Scoring Performance on Computer-Based Patient Simulations: Beyond Value of Information

Stephen M Downs*, Farah Marasigan, Vijoy Abraham†,
Barbara Wildemuth* and Charles P Friedman†
*University of North Carolina at Chapel Hill
†University of Pittsburgh

As computer based clinical case simulations become increasingly popular for training and evaluating clinicians, approaches are needed to evaluate a trainee's or examinee's solution of the simulated cases. In 1997 we developed a decision analytic approach to scoring performance on computerized patient case simulations, using expected value of information (VOI) to generate a score each time the user requested clinical information from the simulation. Although this measure has many desirable characteristics, we found that the VOI was zero for the majority of information requests.

We enhanced our original algorithm to measure potential decrements in expected utility that could result from using results of information requests that have zero VOI. Like the original algorithm, the new approach uses decision models, represented as influence diagrams, to represent the diagnostic problem.

The process of solving computer based patient simulations involves repeated cycles of requesting and receiving these data from the simulations. Each time the user requests clinical data from the simulation, the influence diagram is evaluated to determine the expected VOI of the requested clinical datum. The VOI is non-zero only it the requested datum has the potential to change the leading diagnosis.

The VOI is zero when the data item requested does not map to any node in the influence diagram or when the item maps to a node but does not change the leading diagnosis regardless of it's value. Our new algorithm generates a score for each of these situations by modeling what would happen to the expected utility of the model if the user changes the leading diagnosis based on the results. The resulting algorithm produces a non-zero score for all information requests. The score is the VOI when the VOI is non-zero It is a negative number when the VOI is zero.

INTRODUCTION

Interest in the use of computer based patient simulations in the training and evaluation of clinicians has grown in recent years because computer simulations offer interactivity and evolution of clinical problems over time in a way that is impossible with a paper based test. However, the field is only now developing rigorous methods for measuring a clinician's solution of a computer based case. [1, 2]

Ideally, assessment of performance on simulated clinical cases should measure the quality of the clinician's judgment throughout the evolution of the case. The assessment should be context sensitive so that credit for very expensive or invasive tests, for example, is given only when justified by previously discovered information. This avoids rewarding novices for being too thorough while penalizing experts for their efficiency.[3] Finally, scoring of simulated cases should also be sensitive to the relative seriousness of the misdiagnoses.

PREVIOUS WORK

Value of Information

In 1997, we developed an decision analytic algorithm for scoring clinical simulations based on the decision analytic concept of expected value of information (VOI).[4] Decision analysis compares the relative merits of alternative actions based on the expected value of the possible outcomes of those actions.[5] Decision analytic models represent the alternative courses of action the decision maker may take, the probabilistic relationships between those actions and the possible resulting outcomes, and a quantitative representation of the relative desirability, or utility, of each outcome. The product of the probability and the utility of each outcome, summed over all the possible outcomes, is the basis for comparing alternative actions. The value of any course of action is measured by this expected utility.

Prior to choosing a course of action, a decision maker may collect additional information relevant to a decision. This information may alter the probabilities of certain outcomes and, thereby, change which course of action s/he will choose. For example, a blood test may suggest or exclude a particular diagnosis depending of whether the test is positive or negative. Any information that can potentially change which course of action is best has some potential value that can be measured by the change in the expected value of the decision. We call this change the expected VOI (Chap 5 in [5]), and it is the basis on which our original algorithm scored information requests from the clinician using the case simulation. To calculate the expected VOI, we developed decision models using influence diagrams[6].

In the process of completing a patient simulation, the clinician is presented with a set of presenting symptoms. The clinician evaluates the case by requesting information from the simulation, receiving this information, evaluating it, and requesting additional information. This cycle is repeated until a diagnosis is made. Each turn of the cycle is used as an opportunity to evaluate the clinician's information request.

Influence Diagrams and the Scoring Algorithm

An influence diagram is a directed, acyclic graph, i.e., a set of nodes incompletely connected by directed arcs.[5, 6] The nodes represent three types of variables in the decision model: chance nodes, decision nodes, and a value node. Each chance node represents a random variable and the probability distribution over its sample space. Arcs entering a chance node represent conditioning variables.

