SENSITIVITY ANALYSIS AND THE EXPECTED VALUE OF PERFECT INFORMATION

James C. Felli, Ph.D.
Defense Resources Management Institute
Naval Postgraduate School
Monterey, CA 93943

Gordon B. Hazen, Ph.D.
Department of Industrial Engineering and Management Science
Northwestern University
Evanston, IL 60208

11 June, 1997

Please direct all queries and requests for reprints to James C. Felli, DRMI (Code 64 FL), Naval Postgraduate School, 1522 Cunningham Road, Monterey, CA, 93943-5201.

SENSITIVITY ANALYSIS AND THE EXPECTED VALUE OF PERFECT INFORMATION

ABSTRACT

We examine measures of decision sensitivity that have been applied to medical decision problems. Traditional *threshold proximity* methods have recently been supplemented by *probabilistic* sensitivity analysis, and by *entropy-based* measures of sensitivity. We propose a fourth measure based upon the *expected value of perfect information* (EVPI), which we believe superior both methodologically and pragmatically. Both the traditional and the newly suggested sensitivity measures focus entirely on the likelihood of decision change without attention to corresponding changes in payoff, which are often small. Consequently, these measures can dramatically overstate problem sensitivity. EVPI, on the other hand, incorporates both the probability of a decision change and the marginal benefit of such a change into a single measure, and therefore provides a superior picture of problem sensitivity. To lend support to this contention, we revisit three problems from the literature and compare the results of sensitivity analyses using probabilistic, entropy- and EVPI-base measures.

INTRODUCTION

The effective management of uncertainty is one of the most fundamental problems in decision making. While uncertainty can arise from a plethora of sources and assume a multitude of forms, in this paper we focus on parametric uncertainty within the framework of quantitative medical decision-making models. Such models typically contain several parameters whose values are unknown and must be estimated by the decision maker (e.g., disease incidence rates, the likelihood of drug side effects, the sensitivity and specificity of diagnostic tests). Currently, most medical decision models rely on point estimates for input parameters, although the uncertainty surrounding these values is well-recognized.[1,2] Because the values decision makers (DMs) assign to input parameters greatly determine the output of their models, it is natural that they should be interested in the relationship between changes in those values and subsequent changes in model output. This relationship constitutes the underpinning of a class of analytic procedures collectively referred to as sensitivity analysis (SA).

Common sense dictates that a quantitative model which exhibits large fluctuations in output for relatively small changes in the value of some input parameter is sensitive to the parameter, whereas a model which exhibits small output variation for substantial perturbations is insensitive to the parameter. But sensitive in what respect? Because we are concerned with models designed to identify a preferred course of action, most often a treatment strategy, we must draw an important distinction between *value sensitivity* and *decision sensitivity*. Given some change in the input parameters, value sensitivity refers to a change in the magnitude of a model's optimal value; decision sensitivity, on the other hand, refers to a change in the preferred alternative identified by the model. It is possible for a decision model to simultaneously exhibit

high levels of value sensitivity and little or no decision sensitivity, or vice versa. Many formal SA procedures focus entirely on decision rather than value sensitivity.

As decision modeling tools and the techniques available for their solution have become more sophisticated, the role of SA has become more prominent and DM's requirements of SA have become more well-defined. We examine and critique three measures of decision sensitivity that have been utilized in the medical decision making literature, based upon threshold proximity, probability of a threshold crossing, and entropy. We then propose the use of a new sensitivity indicator based upon the expected value of perfect information (EVPI). We present what we believe to be convincing arguments that EVPI-based sensitivity analysis is the proper way to proceed. In a simple example, we demonstrate how threshold proximity, probabilistic, and entropy-based measures can dramatically overstate problem sensitivity compared to EVPI. The reason is the exclusive focus of the former measures on the likelihood of decision change. In contrast, EVPI considers not only the probability of decision change but also takes into account the payoff differential resulting from such a change. We assert that the same overstatement of sensitivity can occur in real decision problems, and substantiate this assertion by comparing the results of probabilistic, entropy-based, and EVPI-based SAs for a selection of three problems taken from the medical literature.

THRESHOLD PROXIMITY MEASURES

Threshold proximity measures use distance to a threshold as a proxy for decision sensitivity. Figure 1, adapted from Plant *et al.* depicts a two-way sensitivity analysis of this type for model parameters F and Q, the "fold-increase in disease-free survival" and "quality of life

after surgery" for a patient with stage III squamous cell carcinoma of the pyriform sinus.[3] These parameters are valued from 1 to 1.5 and from 0 to 1, respectively. In Figure 1, the (F,Q) parameter space is partitioned into 3 regions, each of which represents a mix of values of F and Q for which a particular treatment alternative is optimal. Boundaries shared by two regions are called *thresholds* and designate indifference between adjacent alternatives.

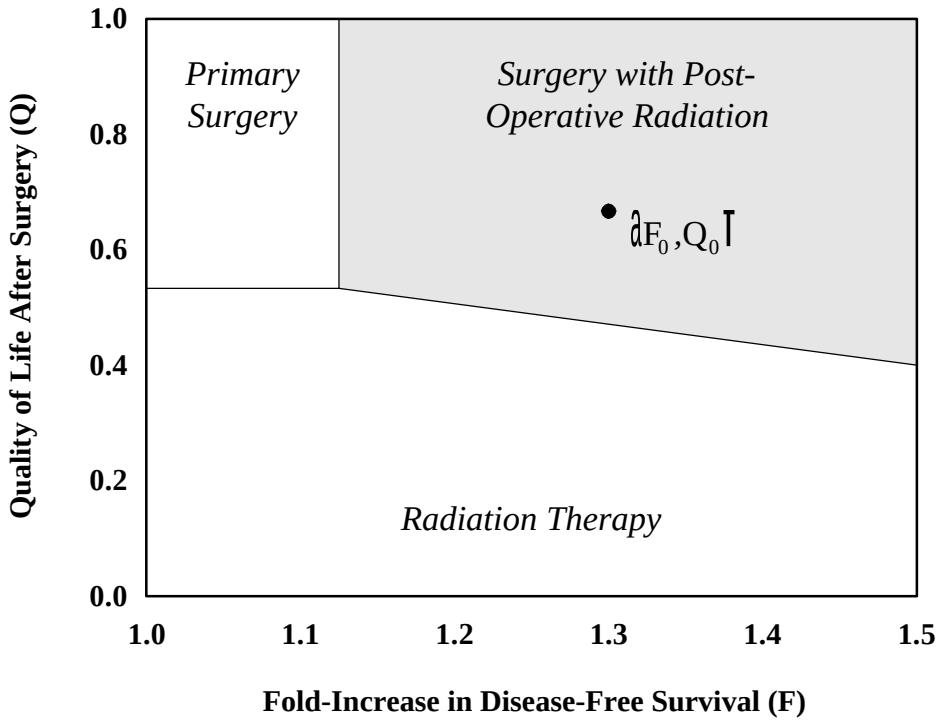

Figure 1. Two-way threshold proximity SA for parameters F and Q.

The DM's *base values*[a] of F and Q determine a base point (F_0,Q_0) in the (F,Q) space, as illustrated in Figure 1, and identify the *base-optimal alternative* to be Surgery with Post-Operative Radiation. By examining the proximity of the base value (F_0,Q_0) to neighboring thresholds and contrasting it with his beliefs about the likely values of F and Q, the DM can get a

feel for how the optimal alternative is likely to change with variations in F and Q. For example, Plante *et al.* believed it quite likely that F could fall below the $F = 1.12$ threshold, thereby shifting the preferred alternative to Primary Surgery. They therefore concluded that the optimality of the Surgery with Post-Operative Radiation alternative was very sensitive to the value of F.

Threshold proximity SAs can be conducted for a single parameter, or for two or more parameters simultaneously. Graphical displays such as the one presented in Figure 1, however, become difficult to construct and interpret for SAs involving three parameters and next to impossible for four or more parameters. This is unfortunate, as one- and two-way SAs may not capture the full sensitivity of the base-optimal alternative to multiple parameters considered simultaneously. It may be the case, for instance, that a decision is insensitive to the variation of some set of parameters individually, but sensitive to their simultaneous variation.[4,5]

The lack of graphical representation for multiparametric SA makes it difficult for the DM to estimate the likelihood of a threshold crossing when several parameters are allowed to vary jointly. Some researchers have advocated the calculation of the distance in parameter space from the point defined by the parameters' base values to the nearest threshold as a numerical proxy for this likelihood.[6,7,8,9,10] The choice of an appropriate distance metric then becomes an issue, as difficulties can arise due to non-commensurable units of measure across parameters, or due to the choice between different but equivalent ways of jointly defining parameters.[11,12] Presently, the lack of established guidelines for differentiating between "sensitivity" and "insensitivity" under threshold proximity SA can result in essentially arbitrary interpretation of analytic results.[2,5] An alternate approach to multiparametric SA, which bypasses these difficulties, is *probabilistic sensitivity analysis*, which we discuss next.

