M a t h S o f t

$\sum$ + $\sqrt{}$ – = × $\int$ ÷ δ

StatSci Division

StatSci Research Report 19

Using GRAPHICAL-BELIEF to predict risk of Coronary Artery Disease

Russell G. Almond

StatSci (a division of MathSoft, Inc.)
1700 Westlake Ave, N., Suite 500
Seattle, WA 98109
`almond@statsci.com`

David Madigan

University of Washington
Department of Statistics

Last Revised 4/13/94

Using GRAPHICAL-BELIEF to predict risk of Coronary Artery Disease

Russell G. Almond, StatSci
David Madigan, University of Washington, Department of Statistics

ABSTRACT

Medical knowledge about risks consists of a combination of structural information about known biological facts and probabilistic or actuarial information about exposures to hazards and recovery rates. While both types of information present significant practical challenges, probabilistic information is especially difficult to use because (1) it requires constant maintenance as new studies provide new data and (2) it usually comes in the form of study results which are not ideally suited for making individual predictions.

The program GRAPHICAL-BELIEF is an environment for building and manipulating complex risk models. Graphical models can store and manipulate both structural and probabilistic knowledge and the model graph—in which nodes represent variables and edges represent relationships—is a natural visual metaphor for the more complex mathematical model. GRAPHICAL-BELIEF provides tools for both model manipulations and maintaining the knowledge bases on which the model is built.

This paper introduces GRAPHICAL-BELIEF through an extended example: a model built from a study of patients with coronary artery disease. It shows how the model can provide valuable information about risk to the patient and value of information for medical tests.

1.0 Introduction

Making appropriate medical decisions is a difficult task. It usually involves at least three steps: (1) making a comprehensive list of available options, (2) evaluating the risks and benefits of those options, and (3) selecting a "best" option from among the alternatives. Complicating this simple model of decision making are options for "testing" which may yield better diagnoses or suggest alternative treatments. Also, searching for more options must itself be considered an option, especially as there exist risks which are so great that they must be addressed immediately (emergencies).

The term *risk* is used for an undesirable consequence which may or may not occur. Note that both the undesirability of the consequence and its probability go into an assessment of risk: we say that something is riskier if either the probability is higher or the consequence is worse. The term *risk* is often more loosely used as a synonym for *probability*; this is because assessing probabilities are a first step towards assessing risks. It follows the medical decision makers require tools for assessing probabilities of risky outcomes.

There are many kinds or risks associated with medical procedures: loss of life, loss of sensory or motor functions, pain, and financial cost. Different patients assess these risks differently. For example, a concert pianist might consider quadraplegia and loss of manual dexterity almost equally bad. Furthermore, it is becoming increasingly common for a patient to express a preference against heroic options for saving his life. Such quality of life assessments can only be made by the patient, and are difficult for the patient to make unless the patient has full information about the possible risks. Recent experiments to get patients more involved in medical decisions (Kasper, Mulley and Wennberg[1992]) have shown that after a more complete disclosure of the risks, many patients opt for less expensive "watchful waiting". Such a shared decision making exercise could be very valuable in helping a patient control chronic risk factors such as excess weight, high blood pressure or cholesterol level, or smoking.

Tools for making medical decisions require methods for assessing risk probabilities. This is a challenging task because it requires assembling the information from diverse sources: findings of many studies, census and demographic information from state and national health agencies, and local experience within a particular clinic or hospital. These data must be integrated into a complete model which can be used for calculations. The calculations themselves are too complex for the physician to perform while on rounds, so they are often "simplified" into tables or simply ignored. Furthermore, the risk model rapidly gets out of date and must be updated with new information.

GRAPHICAL-BELIEF is an environment for building risk models and assessing such risk probabilities. However, it provides more than a simple calculator for risks. It also can help with planning options: finding

1

the value of medical tests and alerting the physician to possible complications or alternative therapies. It also contains tools for explanation, so that the decision maker can understand the impact of various factors on the decision. Finally, it contains knowledge maintenance tools so that the physician can update the knowledge inherent in graphical models.

This paper describes GRAPHICAL-BELIEF through a simple example. The data are taken from a coronary artery disease study of Detrano *et al.*[1989] (from the Murphy and Aha[1992] repository). The example is not meant as an exhaustive analysis of that study but rather as an example of what could be done with a tool like GRAPHICAL-BELIEF. The example provides a context for explaining the current and planned features of GRAPHICAL-BELIEF

2.0 The "Heart1" Model

The Detrano *et al.*[1989] study attempts to find clinical variables to predict the presence of coronary artery disease without an intrusive angiograph. This study measures a number of clinical and test variables for 303 patients referred for to the Cleveland Clinic for coronary angiography. The data are available through the Murphy and Aha[1992] repository.

As the current version of GRAPHICAL-BELIEF can only handle discrete variables, we divided the categorical variables into a number of categories. This gave us the following variables and codings:

Clinical Data:

Age: Coded as: `-50, 50--60, 60+`

Sex: Coded as: `Female, Male`

Rest-Bp: Systolic blood pressure at rest (on admission to the clinic). Coded as: `Low` (less than 120 mm Hg), `Moderate` (120–140 mm Hg) and `High` (more than 140 mm Hg).

Chest-Pain: Patient complaint (Detrano *et al.*[1984] describes coding). Coded as: `Typical-Anginal, Atypical-Anginal, Non-Anginal, Asymptomatic`.

Routine Test Data:

Chol: Serum cholestoral in mg/dl. Coded as: `Normal` (less than 200), `Moderate` (200–250), `High` (250–300) and `Very-High` (more than 300).

Fast-Bsug: Fasting Blood Sugar in mg/dl. Coded as: `>120, <120`.

Rest-Ecg: Electrocardiographic results at rest classified (in the source) as: `Normal, ST-T-Wave` (Having ST-T wave abnormality—T wave inversion and/or ST elevation or depression of greater than 0.05 mV), `Left-vent-Hyper` (showing probable or definite left ventricular hypertrophy by Estes' criteria)

Exercise Data:

Max-Heart-Rate: Maximum heart rate achieved during exercise. Coded as: `Low` (less than 130), `Moderate` (130–160), `High`, (more than 160).

