Uncertainty and Decisions in Medical Informatics[1]

Peter Szolovits, Ph.D.
Laboratory for Computer Science
Massachusetts Institute of Technology
545 Technology Square
Cambridge, Massachusetts 02139, USA

psz@mit.edu

This paper presents a tutorial introduction to the handling of uncertainty and decision-making in medical reasoning systems. It focuses on the central role of uncertainty in all of medicine and identifies the major themes that arise in research papers. It then reviews simple Bayesian formulations of the problem and pursues their generalization to the Bayesian network methods that are popular today. Decision making is presented from the decision analysis viewpoint, with brief mention of recently-developed methods. The paper concludes with a review of more abstract characterizations of uncertainty, and anticipates the growing importance of analytic and "data mining" techniques as growing amounts of clinical data become widely available.

1 Introduction

Uncertainty is the central, critical fact about medical reasoning. Patients cannot describe exactly what has happened to them or how they feel, doctors and nurses cannot tell exactly what they observe, laboratories report results only with some degree of error, physiologists do not understand precisely how the human body works, medical researchers cannot precisely characterize how diseases alter the normal functioning of the body, pharmacologists do not fully understand the mechanisms accounting for the effectiveness of drugs, and no one can precisely determine one's prognosis.

Nevertheless, we must make important, even critical decisions about testing and treatments, and despite our uncertainties about the bases for those decisions, the decisions themselves must be definitive. The patient must choose whether to undergo the recommended surgical treatment, despite remaining doubts about whether it is truly necessary; the physician must decide what test to perform next, despite conflicting desires for various different test results. People are, in fact, often uncomfortable making decisions for which they understand the rationale to be uncertain, and they seek means to relieve these uncertainties. In medicine, for example, one typical approach is *temporizing*: putting off the final decision and hoping that additional information will evolve to make the decision more straightforward. Another is buying additional information by performing less risky or costly tests that help to reduce the uncertainties in the more critical decision.

This paper surveys historical and contemporary approaches taken by medical informatics to dealing with uncertainty and the need to make decisions in the fact of uncertainty. We begin with a characterization of the issues around which research is organized. We then review the probabilistic approach to inference under uncertainty and the decision analytic approach to reaching decisions under uncertainty. We conclude by briefly examining some interesting ideas

1. This paper was published as P. Szolovits. Uncertainty and Decisions in Medical Informatics. *Methods of Information in Medicine*, 34:111–21, 1995

that approximate or abstract away from these classical approaches.

1.1 Breadth of Issues About Uncertainty

To gain a sense of the range of issues that concern people interested in uncertainty in medicine, it is instructive to look over a quick search of Medline articles from the past year that are indexed by the term "uncertainty." The main thrust of these few dozen articles fall into four broad categories: (1) methodologic questions, (2) reducing the uncertainty in specific medical conditions, (3) patients' understanding of and response to uncertainty, and (4) physicians' coping with uncertainty.

a. Methodologic questions

A number of the articles concern analysis of different technical means of representing and reasoning with uncertainty. Examples include a comparison of Bayesian vs. non-linear least squares analysis of data, statistical methods for analysis of longitudinal data, analysis and extensions to Mycin-like certainty factor schemes, impact of error in linkage analysis, mathematical foundations of decision theory, and methods for developing prediction rules.

Other papers discuss the applied psychology of human judgments of uncertainty. For example, one examines the move from psychiatric interviews to diagnostic decisions, another considers feedback from therapy preferences to diagnostic plans, several concern accuracy and bias in human experts' intuitive judgments, and another examines violations of "rationality." One paper argues that women health care providers are less directive than men in the face of uncertainty.

Another set of papers examines methods of expressing uncertainty based on other than numerical measures. For example, several look at uncertainty based on linguistic data, one pursues logical methods in medical reasoning, and one tackles the meaning of the "non-" prefix in nosology.

A few papers concern methods for estimating uncertainty in specific situations. For example, one studies information elicitation for an image understanding system using Bayesian methods, another compares risk assessment by judgment of risks vs. physiological basis of risk estimates, and another looks at the effect of uncertain paternity on pedigree analyses in animal lines. One concerns interobserver variation in endoscopic diagnosis.

b. Reducing Uncertainty

Much of clinical research can be viewed as an attempt to reduce uncertainty in various medical situations by collecting empirical data on the natural course of disease, the accuracy of tests, the effects of interventions, etc. Basic research studies often strive to reduce medical uncertainty as well by providing better explanations of natural phenomena, and making distinctions that create more homogeneous and therefore more predictable patient subpopulations.

Not surprisingly, the Medline search turned up a number of articles of this kind, including ones examining the treatment of asthma, diagnosis of Huntington's disease, acute myocardial infarction in patients with acute thoracic pain, the effectiveness of alternate forms of ECMO (extracorporeal membranous oxygenation), modes of inheritance in breast cancer, and improvements in planning of radiation treatment.

Despite their aim to reduce uncertainty, most articles of this kind naturally fall outside the scope of a bibliographic search on the word "uncertainty," and may never actually mention the word.

For authors of such papers, the desire to reduce uncertainty is tacit and so obvious that it is "beneath mention."

c. How Patients Deal With Uncertainty

As we have asked patients to become more and more actively involved in decisions about their own medical care, we must be sure that they actually understand the situations about which they are expected to reason and decide. Decision theory gives a reasonable normative answer to how one *should* make decisions when all circumstances are understood, the likelihoods of all possible events are known, and the desirabilities of all consequences of actions have been fully explored. But even if all these factors are very well understood—a rare occurrence—patients may not be able to deal with them appropriately. Empirically, people tend to be quite poor at calculating the joint probabilistic consequences of numerous factors that influence likelihoods. They give quantitative utility judgments that are inconsistent from one time to another and they may be unable to estimate "how they would feel" under certain circumstances without actually being in them.

Two of the articles consider how well patients understand their risks and options. One concerns patients' attitudes and expectations in facing genetic counseling. The other examines patients' understanding of the specific risk factors and their consequences in two genetic disorders, retinitis pigmentosa and choroideremia.