PROBABILISTIC MEASURES

Probabilistic SA techniques require the DM to assign a probability distribution to each uncertain parameter, reflecting the likely value of that parameter. This methodology represents a paradigm shift from threshold proximity SA by virtue of its emphasis on the *probability* of a threshold crossing rather than the distance to the threshold.

Because the DM must provide distributions for problem parameters under examination, he must think about these uncertain quantities in some detail before the SA can be performed. In general, this task should not prove cumbersome, as the required level of detail is a natural extension of the thought processes already employed in his initial modeling of the problem. For example, the DM may have estimated the probability of a key event in his decision model. This probability represents the long-term relative frequency of the event in question, about which the DM may have residual uncertainty.[13,14] The DM can select bounds for or construct a confidence interval about his estimate to formalize this second order uncertainty; however, a more complete formalization would be for him to specify, in accordance with his beliefs, a probability distribution over all possible values of the parameter. This latter task must be performed for all parameters for which the DM desires to perform probabilistic SA.

As an illustration, consider again the threshold proximity analysis presented in Figure 1. The base value of the parameter F is 1.3 and the threshold value is 1.12. Numerically, the value 1.3 is physically "close" to the threshold, but is it "close enough" to constitute sensitivity of the base-optimal alternative to F? This is where the DM's beliefs about the behavior of F play a critical role. Given a distribution over the values of F that reflect his beliefs, any questions the

DM might have regarding the likelihood of F obtaining beyond some critical value may be directly addressed probabilistically.

Figure 2 presents two possible distributions for F. Both of these distributions exhibit a most likely value of 1.3, however distribution f_1 reflects a greater uncertainty about the value of F than distribution f_2.

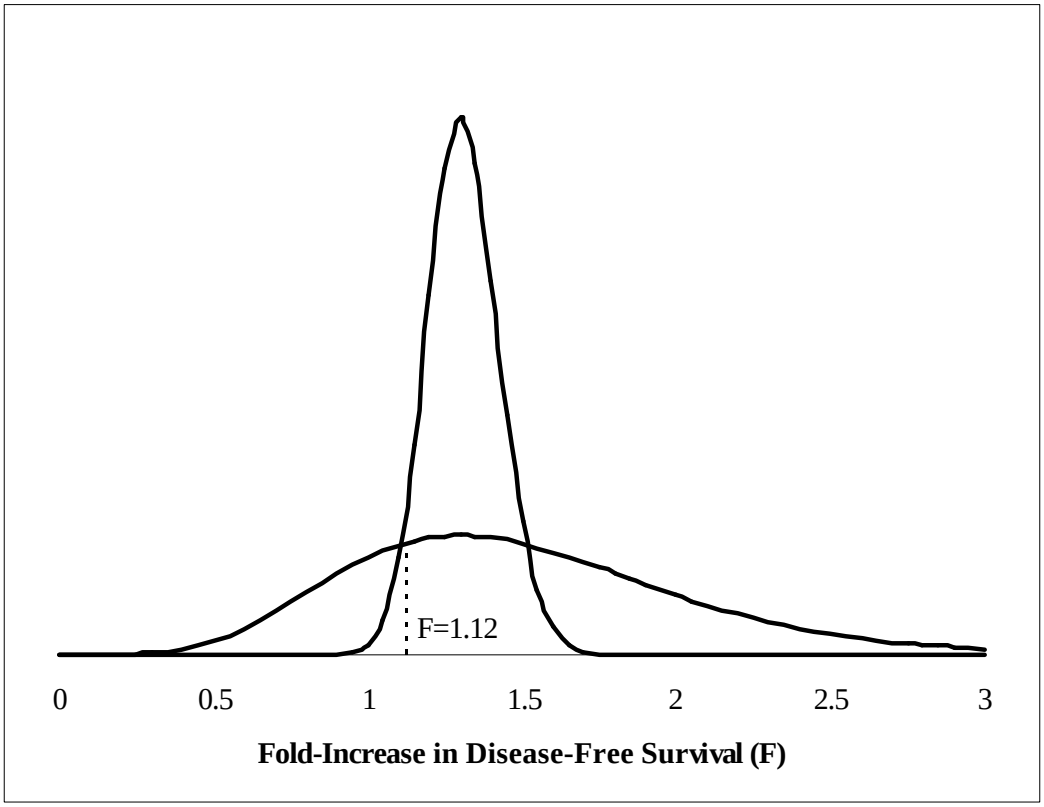

Figure 2. Two possible distributions for the parameter F.

If the DM's beliefs about the behavior of F are represented by distribution f_1, the area under f_1 to the left of 1.12 is the probability that F will take on a value below 1.12. In this case, the probability is fairly large, and this may induce the DM to label the decision sensitive to F, as

did Plante *et al*. If, on the other hand, he felt that distribution f_2 better reflected his beliefs about F, he would examine the area under f_2 to the left of 1.12. That probability is quite small, and could induce him to label his decision insensitive to F.

Probabilistic SA can be extended to accommodate the multiparametric case by the use of a joint distribution over the parameters of interest. The drawback to this approach is that mathematical calculations involving probability distributions can be quite cumbersome, especially when several parameters are considered concurrently. To circumvent some of the mathematical complexity involved in multiparametric probabilistic SA, Monte Carlo methods have been employed to enable the DM to directly model his beliefs about parameter behavior.[15,16,17,18] Repeated simulation of a decision model on a computer enables the DM to estimate critical long term probabilities (e.g., the probability that some alternative is optimal) while allowing all problem parameters of interest to vary according to their distributions.

As a case in point, Doubilet *et al.* proposed a Monte Carlo approach to probabilistic SA in the context of selecting a treatment procedure for a patient with suspected herpes simplex encephalitis (HSE).[18] In their model, the DM's three alternatives were: perform a brain biopsy followed by treatment with vidarabine only if the biopsy is positive (B), forego a brain biopsy and provide treatment with vidarabine (NB/T), or forego both biopsy and treatment (NB/NT). Outcomes were assigned payoff values between the arbitrary endpoints 0 and 1: a payoff of zero was given to the least favorable outcome (death of the patient); a payoff of 1 was given to the most favorable outcome (minimal or no sequalae). Uncertain parameters were characterized by two quantities, a baseline estimate and a bound on the parameter's 95% confidence interval. In addition, the authors assumed the parameters to be independent random variables and that the logit transformation[b] of each parameter was normally distributed, thereby formalizing parameter

distributions for their simulation model. Baseline analysis revealed the NB/T option as the preferred alternative with an expected value of 0.5907. Repeated simulation of the problem on a computer enabled them to make probability statements about the reliability of conclusions drawn from their baseline analysis. Some of their SA results are provided in Table 1.

	Alternatives		
Probabilistic SA (1000 Simulations)	**A_1=B**	**A_2=NB/T**	**A_3=NB/NT**
Mean Expected Value	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%

Table 1. Results from Doubilet *et al.'s* simulation for the HSE problem.

The authors' simulation results demonstrate that the NB/T alternative is the optimal strategy 79.5% of the time, exhibiting an average expected value of 0.563 with a standard deviation of 0.096. The B and NB/NT strategies yield the highest expected value only 18.4% and 2.1% of the time, respectively. In terms of stability, these results imply that when all parameters are allowed to vary according to their distributions simultaneously there is a 20.5% chance that the base-optimal alternative is in fact suboptimal.

The authors' definition of "buys" and "costs" are based upon comparisons of the payoffs of all three strategies. An alternative exhibiting a buy of 0.004 means that the expected value of the alternative exceeded those of all other alternatives by at least 0.004; similarly, an alternative

exhibiting a cost of 0.004 designates that the expected value of that alternative fell short of those of all other alternatives by at least 0.004. The results provided in Table 1 suggest that although the NB/T strategy can be expected to yield the greatest expected value 79.5% of the time, the DM can only be 58.3% confident that it is best by a margin of at least 0.004. This figure is important because it incorporates into the analysis a measure of significance in terms of payoffs: Differences in payoff of 0.004 or more are considered significant by the DM, whereas differences of less than 0.004 are not. The NB/T strategy is optimal 79.5% of the time, however, the expected value of the alternative is *significantly greater* than those of the other strategies only 58.3% of the time. We will discuss this issue in more detail later.

Although Doubilet *et al.* performed a multiparametric SA in which all problem parameters were allowed to vary simultaneously, repeated application of the Monte Carlo method can be employed to provide insight into the relative importance of specific parameter sets provided that only those parameters are allowed to vary with each simulation (with all other parameters held fixed at their base values). When one considers the speed of today's computers and the ready availability of simulation software, the prospect of repeated simulations is not prohibitive.

In addition to the estimation of long term probabilities, probabilistic SA provides a mechanism for the DM to examine output distributions directly, such as the payoff for a single alternative or the difference between payoffs for some pair of competing alternatives.[19,20] Knowledge of the likelihood of each payoff (or payoff difference) over the entire range of possible values enables the DM to better assess the risk of an adverse outcome or, in the case of difference in payoffs between two competing alternatives, select an alternative based upon the likelihood that its payoff will exceed that of its competitor by some specified amount.

