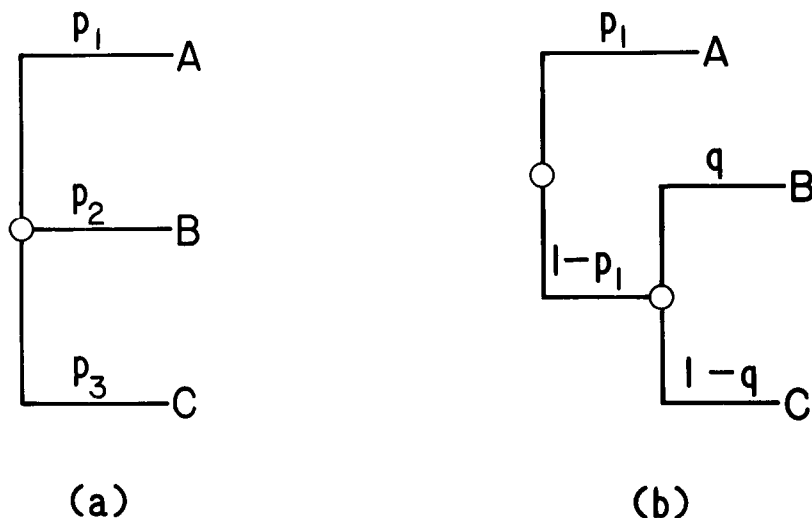


Figure 4. Three-branch chance node (a) and an equivalent sequence of two-branch nodes (b). q , the probability of B occurring given that A does not occur, is equal to $p_2 / (1 - p_1)$.

depends on two more basic ones. In particular, if p_1 is the probability of death from non-HSE and p_2 that of death from vidarabine, and if interactions between the two are insignificant, then the probability of death in such a patient is

$$p_1 + p_2 - p_1 \cdot p_2.$$

Another example (not in this decision tree) of a derived probability is one computed from prior and conditional probabilities using Bayes' theorem [6].

In such cases, since the distributions of two or more derived probabilities or utilities that depend upon a common set of fundamental values are likely to be highly correlated, it is preferable to base the Monte Carlo simulation on the fundamental values. That is, one should specify the distributions about the fundamental values, and carry out the Monte Carlo simulation by random selection from these distributions. In the HSE tree, in each run of the simulation p_1 and p_2 are assigned values randomly from their distributions, and the probability of death in non-HSE patients receiving vidarabine is computed from these by means of the above formula.

Application of the Technique: An Example

As noted earlier, the decision tree in Figure 1 is designed to examine the choice among three management options for a patient suspected of having Herpes simplex encephalitis (HSE). In this section, we demonstrate the application of our technique to this tree. It should be noted that this is meant as an illustrative example only, for two reasons. First, the example is modified from an analysis in which experts were polled for their best estimates of probabilities, but were not asked to provide upper or lower bounds. Second, the tree structure omits one potentially important reason for doing a brain biopsy — to find another treatable cause for the patient's symptoms, such as an occult abscess not identified by the patient's prior workup.

Table 1 provides mean (baseline) and boundary values of the needed probabilities and utilities, based in part on the experts' estimates in the original analysis [2]. Note that one bound is selected for each parameter, and that it need not be of the same type (upper or lower) for all.

The baseline expected utilities of the three options, computed from the baseline values of probabilities and utilities, are listed in the top row of Table 2. The *no biopsy; treat* strategy has the highest expected utility, by a small margin over *brain biopsy*. That is, using our best set of estimates of probabilities and utilities, the optimal strategy is to treat the patient with vidarabine empirically, without prior biopsy.

An Apple II Plus computer was programmed to carry out the probabilistic sensitivity analysis. The program derives the normal distribution associated with each probability and utility (as per Appendix A), and carries out