Interestingly, four papers discuss the negative impact of the anxiety caused by uncertainty itself; it may affect the patient's mood and willingness to cooperate. One pursues this topic in the context of nursing for arthritis patients. Another studies the effect of using euphemisms for describing their disease to patients with cancer. Another discusses truth-telling among American doctors in the care of dying patients.

d. How Physicians Deal With Uncertainty

A few of the papers deal specifically with how physicians themselves cope with the constant uncertainty under which they must make decisions. One paper, for instance, considers which factors influence the doctor in deciding whether to prescribe medication for a patient or not. Another interesting study subjects medical residents to the patient experience (by actually admitting them, incognito, to the wards) to have them learn the patient's view of the confusing health care system.

1.2 Treatment of Uncertainty in Medical Information Science

Although researchers in medical information science (medical informatics) and medical artificial intelligence (AIM) have addressed most of the above concerns to some extent, the bulk of the work in this field has concerned methods for properly representing uncertainty in diagnostic problems and for making inferences and decisions in the face of uncertainty. In addition, a substantial amount of effort has been focused on other methodologic concerns, such as how to elicit models from experts accurately, how to deal with possible inaccuracies in even experts' judgments of uncertainty, and how to take into account the human need to restrict the number and complexity of diagnostic hypotheses.

Efforts to quantify uncertainty in specific medical domains have typically been treated as a separate problem, to be solved by others, though many of the early projects in AIM [22,23,35] did in fact build substantial knowledge bases that included pseudo-probabilistic models of uncertainty (*vide infra)*. The important issues of how patients and doctors deal with the psychological stresses of uncertainty have not received much attention, though preliminary studies such as [10] point to interesting ways of incorporating such concerns into a program's interface.

Medical informatics has changed tremendously over the past few decades, and changes in the approach to uncertainty are probably not the most important advances in the field. In [37] we discuss some of these other important changes. Perhaps the most important is the envisioned role of computer programs in health care. Twenty years ago most researchers expected programs to play the role of expert consultants, sought out by practicing physicians and nurses to evaluate and analyze difficult cases. Today, we instead expect background use of programs to detect potential errors and to recognize treatment opportunities, coupled intimately to evolving comprehensive medical record systems.

2 Probabilistic Models of Uncertainty

From the vantage point of the 1990's, it appears that the most appropriate analysis of uncertainty should be organized around formal probabilistic models, at least as a point of departure. Even if representation and reasoning models radically different from probability theory are adopted, it is essential to be able to compare them with each other, and probability theory serves as an often-useful language of comparison. In addition, probabilistic models have often served, from the earliest work, as the actual representation and inference method for a number of computer programs that have been built to do medical reasoning [16]. Our choice of focus in this paper is not meant to exclude or reject other formalisms such as the Dempster-Shafer theory of evidence [34] or the large body of work based on Zadeh's fuzzy set theory [40].

This point of view is somewhat controversial, because it does assign a primary role to probability theory. The issues are similar to those discussed in the field of artificial intelligence, where many researchers propose a similar role for predicate logic [7].

2.1 The "Idiot Bayes" Formulation

For many diagnostic problems, it is reasonable to assume that there is a single cause of the patient's problems, and that it must be one of a set of known hypotheses. Furthermore, it is often also convenient to assume that the findings or symptoms associated with diseases are conditionally independent. These two assumptions have come to be known by the charming sobriquet "Idiot Bayes," referring to the simplicity, not necessarily the innate quality of the method.

Given an *a priori* probability distribution over a set of possible states of the world (e.g., alternative diagnoses), and a *conditional* probability distribution telling how the outcome of a test depends on the states of the world, we can use the Reverend Bayes' theorem[1] to compute the *a posteriori* probability after we know the outcome of the test. Let $D_1, \ldots, D_n$ be the possible states of the world. We assume that they are exhaustive (i.e., no other states are possible) and mutually exclusive (i.e., only one state can hold). These assumptions assure that

$$\sum_{i=1}^{n} P(D_i) = 1. \qquad (1)$$

The assumption of exhaustiveness requires that the model be complete for the population of problems to which it is to be applied. Violations of this requirement can lead to nonsense re-

[1] Bayes, a nonconformist minister, developed his theory as a formal means of arguing for the existence of God. The role of symptoms in this argument is taken by the occurrence of miracles and other manifestations of God's good works, and the two hypotheses are the affirmation and denial of God.

sults, because—by definition—one of the possible hypotheses must be correct. Unless "other" is a valid hypothesis, a problem inappropriately brought to a program assuming exhaustiveness of its hypothesis set will be inappropriately diagnosed. But "other" is a difficult hypothesis to characterize—what is its prior likelihood, and what are its predicted manifestations? The second assumption, that only one state can hold, is only mildly troubling if only acute disorders are considered. This is because it is actually rather unlikely that two independent acute diseases will strike an individual at the same time. If chronic disorders are also under consideration, however, then the assumption of mutual exclusivity is often erroneous [21], and will tend to make it difficult to identify all of a patient's problems.

The conditional probability of a symptom S given D is simply the probability that S occurs when D occurs. We write $P(S|D_i)$ for the conditional probability of S given that D_i is the true state of the world, and this is equal to $P(S, D_i)/P(D_i)$. Bayes' rule is then:

$$P(D_i|S) = \frac{P(D_i)\,P(S|D_i)}{P(S)} \tag{2}$$

where by the closed-world assumption that S must be caused by one of the D_i,

$$P(S) = \sum_j P(D_j)\,P(S|D_j) \tag{3}$$

When using Bayesian inference where there are many possible independent observations (symptoms) possible, it is common (though not necessarily correct) to assume *conditional independence*. Mathematically, two symptoms, S_1 and S_2 are conditionally independent just in case

$$P(S_1, S_2|D) = P(S_1|D)P(S_2|D) \tag{4}$$

Intuitively, the symptoms are conditionally independent if they are linked only through the diseases that cause them, but are otherwise unrelated.