Probabilistic SA for a realistic problem with many parameters requires computational software such as that available for Monte Carlo simulation. One troublesome question is how best to integrate information gleaned from more than one payoff (or payoff difference) distribution when the decision is not dichotomous. Also, as with threshold proximity measures, the subjectivity surrounding what constitutes decision sensitivity remains at issue: given that *p* is the long run probability that a given alternative is optimal, only the DM can say whether *p* is "sufficiently large" to call the decision "sensitive" to the parameter set under investigation.

ENTROPY-BASED MEASURES

The concept of information *entropy*[21] has been proposed as a basis for a measure of decision sensitivity.[16,22] Given two random variables X and Y, the expected information X yields about Y is given by the *mutual information* (MI),[23] or *cross-entropy* between X and Y, defined mathematically as:

$$MI(Y \mid X) = \sum_{y} \sum_{x} p_{xy} \log_2 \left(\frac{p_{xy}}{p_x p_y} \right)$$

where $p_x = \Pr[X = x]$, $p_y = \Pr[Y = y]$, and $p_{xy} = \Pr[X = x, Y = y]$.

The preferred alternative in a decision model is a function of the model's parameters, so any uncertainty surrounding those parameters is inherited by the preferred alternative. The preferred alternative itself, therefore, can be regarded as a random variable. In this framework, entropy can be employed to quantify the information content a given parameter carries with

respect to an artificial random variable designating the preferred alternative. Critchfield and Willard used this approach to construct a normalized measure of mutual information as follows.[16] Given that a DM's optimal decision is influenced by some parameter ξ and the optimal action identified by the DM's model is A (a function of ξ and the remaining problem parameters), they defined a *mutual information index* (MII) as the mutual information between ξ and A normalized by the mutual information of A with itself:

$$S_{\xi A} \equiv \frac{MI(A \mid \xi)}{MI(A \mid A)}$$

Critchfield and Willard proposed that $S_{\xi A}$ could serve as a viable proxy for decision sensitivity to the parameter ξ because the sensitivity of the base-optimal alternative A to variation in ξ was reflected in the magnitude of the ratio $S_{\xi A}$. The higher the value of $S_{\xi A}$, the more the parameter ξ explained about the variability of A.

To illustrate the merit in this approach to SA, Critchfield and Willard applied their MII to a decision model first presented by Klein and Pauker.[16,24] In this model, the decision under consideration was whether or not to administer anticoagulants to a 25 year old pregnant woman presenting deep vein thrombosis (DVT) during the first trimester of pregnancy. Outcomes were valued between the arbitrary endpoints 0 and 100. A value of zero was given to the least favorable outcome (death of the mother) and a value of 100 was given to the most favorable outcome (survival of mother and infant with no anticoagulant fetopathy). The base-optimal alternative was to administer anticoagulants, which exhibited an expected value of 96.655. Table 2 provides the assumed mean and standard deviation Critchfield and Willard used for each of the

seven independent problem parameters in their reanalysis. Using these values, they determined beta distributions for each parameter and performed an MII-based SA, the results of which are also provided in Table 2.

Parameter ξ	**Mean**	**Std. Deviation**	**$S_{\xi Y}$**
Pr(Pulmonary Embolism)	0.195	0.061	25%
Utility of Adverse Fetal Outcome	90	4.5	23%
Pr(Death | Pulmonary Embolism)	0.28	0.058	5.8%
Pr(Fetopathy)	0.2	0.035	3.6%
Efficacy of Treatment	0.75	0.059	0.80%
Pr(Maternal Bleeding)	0.03	0.0058	<0.1%
Pr(Death | Maternal Bleeding)	0.0025	0.0014	<0.1%

Table 2. Critchfield and Willard's MII results for the DVT problem.

Based on the results illustrated in Table 2, a DM might presume the problem to be sensitive to the probability of pulmonary embolism and the patient's utility of an adverse fetal outcome, as these two parameters contribute highly to the instability of the preferred alternative relative to the other parameters. While it is true that the problem is "more sensitive" to these parameters than to others, it is not clear what this means. This is one of the deficiencies of MII: the subjective call the DM is required to make regarding what is and is not "sensitive" is less transparent than when using probabilistic SA. Each of the two parameters represent in excess of 20% of the instability of the decision, but is that a lot? Because the MII measures MI relative to the entropy of the simulated decision variable A, the MII of the entire parameter set is naturally

100%. This normalization masks the overall sensitivity of the base-optimal alternative to parametric variation. It is not clear from such analysis whether the decision itself is significantly unstable, or whether high $S_{\xi A}$ values are artifacts of the normalization procedure.

The second deficiency of the MII lies in the scope of the SA. Because of the computational difficulties involved in calculating the conditional probability distributions required, the MII has not been utilized to address the relative importance of parameter sets of orders higher than one. In this context, insight into parameter interaction has been sacrificed for more detailed exploration into the information content of a single parameter.

Lastly, the driver of the MII is the probability of a decision change given some information about a parameter, where a change of decision is mandated whenever an increase in expected value is available by selecting another alternative. The magnitude of that increase, however, plays no role in the final sensitivity measure. While it may be the case that some parameter ξ contributes 20+% of the instability of some decision model, it may also be the case that the marginal benefit to be gained from switching alternatives is trivial. In that event, the DM must consider whether it is really worthwhile to change alternatives. If not, can the decision truly be considered sensitive to ξ? To answer this question, we must turn our attention from information *content* to information *value*.

AN INFORMATION VALUE-BASED MEASURE

Consideration of the value of clairvoyance in decision problems engendered the concept of the *expected value of perfect information* (EVPI).[25] Within the context of medical decision making, EVPI has been employed primarily as a diagnostic tool for test selection.[26,27,28,29] We

consider a sensitivity measure based on EVPI to be a logical basis for a measure of decision sensitivity. Methodologically, EVPI represents a natural extension of probabilistic SA by focusing simultaneously on the probability of a decision change and the change in payoff commensurate with such a decision change. Unlike the measures of decision sensitivity examined thus far, however, the calculation of EVPI is consistent with both the maximization of expected value and the economic concept of marginal reasoning.

As an indicator of decision sensitivity, EVPI is perhaps best introduced from a conceptual standpoint. Consider the basic dilemma of a DM faced with a problem characterized by several actions and a single uncertain parameter ξ. The DM wants to select the alternative that will maximize his expected payoff. His most visible course of action is to solve his problem using the value of ξ he feels most likely to obtain, call it ξ_0, and then select the alternative that maximizes his expected payoff given that $\xi = \xi_0$. Suppose that alternative is alternative a_0. Using expected value notation, this means that $\mathrm{E}\left[\mathrm{V}_{a_0} \mid \xi = \xi_0\right] = max_a \, \mathrm{E}\left[\mathrm{V}_a \mid \xi = \xi_0\right]$.

However, due to the uncertain nature of ξ, the DM faces some risk in using $\xi = \xi_0$ to determine the "optimal" alternative. If the value actually obtained by ξ happened to fall to the wrong side of some threshold, then the alternative preferred under $\xi = \xi_0$ would lose optimality and some other alternative would yield a higher expected payoff. By selecting alternative a_0, the DM foregoes any additional payoff he could expect to receive under a different alternative. For any particular value of ξ, this foregone payoff can be expressed as the difference $max_a \, \mathrm{E}\left[\mathrm{V}_a \mid \xi\right] - \mathrm{E}\left[\mathrm{V}_{a_0} \mid \xi\right]$. Averaging over all possible values of ξ yields the DM's average foregone payoff, which is the expected value of perfect information on the parameter ξ:

$$\text{Average foregone payoff} = E_{\xi}\left[max_{a} E\left[V_{a} \mid \xi\right] - E\left[V_{a_0} \mid \xi\right]\right]$$

$$= \text{EVPI}(\xi)$$

The EVPI(ξ) represents the average improvement the DM could expect to gain over the payoff resulting from his selection of alternative a_0 given perfect information on the parameter ξ prior to the time of his decision. It can also be expressed as the product of the probability of a change in the preferred alternative due to variation in ξ (the likelihood of ξ making a threshold crossing) and the average change in payoff resultant from such a decision change (the opportunity loss expected from ξ making a threshold crossing). In this sense, EVPI succinctly encapsulates two pieces of information DM's typically seek from SA. Parameters to which the DM considers his problem sensitive naturally correspond to those parameters with high EVPIs, because those parameters possess *both* great potential to induce a change in the base-optimal alternative *and* afford significant marginal improvement in expected payoff consequent to such a decision change (note the similarity between this added emphasis on marginal improvement in payoff and the notion of "buys" incorporated into Doubilet *et al.*'s probabilistic SA[18]).

As with the case of threshold proximity, probabilistic, and entropy-based SA measures, the DM's final distinction between "sensitivity" and "insensitivity" is subject to his personal beliefs. Unlike those SA measures, however, in the case of an EVPI analysis the distinction is made based upon a comparison of values in the *same units as the problem's payoffs* (as opposed to distance, probability, or some fraction of the entropy of an artificial random variable). EVPI thereby enables the DM to draw conclusions regarding the relative importance of parameter sets based upon direct consideration of the average marginal improvement attributable to a (hypothetical) gain of perfect information. The labels "sensitive" and "insensitive" can therefore

be assigned based upon the DM's appraisal of whether or not perfect information could be expected to yield a *significant improvement* over the payoff he could expect to receive without such knowledge. To assign sensitivity labels to parameter sets based upon EVPI, the DM needs only determine what he considers a minimum level of significant improvement.