Table 1. Data Used for HSE Probabilistic Sensitivity Analysis

	Baseline (mean)	Lower bound	Upper bound
I. Probabilities			
Brain biopsy sequelae			
Death	0.004		0.02
Severe ^a	.01		0.05
Moderate ^a	.03		0.06
BRAIN BIOPSY SENSITIVITY FOR HSE	.95	0.85	
BRAIN BIOPSY SPECIFICITY FOR HSE	.99	0.95	
PROBABILITY OF HSE (IN CLINICALLY SUSPICIOUS PATIENTS)			
Overall	.4	0.05	
Low suspicion	.1		0.25
High suspicion	.6		0.85
SEQUELAE OF HSE			
1. Untreated			
Death	.7		0.8
Severe ^a	.333		.5
Moderate ^a	.5		.8
2. Treated with vidarabine (fractional reduction compared with untreated) ^b			
Death	.37		.5
Severe	.2		.4
Moderate	.2		.4
SEQUELAE OF NON-HSE			
1. Untreated			
Death	.18	0.05	
Severe ^a	.122		.25
Moderate ^a	.139		.25
2. Treated with vidarabine (additional probability compared with untreated) ^c			
Death	.004		.02
Severe	.01		.03
Moderate	.02		.04
II. UTILITIES^d			
Death	0	0	0
Severe sequelae	0.02		0.1
Moderate sequelae	0.8		0.95
Minimal-no sequelae	1	1	1

^aProbabilities of severe or moderate sequelae, given that no worse outcome has occurred.

^bProbability of death with treated HSE (baseline value) = $0.7 - 0.37 \times 0.7 = 0.44$. Similar computations apply for probabilities of severe and moderate sequelae of treated HSE.

^cProbability of death with treated non-HSE (baseline value) = $0.18 + .004 - 0.18 \times .004 = 0.183$. Similar computations apply for probabilities of severe and moderate sequelae of treated non-HSE.

^dThe utility assigned to a patient who has sequelae of both brain biopsy and disease (treated or untreated) is the product of the two individual utilities.

Table 2. HSE Decision Analysis and Probabilistic Sensitivity Analysis

	Strategy		
	Brain biopsy	No biopsy; treat	No biopsy; do not treat
BASELINE EXPECTED UTILITY	0.558	0.566	0.494
PROBABILISTIC SENSITIVITY ANALYSIS (1000 SIMULATIONS)			
Mean expected utility	0.555	0.563	0.490
Standard deviation	0.099	0.096	0.136
Frequency maximum	18.4%	79.5%	2.1%
Frequency buys 0.004	7.2%	58.3%	0.5%
Frequency costs 0.004	4.1%	1.8%	89.5%

the Monte Carlo simulation by repetitively sampling from each distribution and computing the expected utility of each of the three decision options. The running time of the program is 2.5 minutes per 100 simulations.

The results of a probabilistic sensitivity analysis involving 1000 simulations of the decision tree are presented in the lower portion of Table 2. The mean values of the three expected utilities over all runs are almost exactly equal to the baseline expected utilities. (This, as noted earlier, is because the mean value of the distribution about each probability and utility equals the baseline value.) The *no biopsy; treat* strategy had the greatest expected utility in 795 of the 1000 runs, suggesting that we can be 79.5 percent confident that this is the best strategy. However, we can be only 58.3 percent confident that it is best by a margin of at least 0.004 utility units. This figure would be of interest if a difference of four deaths (or their equivalent in severe or moderate sequelae) per 1000 patients were felt to be the smallest clinically important difference. The *no biopsy; do not treat* strategy, on the other hand, is worst by at least this margin in 89.5 percent of runs.

It is notable that the *no biopsy; treat* strategy is optimal in as many as 795 or the 1000 runs, despite the fact that the standard deviations of the expected utilities of the strategies are approximately 0.1, tenfold greater than the difference in mean values of the two best strategies. If the expected utilities of the strategies had been independent of each other, then *no biopsy; treat* would have been superior to *brain biopsy* in considerably fewer of the runs. The observed value of 79.5 percent occurs because the distributions of expected utilities of the three strategies are highly interdependent, since they all depend on a common set of probability and utility estimates. As an example of such interdependence, as the probability of HSE increases or the utility value assigned to moderate sequelae decreases, the expected utilities of all three strategies decrease.

The numbers in Table 2 would vary somewhat from analysis to analysis.

Using binomial sampling we can predict that, in 95 of 100 Monte Carlo analyses (each involving 1000 runs), the *no biopsy; treat* strategy would be optimal in $0.795 \pm 1.96 \times \sqrt{795 \times 205 / 1000^3}$, or 77%–82% of the runs. A similar procedure can be used to estimate the ranges about the other values in Table 2.