Exer-Angina: Exercise induced angina. Coded as: `Yes, No`.

Old-Peak: ST depression induced by exercise relative to rest. Coded as: `Zero, Low` (1 or 2), `High` (more than 2).

Slope-Peak: The slope of the peak exercise ST segment. Coded as: `down` (downsloping), `flat`, `up` (upsloping).

Exer-Thal-Defects: Exercise thallium scintigraphic defects. Coded as: `fixed-defects, normal, reversible-defect`.

Experimental Non-invasive Test (this test was proposed by Detrano *et al.*[1986] as a non-invasive alternative to angiographic tests):

Colored-Floro: Number of blood vessels colored by fluoroscopy. Coded as: `0, 1, 2, 3+`.

Outcome Variables:

Health-State: Coded as: `Healthy, S1, S2, S3, S4`. This coding is supplied with the version of the data in Murphy and Aha[1992] and is not explained in Detrano *et al.*[1989]. Detrano *et al.*[1989] uses angiographic narrowing of greater than 50% as the outcome.

Healthy?: Is the *Health-State* `Healthy`? Coded as: `Yes, No`.

2

The data from this experiment support a wide variety of models with thirteen input variables and one output variable *Health-State*, (the variable *Healthy?* is just a transformation of *Health-State*). Most of these variables have three or more allowable states. If we tried to model all possible interactions, the resulting table would have 933,120 cells: over 3,000 times as many cells as subjects in the study! Obviously, the data cannot support such a complex model, but fortunately, such as complex model is usually not required. Often variables in the study are independent of one and another, or are only indirectly related through a third variable.
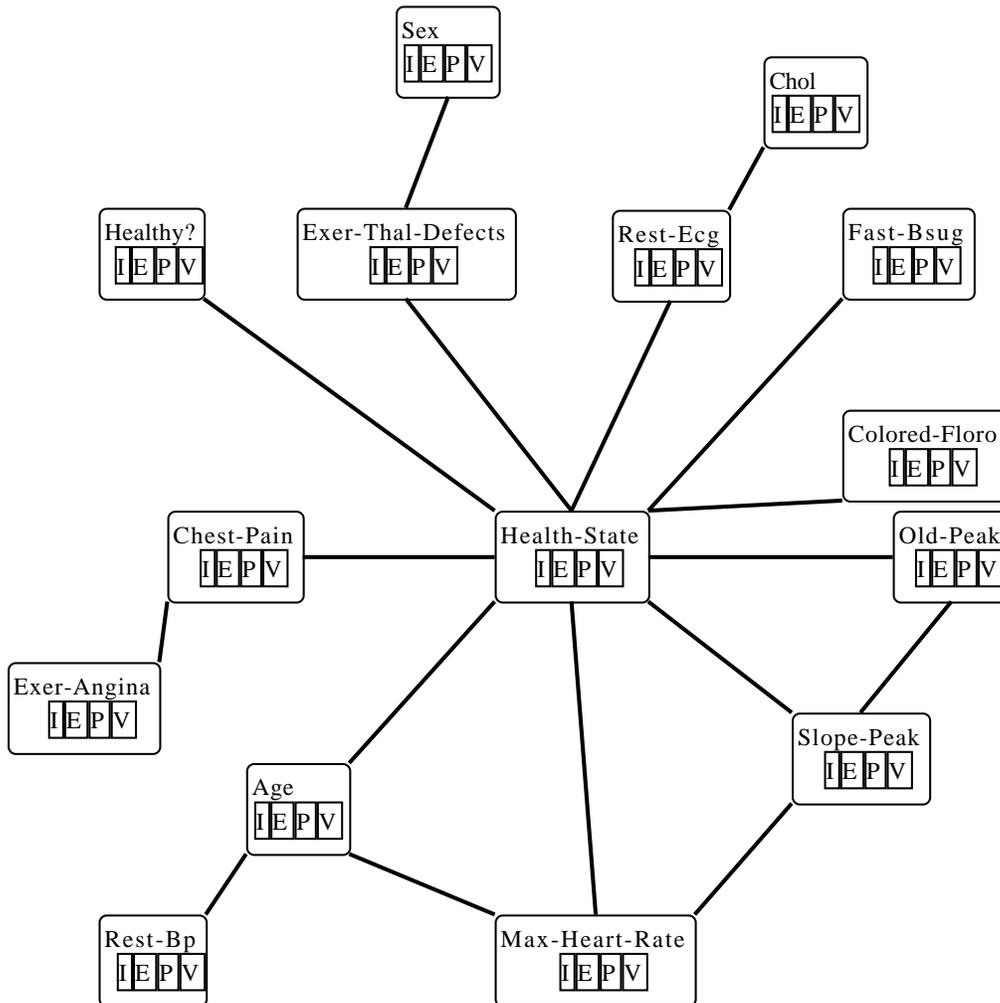

Figure 1. Undirected Graphical Model for Coronary Artery Disease
In this view, nodes (text surrounded with a round box) are variables in the model and the edges connect variables which are directly related. Variables which are not directly connected are conditionally independent, that is they are independent when given the value of the variables that separate them in the graph. Each of the variables has controls for accessing input and output views (Input value, Edit corresponding factor, attach Probe, and View marginal distribution.

A graphical model can take advantage of those independence conditions. A graph, such as figure 1, encodes the structure of a graphical model. Nodes in the graph represent variables and edges connect variables which are directly related. Variables which are not directly connected are conditionally independent; in other words, they relate only indirectly through the intermediate variables on the paths connecting them.

For example, the nodes *Health-State* and *Rest-Ecg* are joined; therefore, information about the resting electrocardiograph reading provides information about the state of coronary artery disease. As *Chol* and

3

Rest-Ecg are joined, the patients serum cholestoral level provides information about the likely result of an electrocardiograph which in turn provides information about the disease state. If we already know the electrocardiograph results, these extra test results become irrelevant.