Although formal distinction between "significance" and "insignificance" remains a point of contention among clinical researchers[30], the idea of identifying a minimum appreciable level of improvement is not new. The specification of minimum significant improvement been championed by advocates of Bayesian statistical approaches to clinical trials, and is likely to become more frequent as Bayesian approaches gain acceptance.[31,32] More recently, a measure of the minimally important difference in patient health status has been proposed as a tool to determine whether a trivial or notable change has occurred in a clinical trial.[33,34,35,36] For an EVPI analysis, we believe that the minimum significant improvement serves as a natural threshold between "sensitivity" and "insensitivity." If V* is the smallest improvement the DM believes significant, then it is reasonable that he assign the label "sensitive" only to those parameter sets ξ with $\mathrm{EVPI}(\xi) \geq V^*$.

The primary difficulty in calculating EVPI (and $\mathrm{MI}(A \mid \xi)$ for that matter) lies in the mathematical complexity inherent in determining expectations with potentially complex parameter distributions. As with probabilistic and entropy-based SA, this difficulty can be overcome by means of computer simulation. The necessary calculations, regardless of the size of the parameter set under consideration, fall well within the purview of contemporary DMs armed with a moderately powerful personal computer and appropriate software. Via simulation, the calculation of EVPI remains tractable for large parameter sets, thereby enabling the DM to perform meaningful and easily interpreted single and multiparametric SAs.

EVPI SENSITIVITY ANALYSIS VIA MONTE CARLO SIMULATION

EVPI values can be calculated in closed form only for problems with very simple or special structure. For most realistic decision problems, EVPI values must be numerically approximated. Because EVPI is an *expected* value, it is natural and relatively straightforward to estimate it using Monte Carlo simulation. In this section, we give a short description of how we used Monte Carlo simulation for this purpose.

In general, consider a decision problem whose payoff V depends on, among other things, a set ξ of parameters and an alternative A. Let $E[V|\xi,A]$ be the DM's expected payoff as a function of ξ and A. Suppose that $\xi_I = \{\xi_i\}_{i\in I}$ is a collection of parameters whose EVPI we wish to calculate and let ξ_I^c be the set of remaining parameters in the problem (i.e., $\xi = \xi_I \cup \xi_I^c$). Let $A^*(\xi_I)$ be the optimal decision as a function of ξ_I, and let A^* be the base optimal alternative. Expressing $EVPI(\xi_I)$ as expected improvement, we have

$$\begin{aligned} EVPI(\xi_I) &= E_{\xi_I,\xi_I^c}\left[E[V|\xi_I,\xi_I^c,A^*(\xi_I)] - E[V|\xi_I,\xi_I^c,A^*]\right] \\ &= E_{\xi_I,\xi_I^c}\left[\text{Improvement using } A^*(\xi_I) \text{ instead of } A^*\right] \end{aligned}$$

The Monte Carlo simulation procedure amounts to repeatedly simulating improvement values and averaging them. Here is a sketch of the general procedure:

MC1: *General Monte Carlo simulation procedure*

1. Repeatedly generate random parameter values ξ_I.

2. For each generated ξ_I

   i. Determine $A^*(\xi_I)$ as the A maximizing $E[V|\xi_I, \xi_I^c, A]$.

   ii. Calculate the improvement achieved by using $A^*(\xi_I)$:

$$\text{Improvement} = E[V|\xi_I, \xi_I^c, A^*(\xi_I)] - E[V|\xi_I, \xi_I^c, A^*]$$

   End For

3. Estimate $EVPI(\xi_I)$ as the average of the calculated improvement values.

To illustrate, consider the DVT problem. The structure of the problem, as provided by Klein and Pauker[24], is provided in Figure 3.

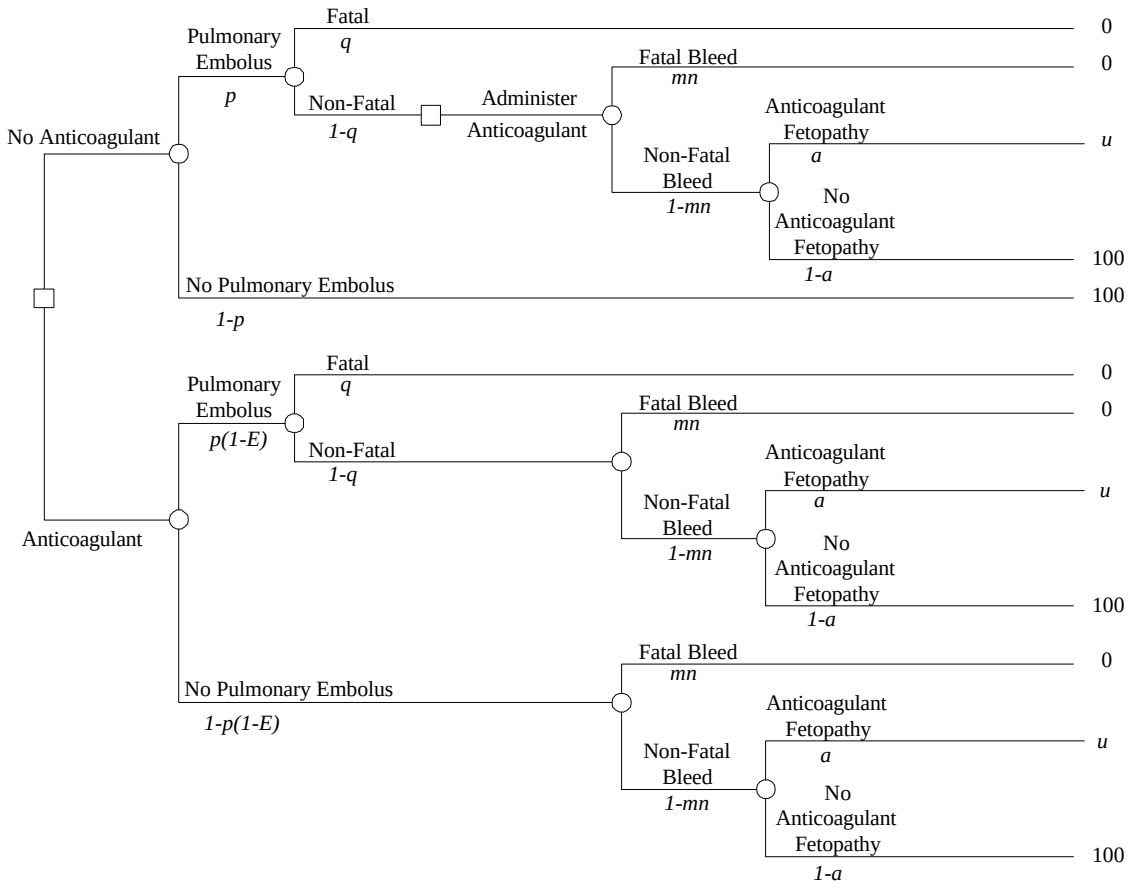

Figure 3. Decision tree for the DVT problem.

There are two alternatives available to the DM: $A = A_1 =$ "Administer anticoagulants" and $A = A_2 =$ "Do not administer anticoagulants". The set of parameters is $\xi = \{p,q,a,m,n,E,u\}$ where:

$$p = \Pr\{\text{Pulmonary Embolism}\}$$

$$q = \Pr\{\text{Death} \mid \text{Pulmonary Embolism}\}$$

$$a = \Pr\{\text{Fetopathy}\}$$

E = Efficacy of Treatment

m = Pr(Maternal Bleeding)^c

n = Pr(Death | Maternal Bleeding)

u = Value of Adverse Fetal Outcome

Rollback of the decision tree produces the following expected payoffs as functions of ξ and A:

$$\mathrm{E}[\mathrm{V} \mid \xi, \mathrm{A} = \mathrm{A}_1] = 100(1 - a)(u - 1)(1 - mn)(1 - p)(1 - E)q$$

$$\mathrm{E}[\mathrm{V} \mid \xi, \mathrm{A} = \mathrm{A}_2] = 100[(1 - a)(u - 1)(1 - mn)(1 - q)p + 1 - p]$$

Suppose we are interested in estimating $\mathrm{EVPI}(p, m)$. Then $\xi_\mathrm{I} = (p, m)$ and $\xi_\mathrm{I}^c = (a, q, n, E, u)$. Using the mean values from Table 2 as base parameter values, we find that $\mathrm{E}[\mathrm{V} \mid \mathrm{A} = \mathrm{A}_1] > \mathrm{E}[\mathrm{V} \mid \mathrm{A} = \mathrm{A}_2]$, so $\mathrm{A}^* = \mathrm{A}_1$. To estimate $\mathrm{EVPI}(\xi_\mathrm{I})$, we would let $\mathrm{A}^*(\xi_\mathrm{I})$ be the optimal decision as a function of ξ_I and apply the general Monte Carlo procedure MC1. The result is $\mathrm{EVPI}(p, m) = 0.0248$.