If subsets of patients at low and high risk for HSE can be defined, then a single strategy might be selected with greater confidence than in the overall group. To examine this, we reran the entire analysis twice, the first assuming a subset at high risk (mean prevalence 0.6, bound 0.85) and the second at low risk (mean 0.1, bound 0.25). The *no biopsy; treat* strategy had the highest baseline expected utility in both of these groups, but the confidence in its optimality (as reflected by the frequency with which its expected utility is greatest) differed considerably in the two. In the high-risk group it was optimal 92.3 percent of the time, but in the low-risk group in only 51.1 percent of runs. This would suggest that in the latter group there is too much uncertainty in the data to rely confidently on the results of the analysis.

Discussion

We have developed a practical method for probabilistic sensitivity analysis in which uncertainties in all probabilities and utilities are considered simultaneously. This method distinguishes between uncertainties inherent to a decision problem, which are represented by chance nodes in the decision tree, and uncertainties surrounding the estimation of the probabilities and utilities. The latter set of uncertainties is usually ignored, or incompletely examined, in conventional decision analysis. Probabilistic sensitivity analysis accommodates them by calculating the probability distributions of the expected utilities, via Monte Carlo simulation. Probability statements about the reliability of the conclusions of the decision analysis can thereby be made.

The method we have presented builds upon the general approach to probabilistic sensitivity analysis described earlier by Pass and Goldstein [8]. By assuming that all distribution functions have parametric form probabilistic sensitivity analysis becomes a practical tool, in that distributions can be fully determined by specifying two values. The choice of the logistic-normal distribution enables us to approximate the full distribution in closed form (Appendix A) from the specified mean and either an upper or lower bound.

We must emphasize that the logistic-normal distribution is an option we have chosen for modeling probabilities and utilities. Others, such as the binomial, triangular, or beta distributions, might be considered, but each of these has important limitations with respect to use in probabilistic sensitivity analysis. Direct simulation from a binomial distribution is time-consuming (even using a microcomputer), and simulation from the normal approximation to the binomial distribution will occasionally yield values of proba-

bilities or utilities outside the permissible range (0–1). Triangular distributions (i.e., ones whose density functions are inverted Vs) are simple bounded distributions. However, they present difficulties when the mean value is very close to 0 or 1, in that the distribution must be very narrow to have an extreme mean value. For example, with a mean value of 0.004 (the value selected for the mortality rate of brain biopsy), a triangular distribution can have a maximum value of at most 0.012, and a 95 percent upper bound of at most 0.01 (the verification of this statement is straightforward, and is not presented). Thus the chosen values of $m = 0.004$ and $b = 0.02$ would not be compatible with a triangular distribution. The beta distribution has a bounded range (0–1) and is often used to model distributions about probabilities, but we know of no simple way to compute the two parameters of a beta distribution from its mean and 95 percent bound (or from any other pair of values that can be estimated subjectively).

There is, of course, a loss of generality in using the logistic-normal model to approximate all distribution functions. The possible values that a given probability or utility can assume may not conform to such a distribution. However, we believe that our model's ease of application outweighs its loss of generality. Furthermore, there is rarely sufficient information concerning the actual distribution of a probability or utility to indicate that it fits poorly to a logistic-normal pattern, or that it fits more closely to another distribution type such as the beta distribution.

Our technique is well suited to computer implementation. A computer program could be written specifically for a single decision tree (as we have done for the HSE tree), or it could be part of a more general program that would apply to any decision tree. For each probability and utility the computer would request as input the mean m and a bound b , as well as the range of permissible values. If the formulas presented earlier cannot be applied (e.g., the term inside the square root sign is negative), the user would be informed that the pair of values m and b is not compatible with a logistic-normal distribution and would be asked to provide a new pair. If the formula can be applied, then the mean μ and the standard deviation s of the associated normal distribution would be computed, and the implicit other bound of the probability or utility would be displayed to the user.

If the user is not satisfied with the implicit bound, he can input new values for m , for b , or for both. For example, after providing the values $m = 0.18$ and $b = 0.05$ for the mortality rate of untreated non-HSE, the user would be informed that these values imply a 95 percent upper bound of 0.42. If he felt this value to be too high, he might provide the new value $b = 0.30$ (while leaving $m = 0.18$), and be informed that the new implicit lower bound is 0.09. This process would continue until the user was satisfied with the values of m , b , and the implicit other bound. After the data entry for all probabilities and utilities is completed, the computer would carry out the simulation by repetitively sampling from each distribution and comput-

ing the expected utilities. The results of the simulation would be presented in a format similar to Table 2.

