

# Uncertainty and the value of diagnostic information, with application to axillary lymph node dissection in breast cancer

Giovanni Parmigiani<sup>\*,†</sup>

*The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore,  
MD, 21205, U.S.A.*

## SUMMARY

In clinical decision making, it is common to ask whether, and how much, a diagnostic procedure is contributing to subsequent treatment decisions. Statistically, quantification of the value of the information provided by a diagnostic procedure can be carried out using decision trees with multiple decision points, representing both the diagnostic test and the subsequent treatments that may depend on the test's results. This article investigates probabilistic sensitivity analysis approaches for exploring and communicating parameter uncertainty in such decision trees. Complexities arise because uncertainty about a model's inputs determines uncertainty about optimal decisions at all decision nodes of a tree. We present the expected utility solution strategy for multistage decision problems in the presence of uncertainty on input parameters, propose a set of graphical displays and summarization tools for probabilistic sensitivity analysis in multistage decision trees, and provide an application to axillary lymph node dissection in breast cancer. Copyright © 2004 John Wiley & Sons, Ltd.

**KEY WORDS:** probabilistic sensitivity analysis; decision trees; sequential decision making; expected utility; early breast cancer treatment

## 1. INTRODUCTION

Assessing uncertainty about decision models is emerging as a critical component of sound medical decision making. Probabilistic sensitivity analysis (PSA) provides a practical, flexible, and well founded approach for representing uncertainty about model inputs and for exploring its implications [1–4]. Monte Carlo simulation permits implementation of PSA in complex

---

\*Correspondence to: Giovanni Parmigiani, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, MD 21205, U.S.A.

†E-mail: gp@jhu.edu

Contract/grant sponsor: NCI; contract/grant numbers: P50CA88843, P50CA62924-05, 5P30CA06973-39  
Contract/grant sponsor: Hecht fund

multiparameter models, while accommodating for non-standard distributions and dependence among input parameters. PSA can be linked to statistical inference procedures, for example by using output from Markov Chain Monte Carlo exploration of posterior distributions [5] as the input of a PSA [6], or via frequentist resampling techniques [7].

With rare exceptions [8], investigation of PSA in decision trees has focused on trees including a single, although possibly multi-branched, decision node. This paper presents tools for probabilistic sensitivity analysis in multistage decision problems, that is problems in which several decisions are made at different points in time, on the basis of different information. An important example in medicine is deciding whether to perform a diagnostic or staging procedure to support subsequent treatment decisions.

In multistage problems, the decision about whether or not to gather certain information has impact on the overall utility via its effect of future decisions. The worthiness of the information can then be quantified by comparing the expected utility of the optimal solutions with and without it [9–12, 8]. The standard version of this quantification, also called ‘value of information’ analysis, relies on knowing with certainty the optimal decision functions downstream. Uncertainty about model inputs, however, determines uncertainty about optimal decisions at all stages of a multistage problem. Because of this, in a PSA of the value of diagnostic information it is important separately to quantify and convey the uncertainty deriving directly from randomness in the input, and the uncertainty deriving from randomness in the optimal future decisions. In clinical settings, this distinction becomes critical, as will be seen, because it has implication for the appropriateness of a treatment.

This article is motivated by the controversy surrounding axillary lymph node dissection (ALND), or the surgical removal of the axillary lymph nodes, a common procedure in early breast cancer patients. ALND is routinely performed as part of a mastectomy, but it is elective in breast conservation patients, for whom it is a distinct procedure from the main surgery. It is used primarily to guide the choice of systemic adjuvant therapy, to provide local control of cancer and to give patients prognostic information for personal use. ALND results in morbidity and minor but permanent problems [13] and requires a significant use of resources [14].

Recent randomized trials have questioned the long held belief that the relative benefits of hormonal versus cytotoxic adjuvant therapies depend on nodal status [15, 16]. The prognostic information conveyed by the results of an ALND, however, may still be useful in decision making, as a patient with a worse prognosis may be more inclined to undergo a more aggressive form of treatment. The role of ALND in local control is also in question. Axillary irradiation (XRT) appears to provide comparable local control and survival benefits in patients with clinically uninvolved lymph nodes [17].

This complex scenario has led to disagreement on the routine use of ALND [18–22], and encouraged the use of decision analyses to evaluate its worthiness [23, 24]. The sources of disagreement reside in part in the relatively large uncertainties that still involve the critical input parameters. This makes ALND a good example for studying the effect of uncertainty on measuring the value of diagnostic information.

Section 2 presents the solution of multistage decision problems in the presence of parameter uncertainty, discusses the type of Monte Carlo simulation that is necessary for probabilistic sensitivity analyses, and proposes a set of summaries and graphical displays to help understanding and communicating uncertainty about the value of diagnostic information. Section 3 applies the methodology of Section 2 to evaluating the diagnostic value of ALND.

