Primer on Medical Decision Analysis:

Part 4—Analyzing the Model and Interpreting the Results

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

This paper is the fourth of a five-part series that describes the principles of construction and evaluation of valid decision models. In this review, the authors describe the key principles of detecting and eliminating structural and programming errors in decision trees (debugging). In addition, they offer guidelines to facilitate the interpretation of analytic results of decision models. *Key words*: decision analysis; expected value; utility; sensitivity analysis; decision trees; probability. **(Med Decis Making 1997;17:142–151)**

The first three parts of this series[1–3] offer practical guidance in building a model that is structurally valid and clinically sensible, and obtaining the best available probabilities and utilities for the model. This paper is about the next step: evaluating the model and interpreting the results. "Folding back," or analyzing the tree (described in detail in introductory texts[4,5]) will give us the expected value for each strategy modeled in the tree, and should tell us which is the preferred strategy.

Sensitivity Analysis

Before the results of folding back the tree can be interpreted, though, an intermediate step is required: sensitivity analysis. Sensitivity analysis is the process of repeatedly folding back the tree using different values for probability and utility variables. There are two main reasons to perform sensitivity analysis. First, it is one of the most useful methods of "debugging," or correcting errors within decision trees. Second, sensitivity analysis is the decision analyst's version of statistical hypothesis testing; that is, it is the primary way decision analysts assess the degree of uncertainty associated with an analytic result. We discuss these two uses in order.

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

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

Debugging the Tree

We use the term "bug" to describe both structural errors (failure to follow the six recommendations set forth in Part 2 of this series[2]) and technical or programming errors that result in the tree formalism incorrectly reflecting the ideas of the modeler. As a great decision-analytic guru and mystic likes to say, "All trees have bugs." This often includes trees that have already been debugged, and it particularly includes the trees of neophytes. Religiously following the principles of sound tree construction will usually result in fewer bugs, but bugs may remain despite your best efforts. Sensitivity analysis is the main tool we use to ferret them out.

We suggest that the process of debugging should start with changing one variable at a time (one-way sensitivity analysis) over its entire range, not just its plausible range. If you have followed the rule of having only two branches after each chance node, it should be possible to evaluate the model for all probability values between the range of 0 and 1. We also suggest, for the purpose of debugging, that you run the model for all utility and disutility values between the ranges of 0 and 1, even though this may occasionally give paradoxical results such as a "negative" expected utility.

For the purpose of debugging, we find it easiest to ignore the specific expected utility values generated by the computer and simply evaluate the results graphically. What will undoubtedly occur when one starts to "debug" is that some of the sensitivity analyses will not "make sense," i.e., they will not correspond to our predictions of what should be hap-