If we observe two different symptoms, it is possible to apply Bayes' rule once to calculate the posterior probability distribution for diseases by using this last formula. This is equivalent to treating the two symptoms together as a single compound symptom, and using that formula to compute the conditional probability of the compound symptom given each disease from the conditional probabilities of the individual symptoms alone. Of course if we have reason to believe that the symptoms are not conditionally independent, we can directly estimate the joint conditional probabilities, which are in that case no longer just the product of the individual ones. Note, however, that if many subsets of a large number of symptoms are conditionally dependent, we must estimate a vast number of joint conditional probabilities.

Instead of forming a single compound symptom from all the ones observed, we can adopt a *sequential Bayesian* inference method [8]. After observing any symptom, say S_l, we use Bayes' rule to compute the posterior probability $P(D_i|S_l)$. Then, we may treat that posterior as a new prior probability, corresponding to the likelihood that the patient has D_i if he is selected from a population of people all of whom have the symptom S_l. We then repeat the same steps for all other observed symptoms, at each step using the posterior probabilities as the new priors for the next step. If conditional independence holds, we can show that this is equivalent to computing conditionals for the compound symptoms, as suggested above. Its advantage is that it allows us to consider doing the diagnostic reasoning step by step, and therefore to reason between steps about which would be the most useful symptom to try to observe next.

Especially in sequential application, it is very convenient to reformulate Bayes' Rule to use odds likelihood ratios rather than direct probabilities. If we compute not only $P(D|S)$ but also $P(\overline{D}|S)$, it is clear that they share the same denominator. If instead of asking the posterior probability of D, we ask for the posterior *odds* of D, we have the odds-likelihood form of Bayes' Rule:

$$O(D|S) = L(S|D)O(D) \tag{5}$$

where the *prior odds* of D is given by

$$O(D) = \frac{P(D)}{P(\overline{D})} = \frac{P(D)}{1 - P(D)} \tag{6}$$

and the *conditional odds* (or *likelihood ratio*) of the symptom given the disease is given by

$$L(S|D) = \frac{P(S|D)}{P(S|\overline{D})} \tag{7}$$

Incidentally, there is something potentially troublesome in a term such as $P(S|\overline{D})$ when D is not a binary variable. If D is k-ary, then it must take on one of the values $d_1, \ldots, d_k$. If, say, d_1 is the distinguished value whose absence we denote by $\overline{D}$, then $\overline{D}$ must correspond to some superposition of the states in which D has the values $d_2, \ldots, d_k$. But then $P(S|\overline{D}) = P\left(S|\overline{d_1}\right) = \sum_{j = 2, \ldots, k} P(d_j)P(S|d_j)$. This expression is not, in general, a constant, but will vary with the relative likelihoods of the *other* possible values of D. Assuming it to be a constant is, in fact, an additional assumption that does not follow from the more typical assumption of conditional independence, but a number of papers in the literature do not appreciate this.

This formulation leads us to talk about "three-to-one odds" instead of "75% probabilities", but has the advantage that successive applications of Bayes' rule consist merely of successive multiplications by the likelihood ratios corresponding to successive observations, if those observations are conditionally independent. A further advantage of the odds-likelihood notation is that people untrained in probabilistic reasoning appear to be somewhat better at being able to estimate odds corresponding to very small and very large probabilities than at estimating the probabilities themselves—though this is a controversial claim. For example, the difference between a 98% or a 99% probability of some event seems small to many people, but actually corresponds to a difference in odds between 50:1 and 100:1.

When conditional dependence is present, we need to use joint likelihood ratios, in exact analogy to joint conditional probabilities. For example, if S_1 and S_2 are not conditionally independent given D, then we must directly estimate $O(S_1, S_2|D)$, which will not, in general, equal $O(S_1|D)O(S_2|D)$.

Pearl [25] points out a very important conceptual use of the likelihood ratio in interpreting uncertain evidence. When evidence is unreliable, the best way to express its significance may be to judge how much more likely it is in the presence of an hypothesis than in its absence. Pearl imagines a drunken neighbor's report of a burglar alarm sounding; even though it is difficult to model just how the neighbor's drunkenness and the alarm sound might interact to cause him to make this report, we may summarize its effect by saying that it is twice as likely to occur if the alarm is in fact sounding than if not. This corresponds to conditional odds of

2:1, which can be used directly in (5) to update the likelihood of the alarm having sounded (and ultimately of a burglary having taken place).

If the odds form of Bayes rule is transformed by taking the logarithm of both sides of the equation, we arrive at a formula that computes the log odds of a hypothesis as the sum of the log likelihood ratios of the observed evidence. In this formulation, each piece of evidence contributes a *weight of evidence* to the hypothesis. Findings that have a likelihood ratio greater than one (i.e., that are more likely to be found if the hypothesis is correct than if not) will contribute a positive weight, whereas those with a likelihood ratio less than one contribute negative weight. Equivocal evidence, which is neither more nor less likely depending on the presence of the hypothesis, contributes nothing.

Many early AIM programs in fact used additive scoring techniques to estimate the likelihood of various hypotheses. The *frequency* measure used in Internist [22], for example, is best interpreted as a scaled log likelihood ratio.

The approach summarized above is the core of most probabilistic reasoning systems. The restrictive assumptions make efficient computation possible and make it relatively easy to acquire the needed probabilities, but also make the models unrealistic for many real-world problems.

2.2 Bipartite Graph Models

Our original discussion of diagnostic reasoning with probabilities made what have been called the "idiot Bayes" assumptions: exactly a single disease, and all symptoms conditionally independent. Graphically, this is represented by a diagram such as that of Fig. 1

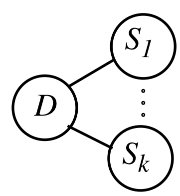

Figure 1. Bayes network for the "idiot Bayes" model. The node D is a random variable whose value is the patient's diagnosis; its possible values form an exhaustive and mutually exclusive set. Each of the nodes S_i represents a possible observation or test result. The S's are all conditionally independent.