Frequently, special payoff structure can allow for a shortcut in the simulation procedure. When all parameters are assumed probabilistically independent and the payoff function is *multilinear* (i.e., linear in each individual parameter), $\mathrm{E}_{\xi_\mathrm{I}^c}[\mathrm{E}[\mathrm{V} \mid \xi_\mathrm{I}, \xi_\mathrm{I}^c, \mathrm{A}]] = \mathrm{E}[\mathrm{V} \mid \xi_\mathrm{I}, \bar{\xi}_\mathrm{I}^c, \mathrm{A}]$, where $\bar{\xi}_\mathrm{I}^c$ is the mean value of ξ_I^c. For example, in the DVT problem with $\xi_\mathrm{I}^c = (a, q, n, E, u)$ we have:

$$\begin{aligned} E_{\xi_I^c}\left[E\left[V \mid \xi_I, \xi_I^c, A = A_1\right]\right] &= E_{a,q,n,E,u}\left[100 \cdot (1 - a)(u - 1)(1 - mn)(1 - p)(1 - E)q\right] \\ &= 100 \cdot (1 - \bar{a})(\bar{u} - 1)(1 - m\bar{n})(1 - p)(1 - \bar{E})\bar{q} \\ &= E\left[V \mid \xi_I, \bar{\xi}_I^c, A = A_1\right] \end{aligned}$$

This special structure allows us to express EVPI more simply as:

$$EVPI(\xi_I) = E_{\xi_I}\left[E\left[V \mid \xi_I, \bar{\xi}_I^c, A^*(\xi_I)\right] - E\left[V \mid \xi_I, \bar{\xi}_I^c, A^*\right]\right]$$

The corresponding "shortcut" Monte-Carlo simulation (MC2) needs only generate random values of the smaller parameter set ξ_I and average over the resulting improvements.

MC2: *Shortcut Monte-Carlo simulation procedure (independent multilinear case)*

1. Repeatedly generate random parameter values ξ_I.
2. For each generated ξ_I
   i. Determine $A^*(\xi_I)$ as the A maximizing $E\left[V \mid \xi_I, \bar{\xi}_I^c, A\right]$.
   ii. Calculate the improvement achieved by using $A^*(\xi_I)$:

$$\text{Improvement} = E\left[V \mid \xi_I, \bar{\xi}_I^c, A^*(\xi_I)\right] - E\left[V \mid \xi_I, \bar{\xi}_I^c, A^*\right]$$

   End For
3. Estimate $EVPI(\xi_I)$ as the average of the calculated improvement values

AN ILLUSTRATIVE EXAMPLE OF SENSITIVITY OVERESTIMATION

Because SA methods based on threshold proximity, probability of a threshold crossing, and entropy ignore the issue of changes in expected payoff, they can dramatically overestimate problem sensitivity. To illustrate this, consider the decision tree in Figure 4.

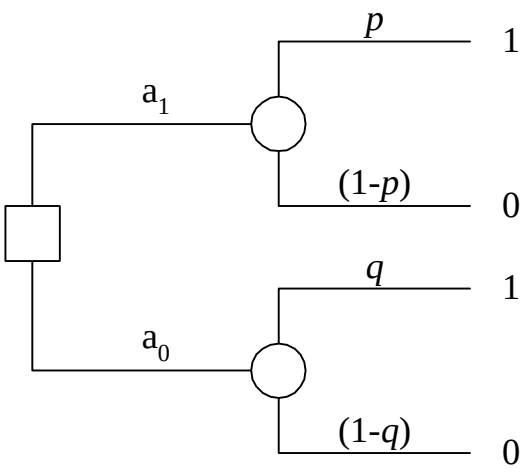

Figure 4. A decision tree example.

If the DM selects alternative a_0, he has a q chance of receiving a payoff of 1 and a $(1-q)$ of receiving a payoff of zero; if he selects alternative a_1, he has a p chance of receiving a payoff of 1 and a $(1-p)$ of receiving a payoff of zero. Let p and q be independent with common distribution symmetric about ½. Some examples of distributions of this type are provided in

Figure 5. We assume that the DM's base values for these parameters are their expected values, so $p_0 = q_0 = ½$. Since both actions have optimal payoff of ½, the decision is a toss-up.

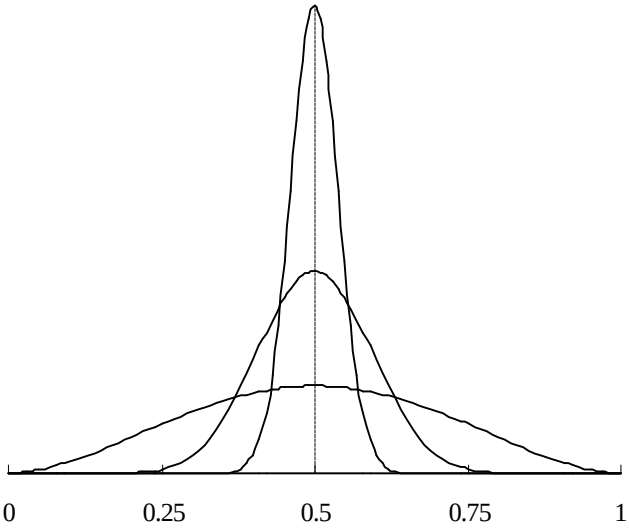

Figure 5. Some sample densities for parameter *p* (and *q*) in Figure 4.

Consider the parameter *p* in Figure 4. We find that threshold proximity SA labels the decision extremely sensitivity to *p*: Since the threshold for *p* is p_0, any perturbation of *p* to the right or left of p_0 will induce the DM to switch between a_0 and a_1. The same conclusion is drawn from probabilistic SA: The probability that a_0 is the optimal alternative is ½. Similar results are obtained under entropy-based SA: Regardless of the common distribution of *p* and *q* (as long as it is symmetric about ½), the value of the mutual information index is $S_{pA} = 1 - 1/2\, ln(2) = 0.279$ (see Appendix A for details). The problem is therefore quite sensitive to *p*, as 27.9% of the instability of the base-optimal alternative can be explained by that parameter. Thus, all three

sensitivity measures classify the problem as sensitive to *p*. (And of course, the identical conclusion obtains for *q*.) Unfortunately for these measures, these conclusions hold *regardless* of how wide or narrow is the common distribution of *p* and *q*.

An EVPI analysis paints quite a different picture. The decision is a toss-up, but some action must be chosen, so suppose it is a_0, If the true value of *p* exceeds its base value ½, then action a_1 should in fact have been chosen; therefore some profit is forgone. The *average foregone profit* (EVPI) resulting when action a_0 is taken depends directly on how great the tendency of *p* is to exceed its base value ½. If *p* is unlikely to be far from ½, then EVPI will be small, and the problem must be labeled as *insensitive* to *p*.

This is where an EVPI analysis demonstrates a clear advantage as a sensitivity measure. If we let $\alpha = \mathrm{E}[p \mid p \geq 1/2]$, then we can show that $\mathrm{EVPI}(p) = (\alpha - 1/2)/2$ (see Appendix B for details). Consequently, as the variability of *p* decreases and α tends toward ½, then EVPI(*p*) tends toward zero. The decision becomes less sensitive to *p* because the marginal benefit of changing from alternative a_0 to a_1 becomes trivial *even though the probability that a_1 offers a higher expected payoff remains constant at 50%*. By virtue of its joint emphasis on the probability of a decision change and the marginal benefit afforded the DM as a consequence of that decision change, EVPI presents a more accurate picture of decision sensitivity.

A COMPARISON OF SENSITIVITY MEASURES

To demonstrate the degree to which sensitivity overestimation can occur in real decision analyses, we investigate probability- and entropy-based SAs taken from the medical decision-making literature and compare their conclusions with those of an EVPI analysis.[d] The problems

examined include: 1) herpes simplex encephalitis, 2) recurrent deep venous thrombosis in pregnancy, and 3) screening for asymptomatic bacteriuria.

Our analyses were performed using a 60 MHz Pentium PC running @Risk, a simulation add-on package for Microsoft Excel. In each of the problems, we assumed all parameters to be independent and used Latin Hypercube sampling to increase the rate of convergence of simulated quantities. We allowed each simulation to run until the percent change in the mean value of all simulated quantities held stable at less than 0.75%.

Herpes Simplex Encephalitis

Using the same distributions and method of Monte Carlo simulation employed by Doubilet *et al.*, we reanalyzed the HSE decision model and performed probabilistic SA, a MII analysis, and an EVPI analysis on a set of six parameters. Although the authors' original model contained 20 parameters, our points regarding the different insights provided by each of these SA methods can be made by limiting our attention to the six probability parameters in the model. Our analytic results are provided in Table 3.