Downloaded from mdm.sagepub.com at Max Planck Society on November 1, 2016

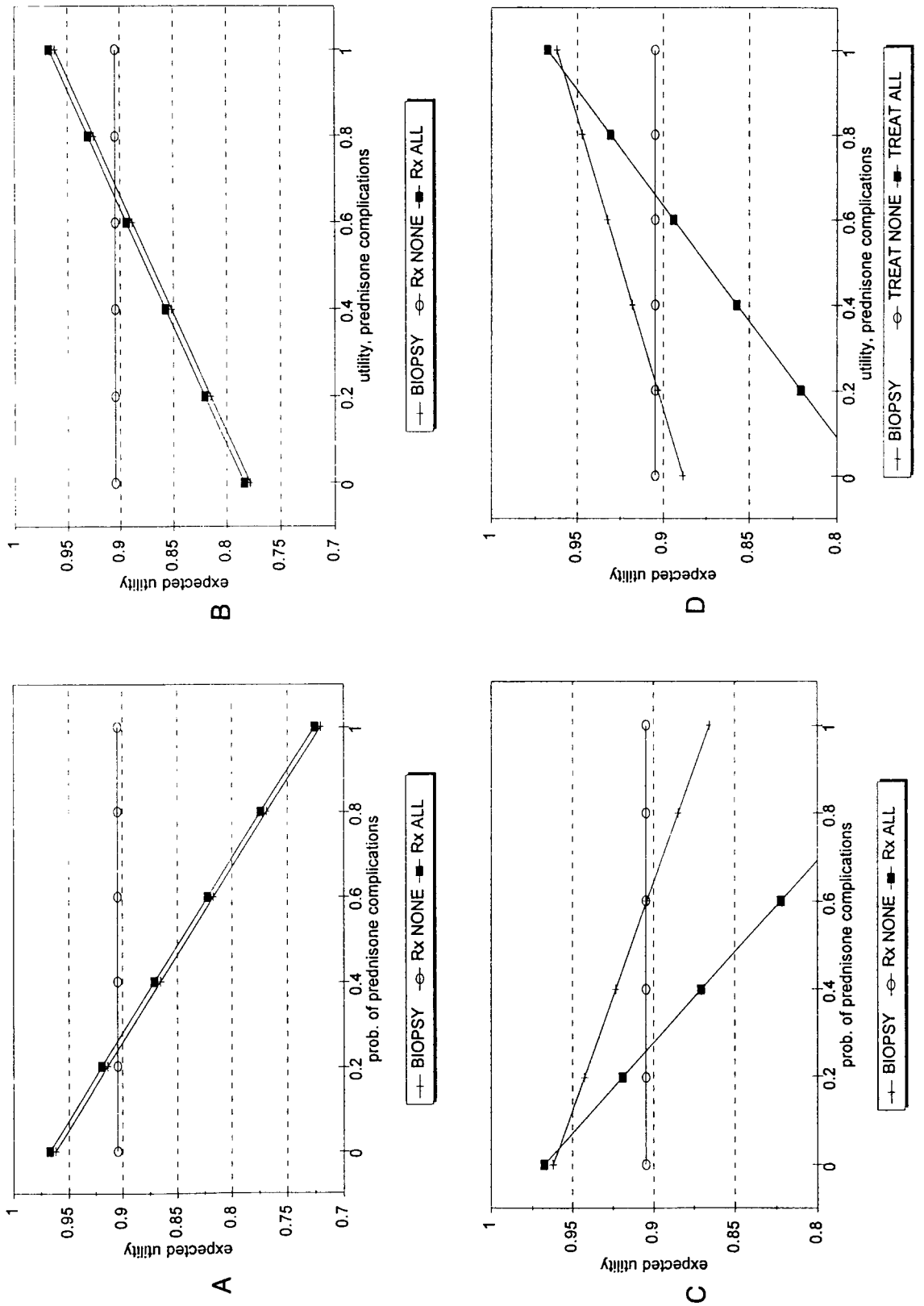


FIGURE 1. One-way sensitivity analyses of the GCA tree. *A* and *B* (*top row*) are sensitivity analyses of the probability and utility of prednisone complications in the presence of a programming error or "bug." *C* and *D* (*bottom row*) show the sensitivity analyses after correction of the error.

Downloaded from mdm.sagepub.com at Max Planck Society on November 1, 2016

Table 1 • Summary of Debugging Tips

1. Perform one-way sensitivity analyses on all variables over their entire ranges (usually 0–1)
2. Evaluate results graphically
3. Evaluate slopes and rank order of strategies at extreme values
4. Do a risk analysis
5. Perform pairwise comparisons of strategies, after changing key variables to give identical expected results
6. Develop consistent nomenclature habits (e.g., start probability, utility, and Markov-state names with the same upper- or lower-case letter)
7. Delete archaic variables and nodes

pening as variables change. When this happens, you have either a new insight or a bug. If you've just built the tree, it's likely to be a bug.

Bugs come in many phyla and species. Providing an exhaustive phylogeny and ontogeny is possible but of doubtful practical value, since there are innumerable ways of building trees wrong and only a few ways of doing it right. Learning how to find bugs, though, is an immeasurably useful skill. The following section illustrates, with examples, the method we've found useful for tracking them down. The giant cell arteritis decision tree we've been using as an example is shown correctly programmed in SMLTREE (Hollenberg JP, Roslyn, NY) or DECISION MAKER (DECISION MAKER, Pratt Medical Group, Boston, MA) format in the appendix. To follow the argument in the next section, you will have to periodically refer to the appendix. Also, notice that the bugs referred to below have been "fixed" in the tree shown in the appendix.

In figure 1, panels *A* and *B* illustrate one-way sensitivity analyses for the giant cell arteritis decision tree we've been using as an example. What's wrong with these figures? If the answer is not obvious, we suggest two strategies for sorting this out: 1) evaluate the slopes of the various strategies, and 2) evaluate the rank order of the strategies at extreme values (usually 0 and 1). Still unsure about what's wrong?

Figure 1*A* shows that the Rx NONE strategy is unaffected by the probability of prednisone complications (slope = 0). This makes sense, since no one is getting prednisone in this strategy. The Rx ALL strategy looks less attractive (negative slope) as the probability of complications rises, as we expect, since everyone gets prednisone in this strategy. When we examine the BIOPSY strategy, though, it's clear that something is wrong. The slope appears to be the same as that of the Rx ALL strategy. This suggests that the probability of prednisone complications is affecting net results just as much in the BIOPSY arm (where 40% of the cohort is getting prednisone) as in the strategy where everyone is treated. That's clearly not right.

The rank ordering of strategies at probability = 0 is plausible: Rx ALL > BIOPSY > Rx NONE. Treating everyone seems like the optimal choice if there are no treatment complications and the test is imperfect. Treating no one seems like the least attractive option. If the probability of incurring a treatment complication is high (probability = 1), less aggressive strategies should be preferred. It is impossible to predict rank order with certainty here, but one might expect that the BIOPSY strategy at some point would be preferred to the Rx ALL strategy. More than twice as many individuals are treated with prednisone in the Rx ALL strategy, prednisone complications are not trivial (utility = 0.75), and treatment itself decreases quality of life (utility = 0.97).

So, analysis of the slope and, to a lesser extent, the rank order suggests that the expected utility of the BIOPSY strategy is too low at higher probabilities of prednisone complications. The Y1 intercept (the Y axis intercept on the left side of the graph) may be correct, but the slope is too negative and the Y2 intercept (the Y axis intercept on the right side of the graph) is probably too low.

In figure 1*B*, the Rx ALL and Rx NONE strategies again behave as predicted. We expect the expected utility to be unaffected by the utility of prednisone complications in the Rx NONE strategy, and to be greatly affected in the Rx ALL strategy. Again, the line has the same slope as that of the Rx ALL strategy, suggesting that the complications of prednisone affect the analytic result as much in the BIOPSY as in the Rx ALL strategy. For some reason, individuals in the biopsy arm are being disproportionately penalized for treatment. At utility = 1.0, the rank order is plausible, but at utility = 0, one would expect the BIOPSY strategy to look better relative to the Rx ALL strategy.

Both sensitivity analyses suggest that there's a bug in the BIOPSY branch, and that it has something to do with the way prednisone complications are expressed. As it happens, the tree builder failed to express the fact that individuals who are biopsy-negative will not get prednisone complications. More specifically, the temporary binding PPREDCMP:=0 is missing at the "Bx_Neg" branch of the BIOPSY strategy. Correcting this oversight yields the results expressed in figures 1*C* and 1*D*, which show, as we predicted, that the BIOPSY strategy has a slope intermediate between the slopes of the other two. The rank orderings in both figures also behave as predicted.