The technique, as we have described it, assumes that the distributions about all probabilities and utilities are independent. (This is unlike the case for expected utilities, which, as noted in the previous section, are interdependent.) This assumption will be violated in many decision trees. For example, in the HSE decision tree two probabilities that are clearly not independent are that of death in non-HSE patients receiving vidarabine and that of death in untreated non-HSE patients. Interdependence can also occur if several probabilities or utilities have been estimated from a single data base; any bias in the data base (e.g., patient selection bias) may tend to inflate or diminish all estimates.

A solution to the problem of interdependence would be to choose correlation coefficients that specify all interrelationships among probabilities and utilities, and to then draw the random values from a multivariate normal distribution (instead of a collection of independent normal distributions). This solution, even if theoretically valid, is impractical. Fortunately, a reasonable compromise between this ideal solution and the assumption of independence is available in many situations. In the HSE example, the compromise resolution is effected by an appropriate choice of "fundamental" probabilities (see Derived Probabilities and Utilities, above). When interdependence results from several quantities being estimated from a common data base, all can be considered to be completely correlated. Thus, in each run of the Monte Carlo analysis, all of these quantities can be simulated from a single, randomly selected quantile of their distribution functions.

One further aspect of our technique — the need to specify 95 percent confidence bounds — raises important questions. While bounds for objectively determined probabilities (those based on observed ratios) can be systematically approximated, how reliable are subjective estimates of bounds? What biases are operative in the subjective estimation of extreme values? Can the estimator truly differentiate a 95 percent bound from, say, a 90 percent or 99 percent bound? In fact, there is evidence that people do estimate extreme values poorly [14]. This does not, however, invalidate the use of our technique, just as known shortcomings in subjective probability estimation more generally [1] do not invalidate conventional decision analyses that employ subjective probabilities.

We have presented the tools needed to carry out probabilistic sensitivity analysis using one type of parametric distribution. We have not attempted to answer all potential questions about the technique. For example, given a decision tree and a set of estimates of means and bounds for all probabilities and utilities, how many simulation runs should be carried out? Are there families of parametric distributions other than the logistic-normal family that can be used conveniently for probabilistic sensitivity analysis? If so, are the results of an analysis, such as those presented in Table 2, sensitive to the

choice of parametric family? If this technique generates interest among decision analysts, questions such as these can lead to fruitful areas for future research.

In summary, we have described an easily applied method for sensitivity analysis that considers variability in all probabilities and utilities simultaneously, and that allows the analyst to make statements to the form: "Available data suggest that, with 80 percent certainty, *no biopsy; treat* is the best strategy," or that "with 58 percent certainty, it is better than the remaining strategies by at least an additional four lives per 1000 patients." In another setting, the smallest clinically important difference may be one year of life expectancy [15]. This ability is achieved at low cost to the analyst: One additional value (an upper or lower bound) must be specified for each probability and utility, and a computer must be available to carry out the heavy computational burden.

Appendix A: Derivation of the Parameters of the Normal Distribution of $\text{logit}(X)$ from the Mean and (Upper or Lower) Bound of X

The text of this paper contains all of the information required to carry out probabilistic sensitivity analysis using logistic-normal distributions. This Appendix, which derives the formulas presented in the text, and which involves calculus, can be skipped without compromising the reader's ability to apply our technique.

Our problem is as follows: given the mean and a bound (i.e., upper or lower bound of the 95 percent confidence range) of a probability or utility X , measured on a scale of 0–1, can we construct a normal distribution on the logistic scale that has a mean and bound on the 0–1 scale equal to the specified values? We here provide a solution to this problem. Let m be the mean and b the lower bound of X , and let μ and s be the mean and standard deviation of the distribution of $\text{logit}(X)$. (The values m and b are known; we seek μ and s). (We present the derivation for the case in which a lower bound is specified; that for the upper bound is similar.)

One equation relating these values is

$$s = (\mu - B)/1.96, \quad (1)$$

where $B = \text{logit}(b)$. This equation follows because B is the 95 percent lower bound of the normal distribution, since

$$P(\text{logit}(X) \leq \text{logit}(b)) = P(X \leq b),$$

from which it follows that B is 1.96 standard deviations below the mean μ .