Graphical models represent the factorization of the joint probability (or belief function) distribution of all variables in the model. For example, the joint probability distribution for the model shown in figure 1 is:

$$
\begin{aligned}
&p(\textit{Health-State, Healthy?})p(\textit{Exer-Thal-Defects}|\textit{Health-State})p(\textit{Sex}|\textit{Exer-Thal-Defects}) \times \\
&\quad p(\textit{Rest-Ecg}|\textit{Health-State})p(\textit{Chol}|\textit{Rest-Ecg})p(\textit{Fast-Bsug}|\textit{Health-State})p(\textit{Colored-Floro}|\textit{Health-State}) \times \\
&\quad p(\textit{Max-Heart-Rate, Slope-Peak}|\textit{Health-State})p(\textit{Old-Peak}|\textit{Slope-Peak, Health-State}) \times \quad\quad (1) \\
&\quad p(\textit{Age}|\textit{Health-State, Max-Heart-Rate})p(\textit{Rest-Bp}|\textit{Age}) \times \\
&\quad p(\textit{Chest-Pain}|\textit{Health-State}|\textit{Chest-Pain})p(\textit{Exer-Angina}|\textit{Chest-Pain}).
\end{aligned}
$$

This expression becomes much easier to understand if we notice its close connection to the graph. Each factor in equation 1 the factors corresponds to a *cliques* of the graph—a set of nodes which are all connected and is not contained by any larger set of nodes which are all connected. The set {*Health-State, Rest-Ecg*} is a clique corresponding to the factor $p(\textit{Rest-Ecg}|\textit{Health-State})$. The clique {*Health-State, Slope-Peak, Max-Heart-Rate*} corresponds to the factor $p(\textit{Max-Heart-Rate, Slope-Peak}|\textit{Health-State})$ (The set {*Health-State, Slope-Peak*} is not a clique because it is contained in a larger connected set of nodes.)

Directed graphical models (Figure 2) provide another way of representing the factorization. The directed model factors the model into conditional probabilities (or effectively conditional belief functions[1]) and uses the direction of the arrows to indicate the direction of the conditioning. An arrow from *Disease* to *Symptom* indicates that we are considering the causal probability of the symptom given the disease. An arrow from *Symptom* to *Disease* indicates that we are considering the diagnostic probability of the disease given the symptom.

The directed graph yields a slightly different factorization:

$$
\begin{aligned}
&p(\textit{Health-State})p(\textit{Healthy?}|\textit{Health-State})p(\textit{Exer-Thal-Defects}|\textit{Health-State})p(\textit{Sex}|\textit{Exer-Thal-Defects}) \times \\
&\quad p(\textit{Rest-Ecg}|\textit{Health-State})p(\textit{Chol}|\textit{Rest-Ecg})p(\textit{Fast-Bsug}|\textit{Health-State})p(\textit{Colored-Floro}|\textit{Health-State}) \times \\
&\quad p(\textit{Max-Heart-Rate, }|\textit{Health-State})p(\textit{Slope-Peak}|\textit{Max-Heart-Rate, Health-State}) \times \quad\quad (2) \\
&\quad p(\textit{Old-Peak}|\textit{Slope-Peak, Health-State})p(\textit{Age}|\textit{Health-State, Max-Heart-Rate})p(\textit{Rest-Bp}|\textit{Age}) \times \\
&\quad p(\textit{Chest-Pain}|\textit{Health-State}|\textit{Chest-Pain})p(\textit{Exer-Angina}|\textit{Chest-Pain}).
\end{aligned}
$$

Two of the factors of equation 1 ($p(\textit{Health-State, Healthy?})$ and $p(\textit{Max-Heart-Rate, Slope-Peak}|\textit{Health-State})$) have now been split apart. This does not change the joint distribution, only the way we express it.

Which direction we chose to place the arrows is mostly a matter of convenience. As long as there are no directed cycles in the graph, the model will correspond to the factorization of a joint probability or belief function model (directed cycles mean that we have entered the same information twice). Pearl[1988] suggests using the direction of causality, although most other authors shy away from the causal interpretation. The influence diagram school, in particular, avoids the causal interpretation and performs many calculations by reversing edges (arcs, Shacter[1986]).

GRAPHICAL-BELIEF internally represents the graphical model via its factorization. In particular, it uses a *directed hypergraph* internally. The directed and undirected models are simply views of this hypergraph. Each modeller can chose what representation to work with. Figure 3 shows a directed hypergraph, each square corresponds to a factor in equation 2. Almond[1994b] describes some of the other advantages of the directed hypergraph representation.

---

[1] Because conditional belief functions present some difficulties (Shafer[1982]), we use effectively conditional belief functions (Xu[1992]) instead. These are simple joint belief functions over all of the variables which are vacuous over the conditioning variables.
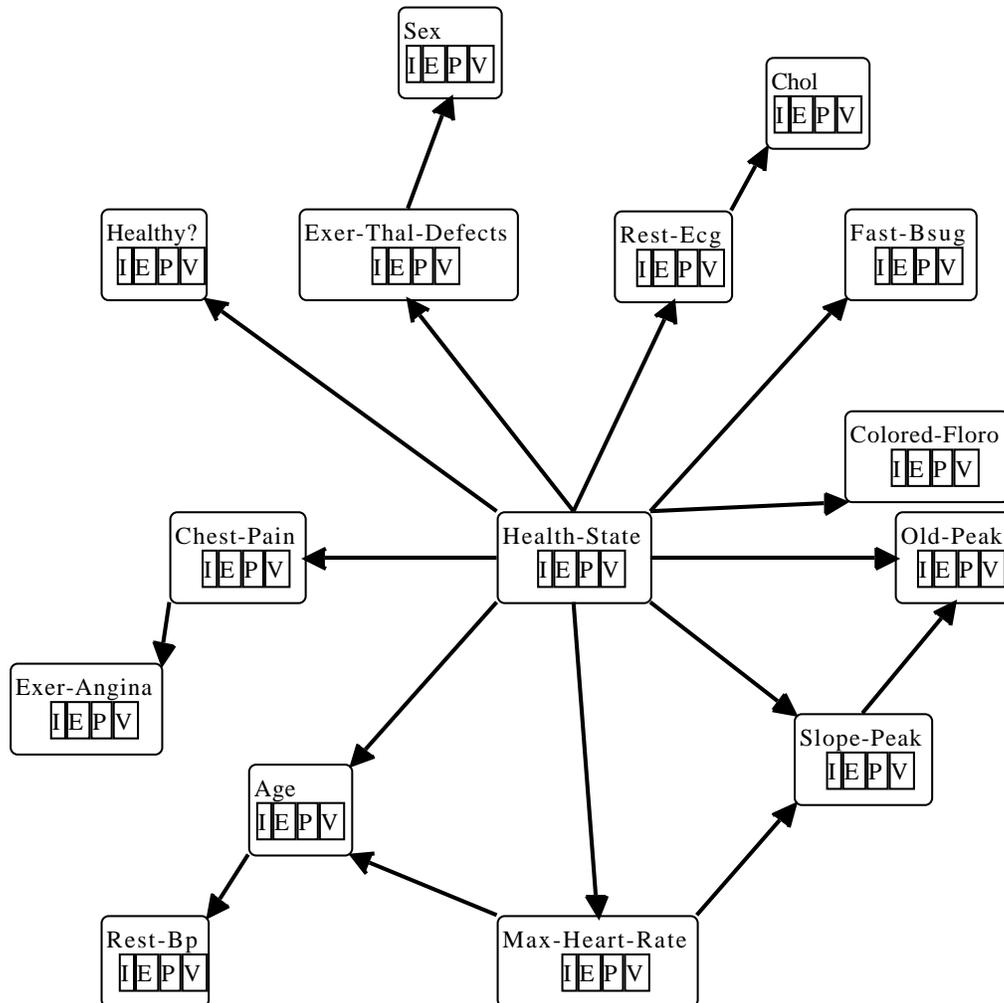
Figure 2. Directed Graphical Model for Coronary Artery Disease
In this view, variables in the model are nodes (text in a round box) and arrows show the direction of
probabilistic conditioning. In particular, the probability distribution for each variable is stored and assessed
conditionally on the value of its parents in the graph.