In general, we often want to consider domains in which there may be more than one disease at once. One well-known and useful model is the bipartite graph, in which any number of diseases may cause any number of symptoms. Any combination of diseases may occur together, and they are probabilistically independent, and the symptoms are also conditionally independent—their likelihood is influenced directly only by the likelihood of the diseases that may cause them. Fig. 2 shows a simple bipartite graph representing four diseases and seven symptoms. A probabilistic interpretation of such a model requires that we estimate or measure conditional probabilities of symptoms based on *all* the diseases that may cause them. When there is only one (e.g., S_3), the situation is similar to what we discussed above. For a symptom with multiple possible causes, however (e.g., S_1), we need to determine the probability of the symptom given each combination of possible states of the diseases that may cause it, e.g., $P(S_1 | D_1, D_2, D_3)$. Given the prior probabilities of the D_i and these relevant conditional probability tables, we may continue to apply a method very much like Bayes' rule to compute the probabilities of each of the D_i after having observed any of the S_j. For large numbers of dis-

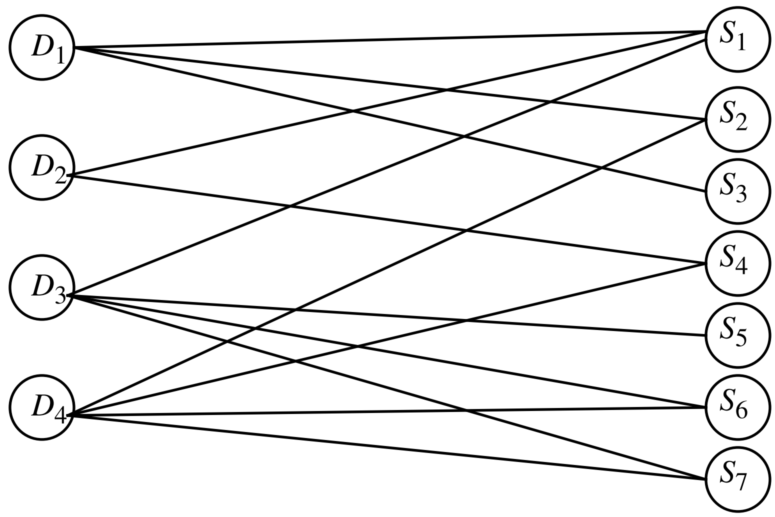

Figure 2. A bipartite graph model showing four diseases and seven symptoms they may cause. Each of the D_i represents the presence or absence or degree of severity of a disease; thus, unlike in the model of Fig. 1, this model can represent the presence of any combination of these diseases. There is no direct probabilistic dependence among the disease nodes. The nodes S_i represent symptoms (or other observables), and may, in general, depend in complex ways on the diseases. As before, the S's are conditionally independent.

eases and symptoms, however, both the sizes of the conditional probability tables and the computation needed grow exponentially and become impractical.

2.3 Noisy Or

One strong assumption, the *noisy-or* assumption [25], can simplify at least the data estimation problem and is widely used. It assumes that the probability that some set of diseases causes a symptom is just the probability that at least one of the diseases does so. The contrapositive is easiest to see: the only way for a symptom to be absent (given this assumption) is if none of its possible causes actually caused it. Thus,

$$1 - P(S_1 | D_1, D_2, D_3) = (1 - P(S_1 | D_1))\ (1 - P(S_1 | D_2))\ (1 - P(S_1 | D_3)) \qquad (8)$$

This assumption is legitimate only if there is no special reason to believe that particular combinations of causes of a symptom make it more or less likely. If appropriate, it reduces the amount of data needed to characterize a conditional probability table for a symptom with k possible causes from 2^k to k, a very significant saving.

There is actually a subtle problem in the formulation of noisy-or above, having to do with how we estimate conditional probabilities like $P(S_1 | D_1)$. The probability that some disease *causes* a symptom is not generally equal to the probability that the symptom is present if the disease is. The latter will generally be a higher number, because it is possible that some other cause of the symptom was also present and actually caused the symptom. When we estimate causal probabilities, we really mean "the probability that S_1 occurs given that D_1 is present, but no other diseases are." This is

$$P_c(S_1 | D_1) = P(S_1 | D_1, \text{all other} D_i \text{ absent}). \qquad (9)$$

The noisy-or approximation then applies (8) using these causal probability estimates for the conditional probabilities.

Another complexity that arises in models such as these is that, because of known but unavoidable incompleteness of the model, a symptom may in fact occur in the absence of any of its possible causes. A simple but apparently adequate approach to this problem has been to estimate an additional (constant) *leak* term for the noisy-or, which represents the probability of the symptom arising from the everpresent but unmodeled background.

Noisy-and is a similar modeling assumption, corresponding to situations in which all of a symptom's causes are necessary for it to appear. For example, a person must be both susceptible to and exposed to a pathogen for its effects to be present; even if both are true, the effect may still fail to arise. There have been proposals for additional simplifying models for how multiple causes interact to provide the probability of a joint effect, though they are used far less often than noisy-or. The most interesting correspond roughly to the causal models of [29].

2.4 More Complex Topologies: Polytrees

There are far more complex causal topologies than bipartite graphs, of course. For example, one disease may predispose to developing another, which, when present together with a third, may cause some syndrome, which in turn may cause a constellation of symptoms. Furthermore, the appearance of some symptoms may depend not only on diseases and syndromes but also on whether other symptoms are present as well. For example, one symptom may mask another.

There is an important special case of these topologies, wherein there is at most one path between any two nodes in the network—such networks are called *polytrees*. This condition assures that any prior knowledge or observation can influence any other node via but one path. Then, a slight generalization of the sequential updating methods described above will successfully calculate the revised likelihood of any node in a time proportional to the diameter of the graph [25], which is very good. Polytrees provide an important and useful class of models, because they permit the aggregation of multiple pieces of evidence, "causal" chains of reasoning, and "explaining away" evidence, yet can be computed efficiently. However, these polytree methods fail as soon as we draw additional links in the graph, making multiple pathways for evidence propagation possible.