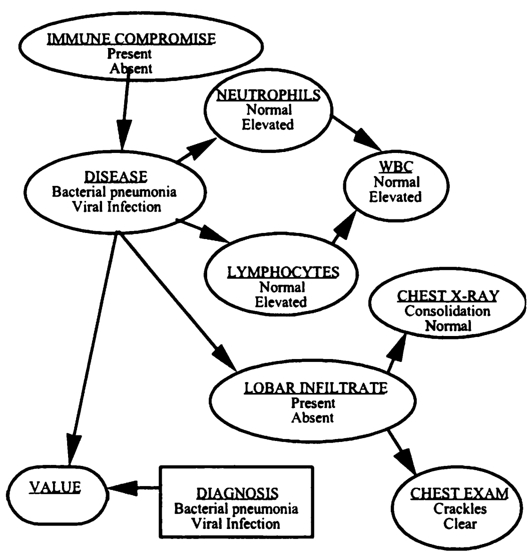

Figure 1. A simplified influence diagram.

A decision node represents the set of alternative actions that may be taken at a given time. Arcs entering a decision node represent information that will be available at the time the decision is made. The value node represents the quantitative value, or utility, placed on the outcome of the decision. Arcs entering a value node come from those variables whose value affects the overall value or utility of the outcome. Figure 1 is a simplified influence diagram illustrating the representing the diagnosis of bacterial pneumonia versus viral infection. We developed influence diagram for four clinical problems in infectious diseases.

Figure 1 also illustrates that some nodes representing findings may be derived from other nodes. In this case, the WBC, or white blood cell count, depends on both the neutrophil and the lymphocyte count. If the clinician requests a WBC without a differential count, the probability that the count will be elevated depends on the probabilities that the neutrophil and the lymphocyte counts are elevated.

Figure 1 also shows how the interdependencies between findings can be represented with hidden "state" variables. The patient may have a LOBAR INFILTRATE in the lung, and this probability depends on the underlying disease. However, the clinician cannot directly observe the presence of a lobar infiltrate. Instead, s/he can observe the presence of crackles on chest exam or consolidation on chest x-ray. These are indirect measures of lobar infiltrate. Because they both measure the same underlying process, albeit imperfectly, they provide partially redundant information. So the VOI from a chest x-ray may be substantially less after the presence of crackles has been detected on clinical exam.

A computer based patient simulation authoring and delivery program was developed using HyperCard. Eight cases, two from each of the four influence diagrams, were developed. The cases were authored by an infectious disease specialist and were based on clinical cases in his experience.

In each simulated case, clinicians are provided a case presentation, containing name, age, sex, and chief complaint of the patient. The clinician then has the opportunity to request information from a hierarchical menu of history, physical examination, and laboratory items. For each request, a log file records the item requested and the value returned by the simulation program to the clinician.

A scoring program was developed which uses influence diagrams to score the clinician's interaction with the patient simulation by determining the expected VOI from each finding the clinician requests. The influence diagrams were evaluated using the algorithms described by Shachter.[7] The algorithms were implemented in C on a Macintosh Quadra 650. For each finding the clinician requests, the algorithm calculates the expected VOI as described below. The result provided by the simulation is used to instantiate the corresponding variables, updating the influence diagram.

Before the clinician requests any findings, the scoring program calculates the expected value for each diagnosis in the decision node of the influence diagram. The highest expected value among the diseases becomes the expected value of the simulation at that point. The program then reads the log files from the simulation program.

A finding in the simulation may correspond to none, one, or many of the nodes in the influence diagram. For each finding the clinician selects, an arc is introduced into the influence diagram going from the node(s) corresponding to the finding to the diagnosis decision node. The expected value of the influence diagram is then recalculated. The difference between the expected value of the influence diagram before and after the arc is introduced is the expected VOI for the finding.

Next the value of the finding given by the simulation is used to update the probability distribution over the diseases in the disease node, using Bayes' theorem, and the node(s) corresponding to the finding is eliminated.

Early Studies

In a pilot study, a convenience sample of third and fourth year students at the University of North Carolina at Chapel Hill and the University of Pittsburgh was recruited to solve the computer based patient simulations. Students completing the simulation were also asked to make a final diagnosis. Trace files were obtained and analyzed by the scoring program. The average VOI for each simulation was calculated. This number was found to correlate with whether the student made the correct diagnosis (Figure 1), suggesting that the measure has predictive validity, but because it is not a perfect correlation, it may offer more information beyond then the final answer alone.

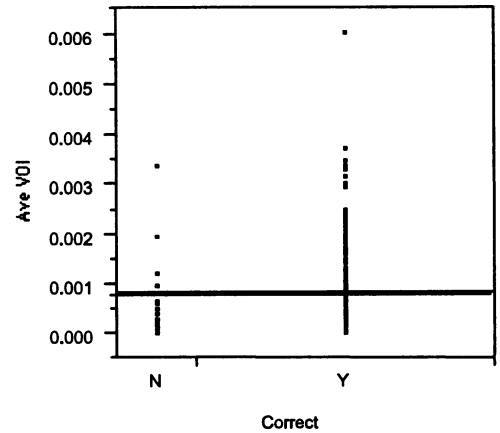

Figure 1. Correlation between average VOI and correct diagnosis

The VOI also measures the quality of the clinician's information gathering during the simulation. Figure 2 shows a graphical representation of the progress of one subject through the simulation. The horizontal axis shows the number of the information item requested during the simulation, and the vertical axis shows the item's expected VOI. This graph shows, at a glance, the points at which the data with positive VOI are obtained and their relationship to the type of data gathered (e.g., history, examination, or laboratory) or points of access to paper or computer-based information sources. In figure 2, for example, the laboratory data requested had the highest expected VOI.