3.0 Using the Coronary Artery Disease Model

The previous section introduced a model for predicting the risk of coronary artery disease based on the
data of Detrano *et al.*[1989]. This section examines what we might do with that model. Section 3.1 looks
at using the model as a risk calculator, possibly in conjunction with treating a patient. Section 3.2 looks
at hypothetical questions, in particular, how the model might drive the selection of medical tests and
procedures based on value of information. Section 3.3 looks at the problem of adding interventions to the
model and discusses the value of controlling random factors. Finally, Section 3.4 discusses what happens
if the model isn't quite right and some way to look at alternatives. Almond[1994a] explains some of the
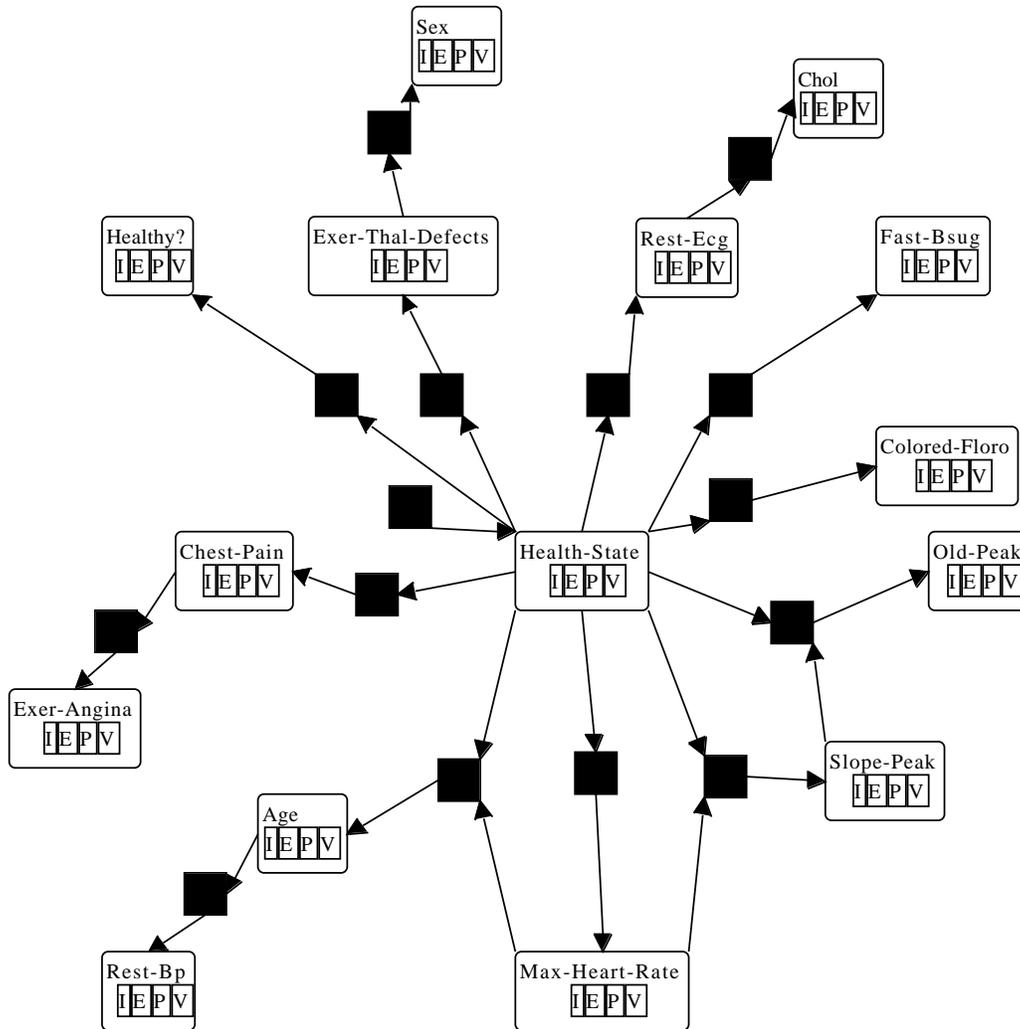motivation behind the user interface design concepts.

Figure 3. Directed Hypergraph Model for Coronary Artery Disease
In this view, nodes (text surrounded with a round box) are variables in the model and factors which make up the joint distributions are the black squares. Arrows going into the squares connect to conditioning variables, arrows going out of the squares connect to consequences.