2.5 Bayes Networks: General Graphs with Undirected Cycles

Bayes nets cannot contain directed cycles, because no node can be part of its own cause.[1] Undirected cycles mean simply that there may be multiple paths of influence from one node to another. The occurrence of such paths makes local propagation schemes fail either by failing to converge (each of two nodes keeps making the other more likely, which in turn makes the first more likely, etc.), or converge to the wrong value (effectively "double"-counting evidence that is accessible by multiple paths).

An appropriate formulation of the probabilistic inference problem in this general case was only worked out in the early 1980's, mostly by Pearl and his students [25], though specialized versions for pedigree analysis were understood by geneticists by the mid-70's [6,19].

Consider a small example of a network of this type, shown in Fig. 3, where for simplicity we will assume that each node is binary: it is either present or absent.

In Bayes nets, it is actually the *absence* of links that is significant. Thus, the above network

[1] At least such is out understanding of the common sense world. When people find themselves tempted to create such cycles, it is usually because they are improperly ignoring time. For example, the amount of money I have tends to influence the amount of money I *will* have, but not the amount I have now!

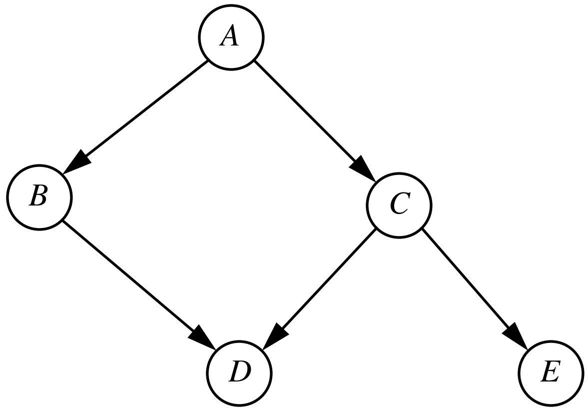

Figure 3. Cooper's MCBN1 network [3] is a simple but general Bayes network. *D* depends on *A* both through *B* and *C*, but is made independent of *A* if *B* and *C* are both known. The absence of arcs denotes independence. The original network represents the dependence of coma (*D*) and headaches (*E*) on metastatic cancer (*A*) via a brain tumor (*C*) and changes in serum calcium (*B*).

says that *E* does *not* depend on anything but C. Therefore, if we know whether C is present, it makes no difference to the likelihood of *E* whether *A, B,* or *D* are present or no.

We can easily compute the joint probability of any particular set of values for the five nodes. For example, the probability that all five are present is given by

$$P(A, B, C, D, E) = P(A)P(B|A)P(C|A)P(D|B, C)P(E|C) \tag{10}$$

Similarly, the probability that *A, C* and *D* are present but *B* and *E* are absent is

$$P(A, \bar{B}, C, D, \bar{E}) = P(A)P(\overline{B}|A)P(C|A)P(D|\bar{B}, C)P(\bar{E}|C) \tag{11}$$

In general, if the nodes of a network are X_i and the set of parent nodes of X is $\pi(X)$, then the probability of any particular set of values for all the nodes in the network is given by

$$P(X_1, \ldots, X_n) = \prod_i P(X_i|\pi_{X_i}) \tag{12}$$

For nodes with no parents, $P(X_i|\pi_{X_i})$ should be read as just $P(X_i)$, the prior probability.

Usually, we are not interested in the probability of some state of the world in which all nodes have specific values, but in a partial state of the world in which only *some* nodes have specific values. We might want to know, for example, what is the probability that the patient has some specific disease even when we do not know the presence or absence of all symptoms, syndromes and other diseases. In this case, we must simply sum over all possible states of those nodes whose value we do not know. Thus, if we do not know the values of the first k of the n nodes, then

$$P(X_{k+1}, \ldots, X_n) = \sum_{X_1, \ldots, X_k} P(X_1, \ldots, X_n) \tag{13}$$

The sum is taken over all possible combinations of values of the nodes $X_1, \ldots, X_k$, and each term of the sum may be computed by (12).

In the simple example above, the probability that *A* is present and *E* is absent (over all possible states for *B, C,* and *D*) is thus given by

$$P(A, \bar{E}) = \sum_{B, C, D} P(A, B, C, D, \bar{E}) \quad (14)$$

and the summation is over the eight terms corresponding to all possible combinations of values of *B, C,* and *D*. By (10), each term requires looking up five conditional probabilities and multiplying them together. Thus, we will perform 8x4=32 multiplications if the computation is done naively.

2.6 Speeding Up the Computation of Probabilities

Fortunately, there are many independencies in most real Bayes networks. These correspond mathematically to being able to factor expressions like (14). For example, one reasonably good factorization of that equation is

$$P(A, \bar{E}) = P(A)\sum_{C} P(C|A)P(\bar{E}|C)\sum_{B} P(B|A)\sum_{D} P(D|B, C) \quad (15)$$

This version of the calculation requires only 9 multiplications, because it recognizes facts such as that the dependence of $\bar{E}$ on *A* through *C* does not depend on *B* or *D*. Such savings become spectacular for very large networks, if the factorization is done well.

Most approaches to solving this problem have viewed it not as a symbolic algebra problem (as presented above) but as a graph-separation problem [18]. Most algorithms attempt to find small subsets of the original graph whose nodes, if known, render other large subsets independent. For example, in MCBN1, above, knowing the value of *C* assures that the probability of *E* does not depend on the values of *A, B,* or *D*, and vice versa. This observation corresponds exactly to the factorization of the algebraic formula in (15), and the correspondence is universal: for every good graph separation there is a good factorization, and vice versa.

Unfortunately, most problems of interest about the general case of Bayes networks are NP-hard, including exact solution of a network and optimal structuring of the computations—optimally factoring a large expression or finding optimal graph separators [4]. Thus, the best algorithms are heuristics that seem mostly to do a very good job, but that do not guarantee the best possible results.