We'll try one more example. Look at figure 2*A*. The slope of the BIOPSY branch is negative. That seems right, since the attractiveness of this strategy should decrease as having a biopsy becomes worse. However, no one gets biopsied in either of the other two arms, so both lines should be horizontal. Yet

Downloaded from mdm.sagepub.com at Max Planck Society on November 1, 2016

the Rx ALL strategy also becomes less attractive as having a biopsy becomes worse. For some reason, individuals in the Rx ALL strategy are being incorrectly penalized for having a biopsy.

The rank order of the strategies looks right for low disutilities: BIOPSY is preferred to Rx ALL, which is preferred to Rx NONE. Since the baseline value for this disutility is very low (0.005), we expect the rank order at or around a value of zero to be the same as that observed in the baseline analysis. However, the rank order at the Y2 intercept is clearly wrong. BIOPSY should be the worst (because no one is biopsied in the other branches), whereas the rank order of the other two strategies should the same as it is at disutility = 0, i.e., Rx ALL should be preferred to Rx NONE, the reverse of what is seen in figure 2*A*.

Again, the sensitivity analysis not only shows us that there is a bug, but also tells us something about where the bug is. The bug has something to do with how the disutility of biopsy is evaluated in the Rx ALL arm. We can even be more specific: we know that the variable expressing disutility of biopsy appears in SUBTREE1, which is the same in all branches. So, there must be something wrong about the way the disutility of biopsy is expressed that is not in SUBTREE1 (otherwise, all strategies would be affected), but is in the Rx ALL strategy. That doesn't leave much: the only thing that happens to the disutility of biopsy that's not in the subtree is in the local bindings. The binding at the Rx ALL branch, assigning a local value of disutility of biopsy of 0, appears at first glance to be correct, but on more careful inspection, the variable is "**duBX**" rather than "**duBx**." If you still can't see the difference, notice that one "x" in "duBx" is capitalized, whereas the other is not. An expression for the disutility of biopsy (duBX) was created during tree construction, never deleted, and incorrectly used in the temporary binding. Thus, "duBX" does not have a local value of 0, as it should, but rather assumes its baseline, unmodified value (0.005). Thus, sensitivity analysis on duBx affects both the Rx ALL and the BIOPSY strategies. Correcting this error leads to figure 2*B*. Here, the slopes of both Rx ALL and Rx NONE are horizontal, and the rank order at the extremes appears to be correct.

Let's assume now that you've run all your one-way sensitivity analyses, and that the slopes and intercepts are behaving as predicted. As a final check to ensure you've programmed the tree correctly, we suggest that you perform a series of pairwise comparisons between strategies. Think about the ways in which pairs of strategies differ, and adjust the model parameters to give identical expected results. Using our GCA tree, for example, let's compare the Rx ALL and the BIOPSY branches. The test strategy should be equivalent to the Rx ALL strategy if the same number of people are treated (sensitivity = 1.0, specificity = 0) and there is no ill effect of testing (the disutility of biopsy is 0). Changing these three values should give us an identical expected utility.

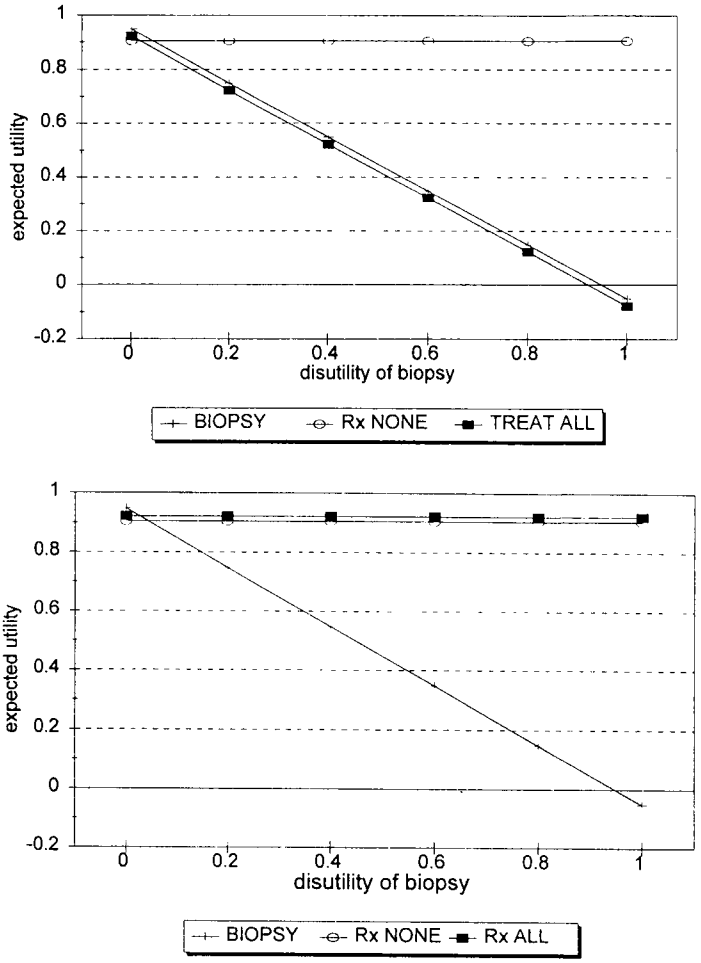


FIGURE 2. Sensitivity analysis of the disutility of temporal artery biopsy. *A* (*above*) shows analytic results in the presence of a "bug." *B* (*below*) shows results after the error has been corrected.

Similarly, testing should be equivalent to treating no one if the same number of individuals are treated (sensitivity = 0, specificity = 1.0) and there is no ill effect of testing (the disutility of biopsy is 0).