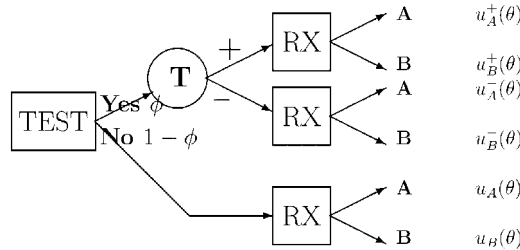


Figure 1. Decision tree for the generic binary diagnostic test problem. Leaves are utilities conditional on knowledge of the model parameters  $\theta$ .

## 2. GENERAL CONCEPTS

### 2.1. Setting

This section introduces basic concepts and definitions for probabilistic sensitivity analysis of multistage trees, using the context and language of diagnostic testing. For conciseness of notation consider the binary, two-stage case. A richer two-stage tree, with multiple treatment arms and test outcomes, is illustrated in the application of Section 3. Figure 1 summarizes the tree structure and the notation. There are two treatment options: A and B, and a diagnostic test with binary outcome  $T$ , either positive ( $T+$ ) or negative ( $T-$ ). The probability of a positive test result for the cohort under consideration is  $\phi$ , that is  $p(T+|\phi) = \phi$ . The parameters describing the medical history of patients after treatment are collectively denoted by the vector  $\theta$ . For example,  $\theta$  could represent the vector of parameters of a survival model incorporating an effect for treatment. Utilities associated with tree branches are indicated as in Figure 1, with the notation emphasizing dependence on the parameters as well as treatment and test outcome.

To simplify the presentation and focus on the diagnostic value of the test, we assume that performing the test does not have any direct effect on the utilities of the treatment branches. Most non-invasive tests based on blood samples or simple tissue extraction are such. Also, we do not consider the cost of the test explicitly. Under these assumptions, when the parameters  $\phi$  and  $\theta$  are known, utilities for the ‘no test’ branch are related to the utilities for the test branch by the relationship

$$u_{RX}(\theta) = \phi u_{RX}^+(\theta) + (1 - \phi) u_{RX}^-(\theta)$$

where RX is either A or B. Then we can fold back the tree and compute the difference in utility between the two arms, which is

$$\Delta^*(\phi, \theta) = \phi \max\{u_A^+(\theta), u_B^+(\theta)\} + (1 - \phi) \max\{u_A^-(\theta), u_B^-(\theta)\} - \max\{u_A(\theta), u_B(\theta)\} \quad (1)$$

The treatment strategy maximizing utility conditional on knowledge of the true parameter values will be called *true optimal strategy*.

### 2.2. Solving the tree when parameters are unknown

In practice, input parameters  $\phi$  and  $\theta$  are often unknown. The set of possible values of  $\phi$  will typically be the interval  $(0, 1)$ , while the set of possible values of  $\theta$  is more general and will

be denoted by  $\Theta$ . Information about all unknowns can be represented by a joint probability distribution  $p(T, \phi, \theta)$ , representing the decision analyst's uncertainty about  $(T, \phi, \theta)$ , formed based on the studies and expert input used in developing the model. This distribution is conveniently factored as

$$p(T, \phi, \theta) = p(\phi)p(T|\phi)p(\theta|T, \phi) \quad (2)$$

The distribution  $p(T|\phi)$  represents variation in test outcomes over individuals in the cohort, while the distributions  $p(\phi)$  and  $p(\theta|T, \phi)$  represent the uncertainty about the model parameters. In a multistage situation it is critical to consider the joint distribution of parameters and test results, as test results contain information about parameters that needs to be factored in the analysis to obtain coherent inferences.

When parameters are unknown, decision making at the treatment stage in the 'test' branch proceeds by evaluating expected utilities, averaging over unknown parameters conditional on the test's outcome. To this end, we use the distribution  $p(\phi, \theta|T) = p(\phi|T)p(\theta|T, \phi)$ . The term  $p(\phi|T)$  can be obtained using Bayes rule from the terms  $p(\phi)$  and  $p(T|\phi)$  in our original factorization (2). Specifically, we evaluate

$$\bar{u}_{RX}^T = \int_0^1 \int_{\Theta} p(\phi|T)p(\theta|T, \phi)u_{RX}^T(\theta) d\theta d\phi \quad (3)$$

where RX is either A or B, and  $T$  is either  $+$  or  $-$ . These quantities can be used to find the treatment strategy that maximizes expected utility for each test outcome, which will be called *current optimal strategy*. Unlike the true optimal strategy, the current optimal strategy is not conditional on knowing the true parameter values. Rather, it represents the best that can be done, in terms of expected utility, given the information available at the time of the decision. It is consistent with the Bayesian principle that decisions at any point should be made by conditioning on all quantities that are known and averaging over all those that are unknown.