Gaussian networks are an interesting special case of efficiently-solvable Bayes networks [32]. In these, instead of nodes taking on discrete values and probabilities being described by conditional probability tables, we assume that all nodes take on continuous values whose probabilities are normally distributed around a mean, with some standard deviation. Conditional probability distributions are, then, multidimensional normals, and it turns out that polytree-like propagation algorithms that adjust the means and standard deviations of their neighbors rapidly solve the inference problem.

2.7 Sampling Methods Yield Approximations

Even an excellent factorization can leave very large networks too hard to compute. For those cases, many approximation algorithms have been developed that work by statistical sampling. Taking our small example again, imagine repeating the following operation a large number of times:

1. Assign + or – to *A* based on choosing a random number between 0 and 1. If the number is $\leq P(A)$, then say *A* is present, otherwise absent.

2. Assign + or – to B, again using a random number, but this time using the threshold given by $P(B|A)$ or $P(B|\bar{A})$, depending on whether A was assigned + or –.
3. Do the same for *C, D,* and *E*.
4. Determine whether the condition of interest (e.g., *A* present and *E* absent) is true in this sample.

The probability of the condition of interest will be approximately the fraction of times in which it was true in Step 4. Larger sample sizes usually lead to more reliable approximations. Alas, the rate at which such sampling methods converge to the true answer may be very slow. Variations on the above have been explored where new samples are biased to be near previous ones; typically, improvements in convergence rate come at the cost of greater likelihood of completely missing the best solution. Repeated simulated annealing methods may help; they perform local search from many scattered starting points, climbing up gradients except for occasionally heading down, hoping to find less local maxima.

The evaluation of large Bayes networks is an area of much current research (e.g., [33]), and many groups are exploring useful applications. Only some are diagnostic. For example, several groups in natural language processing use Bayes nets to disambiguate word senses, parse spoken speech, and predict which journal articles are likely to be of interest to a person based on the previous articles in which they have shown an interest.

2.8 Learning Bayesian Networks from Data

One of the most exciting current research areas is the attempt to learn Bayesian network models directly from data. The *knowledge engineering* problem—how to construct useful models that embody expertise—is known to be very difficult. Being able to derive such models automatically from data would be a great boon to system construction efforts, and might lead to better systems that are not so sensitive to the biases and limitations of their creators. Furthermore, the advent of comprehensive hospital information systems is making available ever-larger collections of actual clinical data, and thus making the direct assessment of interesting probabilistic relationships feasible.

One of the simplest ideas in this vein is to update the likelihoods in an existing structural probabilistic model by adjusting the conditional (and maybe the prior) probabilities based on actual data. Spiegelhalter [36] suggests indicating the model builder's confidence in a probability by expressing it as a ratio of the number of times the indicated event would occur in a number of trials. Thus, a conditional likelihood of 1/2 has the same numerical value as one of 500/1000, but shows far less confidence in the number. Then, actual data from subsequent experience with the system can be used very directly to update the probabilities. One negative sample would change the initial 1/2 estimate to 1/3, but it would change the more confident 500/1000 estimate only to 500/1001, an almost negligible change. As the amount of actual data increases, the model's initial probability estimates are adjusted toward and eventually supplanted by estimates from the data. Although this is very desirable, problems of selection and ascertainment can nevertheless corrupt the model.

More dramatic opportunities are explored by systems that can learn not only the parameters of a Bayesian network but its structure as well. Pearl and Verma [26] exploit statistical evidence of independence among random variables and asymmetries in the independence relations among triples of variables that are causally linked to identify plausible causal structures from empirical evidence. Cooper and Herskovits [4] define a heuristic search algorithm to find the most probable Bayes network structure given a set of data, and show that they can quite reliably recover Bayes networks from data sets generated by statistical sampling from those networks.

Another interesting approach to the problem of learning Bayes networks strives to improve the efficiency of knowledge acquisition not by learning from data but by making it easier for experts to express their judgments. Heckerman introduced probabilistic similarity networks that allow the expert to concentrate on the individual differentiation problems that are crucial to a domain. His system then builds a complete Bayes network by aggregating the locally-expressed judgments [13].

3 Making Decisions Under Uncertainty

Decision analysis [27] provides a normative model of rational decision making, and recommends in all cases that one select that action to take which maximizes expected value among possible alternatives. Two principal methods of expressing decision analyses have been developed, and both are in limited use in clinical decision making. One is the construction and evaluation of *decision trees*, which explicitly represent the branching pathways of choices and chance events that face the decision-maker. The other is the use of *influence diagrams*, which more compactly represent the probabilistic and decision relationships among random variables and decisions in a network form. Both methods rely strongly on *value functions* that specify the relative desirability of all possible outcome states. Constructing such value functions is a very difficult task, both in principle and in practice.

3.1 Decision Trees

As decision makers, we are generally faced with two sources of uncertainty: (1) chance outcomes of natural processes or actions, and (2) decisions that we must make. By the commandment to maximize expected value, the value of a decision that we get to make is the maximum of the values of the alternative choices. When facing nature's chance response to our actions, the expected value of our state is simply the average of the values of each of the possible outcomes, each weighted by the likelihood of that outcome. Evaluating a decision tree is the process of recursively *folding back* values from the leaves toward the root.

A decision tree is typically (though not necessarily) ordered so that the root of the tree is the first decision that must be made, and further down each branch may come additional decisions that arise depending on previous decisions and chance results. Decision trees can, therefore, represent contingent plans, because later decisions may become relevant only if chance outcomes lead to them.