We have illustrated thus far how to determine whether a bug is present. We've also illustrated that sensitivity analysis may give you clues about the nature and location of the bug. Armed with this information, you are still faced with the onerous task of finding and fixing it. There is no simple recipe for doing this consistently or effectively, but we suggest the following strategy. Start where you think the bug might be, and mentally reconstruct the tree. Follow each branch, think about the meaning of the branch, and examine each node name as it arises. Examine each variable as it arises to ensure that the form of expression is correct. When you come to a

Downloaded from mdm.sagepub.com at Max Planck Society on November 1, 2016

subtree, think about the local meaning of each variable within the subtree and check whether each variable has been correctly expressed or modified by the binding expression. When you come to temporary bindings, examine each one in turn to ensure that both the form of the expression and the idea it expresses are correct. Check the variable menu downstream from temporary bindings to ensure that global variable values have been correctly changed by your temporary binding expressions. Repeat this process until you get to the terminal branches. If the bug doesn't turn up, widen the search. Start closer to the root of the tree and repeat. More often than not, you'll end up mentally recreating the entire tree several times before you find the bug.

One final strategy that some analysts employ early in the debugging process is risk analysis. A risk analysis will show how many different outcomes each strategy has, and will report the probability value associated with each outcome. The number of outcomes and the reported frequencies of those outcomes can tip you off to the presence of a bug if they differ from your predictions. For example, in the Rx NONE strategy, we expect three outcomes: no GCA (most frequent), GCA without complications (next most frequent), and GCA with complications. If our risk analysis showed there to be fewer than three outcomes, or if the relative frequencies differed from our predictions, this would probably mean that a bug was present.

The Bugs That Will Not Die: A Taxonomy of Hardy Tree Pests

If there are bugs in your tree whose will to live exceeds your sleuthing ability, determination, and perspicacity, consider the following checklist:

STRUCTURAL PROBLEMS

1. Symmetry problems. Have you forgotten to describe the same clinical events in each branch? If you've used subtrees to describe common outcomes, this is unlikely to have occurred.

2. Linkage problems. Are all common events in separate branches (e.g., treatment efficacy, predictive values of tests) "linked" by subtrees, common variable names, or common expressions (efficacy equations, predictive value expressions for test results conditioned on disease prevalence)?

TECHNICAL (PROGRAMMING) ERRORS

3. Typographical errors. The most persistent bugs fall into this category. Lower-case substituted for upper-case letters, spelling errors, or inconsistent abbreviations (e.g., bug #2, fig. 2), are common problems. Developing consistent nomenclature habits limits this type of error. We suggest you consistently use the same upper- or lower-case letter to start probability, utility, and Markov-state names.

Table 2 • Sensitivity Analyses

Variable	Baseline Value	Plausible Range	Threshold Value*	Sensitive?†
Prevalence of giant cell arteritis (GCA)	0.50	0.00–1.00	NA‡	NA‡
Probability of adverse outcome from GCA	0.12	0.05–0.40	0.86	N
Sensitivity of temporal artery biopsy	0.80	0.58–0.97	0.40	N
Specificity of temporal artery biopsy	1.00	0.90–1.00	0.42	N
Effectiveness of prednisone	0.89	0.60–1.00	NT§	N
Probability of iatrogenic side effects from prednisone	0.19	0.05–0.40	0.04	N
Disutility of biopsy	0.005	0.00–0.03	0.003	Y
Utility of GCA	0.85	0.70–0.95	0.95	Y
Utility of complications of GCA	0.60	0.20–0.85	NT§	N
Utility of taking prednisone	0.95	0.90–1.00	0.87	N
Utility of prednisone complications	0.75	0.60–0.90	0.95	N
Bias to Rx ALL	—	—	—	Y¶
Bias to Rx NONE	—	—	—	Y¶

*The threshold value is the value of the variable at which two strategies are equivalently valued (equal expected utility or quality-adjusted life expectancy or other index of value). This column shows the threshold nearest to the baseline value when more than one threshold exists.

†"Sensitive" here means that a strategy other than the "TEST Rx, IF POSITIVE" is preferred for some value of the variable within the plausible range. Y = yes, the analysis is sensitive to this variable; N = no, the analysis is not sensitive.

‡NA = not applicable. In this analysis, the prevalence of giant cell arteritis is assumed to be 0.50. We are evaluating diagnostic strategies when the clinical features of the patient suggest a pretest probability (prevalence) of 50%.

§NT = no threshold found for this variable.

¶The analysis is insensitive, under the conditions of systematic bias (best-case or worst-case scenario) if the preferred strategy does not change, and sensitive if it does.

Downloaded from mdm.sagepub.com at Max Planck Society on November 1, 2016

4. Wrong variable/node names. Failure to delete archaic node and variable names often results in their reuse when the tree is being edited or rebuilt. Get rid of orphan nodes and unused variables and this type of error won't occur.
5. Errors in temporary bindings. Bindings can be present when they shouldn't be, or absent when they should be there (e.g., bug #1, fig. 1). Equations expressed in temporary bindings may have errors. Faulty bindings are a very frequent source of error.

Evaluating Uncertainty

Let's assume that you've been successful in eliminating all apparent bugs. The next step is to try to generate some meaningful results. Folding back the tree will give you a series of scores indicating the expected value of each alternative. Folding back our giant cell arteritis tree gives us the following results: BIOPSY (expected utility = 0.9435) > Rx ALL (expected utility = 0.9215) > Rx NONE (expected utility = 0.9046). In our baseline analysis, testing looks like the best strategy. Remember, though, that we were uncertain about some of the probabilities and utilities we used in the model. Given that uncertainty, how confident can we be that the testing strategy is really the best one?

We suggest you approach this question in a systematic way by running one-way sensitivity analyses over all ranges of all variables and placing the results in a table like table 2. The first three columns of table 2 are self-explanatory: part 3[3] of this series is about getting baseline values and plausible ranges for input variables. The threshold value (column 4) is the value for that variable at which two strategies have equal analytic results (expected utility, life expectancy, etc.). At values more extreme than the threshold value, a new strategy will be preferred. If there are more than two strategies, some variables may have more than one threshold. If so, report them all in your table. Fill in the last column by determining whether the threshold value falls within the plausible range for that value. If it does, the result is "sensitive" to that variable. If your analysis is insensitive to changes in any single variable within its plausible range, congratulations, you have a fairly robust analysis. More often than not, though, the analysis will be sensitive to one or several variables.