3.1 Entering Information and Risk Calculations

Consider a patient entering a clinic complaining of problems which may indicate coronary artery disease. On admission, the patient's chart containing the patient's age and gender will be available to the clinician and the patients blood pressure is routinely taken. Furthermore, the patient's complaint about chest pain will also be available. These four pieces of information are the "clinical data" set of input variables. They could be used to form the basis of a risk prediction for that patient.

For example, suppose Mr. Jones is a 62 year old man complaining of atypical-Anginal chest pain. His blood pressure on admission was 130 mm Hg. As a clinician[2] enters this information, GRAPHICAL-BELIEF propagates the results throughout the model. The risk of coronary artery disease for this patient is 0.311. The average risk for all patients in the Detrano *et al.*[1989] study was 0.455. Figure 4 shows the change in patient's risk due to this background information.

---

[2] A link between GRAPHICAL-BELIEF and a medical records system would allow GRAPHICAL-BELIEF to automatical extract this information from the patient's chart.
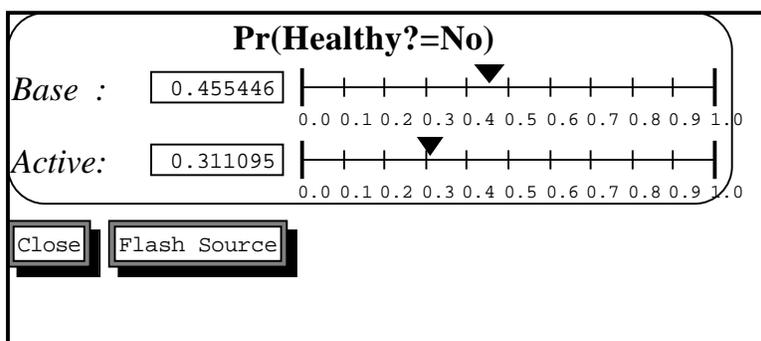
Figure 4. Effect of Clinical Variables on Background Risk
This probe shows the effect of the clinical data: Age: `60+`, Sex: `Male`, Rest-Bp: `Moderate`, Chest-Pain: `Atypical-Anginal`. The Base state is the average risk from all patients and the Active state is the specific reading for Mr. Jones. The "Flash" button flashes the Healthy? node in the graphical model helping the user keep track of the context for the probe.

We can add more information to this model as it arrives. If the clinician ordered a cholestoral screening, we could receive the information that the cholestoral concentration was moderate (225 mg/dl) and that the fasting blood sugar was > 120 mg/dl. To set these measurements we open a set-input dialog box for the appropriate variables (Figure 5 shows a typical dialog). After entering these data, the risk drops to .307.

We have been focusing on just the one outcome variable. That is not a restriction on the system. We could just as simply look at two or more outcome variables. This would be especially important in situations where there was more than one candidate diagnosis.

3.2 Value of Information: Testing

The previous section described how we would add known findings about a patient. GRAPHICAL-BELIEF is not restricted to known findings, we can enter hypothetical findings as well. This enables GRAPHICAL-BELIEF to answer complex questions, such as: "What is the value of a given medical test?"

As an example, look at the colored fluoroscopy test recommended by Detrano *ea*[1986]. There are four possible readings from this test: `0`, `1`, `2` or `3+` colored vessels. Given the information already entered about Mr. Jones (*Age*, *Sex*, *Rest-Bp*, *Chest-Pain*, *Chol* and *Fast-Bsug*) his risk of coronary artery disease is 0.308. If the result of the test is `0`, the probability that the patient has coronary artery disease drops to 0.142. If the results of the test is `1`, then the risk rises to .519. For results of `2` and `3+` the risk is .731 and .736.

If we back up to the point before we entered the serum cholestoral results, we can calculate the value of cholestoral screening as an alternative test. Before entering the cholestoral reading, the risk was 0.311. If the cholestoral was `Normal`, the risk would drop to 0.295. If the cholestoral was `Moderate`, the risk would be 0.302. If the cholestoral was `High` or `Very-High`, the risk would be 0.331 or 0.319. Obviously, this indicant is much less powerful. On the other hand, cholestoral is not a very expensive test and is often routinely administered. Therefore, it may still be considered cost effective.

Thus, colored fluoroscopy seems to be an effective test (thus confirming the results of the Detrano *et al.*[1986] study which proposed it), which cholesterol does not seem to be a strong indicator. To really judge a tests effectiveness, we must also assess the probability that the test will provide a useful distinction. In particular, a test which is a good indicator of a condition but rarely has a positive result does not have much effective discriminating power. A "goodness of test" metric which combines the discriminating power of the test against the probability of a positive result is the *expected weight of evidence*. (Good and Card[1977], Glasziou and Hilden[1989]).

The *weight of evidence* that a test result T_j provides for a hypothesis D (for disease) is

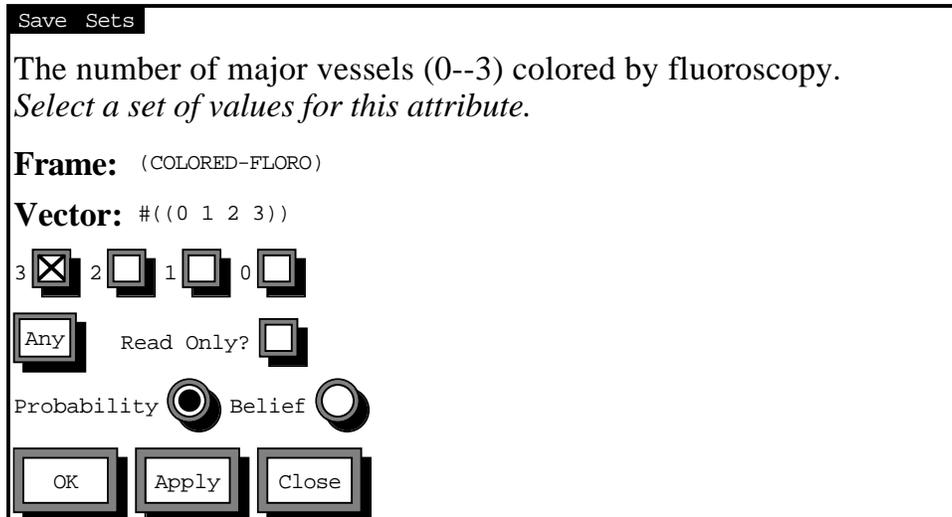$$W(D:T_j) = log \frac{P(T_j|D)}{P(T_j|\bar{D})}$$

7

Figure 5. Dialog for Setting Input Value

This dialog sets the value of the Colored-Floro variable. One or more possible values can be set for this measurement. Clicking on "OK" sets the value and closes the dialog, clicking on "Apply" sets the value and leaves the dialog open for further changes. Clicking on "Close" closes the dialog, leaving the value unset or set to the last application of apply. The "Any" button corresponds to no information about the variable being set.

where $\bar{D}$ is the negation of the hypothesis D. The *Colored-Floro* test yields the following weights of evidence:

0	-0.81	1	0.91	2	1.57	3+	1.76 ,

while the *Chol* test yield these weights of evidence:

Normal	-0.007	Moderate	-0.004	High	0.009	Very-High	0.004 .