If $m = 0.5$, then the distribution of X is symmetric, so that $\mu = 0$, and s is easily computed from equation (1) to be $-B/1.96$. For any other value of m it is not the case that $\mu = \text{logit}(m)$, and we proceed as follows.

Letting $F(x)$ represent the distribution function of X , its mean m is given by the formula

$$m = \int_0^1 x F'(x) dx. \quad (2)$$

Because X is logistic-normal, its distribution function is (by definition) given by $F(x) = \Phi_{\mu,s}(\text{logit}(x))$, where $\Phi_{\mu,s}$ is the normal distribution function with mean μ and standard deviation s . (This relationship is illustrated in Figure 3.) In view of this, and using the change of variables $z = \text{logit}(x)$ [or, equivalently, $x = e^z/(1 + e^z)$], equation (2) can be rewritten as

$$m = \int_{-\infty}^{+\infty} \frac{e^z}{1 + e^z} \phi_{\mu,s}(z) dz, \quad (3)$$

where $\phi_{\mu,s}(z)$ is the probability density function of the normal distribution of mean μ and standard deviation s .

We need to solve equations (1) and (3) for μ and s . The integral in equation (3) cannot be solved explicitly. However, the function $e^z/(1 + e^z)$ can be well approximated by a zero-mean normal distribution of suitable standard deviation t [10], as seen in Figure 5. The two points (other than 0) at which $e^z/(1 + e^z)$ and $\Phi_{0,t}(z)$ are equal depend on the choice of t . In this setting, we choose t so that the functions are equal at the quantile corresponding to m . [We do this so that the two functions are equal near the maximum of $\phi_{\mu,s}(z)$, the term by which $e^z/(1 + e^z)$ is multiplied in the integrand in equation (3).] From Figure 5, this value of t is the one for which $\Phi_{0,t}(\text{logit}(m)) = m$. Noting that $\Phi_{0,t}(\text{logit}(m)) = \Phi_{0,1}(\text{logit}(m)/t)$, it follows that

$$t = \frac{\text{logit}(m)}{\Phi^{-1}(m)},$$

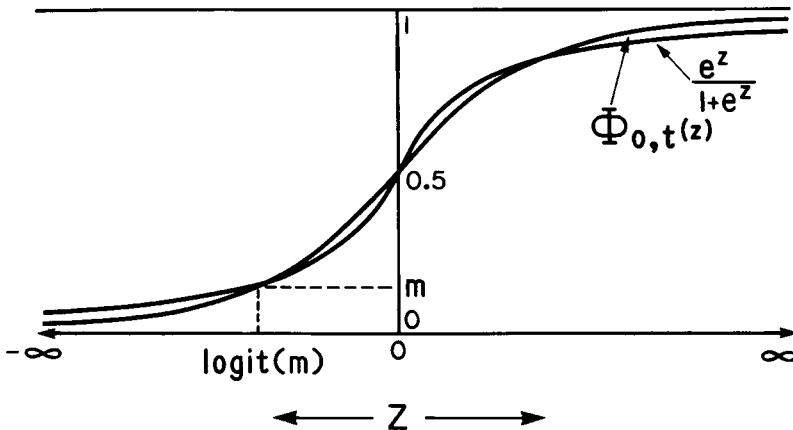


Figure 5. Close approximation of $e^z/(1 + e^z)$ and the zero-mean normal distribution function $\Phi_{0,t}(z)$. The crossing points of the curves depend on the standard deviation t . By suitable choice of t (see Appendix A), the two functions are equal at the m^{th} quantile, where m is the mean value of the probability or utility in question.

where $\Phi = \Phi_{0,1}$ and Φ^{-1} is the inverse of Φ . Using $\Phi_{0,1}(z)$ as an approximation for $e^z/(1+e^z)$ in equation (3), the right side of that equation becomes the convolution of two normal distribution functions [16]. It follows that

$$m = \Phi\left(\frac{\mu}{\sqrt{s^2 + (\text{logit}(m)/\Phi^{-1}(m))^2}}\right). \quad (4)$$