Even if your analysis is robust to changes within a single variable, though, it may not be robust to changes in multiple variables, so the next step is multi-way sensitivity analysis. We suggest that you choose sets of two variables, starting with the variables to which the analysis seems most sensitive, and calculate threshold values for each strategy. Calculating thresholds will result in a graph that looks like figure 3. The region at the lower right, at which the disutility of biopsy is high and the utility of giant cell arteritis is low, consists of pairs of values that give an analytic result favoring the Rx ALL strategy. Conversely, the upper left region favors the BIOPSY strategy. The threshold line dividing the two regions consists of pairs of values at which the analytic results are exactly the same for the two strategies. The "x" represents the baseline value for both variables, and the box encloses the range of clinically plausible values.

In figure 3, it is possible to find a plausible set of values for these two variables at which the Rx ALL strategy is preferred, but this will occur only when the disutility for biopsy and the utility for giant cell arteritis are simultaneously close to the extremes of their plausible ranges. Though this is possible, it is unlikely. Exactly how unlikely we can't say, unless we know something about the probability distribution of variables within the plausible range, information that is not commonly available.

Because our giant cell arteritis (GCA) tree considers more than two strategies, there may be more than one threshold. Thus, were we so inclined, we could redraw figure 3 with additional threshold lines comparing additional pairs of strategies.

Software packages usually allow sensitivity analysis for three variables as well as two. Figure 4 illustrates a three-way analysis for the disutility of biopsy, the utility of giant cell arteritis, and the prevalence of giant cell arteritis. This graph shows that at a prevalence of 0.25 there are no plausible values for the other two variables at which the Rx ALL strategy is preferred. As prevalence rises, it is increasingly likely that combined values for the other two variables will yield a result favoring the Rx ALL strategy. At a prevalence of 0.75, the Rx ALL strategy is almost certain to be preferred. This coincides with clinical intuition, which suggests that testing is likely to be of greatest value at intermediate disease prevalence rates.

There is no rule about which variables should be examined in two- and three-way sensitivity analyses. In general, though, variables that seem important in one-way analyses should be carefully evaluated in multi-way analyses.

As a final sensitivity analysis, we recommend evaluating the model under "best-case" and "worst-case" scenarios (analysis of extremes). When evaluating two strategies, set all the variables at the extremes of their plausible ranges to favor the first strategy. Then, set all the variables at the opposite extremes to favor the other strategy. For example, if we were comparing only the BIOPSY and the Rx ALL strategies, we would first set all the variables to favor the Rx ALL strategy as follows: high prevalence of

Downloaded from mdm.sagepub.com at Max Planck Society on November 1, 2016

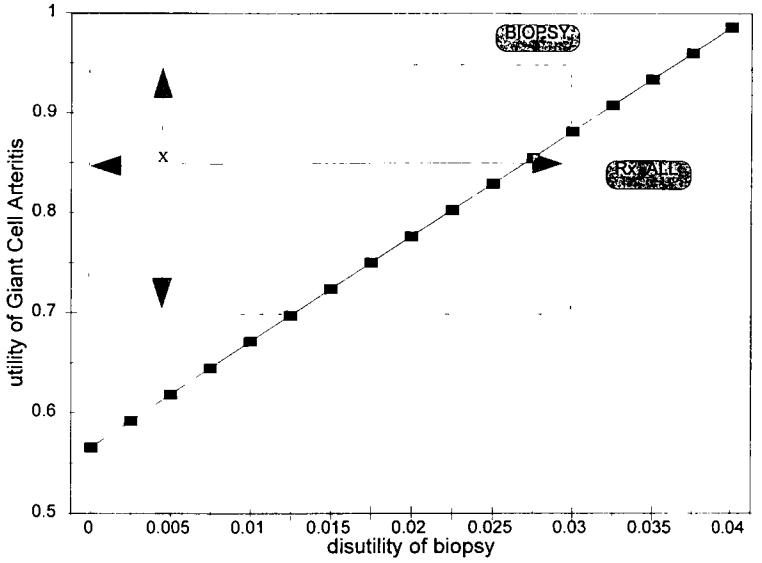


FIGURE 3. Two-way sensitivity analysis evaluating the effect of simultaneously changing the utility of living with giant cell arteritis and the disutility of temporal artery biopsy. Pairs of values below and to the right of the line yield analytic results favoring the Rx ALL strategy, whereas values above and to the left favor the BIOPSY strategy. The "x" marks the baseline values for these two variables, and the dotted box encloses all potential pairs of values that fall within the plausible range.

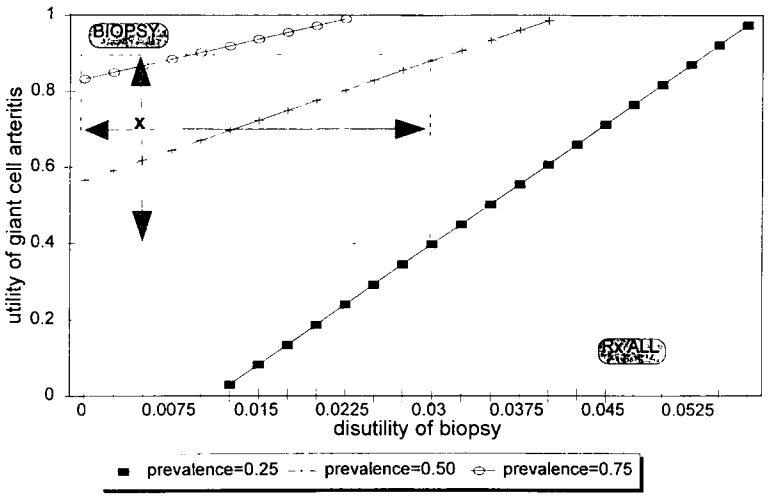


FIGURE 4. Three-way sensitivity analysis evaluating the effect of simultaneously changing the disutility of temporal artery biopsy, the disutility of living with giant cell arteritis, and the prevalence (pretest probability) of giant cell arteritis.