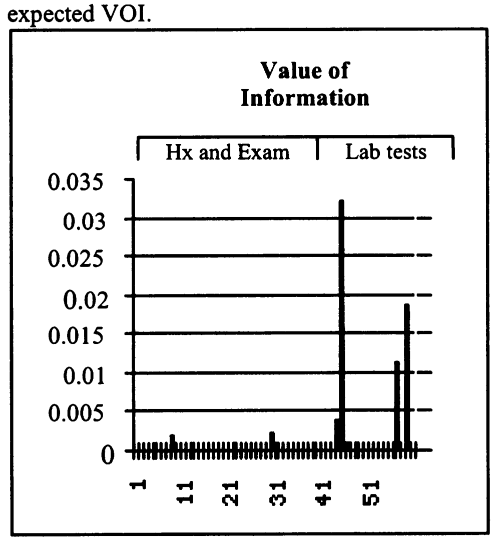

Figure 2: Graphical output of the scoring program.

However, the VOI was non-zero for only a minority (<10%) of the information items requested. The average number of findings requested per student was 64 (range 41-116). Of the findings requested, on average 5 findings (95% CI: 4, 6) had a non-zero

VOI. Because the so few of the items requested had a positive value of information even when they were relevant to the clinical problem, we were concerned that we were missing important information about the other data items requested.

NEW INFLUENCE DIAGRAM SCORING ALGORITHM

We augmented our scoring algorithm to generate a score for requested clinical data items that have zero VOI. The VOI is zero whenever the information requested will not change the leading diagnosis, i.e., the diagnosis with the highest expected value. Our strategy was to calculate the expected effect on the value of the decision model if the leading diagnosis were to change.

When the user requests a piece of information, three states may occur: (1) The requested item has a VOI; (2) the item requested my have no corresponding node in the influence diagram; (3) The item requested corresponds to a node in the influence diagram, but has a zero expected VOI because no possible value of the item would change which diagnosis has the highest expected utility. In the first case we have retained the expected VOI as the score. The algorithm was modified as described below to accommodate the latter two situations.

Requested Data do not map to the Influence Diagram

When the item requested has no corresponding node in the influence diagram, the score is the average expected utility across all diagnoses in the Diagnosis node minus the utility of the best diagnosis.

Algorithm: At the time that the influence diagram is solved by Shachter's algorithm[7], the expected utilities of all of the diagnoses in the diagnosis node are calculated. Ordinarily, only the expected utility of the diagnosis with the highest expected utility is used. In the new algorithm, we calculate the average of the expected utilities across all of the diagnoses, then subtract the expected utility of the diagnosis with the highest utility. This results in a negative score.

Rationale: It is presumed that the clinician requests information with a diagnostic strategy in mind. If s/he requests an item that has no corresponding node in the influence diagram, it is assumed that s/he has a complete misunderstanding of the clinical problem relative to the influence diagram. In this case the clinician is equally likely to make any diagnosis. The expected change in utility is the expectation across all possible diagnoses.

Requested Data Map to Influence Diagram but Have Zero VOI

When a requested item corresponds to a node in the influence diagram, but has a zero expected value of information, it is because no possible value of the item would change which diagnosis has the highest expected utility. In this case, the score is calculated by assuming the clinician will change her diagnosis if the results of the request do not favor the diagnosis. If the result favors the diagnosis, we assume the clinician will not change the diagnosis. If the result does not favor the leading diagnosis, we assume s/he will take the next best diagnosis. The score is the average (expected) utility of the best or second best diagnosis (which ever is supported by the result) weighted by the probability of the result.

Algorithm: To calculate this score, we use a slight modification of Shachter's algorithm[7]. In the last steps in the algorithm, the influence diagram is reduced to the model shown in Figure 3.

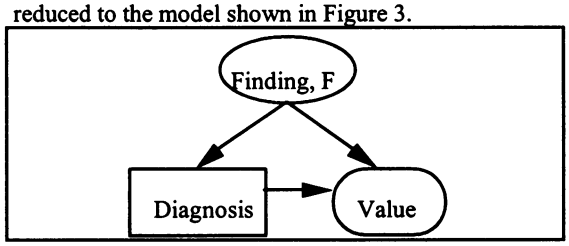

Figure 3. The configuration of the influence diagram in the last steps of Schachter's algorithm.