Taking the average of these weights of evidence yields a summary of the information yield of the test. The *expected weight of evidence* is a weighted average of the weights of evidence, taking as weights the probability of the test result under the disease. Thus:

$$EW(D : T) = \sum_{j \in \mathrm{allresults}} W(D : T_j)\mathrm{P}(T_j|D)$$

In this case the expected weight of evidence for the *Colored-Floro* test is 0.524 and the expected weight of evidence for the the *Chol* screening is 0.00204. Madigan and Almond[1994] describe ways in which GRAPHICAL-BELIEF can exploit the structure of the graph when calculating the expected weight of evidence.

Finally, note that we still need to balance the cost of obtaining the information into this procedure. Although Cholestoral screening is far less informative than Colored Fluoroscopy, it is also less expensive and less invasive. To make a final decision on which of these tests to perform, we would need to take into account the competing risks of misdiagnosis, cost and invasiveness to the patient. Glasziou and Hilden[1989] (also Almond and Madigan[1994]) discuss extending weight of evidence to incorporate misclassification costs and cost of testing.

3.3 Value of Control: Treatment

Such hypothetical calculations can also be used to assess the value of interventions. Just as GRAPHICAL-BELIEF can calculate the expected information yield from test results, it can calculate the expected effect of a treatment.

Suppose that Mr. Jones had a very-high serum cholestoral in addition to the clinical variables mentioned in Section 3.1. If he could lower his serum cholestoral, that would lower his risk of coronary artery disease. Using the calculations from GRAPHICAL-BELIEF we can see that he would lower his risk from 0.319 to 0.295. As this reduction is rather small, it is probably not worth spending too much energy trying to control his cholestoral.

3.4 Sensitivity Analysis

All of the above calculations are based on a statistical model, and statistical models are only approximations of the truth. If the model was changed a little bit, it may effect our conclusions a great deal, or it may effect them only a little bit. Assessing the effect of small changes in the model is a process known as *sensitivity analysis*.

For example, Detrano *et al.*[1989] only included patients at high risk for coronary artery disease (*i.e.,* patients who were admitted to the clinic for angiography) in the study. If the clinician felt that the risk factors were lower at her clinic, she could study the sensitivity of model's predictions to the overall rate of coronary artery disease.

One of the factors in the model is the observed distributions over the variable *Health-State* for patients in the study. Because GRAPHICAL-BELIEF stores the data for that distribution, opening an editor for that factor reveals the data from the study. Figure 6 shows the initial distribution of patients in the study over various values of *Health-State.* Note in particular that 165 patients, slightly over half, were Healthy.
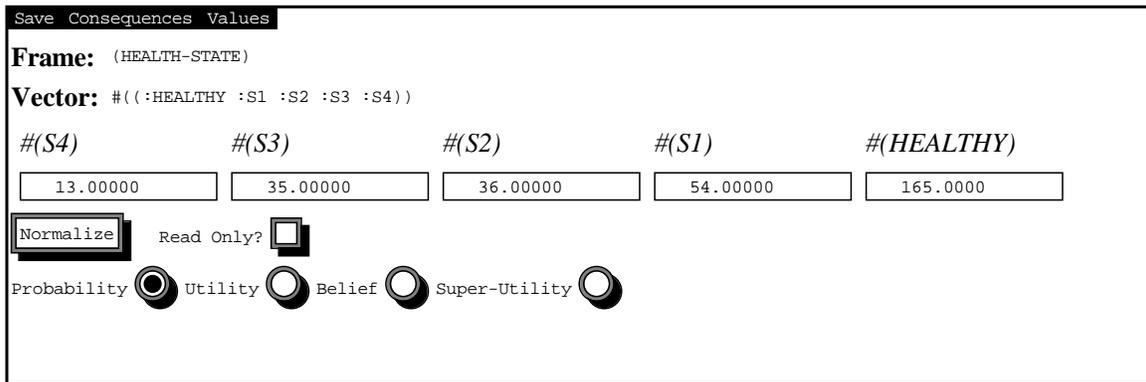
```
 Save  Consequences  Values

Frame:  (HEALTH-STATE)

Vector:  #((:HEALTHY :S1 :S2 :S3 :S4))

  #(S4)            #(S3)            #(S2)            #(S1)            #(HEALTHY)

   13.00000          35.00000          36.00000          54.00000          165.0000

 Normalize      Read Only?

Probability  O  Utility  O  Belief  O  Super-Utility  O
```

Figure 6. Value Editor for Factor over Health-State
This gadget provides a way of inputting and viewing a probability distribution. The numbers show the distribution of the Health-States among the training data. Note that the distribution has been left as a raw number of cases. The user can manually normalize the distribution by pressing the "normalize" button, or GRAPHICAL-BELIEF *will automatically normalize when calculating marginal distributions.*

For a clinician considering whether or not to refer a patient for angiography, the percentage of patients with coronary artery disease should be much less. To study the sensitivity of our conclusions to this number, we increase the number associated with healthy patients from 165 to 365. This increase the baseline proportion of healthy patients from 0.545 to 0.725. Looking at Mr. Jones, we see that his risk drops from 0.311 to 0.170.