GCA, low probability of adverse outcome from GCA, low sensitivity and specificity of biopsy, high effectiveness of prednisone, and so on. If there are more than two strategies, favor each strategy in turn by setting all the variables to the extremes that favor that strategy. If changing a variable doesn't improve expected outcomes for the favored strategy, but simply penalizes one of the other two strategies (e.g., changing the sensitivity and specificity of temporal artery biopsy in an analysis biased toward the Rx ALL strategy), leave the variable at its baseline value for the "biased" analysis.

The last two rows of table 2 yield the results of our "biased" analysis. Biasing toward Rx NONE or Rx ALL makes a difference. Under conditions of systematic bias (best-case or worst-case scenario), each strategy can become the preferred one.

This completes the set of sensitivity analyses we would recommend for a simple, beginner's model. There are more sophisticated ways of evaluating the overall uncertainty in the model,[6–8] but these approaches are beyond the scope of this paper.

Interpreting the Results

A decision analysis has three possible outcomes: 1) strategy A is the best one; 2) the choice between two (or more) strategies is a "toss-up" or a "close call"; 3) we don't know. The baseline analysis will almost always give us a strategy whose score (expected utility, quality-adjusted life expectancy) is numerically the highest. However, the difference between the best strategy and the next-best strategy may be very small. Alternatively, one strategy may be clearly better, but there is so much uncertainty that a clear winner cannot be declared.

First, let's consider the issue of uncertainty. How much uncertainty is too much? At the one extreme, an analysis may be insensitive to all one-way and multi-way analyses. Even systematically biasing the

Downloaded from mdm.sagepub.com at Max Planck Society on November 1, 2016

analysis does not change the baseline result. Under these circumstances, the uncertainty is small, the analysis very robust, and the preferred option quite clear. At the other extreme, the analysis may be sensitive to small changes in one or several variables within the clinically plausible range. A high degree of uncertainty clearly attaches to this analytic result.

Most analyses fall between these extremes. Under these circumstances, we recommend that you systematically review the one-way and multi-way analyses. Find the variables to which the analysis is sensitive, and refer back to the literature from which they were derived. What is the quality of the evidence that underlies the quantitative estimates of probability and utility? How much variation is there in the available data?

There is an unavoidably subjective element in interpreting the results of a decision analysis, particularly this type of analysis, that precludes calculation of the overall uncertainty in the analytic result. Thus, you will have to make a critical judgment, based on the sensitivity analyses and the quality of the evidence, about whether the level of uncertainty in the analysis is low enough that you can declare a clear winner. If the uncertainty is too high, you will have to conclude that the state of the evidence does not permit a firm conclusion. At the very least, you will be able to highlight the central issues in the decision problem, and determine which variables require further empirical evaluation.

What about the magnitude of the gain? How much of a gain is "clinically" as opposed to "numerically" significant? Decision-analytic purists might argue that this question is immaterial. If you have captured all the dimensions of the decision problem in your analysis, the analysis will give you the very best solution. How much better it is than the next-best solution is unimportant.[9] Purists, though, may need reminding that even very sophisticated analyses usually overlook some of the relevant facets of a decision problem. Purely clinical analyses overlook cost. Nearly all analyses ignore the relative "riskiness" of the strategies under consideration.[10–12] Individual preferences may vary with time or experience, which may not be reflected in the analysis.[13] Clinical events that have small effects on quality of life, such as undergoing a test, may also not be represented in the model. Finally, outcomes that result from medical interventions may be valued differently by patient and physician than outcomes that occur as the result of an underlying disease process, even if the outcomes are identical.[14] Thus, very small gains should be interpreted with caution, even if the analytic result appears to be robust.

But how small a gain is small? If outcomes are expressed as "expected utilities," there is no general, a priori answer to this question. Because outcomes are specific to the decision problem, with a unique time frame and set of outcomes, units of "expected utility" vary in value from analysis to analysis. Interpreting information about outcomes characterized in terms of life expectancy, or quality-adjusted life expectancy, is more straightforward. Some authors have suggested that a life-expectancy gain of two months is significant, since it corresponds to risk reductions observed in clinical trials widely judged to have clinically significant outcomes.[15] Gains of six months or more would probably be considered significant by most analysts, and are produced by interventions such as smoking cessation (13 months),[16] coronary bypass for severe three-vessel disease (10.8 months),[17] treatment of postmenopausal women with estrogen replacement (10.3 months),[18] and cholecystectomy in asymptomatic diabetic patients (6.1 months).[19] Gains of a few days to a few weeks are usually,[16,20,21] though not invariably,[13,22,23] considered "toss-ups."

Concluding that a "toss-up" exists does not mean you've wasted your time. Knowing that two strategies are more or less equivalent is as useful as knowing which one is the better.[9] You know something you didn't know at the outset: that there is no major loss or gain in choosing either of the equivalent strategies. You also know that making the decision based on criteria not explicitly represented in the model is probably legitimate.

Postanalytic Considerations

Once you've done your best to interpret the analytic results generated by your model, there are some additional issues that you will want to consider before you announce your freshly minted clinical policy to the world. Most of these don't make it into formal models, as we've discussed above. The first is the economic factor. If two strategies are a "toss-up" on clinical grounds, but one is substantially less costly, that strategy is clearly the more attractive one.