In the 'no test' branch, treatment arms are compared based on expected utilities, averaging over both parameters and test outcomes. To this end we use the joint distribution in equation (2). Specifically, we evaluate

$$\begin{aligned} \bar{u}_{RX} &= \int_0^1 \int_{\Theta} \sum_T p(\phi)p(T|\phi)p(\theta|T, \phi)u_{RX}^T(\theta) d\theta d\phi \\ &= \int_0^1 \int_{\Theta} p(\phi)[\phi p(\theta|T+, \phi)u_{RX}^+(\theta) + (1 - \phi)p(\theta|T-, \phi)u_{RX}^-(\theta)] d\theta d\phi \end{aligned} \quad (4)$$

There are eight possible configurations of solutions for the three decision nodes at the treatment stage of the tree, depending on the rank order of the  $\bar{u}$ 's. For the purpose of our analysis we only need to distinguish two types: configurations in which the current optimal treatment is always the same, and configurations in which it varies with the test result. Without losing generality we can focus on two specific cases:

*Case A:* A is always preferred;

*Case B:* B is preferred if the test is negative and if it is not performed.

In case B the 'test' arm is preferred to the 'no test' arm if the expected utility of the 'test' arm, call it  $\bar{u}$ , is greater than the expected utility  $\bar{u}_B$  for the 'no test' arm. Computationally,

$\bar{u}$  is obtained by averaging over all unknowns, and assuming that the current optimal strategy will be followed once the treatment node is reached. Therefore

$$\begin{aligned} \bar{u} &= \int_0^1 \int_{\Theta} \sum_T p(\phi) p(T|\phi) p(\theta|T, \phi) u_{\text{current optimal}}^T(\theta) d\theta d\phi \\ &= \int_0^1 \int_{\Theta} p(\phi) [ \phi p(\theta|T+, \phi) u_A^+(\theta) + (1 - \phi) p(\theta|T-, \phi) u_B^-(\theta) ] d\theta d\phi \end{aligned} \tag{5}$$

The expected value of diagnostic information is the difference

$$\bar{\Delta} = \bar{u} - \bar{u}_B = \int_0^1 \int_{\Theta} p(\phi) p(\theta|T+, \phi) \phi [ u_A^+(\theta) - u_B^+(\theta) ] d\theta d\phi \tag{6}$$

This procedure provides a general way of evaluating the expected value of diagnostic information in the presence of parameter uncertainty. Comparing  $\bar{\Delta}$  to the cost of the test gives a cost-effectiveness evaluation.

We assumed that there is no loss of utility from the test. Therefore, when parameters are known, the information provided by the test may either increase the average utility in the cohort, if therapy may be changed by the result or the test, or leave it unchanged if the strategy is also left unchanged. Mathematically, this is a simple corollary of the fact that  $\Delta^*(\phi, \theta)$  cannot, by construction, be negative. The connection between this mathematical property of sequential decision making and the value of information has been part of the foundations of Bayesian decision theory since its inception [9, 12, 25].

When the parameters are not known, it is still the case that the information provided by the test will not decrease the average utility in the cohort, as long as the distribution of the unknown parameters is correctly updated upon knowing the results of the test, and the test has no direct effect on utility. In two-stage binary trees for diagnostic tests, this is true if  $\bar{\Delta} \geq 0$  in cases A and B, as all other proceed along the same lines. In case A,  $\bar{\Delta} = 0$ . In case B, because therapy A is preferred to therapy B if the test is positive, we have

$$\bar{u}_A^+ - \bar{u}_B^+ = \int_0^1 \int_{\Theta} p(\phi|T+) p(\theta|T+, \phi) (u_A^+(\theta) - u_B^+(\theta)) d\theta d\phi > 0$$

Substituting  $p(\phi|T+) = p(\phi) p(T+|\phi) / p(T+) = p(\phi) \phi / p(T+)$  and canceling the term  $p(T+)$  gives

$$\int_0^1 \int_{\Theta} p(\phi) \phi p(\theta|T+, \phi) (u_A^+(\theta) - u_B^+(\theta)) d\theta d\phi > 0$$

that is  $\bar{\Delta} > 0$ . A similar argument holds for two-stage non-binary trees.

### 2.3. Monte Carlo Calculations

To solve the tree and implement a probabilistic sensitivity analysis by Monte Carlo methods, we need to draw  $M$  samples

$$(\phi_1, \theta_1^+, \theta_1^-), \dots, (\phi_M, \theta_M^+, \theta_M^-) \tag{7}$$

where each  $\phi_m$ ,  $m = 1, \dots, M$ , is drawn from  $p(\phi)$ , each  $\theta_m^+$  from  $p(\theta|T+, \phi_m)$ , and each  $\theta_m^-$  from  $p(\theta|T-, \phi_m)$ . Samples in (7) do not need to be independent. For example, MCMC output after convergence is appropriate. We assume also that the Monte Carlo sample size  $M$  is large enough that all the expectations necessary for decision analysis are closely approximated by the corresponding Monte Carlo averages.