We can also assess the value of the colored fluoroscopy under this new model. To do this we again temporarily set the value of the *Colored-Floro* value to some hypothetical values. This time, we use another feature of GRAPHICAL-BELIEF and group all of the positive outcomes of this test together. In this case the risk associated with a result of 0 is 0.083 and the risk associated with a result of 1 or more is 0.405; compare these numbers to the new baseline risk of 0.170. Thus we can see that the colored fluoroscopy test is still rather powerful; on the other hand, it may no longer be cost effective, because the patient has an overall lower probability of coronary artery disease.

4.0 Building Graphical Models

We built the the model used in this paper in five stages: (1) breaking the continuous variables into discrete variables, (2) construction of an undirected model structure, and (3) calculating the conditional distributions.

As GRAPHICAL-BELIEF does not yet contain facilities for handling continuous variables, we broke the continuous variables at their median, assigning every continuous variable two categories: High and Low. We then fit the structure of the model. After fitting the structure, we decided to redivide some of the variables into more than two categories. We really should have refit the model structure after changing the breakpoints, however, as this model is meant to be illustrative rather than accurate, we have not made that improvement. A better approach would be to directly model conditional variables (as in Lauritzen[1992]), this is planned for a future version of GRAPHICAL-BELIEF

A model selection procedure based on the Occam's Window (OW) algorithm of Madigan and Raftery [1994] produced the model of Figure X. We used the program CoCo Badsberg[1991] to provide approximate posterior model probabilities and used the OW algorithm to direct a search of model space for the model with the highest posterior model probability. Note that the analysis we present here, contrary to the suggestions of Madigan and Raftery [1994], does not account for model uncertainty.

To calculate the table of margins, we used StatSci's S-PLUS package. It returned the margins in the form of two and three way tables. We fit one table for each clique in the graph of Figure 1. Therefore, we had one margin over the the the variables *Health-State* and *Rest-Ecg* and another margin over the variables *Health-State*, *Slope-Peak*, and *Max-Heart-Rate*.

One problem that arose was that some of the tables had zeros in them. This presents a difficulty because putting a zero in the cell of a probability distribution indicates that a particular configuration *never* happens. One distinct advantage of the Bayesian approach over the classical approach is that prior knowledge can be used to distinguish between an event which is rare and hence not observed in the sample and something which cannot happen. In this example, rather than go through the exercise of specifying a prior, we applied the trick of adding $1/2$ to all entries in a table which had a zero. This is equivalent to the Jeffreys' prior for the parameter of a binomial distribution. Madigan, Gavrin and Raftery[1994] describe a mechanism for assessing prior distributions via eliciting effective data.

Again, we emphasize that the model used in this paper is meant to be illustrative rather than accurate. Future versions of GRAPHICAL-BELIEF will contain more extensive model development tools.

5.0 Possible Applications

Routine clinical tasks contain many formal and informal estimates of risk. Diagnosis involves identifying which possible conditions involve the most risk to the patient and what new tests or findings could best improve the estimates of those risks. Treatment decisions involve balancing the efficacy of the treatment with the risks of complications and side effects. Although medical professionals can use informal models of risk for routine decisions, the formal models have many advantages: (1) External reviewers can examine and validate them. (2) They are easier to update and maintain as new data and knowledge become available. (3) They form a vehicle for transferring knowledge (*e.g.,* from a specialist to a primary care provider). (4) They yield insight into the system they model.

The knowledge to produce risk predictions exists. What is lacking is the tools to extract and develop those models from experts and databases and apply them in clinical settings. GRAPHICAL-BELIEF will fill that gap.

Although building models is an intensive effort, it yields both insight into the problem studied and a risk prediction model which can be used in a number of different settings, such as:

1. *Patient Education Software.* Teaching aids could deliver patient specific risk predictions to teach about risky behaviors and to tailor educational materials.

2. *Clinical Practice Guidelines.* Good models of patient risk would be invaluable to study groups establishing clinical practice guidelines.

3. *Medical Information Systems.* Graphical models can validate data entry (Madigan *et al.*[1994]) and alert clinical staff to unusual and unlikely conditions.

4. *Medical Research Guidance.* GRAPHICAL-BELIEF can estimate both risks and uncertainty about risks; areas where the risk uncertainty was great indicate the need for more information. Almond[1993a] suggests a procedure for tracing uncertainties about risks, finding fertile areas for future research efforts.

Current emphasis on cost containment has focused on two issues: (1) eliminating procedures whose benefits don't justify the costs, and (2) moving tasks which were formerly done by a specialist in a hospital to the primary care provider in an outpatient clinic. GRAPHICAL-BELIEF offers help with both those issues: (1) it provides a way to measure the potential benefits of a proposed procedure and (2) it provides a way to transfer expertise from a specialist to a primary care provider, in particular, it will allow a primary care provider to anticipate the information needs of the specialist.

6.0 What's Next For GRAPHICAL-BELIEF

GRAPHICAL-BELIEF is a general purpose environment for building an manipulating graphical models. It is built on top of the BELIEF inference engine which handles the basic computations. The BELIEF engine will be available separately so that other software can use embedded graphical belief model intelligence. The GRAPHICAL-BELIEF environment provides additional visualization, model construction and maintenance tools. Domain specific toolkits, such as the medical research toolkit, will customize the GRAPHICAL-BELIEF interface for specific classes of users.
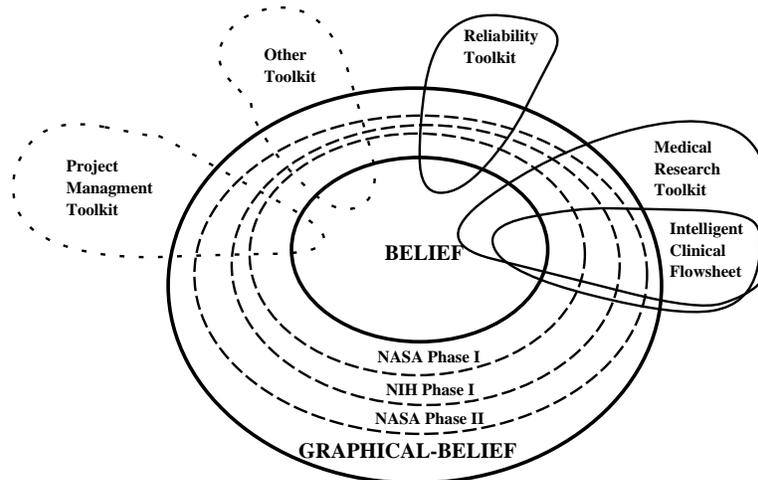

Figure 7. BELIEF, GRAPHICAL-BELIEF *and toolkits*

Figure 7 describes the growth of the GRAPHICAL-BELIEF software. Starting from the BELIEF inference engine, system has grown under Phase I funding from NASA and NIH. The NASA Phase II research contract, "Graphical Belief Models in System Reliability" (NAS 9-18908) is supporting the further development of GRAPHICAL-BELIEF as well as the construction of a toolkit for system level reliability analysis (Section E.1.2 summarizes the NASA research.) This prior development will provide a substantial base for the proposed research.