	Pr(Alternative A_i is Preferred)				
Parameter ξ	**A_1=B**	**A_2=NB/T**	**A_3=NB/NT**	**$S_{\xi Y}$**	**EVPI(ξ)**
Pr(Death)	0	1	0	5.80%	0
Pr(Severe Sequalae)	0	1	0	15.26%	0
Pr(Moderate Sequalae)	0	1	0	3.27%	0
Sensitivity of Brain Biopsy	0	1	0	2.74%	0
Specificity of Brain Biopsy	0	1	0	0.06%	0
Overall Pr(HSE)	0	0.9672	0.0328	30.07%	4.75E-05
All parameters jointly	0.2268	0.7515	0.0217	n/a	1.61E-04

Table 3. Probabilistic SA, MII SA, and EVPI analysis of the HSE model. (Analysis is limited to probability parameters only.)

The probabilistic SA clearly illustrates that when the variation is limited to a single probability parameter, only Overall Pr(HSE) induces deviation from the base-optimal alternative, A_2. These results also point to a high degree of parameter interaction, as evidenced by the marked difference in the probability that alternative A_2 is preferred when any one parameter is allowed to vary (96.72%) versus when all six are allowed to vary concurrently (75.15%). Under probabilistic SA, the DM would be motivated to label the problem insensitive to all probability parameters singly, yet sensitive to all probability parameters jointly.

The MII shows that 30% and 15% of the uncertainty surrounding the preferred alternative is explained by the uncertainty surrounding Overall Pr(HSE) and Pr(Severe Sequalae), respectively. This latter value may or may not be sufficiently high to cause a DM to declare the problem sensitive to Pr(Severe Sequalae); however, when considering all the $S_{\xi Y}$ values, there is

a natural breakpoint between the values 30% and 15% and the other values. It is likely, therefore, that a DM would label the problem sensitive to both of these parameters individually.

The EVPI analysis confirms the probabilistic SA result that only the Pr(HSE) induces a decision change; however, the EVPI is so small as to reflect insensitivity to the parameter. The DM's average benefit from switching alternatives in response to perfect foreknowledge on $Pr(\text{HSE})$ is 4.75×10^{-5} – a mere 0.008% improvement over the payoff of 0.5907 expected from the base-optimal alternative. When all parameters are allowed to vary simultaneously, the EVPI for the entire parameter set is only 1.61×10^{-4}, representing an average marginal improvement of less than 0.03% of the baseline expected payoff. This marginal benefit for deviating from the base-optimal NB/T strategy is so trivial that the DM must conclude the model to be insensitive to variations in the entire parameter set. This insensitivity is not captured by the probabilistic and MII-based SAs because these methods ignore payoff changes accompanying decision changes, and it is the triviality of these payoff changes that make the HSE problem insensitive to its parameters.

Deep Venous Thrombosis

This analysis is based upon Critchfield and Willard reanalysis of Klein and Pauker's DVT model.[16] In our investigation of decision sensitivity, we employed the same distributions for input parameters as Critchfield and Willard. Our results are provided in Table 4.

	Pr(Alternative A_i is Preferred)			
Parameter ξ	**A_1=Anticoagulant**	**A_2=No Anticoagulant**	**$S_{\xi Y}$**	**EVPI(ξ)**
Pr(Pulmonary Embolism)	0.9775	0.0225	24.27%	6.71E-03
Utility of Adverse Fetal Outcome	0.9939	0.0061	22.75%	2.54E-03
Pr(Death | Pulmonary Embolism)	0.9997	0.0003	6.14%	2.95E-05
Pr(Fetopathy)	1	0	3.85%	0
Efficacy of Treatment	1	0	0.81%	0
Pr(Death | Maternal Bleeding)	1	0	0.007%	0
Pr(Maternal Bleeding)	1	0	0.001%	0
All parameters jointly	0.9185	0.0815	n/a	5.89E-02

Table 4. Probabilistic SA, MII SA, and EVPI analysis of the DVT problem.

As illustrated in Table 4, the parameters with the highest $S_{\xi A}$ values are the probability of pulmonary embolism and the patient's utility of an adverse fetal outcome. Because both of these parameters contribute highly to the instability of the preferred alternative relative to the other problem parameters, the DM would likely classify the problem sensitive to these two parameters and insensitive to the remaining five. However, in this case, the large $S_{\xi A}$ values are misleading because the entropy of the artificial decision variable A used to normalize the $MI(A|\xi)$ values is itself small. While it is true that the problem is "more sensitive" to variation in the probability of pulmonary embolism than any other parameter, it is a consequence of the normalization process that makes it appear *significantly* sensitive to the parameter. In fact, not only is the problem insensitive to the probability of pulmonary embolism, but to the entire parameter set. This is corroborated by the results of a probabilistic SA, but more dramatically illustrated by the results of an EVPI analysis.

When only the probability of pulmonary embolism is allowed to vary, the probabilistic SA in Table 4 shows that the base-optimal alternative to administer anticoagulants remains optimal 97.75% of the time. This implies the problem's insensitivity to the probability of pulmonary embolism. When all problem parameters are allowed to vary simultaneously, the base-optimal alternative remains optimal 91.85% of the time. In this case, however, problem insensitivity to the entire parameter set is not quite so clear.

From the EVPI analysis provided in Table 4, the expected marginal benefit from perfect foreknowledge on the values of all problem parameters is only 0.0589. This means that, given clairvoyance on all values of all problem parameters, the DM stands to increase the patient's expected payoff by less than one tenth of one percent of the expected payoff of 96.655 he would receive by simply choosing the base-optimal alternative. This marginal benefit is trivial and provides little incentive for the DM to consider the problem sensitive to the values of its input parameters.

Our conclusion is that the DVT problem is insensitive to the set of input parameters, and therefore to all combinations of those parameters. As was the case with the HSE problem, the sensitivity of the DVT problem was overstated by entropy-based and probabilistic SA because of their sole emphasis on the probability of a decision change.

Asymtomatic Bacteriuria

Wadland and Plante performed a decision analysis on whether or not to screen for asymptomatic bacteriuria in pregnancy.[37] Their baseline analysis revealed the optimal alternative to be to screen, which yielded an expected cost of $99.39 per women. Using threshold proximity SA, the authors identified the alternative to be insensitive to all nine of the

parameters they investigated. Based on the authors' commentary, their choice of parameter values, and the results of their SAs, we fit distributions to the nine problem parameters and then used these distributions to perform probabilistic SA, MII, and EVPI analyses.[38] The results of our analyses are provided in Table 5.

			Pr(Alternative A_i is Preferred)			
Parameter ξ	**Base Value**	**Threshold**	**A_1=Screen**	**A_2=Don't Screen**	**$S_{\xi Y}$**	**EVPI(ξ)**
Pr(Asymptomatic Bacteriuria)	0.06	0.02	0.9805	0.0195	2.25%	0.0793
Pr(Pyelonephritis | Asymptomatic Bacteriuria)	0.3	0.13	0.9622	0.0378	2.47%	0.2496
Cost of Screening test	12	26	1	0	0.08%	0
Cost of Therapy for Positive Screening Test	5	210	1	0	0.06%	0
Efficacy of Therapy in Preventing Pyelonephritis	0.8	0.38	0.8769	0.1231	1.51%	1.2345
Cost of Hospital Day	352	77	0.9890	0.0110	1.09%	0.0285
Days of Hospitialization for Pyelonephritis	7	2.2	0.9970	0.0030	0.82%	0.0063
Sensitivity of Screening Test	0.981	0.35	1	0	< 0.005%	0
Specificity of Screening Test	0.938	0.48	0.9872	0.0128	3.08%	0.0164
All parameters jointly	n/a	n/a	0.6898	0.3102	100%	13.1550

Table 5. Results of probabilistic SA, MII SA, and EVPI analysis for the asymptomatic bacteriuria problem.

The probabilistic SA results in Table 5 suggest that the asymptomatic bacteriuria problem is insensitive to (individual) variation in all but one of its parameters. The exception is the efficacy of therapy in preventing pyelonephritis, whose variation can be expected to induce a change in the optimal alternative in 12.31% of cases. When all parameters are allowed to vary concurrently, the base-optimal alternative loses optimality 31.02% of the time, clearly illustrating the sensitivity of the optimal alternative to joint variation in the entire parameter set. The MII

measure can say nothing about total parameter sensitivity. Concerning individual parameters, none of the MII values seem large enough to raise sensitivity concerns, the largest value being only 3.08%.

The results of our EVPI analysis concur with probabilistic SA that the efficacy of pyelonephritis treatment is the single most important parameter. However, because perfect information regarding this parameter's value reduces expected cost by only $1.23 (1.24% of the base optimal cost), it is doubtful that the DM would consider the problem sensitive to the efficacy of treatment of pyelonephritis. On the other hand, the expected marginal benefit from clairvoyance on the values of all problem parameters simultaneously is $13.16, a 13.24% cost reduction. Here is an instance in which EVPI labels a problem sensitive to its entire parameter set but not sensitive to any individual parameter. Clearly parameter interactions are important in this case.

In this example, all three sensitivity measures broadly agree. Probabilistic SA and EVPI label the problem sensitive to its entire parameter set, and they, along with MII, detect no individual parameter sensitivity. The exception is the 12.31% probability of decision change which probabilistic SA attaches to the efficacy of therapy in preventing pyelonephritis. We feel that this is yet another instance of sensitivity overestimation due to inappropriate focus on threshold proximity.