Using (7) we can evaluate all  $\bar{u}$ 's, determine the current optimal strategy, and evaluate the expected value of diagnostic information  $\bar{\Delta}$ . Specifically, (3) is evaluated by

$$\bar{u}_{\text{RX}}^T \approx \frac{1}{M} \sum_{m=1}^M u_{\text{RX}}^T(\theta_m^T) \quad (8)$$

where again RX is either A or B, and  $T$  is either  $+$  or  $-$ . Similarly, for the no-test arm, (4) is evaluated by

$$\bar{u}_{\text{RX}} \approx \frac{1}{M} \sum_{m=1}^M (\phi_m u_{\text{RX}}^+(\theta_m^+) + (1 - \phi_m) u_{\text{RX}}^-(\theta_m^-)) \quad (9)$$

In case B the expected value of information (6) is evaluated by

$$\bar{\Delta} \approx \frac{1}{M} \sum_{m=1}^M \Delta(\phi_m, \theta_m^+) = \frac{1}{M} \sum_{m=1}^M \phi_m (u_{\text{A}}^+(\theta_m^+) - u_{\text{B}}^+(\theta_m^+)) \quad (10)$$

A critical aspect of Monte Carlo approaches to multistage problems is that multiple (in this case two) sets of samples of  $\theta$  are necessary, one for each possible value of the test outcome.

Because the  $\bar{u}$ 's and  $\bar{\Delta}$  are non-linear functions of the parameters  $\phi$  and  $\theta$ , it is recommended to solve the tree by evaluating (3), (4), and (6), rather than using best estimates of  $\phi$  and  $\theta$  in a deterministic way.

#### 2.4. Measuring uncertainty about the value of diagnostic information

In a probabilistic sensitivity analysis, it is interesting to study how unknown parameters affect the difference between the two arms. This difference depends on parameters as well as the treatment strategy. This suggests to consider two variants:  $\Delta^*(\phi, \theta)$ , as defined in (1), assumes that patients in the 'test' arm will be treated according to the true optimal strategy for the specified  $\phi$  and  $\theta$ ; and  $\Delta(\phi, \theta)$  assumes that patients in the 'test' arm will be treated according to the current optimal strategy, independent of  $\phi$  and  $\theta$ . In case A,  $\Delta$  is always 0 for all values  $\phi$  and  $\theta$ . In case B, we have

$$\Delta(\phi, \theta) = \phi(u_{\text{A}}^+(\theta) - u_{\text{B}}^+(\theta)) \quad (11)$$

The functions  $\Delta$  and  $\Delta^*$  measure different aspects of the value of diagnostic information as a function of the unknown parameters. Understanding uncertainty about the value of the diagnostic test requires joint consideration of both. Our suggestions for a basic description of uncertainty is to (a) graph the joint distribution of  $\Delta$  and  $\Delta^*$ , and (b) compute the three summary quantities

$$\text{PC} = P(\Delta^* = \Delta)$$

$$\text{PD} = P(\Delta^* > 0)$$

$$\text{PE} = P(\Delta > 0)$$

Naturally, other quantities may also be of interest, but these can provide a minimal core set.

PC is the probability that the current best and the true best treatment strategies are the same, or, in other words, that the current treatment strategy is correct. Because the value of the diagnostic test depends on the use that will be made of the information, it is critical to quantify uncertainty on its appropriateness. However, this alone is not sufficient. In problems with multiple options, there are situations in which the current best is only correct for a small fraction of the simulated values, but the true best is very similar to it over a broad range of simulated values.

PD is the probability that the test would have diagnostic value if the true parameter values were known. This quantity reflects the confidence with which we may say that treatment should be based on this test, no matter what the true strategy is. Computation of  $\Delta^*$  is predicated on the impossible requirement of knowing the true parameter values and underestimates uncertainty compared to what can be expected in a clinical implementation of the solution. Therefore, PD is of interest for technology assessment and research planning rather than clinical decisions to be made in the immediate future.

PE is the probability that the current best strategy does more good than harm, or, we could say, is ethical. This is a quantity of direct clinical interest that measures the confidence with which we can recommend the test and subsequent therapy decision as a package. It is based solely on the knowledge currently available. It does not, however, fully answer the question of whether the test is of value. First,  $\Delta$  is identically zero in Case A even though there may be substantial uncertainty about whether all patients should get treatment A. Also  $\Delta$  could be positive in cases in which the test should not be used, if the true parameters rank the treatments in the 'no test' arm in the opposite way as the current optimal strategy.

### 3. AXILLARY LYMPH NODE DISSECTION IN BREAST CANCER

#### 3.1. Model assumptions and data sources

This section uses the model structure of Parmigiani and colleagues [24], updating the values of the parameters wherever appropriate, and developing an assessment of uncertainty for each. Consider a cohort of patients who are 60 years old at diagnosis, have an ER positive primary tumor of size between 0.5 and 1 cm, have a negative clinical exam of the axilla, and have elected to receive breast conservation therapy. Nodal status information resulting from ALND is classified in four categories, as described in Figure 2. Also consider four alternatives for adjuvant therapy: Tamoxifen (T); combination chemotherapy (CT); combination chemotherapy followed by Tamoxifen (CT+T); and no adjuvant systemic therapy (noRX). All suitable adjuvant combination chemotherapy regimens are assumed to be equivalent, both in terms of long-term outcome and effects on quality of life. In practice, differences may exist, but are likely to be small and it is unlikely that the relative risk reduction would be affected by nodal status.