When calculating the expected utility across values of the requested item, we first determine whether the utility of the best diagnosis increases or decreases for each value of the item. If it increases, we average in the utility of the diagnosis with the highest utility. If it decreases, we average in the utility of the next best diagnosis. The pseudocode below summarizes the algorithm.

The difference between the expected utility of the diagram before and after the introduction of the arcs (sum - Last_Value) is the new score for the information request. It will be less than zero.

Rationale: It is assumed that the clinician understands the relationship between the item requested and the competing diagnoses, and that s/he is looking for confirmation or refutation of the leading diagnosis. Furthermore, we assume that if the result of the item requested does not support the leading diagnosis, the clinician will select one of the competitors. (Otherwise, decision theory would argue not to get the item, hence the zero value of information.)

```
Last_Value = the value of the diagram at the last
cycle.
Last_Dx = the leading diagnosis at the last cycle.
sum := 0;
For all F {possible values of item requested}
        if u(Last_Dx,F)>=Last_Value then
        sum:=sum+u(Last_Dx,F)*P(F)
    else
        Best_Dx:=
        the diagnosis with the second highest utility;
        sum:= sum+u(Best_Dx,F)*P(F);
        ;
    ;
;
return sum; {the modified expected value of the
diagram}
```

Figure 4. Pseudocode for calculating the score of an information request that maps to the influence diagram but has zero VOI.

DISCUSSION

We developed a method for assessing performance on computer based clinical simulations using the decision analytic concept of expected VOI. The use of decision theory to assess performance offers several advantages. Because the theory uses probabilistic inference, relationships in the model can be derived from the scientific literature whenever such data exist. Fundamental epidemiological concepts such as prevalence, sensitivity, and specificity are explicitly represented in the influence diagrams.

The use of utilities can potentially take into account not only the sensitivity and specificity of a finding, but also the risks and costs of a test and the risks and costs of possible misdiagnoses. All of these concepts are weighed into the expected VOI measure.

VOI as a score has face-validity and, to the extent that it correlates with making the correct diagnosis, it has predictive validity. However, VOI by itself is not sufficiently sensitive to dissect each step of the diagnostic process because most information requests in the simulation have zero VOI. The modified algorithm described here provides a score for information requests that have zero VOI. The score is negative and lower when the information requested is irrelevant to the diagnostic problem and when the cost of misdiagnosis is higher.

With the new metric the diagnostic process can be dissected. It becomes possible to compare the information seeking behavior of experts and novices or to examine the effect of external information sources on data gathering, for example.

We have extended our decision analytic approach to include a model of clinicians solving computer-based clinical simulations. Data requests that have no VOI in the classic decision analytic model contain information that can be inferred from the relationship between the item requested and the decision model of the scoring program. The resulting algorithm provides a score for every information request the user makes. The scoring algorithm has face validity, but it's value in the study of clinical decision making processes as well as certifying examinations will require further study.

Acknowledgments

This study was supported by Grant R01-LM04843 from the National Library of Medicine.

REFERENCES

1.Friedman, C. and S. Downs, *Alternatives to current practice: Decision theoretic methods,* in *Computer Based Examinations for Board Certification,* Mancall, et al, Editors. 1996, American Board of Medical Specialties: Everston.

2.Swanson, D., I. Norcini, and L. Grosso, *Assessment of clinical competence: written and computer based simulations.* Assess Eval Higher Educ, 1987. **12**: p. 220-246.

3.Newbie, D., J. Hoare, and A. Baxter, Patient management problems: issues of validity. *Med Educ*, 1982. 16:137-42.

4. Downs SM, Friedman CP, Marasigan F, Gartner G, A decision analytic method for scoring performance on Computer-based patient simulations, in *Proceedings of the 1997 American Medical Informatics Association Fall symposium* , Nashville, TN. Masys DR, Ed. 1997, Hanley and Belfus, Inc, Philidelphia.

5.Weinstein, M. and H. Fineberg, *Clinical Decision Analysis*. 1980, Philadelphia: WB Saunders Co.

6.Howard, R. and J. Matheson, *Influence Diagrams,* in *The Principals and Applications of Decision Analysis,* R. Howard and J. Matheson, Editors. 1981, Strategic Decisions Group: Menlo Park, CA. p. 719-62.

7.Shachter, R., *Evaluating influence diagrams.* Operations Research, 1986. 34(6): p. 871-82.