Rearranging terms and replacing $\text{logit}(m)$ with M , one obtains

$$\mu = \Phi^{-1}(m) \cdot \sqrt{s^2 + (M/\Phi^{-1}(m))^2}. \quad (5)$$

Substituting the expression for s given by equation (1) into the above equation, squaring both sides, and letting $E = 1.96/\Phi^{-1}(m)$ yields the following equation for μ :

$$(1 - E^2)\mu^2 - 2\mu B + B^2 + E^2 M^2 = 0. \quad (6)$$

If $m \neq 0.025$ or 0.975 , then $E \neq 1$, so that equation (6) is a quadratic equation for μ . Solving by conventional means,

$$\mu = \frac{B \pm E \cdot \sqrt{B^2 - M^2 + E^2 M^2}}{1 - E^2}. \quad (7)$$

When m is between 0.025 and 0.975 , the quantity inside the radical is always positive, so that there are two possible values for μ . Because equation (6) was obtained by squaring equations (1) and (5), one of the two values of μ might not satisfy (1) or (5), but instead satisfy the negative of (1) or (5) (i.e., left-side = negative of right-side). In fact, it is easily shown that the only acceptable value of μ [i.e., the only one that satisfies (1) and (5)] is

$$\mu = \frac{B - E \cdot \sqrt{B^2 - M^2 + E^2 M^2}}{1 - E^2}. \quad (8)$$

When m is less than 0.025 or greater than 0.975 , the quantity inside the radical can be negative, zero, or positive. If negative, then the distribution cannot be approximated using a logistic-normal model. If zero, then equation (7) yields a single value for μ ; this value is acceptable unless $m < 0.025$ and $b > 0.5$, or $m > 0.975$ and $b < 0.5$. Finally, if the quantity inside the radical is positive, then equation (7) yields two values for μ ; neither is acceptable if $m < 0.025$ and $b > 0.5$, or if $m > 0.975$ and $b < 0.5$; only the one given by equation (8) is acceptable if $m < 0.025$ and b is a lower bound, or if $m > 0.975$ and b is an upper bound; both are acceptable otherwise. In the last situation, in which both values are acceptable, we would use the value given by equation (8); while somewhat arbitrary, this value is closer to B and thus yields a narrower distribution.

If $m = 0.025$ or 0.975 , then $E = 1$, so that equation (6) is a linear equation for μ , whose solution is

$$\mu = (B^2 + M^2)/2B. \quad (9)$$

This value is acceptable if $m = 0.025$ and $b < 0.5$, or if $m = 0.975$ and $b > 0.5$, unacceptable otherwise.

When there is an acceptable root μ given by equation (8) or (9), the standard deviation s is obtained directly from equation (1). When there is no acceptable root [i.e., if $m \leq 0.025$ and $b > 0.5$, or if $m \geq 0.975$ and $b < 0.5$, or if the quantity inside the radical in equation (8) is negative], then m and b cannot be fitted to a logistic-normal distribution.

Based on the same reasoning used to justify equation (1), the implicit upper bound is the value (between 0 and 1) whose logit transform is $\mu + 1.96s$. That is,

$$\text{Implicit upper bound} = \frac{e^{(\mu + 1.96s)}}{1 + e^{(\mu + 1.96s)}}. \quad (10)$$

If b had been chosen to be the 95 percent upper bound, then the implicit lower bound would be given by an equation similar to (10), but with $\mu + 1.96s$ replaced by $\mu - 1.96s$.

Equations (1) and (8) [or (1) and (9) if $m = 0.025$ or 0.975] provide closed-form expressions for μ and s in terms of m and b . Because the derivation of equations (8) and (9) relies upon an approximation — replacing $e^z/(1 + e^z)$ with a zero-mean normal distribution function — we tested our results empirically. For example, we set $m = 0.4$ and $b = 0.05$, and obtained $\mu = -0.51$ and $s = 1.24$ from equations (1) and (8). We then randomly selected 1000 values from the normal distribution of mean -0.51 and standard deviation 1.24 , and took the inverse logit transform ($e^z/(1 + e^z)$) of each. The resulting 1000 numbers had a mean of 0.409 , very close to the desired mean of 0.4 . Furthermore, 2.2% of the numbers were less than 0.05 , close to the desired 2.5% . Similar excellent results were obtained over a wide variety of choices of m and b , including the extreme choice of $m = 0.001$ and $b = 0.0002$.