Solving the tree requires (a) a probability distribution of nodal involvement, the vector equivalent of  $\phi$ , and (b) a probability distribution of survival times by nodal involvement category, adjuvant therapy, and ALND status, whose parameters correspond to  $\theta$ . Data sources were chosen seeking generalizable registry or cohort data for describing the natural history

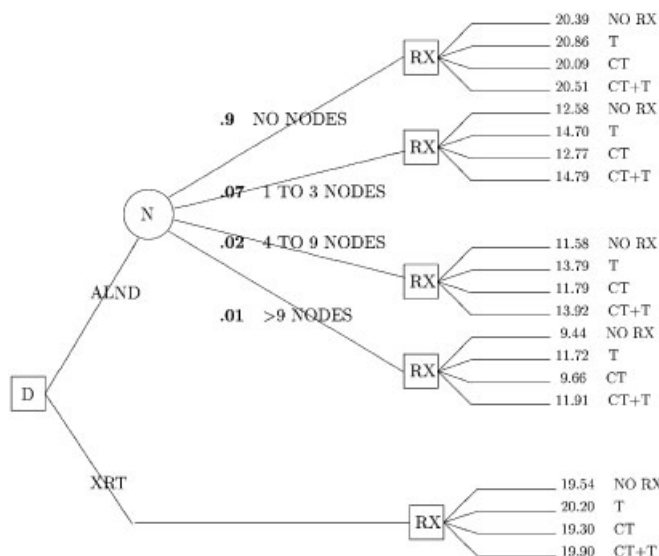


Figure 2. Decision tree for the ALND application. Values attached to treatment branches are averages of terminal utilities in QALYs. Values on ALND outcome branches are average probabilities of nodal involvement categories 'NODES' refers to lymphnodes found to be positive. The current optimal strategy is T if XRT, and if ALND and no positive nodes are found, and CT+T otherwise.

of disease under standard treatment, and meta-analyses for describing effects of treatments. All inputs are based on primary data or published studies. Probability distributions describing parameter uncertainty reflect inferential uncertainty, with minimal additional subjective input. Real valued parameters are assigned normal distributions, while probabilities and percentage of risk reduction are assigned beta distributions, and probability vectors are assigned Dirichlet distributions. Parameters of these distributions were based either on the reported estimates and standard deviations in the original studies, or on the estimates obtained directly from source data.

The probability distribution of the number of positive nodes for the given tumor size, age, and ER status is based on two components. First we estimate probabilities of nodal involvement by category using data from the NCI SEER cancer registries [26], considering both clinically negative and clinically positive cases. Estimates of the probabilities of true-negative and false-negative physical examinations [22] permit conditioning on a clinically negative axilla, via Bayes' rule. All parameters estimated using SEER data were handled using Dirichlet posterior distributions obtained from uniform priors [27]. Because of the very large sample size of the SEER registry, sensitivity to the choice of priors is minimal in this case.

Survival distributions were also derived by combining several elements. Survival of patients with positive nodes receiving chemotherapy was estimated using a multivariate Cox regression model with age, ER status, primary tumor size, and nodal involvement as prognostic variables. Patient level data for this purpose was drawn from four studies conducted within the CALGB cooperative group. Study numbers are 7581, 8082 and 8541 [28–30]. The simulation relative



Table I. Effects of adjuvant treatments. Entries are mortality reductions for alternative adjuvant therapies, compared to no adjuvant treatments. Values in parentheses are the corresponding standard deviations. All values refer to women older than 50 years of age.

	ER+ (per cent)	ER- (per cent)
No RX	0	0
T	32 (10)	0
CT	9 (5)	17 (6)
CT+T	40 (11)	11 (4)

to the Cox proportional hazard component of the model, kept the baseline hazard fixed, and drew only the coefficients at random.

The maximum followup in these CALGB studies is 19 years. Rates beyond 19 years were extrapolated based on a constant hazard rate (set to the average hazard rate in years 15–19). The hazard rate in the first 19 years is not constant, accounting for greater hazard in the first few months following adjuvant treatment. CALGB trials involved only node-positive women. To derive a survival curve for node-negative women, the following strategy was followed: (a) For node-negative women with primary tumors larger than 1 cm, results from the EBCTCG overview [15] were used to estimate an adjustment ratio between node-positive and node negative patients receiving chemotherapy. This ratio is 0.42 for the first 4 years and 0.61 for the remainder [15]. (b) For node-negative women with primary tumor smaller than 1 cm, the results of Rosen and colleagues [31] were used to derive an adjustment. Mortality from causes other than breast cancer was taken from US life tables in Splus [32].