Both BELIEF and GRAPHICAL-BELIEF have served as platforms for research into graphical modelling technology. Our scientists and their collaborators are researching subjects such important topics as *test selection* (Madigan and Almond[1994]), *Knowledge Based Model Construction* (Almond[1994b]), *Explanation* (Madigan, Mosurski and Almond[1994]) and *Model Uncertainty* (Madigan and Raftery[1994]). Despite its development towards a commercial product, it will remain a platform for experimentation with state of the art technology.

A vital and dynamic research community has sprung up around graphical modelling. It has prompted interaction between statisticians and computer scientists working on artificial intelligence and combines the strengths of both approaches. Graphical modelling has many exciting potential applications; the role of GRAPHICAL-BELIEF will be to bring this exciting technology to many new users.

7.0 References

Almond, Russell G. [1990]. *Fusion and Propagation in Graphical Belief Models: An Implementation and an Example.* Ph.D. dissertation and Harvard University, Department of Statistics Technical Report S-130. To be published as a monograph from Chapman and Hall.

Almond, Russell G. [1991a]. "Building Blocks for Graphical Belief Models." *Journal of Applied Statistics,* **18**, 63–76.

Almond, Russell G. [1991b]. "Belief Function Models for Simple Series and Parallel Systems." Personal Technical Report A-6. Technical Report 207, University of Washington, Department of Statistics.

Almond, Russell G. [1992]. "ElToY: A tool for Education and Elicitation," Documentation for ElToY system, distributed through statlib@stat.cmu.edu.

Almond, Russell G.[1994a]. "Using Histories to Compare Models" StatSci Research Report 17. (Submitted for publication).

Almond, Russell G.[1994b]. "Hypergraph Grammars for Knowledge Based Model Construction." StatSci Research Report 23 (Submitted for publication).

Badsberg, J.H. [1991]. A guide to CoCo. *Research Report R 91-43*, Institute for Electronic Systems, University of Aalborg.

Detrano, Robert, Andras Janosi, Walter Steinbrunn, Matthias Pfisterer, Johann-Jakob Schmid, Sarbjit Sandhu, Kern H. Guppy, Stella Lee, and Victor Froelicher [1989]. "International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease." *American Jouran of Cardiology,* (**64**) 304–310.

Detrano R, Salcedo EE, Hobbs RE, and Yinanikas J.[1986]. "Cardio cinefluoroscopy as an inexpensive aid in the diagnosis of coronary artery disease. *American Journal of Cardiology*, (**57**) 11041–1046.

Detrano R, Yinanikas J, Salcedo EE, Rincon G, Go RT, Williams G, and Leatherman J [1984]. "Bayesian probability analysis: a prospective demonstration of its clinical utility in diagnosis coronary disease." *Circulation*, (**69**) 541–547.

Draper, David [1994]. "Assessment and Propagation of Model Uncertainty." To be read before the Royal Statistical Society (Series B).

Geiger, Dan and Heckerman, David[1992]. "Advances in Probabilistic Reasoning." In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface,* Interface Foundation of North America, pp 22-29.

Glasziou, P. and J. Hilden[1989]. "Test Selection Measures." *Medical Decision Making* (**9**), 133-141.

Kasper, Joseph F., Mully, Jr. Albert G., and Wennberg, John E.[1992]. "Developing Shared Decision-Making Programs to Improve the Quality of Health Care," *QRB*, June, 1992, pp 183–190.

Madigan, David[1992a]. "Approaches to Explanation in Bayesian Networks," Statistical Science Research Report #8.

Madigan, David[1992b]. "Temporal Reasoning with Probabilities: A Review," Statistical Science Research Report #7.

Madigan, David and Almond, Russell G.[1994]. "Test Selection Strategies for Belief Networks" StatSci Research Report 20. (submitted for publication)

Madigan, David and Raftery, Adrian[1994]. "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window." To appear in *JASA*.

Madigan, David, Mosurski, Krzysztof and Almond, Russell G.[1994]. "Explanation in Belief Networks." Under revision for publication.

Madigan, David, Jonathan Gavrin and Adrian E. Raftery [1994]. "Enhancing the Predictive Performance of Bayesian Graphical Models." (Submitted for publication).

Murphy, P.M. and Aha, D.W.[1992]. *UCI Repository of Machine Learning Databases.* Online database maintained at the Department of Information and Computer Science, University of California, Irvine, CA.

Martz, H. F. and Waller, R. A. [1990]. "Bayesian Reliability Analysis of Complex Series/Parallel Systems of Binomial Subsystems and Components." *Technometrics* (**32**), pp 407–106.

Shachter, R. D.[1986]. "Evaluating Influence Diagrams." *Operations Research,* **34**, 871–82. Reprinted in Shafer and Pearl[1990].

Shafer, Glenn [1982]. "Belief Functions and Parametric Models." *Journal of the Royal Statistical Society, Series B,* **44**, 322-352.

Speed, T. P. [1985]. "Probabilistic Risk Assessment in the Nuclear Industry: WASH-1400 and Beyond." *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer,* Le Cam, Lucien M. and Olshen, Richard A. (eds.), **1**, 173-200.

Pearl, Judea [1988]. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, California.

Xu, Hong [1992]. "A Decision Calculus for Belief Functions in Valuation-Based Systems." in Dubois, D, Wellman, M.P., D'Ambrosio, B. and Smets, Ph. (eds) *Proceedings of the 8th Conference of Uncertainty in Artificial Intelligence, San Mateo, California.* Morgan Kaufmann, 352-359.