CONCLUSIONS

The stability of the preferred alternative identified by quantitative models to parametric variation has classically been addressed by threshold proximity SA. However, increased

sophistication of decision analytic techniques has altered the nature of the fundamental question driving SA. It is no longer sufficient to ask "How far?"; a more appropriate concern is "How likely and to what effect?"

Two methods of SA, based upon probabilistic threshold analysis and entropy, have been proposed to meet this new sensitivity question and have been employed in the analysis of real decision problems. Both of these methods provide the DM with useful information in the case of single parameter SA, one in terms of the distributional characteristics of each alternative, the other in terms of each parameter's explanation of overall decision instability. In the multiparametric case, however, both methods become unwieldy, and only probabilistic SA can provide the DM with insight into the likelihood that the base-optimal alternative is long term optimal when all parameters vary simultaneously. Neither directly address the issue of the payoff foregone by the DM in cases where some alternative other than the base-optimal offers a higher payoff due to parametric variation.[e]

In order to integrate "How likely?" with "To what effect?", we have proposed that EVPI be employed as a measure of decision sensitivity. We have demonstrated, by means of a simple example, that proximity threshold, probabilistic, and entropy-based SA can dramatically misrepresent decision sensitivity. By our re-examination of three solved problems from the literature, we have illustrated that the probabilistic and entropy-based SAs performed by the authors have in fact overestimated the sensitivity of these problems to their parameters. These results are consistent with the conclusions we present in a forthcoming study of 25 threshold proximity SAs. We attribute this exaggeration of problem sensitivity to the underlying focus on the probability of a decision change without consideration of the associated expected payoff changes.

Our conclusions echo the position enunciated a decade ago by von Winterfeldt and Edwards, who formulated what they called the *flat maxima principle*.[39] In essence, the flat maxima principle states that, upon the elimination of ordinally and cardinally dominated alternatives, moderately sized errors in parametric assessments do not produce large changes in expected payoffs. Von Winterfeldt and Edwards considered the optimal payoff as a function of parameters ξ, and noted for each ξ the maximum of the payoff functions associated with decision alternatives. Because maximization tends to produce concave functions, the optimal payoff function tends to be flat near the baseline value ξ_0 of ξ. Consequently, variation of ξ about ξ_0 will not usually produce dramatic payoff changes, and decision problems will tend to be payoff insensitive to their parameters. While this flatness in the optimal payoff function does not influence probabilistic or entropy-based sensitivity measures, its influence is clearly seen in small marginal benefits accompanying a decision change which, in turn, drive down EVPI values.

Because EVPI represents the DM's average improvement over the base-optimal expected payoff commensurate with an (hypothetical) acquisition of perfect information prior to the point of decision, it is inherently meaningful to medical practitioners trained to think in terms of marginal benefit. The "worth" of the benefit gained is subjective, yet we believe that DMs are better suited to make marginal value judgments than to assess whether 1.3 is "close" to 1.12 or whether an 8.15% chance of a change in the optimal alternative is "high" or not. In this sense, EVPI provides a more natural environment for the DM to frame subjective assessments about the impact of parametric variation on the base-optimal alternative. Although EVPI quantifies the average marginal benefit associated with the gain of *perfect* information, it nonetheless provides a valuable upper limit of the value of information gained from sampling or diagnostic testing.

With regard to implementation, a multiparametric EVPI analysis is tractable (and easily interpreted) for large parameter sets using readily available computing resources and commercially available software. By virtue of our reanalysis of problems from the literature, we have established that multiparametric EVPI analyses are within the range of contemporary DMs and, hopefully, have sparked their interest in performing such analyses on decision problems of their own.

REFERENCES

1. Habbema JDF, Bossuyt PMM, Dippel DWJ, Marshall S, Hilden J. Analysing clinical decision analyses. *Statistics in Medicine*, 1990; 9: 1229-1242.

2. Sonnenberg FA, Roberts MS, Tsevat J, Wong JB, Barry M, Kent DL. Toward a peer review process for medical decision analysis models. *Medical Care*, 1994; 32 (7): JS52-JS64.

3. Plante DA, Piccirillo JF, Sofferman RA. Decision analysis of treatment options in pyriform sinus cancer. *Medical Decision Making*, 1987; 7 (2): 74-83.

4. Felli JC, Hazen GB. Information value and decision sensitivity. *International Engineering Consortium Annual Review of Communications*, 1996; 49: 243-253.

5. O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care*, 1994; 32 (2): 150-163.

6. Buckley JJ. A mathematical programming approach to sensitivity analysis in single-stage decision making. *Decision Sciences*, 1988; 19: 211-232.

7. Evans JR. Sensitivity analysis in decision theory. *Decision Sciences*, 1984; 15: 239-247.

8. Fishburn PC, Murphy AH, Isaacs HH. Sensitivity of decisions to probability estimation errors: a reexamination. *Operations Research*, 1968; 16: 254-267.

9. Isaacs HH. Sensitivity of decisions to probability estimation errors. *Operations Research*, 1963; 11: 536-552.

10. Starr MK. A discussion of some normative criteria for decision-making under uncertainty. *Industrial Management Review*, 1966; 8: 71-78.

11. Rios Insua D, French S. A framework for sensitivity analysis in discrete multi-objective decision-making. *European Journal of Operational Research*, 1991; 54: 176-190.

12. Schneller GO, Sphicas GP. On sensitivity analysis in decision theory. *Decision Sciences*, 1985; 16: 399-409.

13. Brown RV. Assessment uncertainty technology for making and defending risky decisions. *Journal of Behavioral Decision Making*, 1990; 3 (3): 213-228.

14. Brown RV, Ulvila JW. 'Does a reactor need a safety backfit? Case study on communicating decision and risk analysis information to managers'. *Risk Analysis*, 1988; 8 (2): 271-282.

15. Connelly DP, Willard KE. Monte Carlo Simulation and the Clinical Laboratory. *Archives of Pathology and Laboratory Medicine*, 1989; 113 (7): 750-757.

16. Critchfield GG, Willard KE. Probabilistic analysis of decision trees using Monte Carlo simulation. *Medical Decision Making*, 1986; 6 (2): 85-92.

17. Critchfield GG, Willard KE, Connelly DP. Probabilistic analysis methods for general decision models. *Computers and Biomedical Research*, 1986; 19: 254-265.

18. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic Sensitivity Analysis Using Monte Carlo Simulation. *Medical Decision Making*, 1985; 5 (2): 157-177.

19. Dittus RS, Klein RW. Simulating medical decision trees with random variable parameters. *Proceedings of the 1992 Winter Simulation Conference.* Swain JJ, Goldsman D, Crain RC, Wilson JR. (Eds.), 1992: 1050-1056.

20. Dittus RS, Roberts SD, Wilson JR. Quantifying uncertainty in medical decisions. *Journal of the American College of Cardiology*, 1989; 14(3): 23A-28A.

21. Shannon CE. A Mathematical Theory of Communication. *Bell System Technical Journal*, 1948; 27: 379-423.

22. Genrich JF, Sommer FS. Novel approach to sensitivity analysis. *Journal of Petroleum Technology*, 1989: 930-932, 983-985.

23. Soofi ES. Capturing the intangible concept of information. *Journal of the American Statistical Association*, 1994; 89 (428): 1243-1254.

24. Klein K, Pauker SG. Recurrent deep venous thrombosis in pPregnancy: analysis of the risks and benefits of anticoagulation. *Medical Decision Making*, 1981; 1 (2): 181-202.

25. Howard RA. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 1966; SCC-2 (1): 22-26.

26. Gorry GA, Barnett GO. Experience with a model of sequential diagnosis. *Computers and Biomedical Research*, 1968; 1: 490-507.

27. Gorry GA, Kassirer J, Essig A, Schwartz W. Decision analysis as the basis for computer-aided management of acute renal failure. *American Journal of Medicine*, 1973; 55: 473-484.

28. Heckerman D, Horvitz E, Middleton B. An approximate nonmyopic computation for value of information. *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, 1991: 135-141.

29. Heckerman D, Horvitz E, Nathwani B. Toward normative expert systems: the Pathfinder project. Technical Report KSL-90-08, Medical Computer Science Group, Section on Medical Informatics, Stanford University, Stanford, CA, 1990.

30. Feinstein AR. Clinical Epidemiology, The Architecture of Clinical Research. Philadelphia: W. B. Saunders Company, 1985.

31. Spiegelhalter DJ, Freedman LS, Parmer MKB. Bayesian approaches to randomized trials. *Journal of the Royal Statistical Association*, 1994; *(A)* 157: 357-416.

32. Spiegelhalter DJ, Freedman LS. Bayesian approaches to clinical trials. Bayesian Statistics 3. Bernardo JM, Degroot MH, Lindley DV, Smith AFM (Eds.). New York: Oxford University Press, 1988: 453-477.

33. Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. *Journal of Clinical Epidemiology*, 1996; 49 (11): 1215-1219.