Survival curves of women undergoing adjuvant therapies other than chemotherapy were derived from the previous analysis together with the results of EBCTCG overviews [15, 33, 16]. The results among node-positive patients receiving chemotherapy were used as the baseline and then adjusted in accordance with the risk ratios reported in Table I, to generate survival curves for all treatment options. We assumed that none of the effects depend on nodal status [15, 16]. Recent evidence from the NSABP B-14 [34] trial also shows a risk reduction for node-negative patients that is comparable with that of previously observed node-positive patients.

There are a total of 15 parameters related to the prognostic model that are being drawn at random. The distributions of the parameters within the SEER analysis and within the Cox proportional hazard model are dependent, but the two sets are independent of each other.

Decisions about adjuvant therapy depend on survival benefit as well as loss of quality of life (QOL). We assumed the QOL adjustments described in Table II. While these choices do not reflect every woman's experience, they constitute a plausible illustration, consistent with clinical expertise as well as published estimates [35]. To focus on the diagnostic value of ALND, we use the same QOL adjustment of 0.995 for the negative impact of both ALND and XRT.

### 3.2. Probabilistic sensitivity analysis

Figure 2 shows the decision tree, the expected quality adjusted life years for each arm, and the expected proportions of ALND outcomes. The 'current best' adjuvant treatment option

Table II. Illustrative quality adjustment factors, by treatment and month since the beginning of treatment. Adjustment for either ALND or XRT is applied for each subsequent year over and above any adjustment for adjuvant systemic therapy. Quality adjustments for CT+T (not shown in the table) are obtained using the adjustment factor for C for the first six months and both adjustments thereafter.

Month	1–6	7–12	13–24	25–60	> 60
No treatment	1	1	1	1	1
Tamoxifen	0.99	0.99	0.99	0.99	1
Chemotherapy	0.6	0.9	0.99	0.99	0.99
ALND	0.995	0.995	0.995	0.995	0.995
XRT	0.995	0.995	0.995	0.995	0.995

is to administer Tamoxifen to XRT patients, whose nodal status is unknown, and to those ALND patients who are node-negative, and to administer chemotherapy plus Tamoxifen to ALND patients who are node-positive. The difference between the expected quality adjusted life length in the ALND and XRT arms is about 3 days.

There is uncertainty on the correct adjuvant therapy, with  $PC = 0.33$ , meaning that the current best strategy coincides with the correct one in about one third of the simulated values. There is also concern on whether the ALND arm will do better than the XRT arm under the chosen strategy, with only  $PE = 52$  per cent of simulated parameter sets leading to a better outcome. Finally the probability that ALND has diagnostic value is  $PD = 45$  per cent.

Overall, uncertainty is substantial, suggesting that the results of the decision analysis, at least for this particular patient type, should not be used in policy and clinical decision making without explicit consideration of the uncertainty. From the point of view of an individual patient, the decision model, which focusses on quantifiable endpoints, may not consider other aspects of the decision, such as the desire to know nodal status even though this information may not change her course of action. In presence of so much uncertainty about the recommendation emerging from the decision model, these additional factors may become decisive.

A more detailed representation of uncertainty is provided by the joint distribution of  $\Delta$  and  $\Delta^*$ . Both quantities represent the difference between the average QALYs in the ALND and XRT cohorts, with one critical difference.  $\Delta$  is calculated assuming that the adjuvant therapy choices are those implied by the solution of the tree in Figure 2, while  $\Delta^*$  is calculated by rederiving the optimal adjuvant therapy choice for each draw of simulated parameters. In a two-stage decision such as this, ignorance of the correct value of the model parameters implies ignorance of the correct decision at the second stage. This translates into uncertainty about the decision at the first stage as well.  $\Delta$  incorporates this uncertainty, and realistically models the actual state of knowledge, as parameter values will not be known when the adjuvant therapy decision needs to be made.

Figure 3 shows the joint distribution of  $\Delta$  and  $\Delta^*$ , based on 900 simulated values of the model parameters. Because of the different implications of different parts of the  $(\Delta, \Delta^*)$  plane for decision making, it is useful to break down the distribution into four regions:

- $\Delta < 0, \Delta^* = 0$  The test is of no diagnostic value, but under the current best adjuvant strategy it would lead to a decrement of utility for the cohort;
- $\Delta > 0, \Delta^* = 0$  The test is of no diagnostic value, but under the current best adjuvant strategy it would lead to an increment of utility for the cohort;

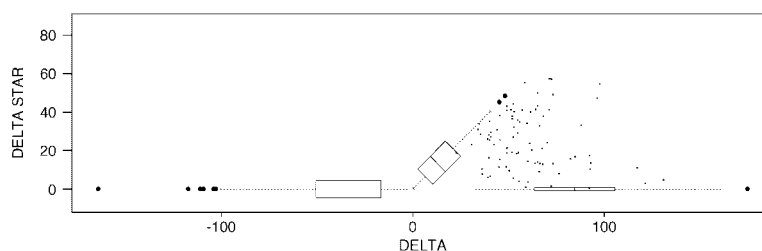


Figure 3. Joint distribution of  $\Delta$  and  $\Delta^*$ , in days, after 900 simulations of input parameters. Areas of boxes in boxplots are proportional to the frequencies of simulated points along the corresponding lines.