Appendix B: Chance Nodes with More than Two Branches

The following is a practical approach (though not precise from a theoretical standpoint) to Monte Carlo simulation involving a multiple-branch chance node with branch probabilities p_1, p_2, \dots, p_n . For each probability p_i , the mean m_i and a bound b_i are specified. (The means must sum to 1.) Each p_i is assumed to be logistic-normal, and proceeding as in Appendix A, the mean μ_i and the standard deviation s_i of the normal distribution representing $\text{logit}(p_i)$ is derived. In each run of the simulation, random selection from these distributions yields a value q_1 as a preliminary estimate of p_1, q_2 of

p_2 , and so on. The values q_1, q_2, \dots, q_n are then normalized to a sum of 1. That is, the value used to estimate the i_{th} probability p_i is

$$\frac{q_i}{q_1 + q_2 + \dots + q_n}.$$

Although each p_i is estimated by an expression whose numerator comes from a distribution that represents p_i and whose denominator has a mean value of 1, the result does not necessarily have the mean and bound specified for p_i . This is because the mean of a quotient of two distributions is not in general the quotient of the means. However, empirical testing of this approach with four-branch chance nodes has confirmed that it performs well over a wide variety of specified values for the mean and bound at each branch.

A theoretically more appealing approach to multiple-branch chance nodes would be to assume that the $(n-1)$ values

$$z_i = \log \frac{p_i}{p_n}$$

(for $i = 1, 2, \dots, n-1$) are $(n-1)$ -variate normal. The probabilities are then derived from z_1, z_2, \dots, z_{n-1} by

$$p_i = \frac{e^{z_i}}{1 + e^{z_1} + e^{z_2} + \dots + e^{z_{n-1}}}.$$

The practical barrier to carrying out such an approach is the problem of determining the parameters of the $(n-1)$ -variate normal distribution so that each of the probabilities p_i has the desired characteristics (i.e., mean and bound).

Notes

1. In fact, any confidence range (e.g., 75%, 90%) could be used, with only minor modifications to the formulas in Appendix A.
2. $\Phi^{-1}(m)$ is the number x (between $-\infty$ and $+\infty$) for which Φ , the normal distribution function of mean 0 and standard deviation 1, evaluated at x equals m . E.g., $\Phi^{-1}(0.5) = 0$; $\Phi^{-1}(0.84) = 1$; $\Phi^{-1}(0.025) = -1.96$.

References

1. Tversky A, Kahneman D: Judgment under uncertainty. Heuristics and biases. *Science* 185:1124-1131, 1974
2. Braun P: The clinical management of suspected Herpes virus encephalitis. A decision-analytic view. *Am J Med* 69:895-902, 1980
3. Whitley RJ, Soong S, Dolin R, et al: Adenine arabinoside therapy of biopsy-proved Herpes simplex encephalitis. *N Engl J Med* 297:289-294, 1977
4. Adams H, Miller D: Herpes simplex encephalitis. A clinical and pathological analysis of 22 cases. *Postgrad Med J* 49:393-397, 1973

5. Barza M, Pauker SG: The decision to biopsy, treat, or wait in suspected Herpes encephalitis. *Ann Intern Med* 92:641-649, 1980
6. Weinstein MC, Fineberg HV, Elstein AS, et al: *Clinical Decision Analysis*. Philadelphia: Saunders, 1980
7. Plante DA, Pauker SG: Enterococcal endocarditis and penicillin allergy. *Med Decis Making* 3:81-109, 1983
8. Pass TM, Goldstein LP: A computerized aid for medical cost-effectiveness analysis (abs). *Med Decis Making* 1:465, 1981
9. Raiffa H. *Decision Analysis. Introductory Lectures on Choices Under Uncertainty*. Reading, MA: Addison-Wesley, 1968
10. Aitchison J, Begg CB: Statistical diagnosis when basic cases are not classified with certainty. *Biometrika* 63:1-12, 1976
11. Aitchison J: The statistical analysis of compositional data. *J R Stat Soc B* 44:139-177, 1982
12. Hastings NAJ, Peacock JB: *Statistical Distributions*. London: Butterworth, 1975
13. Colton T, *Statistics in Medicine*. Boston: Little Brown, 1974
14. Stael von Holstein C-AS. Two techniques for assessment of subjective probability distributions. An experimental study. *Acta Psychol (Amst)* 35:478-494, 1971
15. Levey AS, Pauker SG, Kassirer JP: Occult intracranial aneurysms in polycystic kidney disease. When is cerebral arteriography indicated? *N Engl J Med* 308: 986-994, 1983
16. Chung KL: *A Course in Probability Theory*. New York: Harcourt, Brace, and World, 1968