34. Redelmeier DA, Guyatt GH, Goldstein RS. On the debate over methods for estimating the clinically important difference. *Journal of Clinical Epidemiology*, 1996; 49 (11): 1223-1224.

35. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements. *Archives of Internal Medicine*, 1993; 153: 1337-1342.

36. Wright JG. The minimal important difference: Who's to say what is important? *Journal of Clinical Epidemiology*, 1996; 49 (11):1221-1222.

37. Wadland WC, Plante DA. Screening for asymptomatic bacteriuria in pregnancy *Journal of Family Practice*, 1989; 29 (4):372-376.

38. Felli JC, Hazen GB. Do sensitivity analyses really capture problem sensitivity? An empirical analysis based on information value. Working Paper. Defense Resources Management Institute, Naval Postgraduate School, Monterey, CA, 1997.

39. Von Winterfeldt D, Edwards W. Decision Analysis and Behavioral Research. Cambridge University Press, 1986.

40. Bernardo JM, Smith AFM. Bayesian Theory. West Sussex, England: John Wiley & Sons, Ltd., 1994.

APPENDIX A

If we consider the decision tree presented in Figure 3, we note that alternative a_0 gives a q chance at payoff 1 and alternative a_1 gives a p chance at payoff 1. The parameters p and q are uncertain. This results in the following expected values:

$$\mathrm{E}\left[\mathrm{V}_{a_0} \mid p,q\right] = q \qquad \mathrm{E}\left[\mathrm{V}_{a_1} \mid p,q\right] = p \qquad \mathrm{E}\left[\mathrm{V}_{a_0}\right] = \mathrm{E}[q] = \bar{q} \qquad \mathrm{E}\left[\mathrm{V}_{a_1}\right] = \mathrm{E}[p] = \bar{p}$$

Assume that p and q are independent, identically distributed random variables with symmetric densities about the common mode ½. Then the joint density of (p,q) is radially symmetric about (½,½) and $Pr(p \geq q) = ½$. We can now calculate $\mathrm{S}_{p\mathrm{A}}$.

Let A be the optimal action as a function of p and q and denote $Pr(\mathrm{A} = \mathrm{a}_k)$ by $Pr(\mathrm{a}_k)$ for $\mathrm{k} = 0,1$ Then $Pr(\mathrm{a}_1) = Pr(\mathrm{E}\left[\mathrm{V}_{a_1} \mid p,q\right] \geq \mathrm{E}\left[\mathrm{V}_{a_0} \mid p,q\right]) = 1/2$ and we have:

$$\begin{aligned}
\mathrm{MI}(\mathrm{A} \mid p) &= \mathrm{E}_{\mathrm{A},p}\left[\log_2\left(\frac{Pr(\mathrm{A} \mid p)}{Pr(\mathrm{A})}\right)\right] \\
&= \mathrm{E}_p\left[\mathrm{E}_{\mathrm{A}}\left[\log_2\left(\frac{Pr(\mathrm{A} \mid p)}{Pr(\mathrm{A})}\right) \,\middle|\, p\right]\right] \\
&= \left(\frac{1}{\ln(2)}\right)\mathrm{E}_p\left[\mathrm{E}_{\mathrm{A}}\left[\ln\left(\frac{Pr(\mathrm{A} \mid p)}{Pr(\mathrm{A})}\right) \,\middle|\, p\right]\right] \\
&= \left(\frac{1}{\ln(2)}\right)\mathrm{E}_p\left[Pr(\mathrm{a}_0 \mid p)\ln\left(\frac{Pr(\mathrm{a}_0 \mid p)}{Pr(\mathrm{a}_0)}\right) + Pr(\mathrm{a}_1 \mid p)\ln\left(\frac{Pr(\mathrm{a}_1 \mid p)}{Pr(\mathrm{a}_1)}\right)\right] \\
&= \left(\frac{1}{\ln(2)}\right)\mathrm{E}_p\left[(1-\mathrm{F}(p))\ln\left(\frac{1-\mathrm{F}(p)}{1/2}\right) + \mathrm{F}(p)\ln\left(\frac{\mathrm{F}(p)}{1/2}\right)\right] \\
&= \left(\frac{1}{\ln(2)}\right)\mathrm{E}_p\left[(1-\mathrm{F}(p))\ln(1-\mathrm{F}(p)) + \mathrm{F}(p)\ln(\mathrm{F}(p))\right] + \ln(2)
\end{aligned}$$

Here, we are letting F designate the common cumulative density function of *p* and *q*. By a well-known result, the random variables $F(p)$ and $1-F(p)$ have uniform(0,1) distributions. Therefore, for $U \sim \text{Uniform}(0,1)$, we have:

$$MI(A \mid p) = \left(\frac{1}{\ln(2)}\right)\left(2 \cdot E_U[U \ln(U)] + \ln(2)\right) = 1 + \left(\frac{2 \cdot E_U[U \ln(U)]}{\ln(2)}\right)$$

Integration yields $E_U[U \ln(U)] = -1/4$, therefore:

$$MI(A \mid p) = 1 + \left(\frac{-1}{2\ln(2)}\right)$$

The normalization factor for S_{pA} is the entropy of A, or $MI(A \mid A)$. We calculate this quantity to be:

$$MI(A \mid A) = E_A\left(\log_2\left(\frac{1}{Pr(A)}\right)\right) = \left(\frac{1}{\ln(2)}\right)\left(Pr(a_0)\ln\left(\frac{1}{Pr(a_0)}\right) + Pr(a_1)\ln\left(\frac{1}{Pr(a_1)}\right)\right)$$

Since $Pr(p \geq q) = ½$, this implies that $Pr(a_0) = Pr(a_1) = ½$. Therefore $MI(A \mid A) = 1$ and we have that:

$$S_{pA} = \frac{MI(A \mid p)}{MI(A \mid A)} = 1 + \left(\frac{-1}{2\ln(2)}\right)$$

which implies that *p* contributes 27.87% of the instability of the base-optimal alternative regardless of the variability of its distribution about ½.

APPENDIX B

Consider the decision tree presented in Figure 3 and recall that a_0 is the base-optimal alternative and that $E[V_{a_0} \mid p,q] = q$ and $E[V_{a_1} \mid p,q] = p$. Since we are concerned with determining EVPI(p), all other parameters are held fixed at base values, so $E[V_{a_0} \mid p] = \bar{q}$ and $E[V_{a_1} \mid p] = p$. We are now in a position to calculate EVPI(p).

$$\begin{aligned} \text{EVPI}(p) &= E_p\left[max_a\, E[V_a \mid p] - E[V_{a_0} \mid p]\right] \\ &= E_p\left[max_a\, E[V_a \mid p] - \bar{q}\right] \\ &= E_p\left[max(p - \bar{q}, 0)\right] \end{aligned}$$

The term $p - \bar{q}$ is positive only when $p > \bar{q} = ½$. Thus, for F(p) the cumulative density function of p, we have that:

$$\begin{aligned} \text{EVPI}(p) &= \int_{1/2}^{1} (p - 1/2)\, dF(p) \\ &= \int_{1/2}^{1} p\, dF(p) - \frac{(1 - F(1/2))}{2} \\ &= (1 - F(1/2))\left(\int_{1/2}^{1} p\, dF(p \mid p \geq 1/2) - 1/2\right) \end{aligned}$$

Letting $\alpha = \int_{1/2}^{1} p\, dF(p \mid p \geq 1/2)$, the expected value of p given that $p \geq ½$, and recognizing that p's symmetry about its mode of ½ implies that $F(1/2) = 1 - F(1/2) = ½$, we have that:

$$\text{EVPI}(p) = (1 - 1/2)(\alpha - 1/2) = \frac{(\alpha - 1/2)}{2}$$

ENDNOTES

[a] The base value of a parameter is the value the DM uses in the initial, baseline analysis of his decision problem.

[b] The logit transformation is $logit(x) = log(x/(1-x))$. This assumes that the parameters values range over [0,1]. For parameters with values ranging over an arbitrary interval [α, β], the transformation is $logit((x-\alpha)/(\beta-x))$.

[c] The authors use a single parameter to designate the Pr(Fatal Bleed), although they provide information regarding the P(Bleed) and Pr(Fatal Bleed | Bleed). We use the latter two probabilities as parameters.

[d] As we will present a comparison of EVPI-based and threshold proximity SA in a forthcoming study, we limit our attention in this work to probabilistic and entropy-based SA.

[e] Mutual information MI(A|ξ) can in fact be interpreted as addressing foregone payoff, but in a modified decision problem. The quantity MI(A|ξ) is actually equal to EVPI(x) in a modified decision problem in which the objective is to choose an estimate of the entire probability distribution of A, where the estimate is valued according to a logarithmic scoring rule.[40] One might, however, question the reasonableness of replacing the original decision problem by a new one in which alternatives and payoffs are only weakly related to the alternatives and payoffs in the original problem. Why calculate a form of EVPI(ξ) in an artificially modified decision

problem when one can calculate a more meaningful EVPI(ξ) in the original problem (and usually at less computational cost)?