- $\Delta = \Delta^* > 0$  The test is of diagnostic value, the current best adjuvant strategy is correct and administering the test would lead to an increment of utility for the cohort;
- $\Delta > \Delta^* > 0$  The test is of diagnostic value, the current best adjuvant strategy is incorrect but still leads to an increment of utility for the cohort.

Of these regions, three are lines, so distributions of simulated values are represented in Figure 3 as boxplots of size proportional to the frequencies of simulated points along the corresponding lines. Overall, the variation of  $\Delta$ , attributable to lack of knowledge about the model parameters, far outweighs the mean value of  $\Delta$ .

The fraction of points on the diagonal boxplot of Figure 3 is PC. The fraction of points above zero is PD and that to the right of zero is PE. Points in  $\Delta > \Delta^* \geq 0$  are to the right of the diagonal line and correspond to parameter values for which ALND is of greater value under the current optimal strategy than it is under the optimal strategy for the given  $\theta$ . For example, it is possible that the optimal strategy for the given  $\theta$  may be to administer Tamoxifen plus chemotherapy to all nodal status categories, and therefore to women in the XRT are as well. This means that  $\Delta^*$  has no information component. The current best adjuvant strategy, however, may have an information advantage. In this case, by administering chemotherapy to all women, while reducing the difference between the two arms, we would increase QALYs in both.

#### 4. DISCUSSION

This article presents a solution strategy for multistage decision problems in presence of uncertainty on input parameters, and proposes a graphical display and summarization tools for probabilistic sensitivity analysis in multistage decision trees. For simplicity of exposition, results are developed in terms of the simplest two-stage binary decision tree. All the development of Section 2 can be straightforwardly rewritten in terms of trees with non-binary nodes. This situation is in fact illustrated in Section 3. Extensions to more than two stages can proceed along the same lines conceptually, but rapidly become much more convoluted.

We focused on the value of diagnostic information and thus assumed that the test does not have any direct effect on the utilities of the treatment branches. Also, we do not consider the cost of the test explicitly. In practical application, the test may affect utility in ways that are controlled by unknown parameters, and cost distributions may also contain unknown

parameters. The approach based on the joint distribution of  $\Delta$  and  $\Delta^*$  is useful in those circumstances as well. There may not be accumulation of values along the axes of Figure 3, and some values of  $\Delta^*$  may be negative. The general interpretation of the graph and of the quantities PC, PD and PE will, however, remain valid.

We provided an application to axillary lymph node dissection in breast cancer. To keep the focus on measuring the diagnostic value of information about nodal involvement, we did not consider potential direct effects of ALND on survival, which could occur via differential local control compared to XRT. These effects, however, are not likely to be strong [15], so our analysis does provide estimates that are of clinical interest. In summary, the uncertainty analysis presented in this section highlights the following points: there is substantial uncertainty on whether the course of action chosen by the decision tree is indeed the best—the probability of that being the case is about one third. There is also a high (48%) probability that ALND is worse than XRT for the cohort considered. Finally the probability of the parameter configurations for which ALND carries valuable diagnostic information is PD = 45 per cent. This high level of uncertainty occurs even though the data sources consist of large and well conducted studies or registries, and emphasizes the importance of systematically accounting for and communicating uncertainty in decision analysis.

#### ACKNOWLEDGEMENTS

Work partly supported by NCI grants P50CA88843, P50CA62924-05, 5P30CA06973-39 and by the Hecht fund. Don Berry, Bob Clemen, Lurdes Inoue, Teddy Seidenfeld and the reviewers provided useful comments on earlier drafts.

#### REFERENCES

1. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. *Medical Decision Making* 1985; **5**(2):157–177.
2. Critchfield GC, Willard KE. Probabilistic analysis of decision trees using Monte Carlo simulation. *Medical Decision Making* 1986; **6**(2):85–92.
3. Critchfield GC, Willard KE, Connelly DP. Probabilistic sensitivity analysis methods for general decision models. *Computers in Biomedical Research* 1986; **19**(3):254–265.
4. Parmigiani G. *Modeling in Medical Decision Making*. Wiley: Chichester, 2002.
5. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice*. Chapman & Hall: London, 1996.
6. Parmigiani G, Samsa GP, Ancukiewicz M, Lipscomb J, Hasselblad V, Matchar DB. Assessing uncertainty in cost-effectiveness analyses: application to a complex decision model. *Medical Decision Making* 1997; **17**: 390–401.
7. Pasta DJ, Taylor JL, Henning JM. Probabilistic sensitivity analysis incorporating the bootstrap: an example comparing treatments for the eradication of helicobacter pylori. *Medical Decision Making* 1999; **19**(3): 353–363.
8. Clemen RT, Reilly T. *Making Hard Decisions*. Duxbury: Pacific Grove, 2001.
9. Raiffa H, Schleifer R. *Applied Statistical Decision Theory*. Harvard University Press: Boston, 1961.
10. Lusted LB. *Introduction to Medical Decision Making*. Thomas: Springfield, 1968.
11. Weinstein MC, Feinberg H, Elstein AS, Frazier HS, Neuhauser D, Neutra RR, McNeil BJ. *Clinical Decision Analysis*. Saunders: Philadelphia, 1980.
12. DeGroot MH. Changes in utility as information. *Theory and Decision* 1984; **17**:287–303.
13. Warmuth MA, Bowen G, Prosnitz LR, et al. Complications of axillary lymph node dissection for carcinoma of the breast: a report based on a patient survey. *Cancer* 1998; **83**(7):1362–1368.
14. Orr RK, Col NF, Kuntz KM. A cost-effectiveness analysis of axillary node dissection in postmenopausal women with estrogen receptor-positive breast cancer and clinically negative axillary nodes. *Surgery* 1999; **126**(3): 568–576.
15. Early Breast Cancer Trialists Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. *Lancet* 1992; **339**(1–15):71–85.