The second issue is risk: if two strategies are a close call, but one strategy is riskier (has a greater chance of adverse outcomes), the less risky strategy may be preferred by most patients. Remember that decision models will yield an average gain for a cohort of like individuals. But an average can be arrived at in several ways: a small gain for everyone and a mixture of larger gains and losses will yield the same result. The distribution of gains and losses is not reflected in the analytic results.

For example, let's say your decision model compares a medical option and a surgical option, and the latter has an immediate, nontrivial risk of perioperative death. If your decision model shows that the two strategies are formally equivalent, choosing

Downloaded from mdm.sagepub.com at Max Planck Society on November 1, 2016

the surgical option entails taking a risk of a short-term adverse outcome (death) to achieve a better long-term outcome, if one survives, than that achieved by the medical option. Real patients may prefer the less risky decision. Conversely, real patients may prefer to be screened for cancer, even if the expected gain is trivial, because screening minimizes the risk of an adverse outcome.

The third issue is the ethical consequences of each decision. Critics of decision analysis have argued that there are potential ethical problems in the application of decision analysis, because some patients may be exposed to great losses so that others may achieve gains.[24,25] This is something to think about when interpreting your analysis: are there more "big losers" in your winning strategy than in other strategies? Running a risk analysis will give you some idea of the distribution of gains and losses in the different strategies, and may help you to evaluate the importance of the second and third factors.

Fourth is the issue of time. If you build a simple (e.g., non-Markov) model, you will probably adopt a time frame that is shorter than the life expectancy of the patients you're considering. Are there events beyond your time frame that might affect which strategy is preferred? How does the passage of time affect the efficacy of your intervention? How will time affect the perception of health outcomes? Is there an "adaptation" effect,[13] or are the deleterious effects of the disease or the treatment worse as time passes?

The fifth issue is the interests of others. Most analytic models characterize outcomes from the patient's point of view. Illness and death, however, have a profound impact on family and friends, doctors, the health care system, and society. No decision-analytic model fully characterizes all of the important social and economic dimensions of a health problem. Even full economic evaluations carried out from a societal perspective overlook important dimensions of real decision problems. We suggest that you carefully consider these five issues before you declare a winner, particularly when the difference between two strategies is small.

References

1. Detsky AS, Naglie G, Krahn MD, Naimark D, Redelmeier DA. Primer on medical decision analysis: Part 1—getting started. Med Decis Making. 1997;17:123–5.
2. Detsky AS, Naglie G, Krahn MD, Redelmeier DA, Naimark D. Primer on medical decision analysis: Part 2—building a tree. Med Decis Making. 1997;17:126–35.
3. Naglie G, Krahn MD, Naimark D, Redelmeier DA, Detsky AS. Primer on medical decision analysis: Part 3—estimating probabilities and utilities. Med Decis Making. 1997;17:136–41.
4. Sox H, Blatt MA, Higgins MC, Marton KI. Medical Decision Making. London, England: Butterworth & Co., 1988.
5. Weinstein MC, Fineberg HV. Clinical Decision Analysis. Philadelphia, PA: W. B. Saunders, 1980.
6. Critchfield GC, Willard KE. Probabilistic analysis of decision trees using Monte Carlo simulation. Med Decis Making. 1986; 6:85–92.
7. Doubilet P, Begg CB, Weinstein MC, et al. Probabilistic sensitivity analysis using Monte Carlo simulation. Med Decis Making. 1985;5:157–77.
8. Willard KE, Critchfield GC. Probabilistic analysis of decision trees using symbolic algebra. Med Decis Making. 1986;6:93–100.
9. Kassirer JP, Pauker SG. The toss-up. N Engl J Med. 1981;305: 1467–9.
10. Feinstein AR. The "chagrin factor" and qualitative decision analysis. Arch Intern Med. 1985;145:1257–9.
11. Gafni A, Torrance GW. Risk attitude and time preference in health. Manage Sci. 1984;30:440–51.
12. Krahn M, Gafni A. Discounting in the evaluation of health care interventions. Med Care. 1993;31:403–18.
13. Krahn MD, Mahoney JE, Eckman MH, et al. Screening for prostate cancer: a decision analytic view. JAMA. 1994;272: 781–6.
14. Cohen BJ, Pauker SG. How do physicians weigh iatrogenic complications? J Gen Intern Med. 1994;9:20–3.
15. Naimark DM, Naglie G, Detsky AS. The meaning of life expectancy: what is a clinically significant gain? J Gen Intern Med. 1994;9:702–7.
16. Tsevat J, Weinstein MC, Williams LW. Expected gains in life expectancy for coronary heart disease risk factor modifications. Circulation. 1991;83:1194–201.
17. Wong JB, Sonnenberg FA, Salem DN, et al. Myocardial revascularization for chronic stable angina. An analysis of the role of percutaneous transluminal coronary angioplasty based on data available in 1989. Ann Intern Med. 1991;113: 852–71.
18. Zubialde JP, Lawler F, Clemenson N. Estimated gains in life expectancy with use of postmenopausal estrogen therapy: a decision analysis. J Fam Pract. 1993;36:271–80.
19. Friedman LS, Roberts MS, Brett AS, et al. Management of asymptomatic gallstones in the diabetic patient. A decision analysis. Ann Intern Med. 1988;109:913–9.
20. Richardson WS, Detsky AS, and the Evidence-Based Medicine Working Group. Users' guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of the study valid? JAMA. 1995;273:1292–5.
21. Richardson WS, Detsky AS, and the Evidence-Based Medicine Working Group. Users' guides to the medical literature. VII. How to use a clinical decision analysis. B. Results and applicability. JAMA. 1995;273:1610–3.
22. Krahn M, Detsky AS. Should Canada and the United States universally vaccinate infants against hepatitis B? A cost–effectiveness analysis. Med Decis Making. 1993;13:4–20.
23. Disch DL, Greenberg ML, Holzberger PT, et al. Managing chronic atrial fibrillation: a Markov decision analysis comparing warfarin, quinidine, and low-dose amiodarone. Ann Intern Med. 1994;120:449–57.
24. Deber R, Goel V. Using explicit decision rules to manage issues of justice, risk, and ethics in decision analysis: when is it not rational to maximize expected utility? Med Decis Making. 1990;10:181–94.
25. Brett AS. Hidden ethical issues in clinical decision analysis. N Engl J Med. 1982;305:1150–2.