The decision tree in Fig. 1 represents the sequence of decisions faced by a patient who has a very badly infected foot that has turned gangrenous. The immediate choices are to amputate the foot or to treat medically. Depending on the outcome of the medical treatment, the patient may face the choice of subsequently having to amputate the whole leg or continuing with medical treatment. For the (fictitious) numbers used in the figure, the best decision is to treat medically, but then to amputate the leg if the gangrene gets worse after medical treatment. Normally, one would not simply make such a conclusion and act on it, because both the probability estimates and the utilities are tenuous. Instead, one would typically undertake a *sensitivity analysis*—to see whether the decision is robust against small changes in various parameters. For example, how much would we have to raise the value of life without a foot before an immediate foot amputation would appear preferable to medical treatment and the risk of losing a whole leg. How much more likely would death after initial medical treatment need to be before we (or the pa-

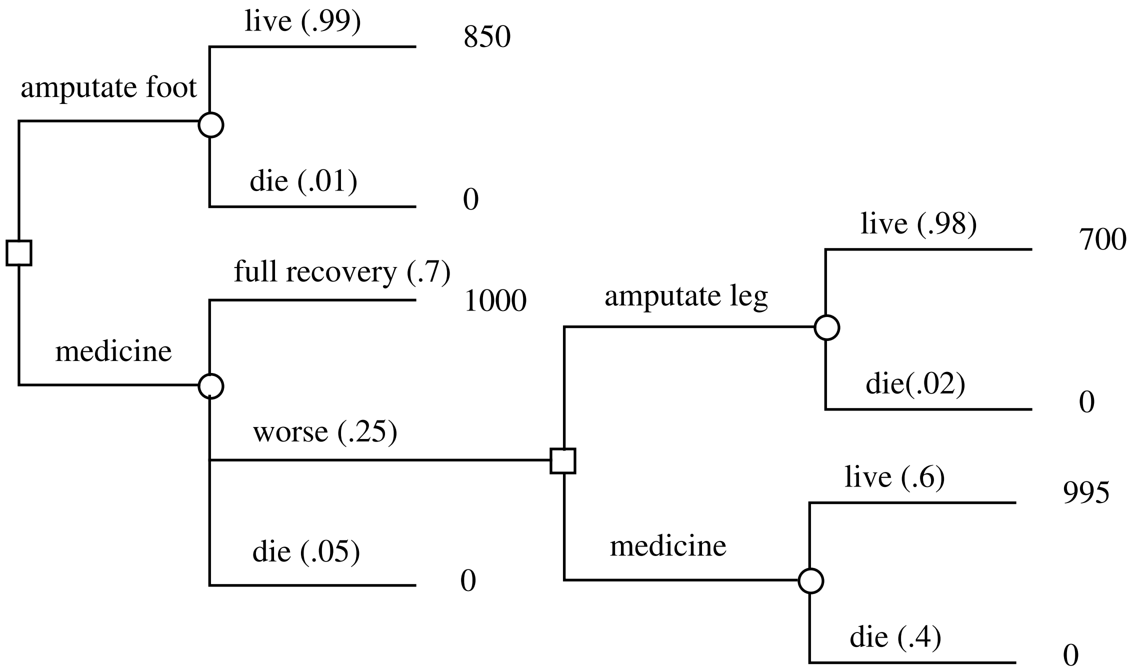

Figure 4. Decision tree for a patient facing possible amputation of his foot or leg as a result of gangrene. Squares represent decisions to be made, and circles the chance nodes that lead to different outcomes with different probabilities. Each branch of the tree terminates in a specific, numeric value, in this case measured on a 0–1000 scale, where 0 represents death, 1000 represents regained full health, and intermediate numbers the relative desirability of different health states. For the example patient, life after having lost his foot to amputation is worth 850 points on this 1000-point scale, whereas life after having lost his whole leg is worth 700. The example is a simplified version of one of the earliest clinical uses of decision analysis, on the clinical decision making service at New England Medical Center [24].

tient) would prefer immediate amputation? Is the decision sensitive to the estimate of the utility of recovery after two courses of medical treatment? In the real world, such sensitivity analyses often suggest that critical additional data be gathered to help resolve the uncertainty.

For very large decision problems, where the decision tree is quite deep, it may be possible to calculate an approximate value for a node in the tree without folding back the actual value by descending through the whole tree to its leaves. Consider, for example, a diagnostic problem wherein the decisions to be made consist of which of a large set of available diagnostic tests to perform. For k tests, there are 2^k ways to order them (assuming no test is done more than once), and for large k, this leads to an enormous tree. In the context of a sequential Bayesian diagnostic program, Gorry [9] observed that the problem can be solved heuristically by looking only one step down each branch of the tree, choosing the best apparent choice, and then further expanding the chosen branch only after the results of the test are known. In place of the full utility of each choice (which would include the likelihoods of all possible subsequent test results and the possible benefits of the best treatments to undertake given all the reachable test results), his program substituted the (negative) *entropy* of the probability distribution over the set of diseases under consideration. Because entropy is a measure of the information content of a distribution, this heuristic leads the program to make choices of tests that are most likely to lead to the most informative probability distributions.

3.2 Influence Diagrams

An influence diagram [15] is a Bayes network with the addition of decision nodes and a value node. Arcs into a decision node represent the information available at the time the decision is made, and that same information is also available to all subsequent decisions. The value node defines a function of those nodes with arcs connecting to it, which represents the overall utility of the problem. Solving an influence diagram corresponds to selecting for each decision node a *decision policy* that tells the optimum choice to make given any possible set of informational inputs to the decision node. Shachter first demonstrated a solution method [31] that does not need to transform the influence diagram into an alternative representation in order to solve it.

Influence diagrams are considerably more compact representations of large decision problems than decision trees, because they do not explicitly expand the consequences of individual choices or chance occurrences. Instead, each arc carries a matrix of probabilistic relationships between all possible states of the source and target of the arc. Influence diagrams are equivalent to fully symmetric decision trees, and this can be both an advantage and disadvantage. It assures that no unintended asymmetry arises in the analysis of a complex decision problem. However, it makes it impossible (within the representation) to simplify the model by omitting branches that have very low probability or that don't significantly influence the utility model. Some practitioners also find the more concrete representation of the decision tree easier to understand. Because influence diagrams and decision trees are interconvertible, it is also possible to build hybrid modeling tools that permit viewing a decision problem from either vantage point.

3.3 Utilities and Time

The most challenging problems that currently attract research attention involve means of modeling utilities and techniques for dealing with various aspects of time; often the two issues must be addressed in common.