16. Early Breast Cancer Trialists Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *The Lancet* 1998; **351**:1451–1467.
17. Early Breast Cancer Trialists Collaborative Group. Effects of radiotherapy and surgery in early breast cancer. *New England Journal of Medicine* 1995; **333**:1444–1455.
18. Cady B. Is axillary lymph node dissection necessary in the routine management of breast cancer? No. *Important Advances in Oncology* 1996; 251–256.
19. Moore MP, Kinne DW. Is axillary lymph node dissection necessary in the routine management of breast cancer? Yes. *Important Advances in Oncology* 1996; 245–250.
20. Silverstein MJ, Gierson ED, Waisman JR, Senofsky GM, Colburn WJ, Gamagami P. Axillary lymph node dissection for T1a breast carcinoma. Is it indicated? *Cancer* 1994; **73**(3):664–667.
21. Epstein RJ. Routine or delayed axillary dissection for primary breast cancer? *European Journal of Cancer* 1995; **31A**(10):1570–1573.
22. Recht A, Houlihan MJ. Axillary lymph nodes and breast cancer: a review. *Cancer* 1995; **76**(9):1491–1512.
23. Velanovich V. Axillary lymph node dissection for breast cancer: a decision analysis of T1 lesions. *Annals of Surgical Oncology* 1998; **5**(2):131–139.
24. Parmigiani G, Berry DA, Winer EP, Tebaldi C, Iglehart JD, Prosnitz L. Is axillary lymph node dissection indicated for early stage breast cancer—a decision analysis. *Journal of Clinical Oncology* 1999; **17**(5): 1465–1473.
25. Ramsey FP. Weight or the value of knowledge. *British Journal for the Philosophy of Science* 1990; **41**:1–4.
26. National Cancer Institute: Surveillance, Epidemiology, and End Results (SEER) Program. Seer homepage, 1997; <http://www-seer.ims.nci.nih.gov>.
27. Gelman A, Carlin J, Stern H, Rubin D. *Bayesian Data Analysis*. Chapman & Hall: London, 1995.
28. Wood W, Weiss R, Tormey D, Holland J, Henry P, Leone L, et al. A randomized trial of CMF versus CMFVP as adjuvant chemotherapy in women with node-positive stage ii breast cancer: a CALGB study. *World Journal of Surgery* 1985; **9**:714–718.
29. Perloff M, Norton L, Korzun A, Wood W, Carey R, Gottlieb A, et al. Postsurgical adjuvant chemotherapy of stage ii breast carcinoma with or without crossover to a non-cross-resistant regimen: a cancer and leukemia group b study. *Journal of Clinical Oncology* 1996; **14**:1589–1598.
30. Wood W, Budman D, Korzun A, Cooper M, Younger J, Hart R, et al. Dose and dose intensity of adjuvant chemotherapy for stage ii, node-positive breast carcinoma. *New England Journal of Medicine* 1994; **330**: 1253–1259.
31. Rosen P, Groshen S, Saigo P, Kinne D, Hellman S. A long term follow-up study of survival in stage I (T<sub>1</sub>N<sub>0</sub>M<sub>0</sub>) and stage II (T<sub>1</sub>N<sub>1</sub>M<sub>0</sub>) breast carcinoma. *Journal of Clinical Oncology* 1989; **7**:355–366.
32. Thernau TM. *A Package for Survival Analysis*. S. Mayo Foundation: Rochester, MN, 1996.
33. Early Breast Cancer Trialists Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *The Lancet* 1998; **352**:930–942.
34. Fisher B, Dignam J, Bryant J, DeCillis A, Wickerham DL, Wolmark N. Five versus more than five years of tamoxifen therapy for breast cancer patients with negative lymph nodes and estrogen receptor-positive tumors. *Journal of National Cancer Institute* 1996; **88**:1529–1542.
35. Gelber RD, Goldhirsch A, Cavalli F. Quality-of-life-adjusted evaluation of adjuvant therapies for operable breast cancer. *Annals of Internal Medicine* 1991; **114**:621–628.