Glossary

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

Downloaded from mdm.sagepub.com at Max Planck Society on November 1, 2016

Best-case/worst-case scenario: A best-case scenario consists of setting all the variables at the extremes of their plausible ranges to favor a single strategy. A worst-case scenario consists of setting all of the variables so that another strategy is favored, or so that the first strategy appears as unattractive as possible.

Markov model: A decision-analytic model that characterizes the prognosis of a cohort of patients by assigning them to a fixed number of health states and modeling transitions among those states.

Bug: A structural or programming error in the tree.

Disutility: The disutility of a health state represents the negative impact on quality of life associated with the state. The disutility of a health state is, by convention, one minus its utility.

Robust: An analysis is robust if the qualitative conclusion (e.g., that therapy A is better than therapy B) is insensitive to the uncertainties in the analysis, such as quantitative estimates of probabilities or utilities.

Linkage: The explicit relationship (by the use of bindings or algebraic expressions) among probabilities or utilities in the various branches of the tree that ought to be related (e.g., the probabilities of a bad outcome with and without treatment).

Symmetry: The consistent representation of events in all strategies considered in the model. Events that occur in one strategy are represented in the same way in other strategies. The construction of symmetrical models is facilitated by using subtrees (see below).

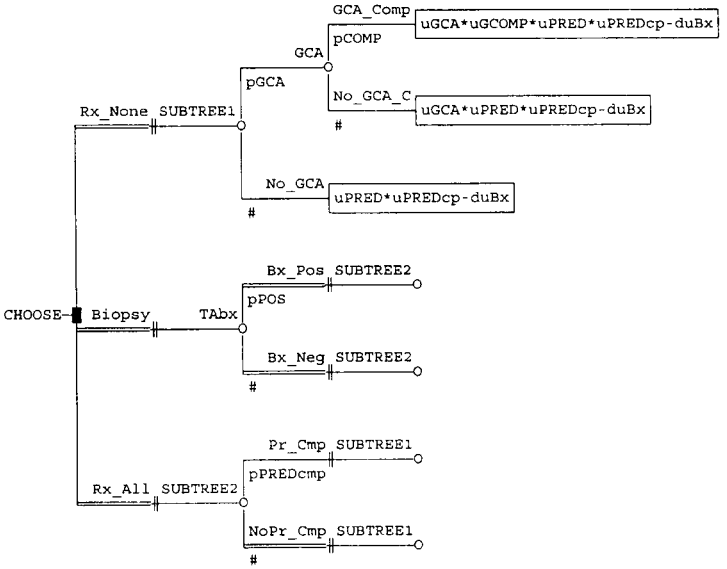
Subtree: A portion of the decision model that is repeated in various places throughout the tree. In SMLTREE or DECISION MAKER, the programmer can use the LINK function to copy subtrees at various locations.

Global values: This expression is related to SMLTREE and DECISION MAKER and refers to the quantitative estimates for all variables found in the variable list. These values are then applied throughout the tree at all times except where temporary bindings override them.

Temporary bindings: Reassigned values of quantitative estimates for specific variables that override the global bindings at various points throughout the tree. This function is particularly useful when subtrees are placed throughout the tree but quantitative estimates of the variables must differ at the various locations.

APPENDIX

How the decision tree referred to in Parts 2 and 3 of this series[2,3] is depicted in SMLTREE or DECISION MAKER, complete with local bindings and subtrees. Students should be able to replicate the tree using either of these programs.




```
Bindings from CHOOSE to Rx_None:
   duBx      := 0
   uPRED     := 1
   uPREDcp   := 1

Bindings from CHOOSE to Biopsy:
   pPOS      := SENS*pGCA+(1-SPEC)*(1-pGCA)

Bindings from TAbx to Bx_Pos:
   pCOMP     := (1-eCOMP)*pCOMP
   pGCA      := SENS*pGCA/(SENS*pGCA+(1-SPEC)*(1-pGCA))
   uGCA      := 1

Bindings from TAbx to Bx_Neg:
   pGCA      := ((1-SENS)*pGCA)/((1-SENS)*pGCA+SPEC*(1-pGCA))
   pPREDcmp  := 0
   uPRED     := 1

Bindings from CHOOSE to Rx_All
   pCOMP     := (1-eCOMP)*pCOMP
   uGCA      := 1
   duBx      := 0

Bindings from SUBTREE2 to NoPr_Cmp:
   uPREDcp   := 1
```

Variable	Tree Notation	Global Value in Tree
Prevalence of giant cell arteritis	pGCA	0.50
Probability of complications of giant cell arteritis without treatment	pCOMP	0.12
Sensitivity of temporal artery biopsy	SENS	0.80
Specificity of temporal artery biopsy	SPEC	1.00
Probability of a positive temporal artery biopsy	pPOS	—
Efficacy of prednisone in reducing the frequency of giant cell arteritis complications	eCOMP	0.89
Probability of complications due to prednisone	pPREDcmp	0.19
Disutility* of temporal artery biopsy	duBx	0.005
Utility of prednisone therapy	uPRED	0.97
Utility of prednisone complications	uPREDcp	0.75
Utility of giant cell arteritis complications (blindness)	uGCOMP	0.60
Utility of having giant cell arteritis	uGCA	0.85

*Disutility for a given health state = (1 − utility).

Downloaded from mdm.sagepub.com at Max Planck Society on November 1, 2016