The classic problem in reasoning about individual utilities is that they are so difficult to quantify. According to decision theory, standard lottery techniques must be able to determine a numerical utility scale for all possible circumstances for a rational decision-maker, assuming only very reasonable constraints. Thus, if we assume that the ends of the utility scale are defined by the state of good health (worth 1000, in our amputation scenario) and death (0), then any other intermediate state can be assigned a value in the following manner: Offer the decision-maker a choice between that state for certain or a gamble in which the probability of getting good health is p and the probability of death is (1-p). If we vary p, there should be a value p^* for which the decision-maker is indifferent to whether he chooses the certain outcome or the gamble. Then, 1000*p is the value of the state in question. Despite the attractiveness of this procedure, people find it difficult to make such judgments reliably, they produce judgments that differ significantly depending on just how the question is phrased, and they may be unable to estimate the value of a hypothetical state until they are actually faced with the decision. We address the consequences of some of these problems near the end of the current paper, but unfortunately there are no satisfactory general solutions to this problem.

In applying decision analysis to clinical medicine, doctors have found one instance of the utility problem to be quite common: many decisions may affect the complete future course of a patient's life, so there is no short time horizon over which one can assess the benefits and defects of a course of action. For example, a decision to perform coronary bypass surgery now will affect the patient's quality of life for many years to come, and will also alter the expected likelihood of various contingencies of interest such as restenosis. To model the value of a current intervention thus requires modeling all its possible long-term consequences. One successful approach to this problem is to assume that the future course of a patient's state is determined by a Markov process, and that the intervention sets the parameters of that process [1]. If each state has a given value, then the long-term value of a decision is calculated by integrating the value of each state times the probability of being in it over a long enough time to encompass the maximum expected life of the patient. This approximation assumes that the critical decision itself is made now, and the rest of the analysis merely explores its consequences over time.

Berzuini *et al.* [2] construct general Bayes networks whose structure reflects repeating episodes, each depending on a set of global patient and process-specific parameters. The beauty of this approach is that the usual inference mechanisms of Bayes networks serve both to estimate the parameters from observed episodes and to predict future behavior of the system.

More general temporal models currently being explored consider cases in which there is not a single decision to be made but a recurring sequence of decisions, in which the optimal strategy may well be to make different decisions in the future than at present. Leong [20] casts the problem in the framework of semi-Markov decision models and is building tools to support construction and evaluation of such models.

4 Approximation and Abstraction in Models

The full probabilistic and decision analytic framework for reasoning about uncertainty is very attractive and has a long history of advocacy and analysis (e.g., [17,30]). Nevertheless, applying these ideas in a straightforward manner requires accurate elicitation of many numeric probabilities and utilities. The difficulty of doing this begs for practical or conceptual simplification.

Early AIM and AI programs introduced a large variety of scoring schemes that were thought, at the time, to be simpler or more attractive than probability theory. In retrospect, however, many of these have been shown to be equivalent to standard probability theory, with perhaps a few additional assumptions or approximations (e.g., [12] concerning Mycin, and the discussion of log likelihood ratios, above, for Internist). Some of the schemes were originally introduced simply because the methods of Bayes networks were unknown, yet the need for chains of probabilistic inference was critical (e.g., the inference scheme of Prospector [5]). The evolution of the powerful Bayes network formalism has obviated the need for many such scoring schemes, and has provided at least a background theory against which to measure any other proposals.

4.1 Doing Without Numbers

One exciting line that has run through the past several decades of research on uncertainty attempts to reason with a version of probability theory or decision theory that is more abstract than the numerically-expressed theories we ordinarily consider. Of course this is not in any sense a departure from the foundations of the theories, which recognize structure as having paramount importance over numerical details, but it puts emphasis properly back onto these more fundamental ideas. In Bayes networks, we have remarked that it is the presence or absence of an arc between two nodes that carries the most vital information—about probabilistic dependence or independence of variables. Similarly, modelers using influence diagrams emphasize the importance of the structure of the diagram, compared to which the numerical details are secondary [14].

Abstraction that completely disregards numerical knowledge and exploits only the structural relationships can be quite powerful and easy to use. Reggia *et al.* developed a set covering diagnostic model [28] that uses the bipartite graph representation but without regard for actual probabilities. A diagnostic solution in this sense is a minimal set of diseases that can cause all of the observed symptoms. By Occam's razor, the simplest explanation of a problem is the best, and the set cover algorithm equates simplicity with the smallest number of hypothesized disorders. This corresponds roughly to assuming that every disease is equally likely, that all symptoms combine causal influences on them by noisy-or, and that the disease-to-symptom link strengths are all equal. In a domain rich in interconnections and alternative explanations of many phenomena, and when the amount of noise in both the models and observations is low, this can be an effective and efficient diagnostic scheme. Improved heuristic search methods and representations for partial solutions can gain further dramatic efficiency improvements [38,39].

4.2 Other Abstract Characterizations of Decision Analytic Models

Restricting reasoning to only the structure of influence diagrams yields models that may be too weak to make many decisions of interest. A few research efforts have explored intermediate levels of representation that give more than simply structure but demand less than full numbers. Wellman [38] defines qualitative influences among random variables in terms of stochastic dominance of their distribution functions, and defines a qualitative version of synergy and anti-synergy among joint influences on a variable. Because these definitions capture only the direction of influence, not its magnitude, the resulting models are clearly not useful to resolve detailed tradeoffs among competing decision policies. Inferences with the qualitative models can, however, derive the kinds of qualitative results needed in planning or critiquing plans. For example, Wellman's planner is able to reject large regions of the space of all treatment plans by recognizing that all of the rejected plans are dominated by better plans still under consideration. Similarly, the method can make some of the sorts of "common sense" inferences useful in critiquing hand-crafted decision analytic models: e.g., don't perform a test whose result does not influence the decision [39]. In a similar spirit, Halpern has explored axiomatizations of qualitative likelihood that may correspond to human intuitions [11].

People obviously do not use full-blown decision analysis to make decisions about their daily lives, yet they often act in quite rational ways. This suggests that there are many as yet unexplored abstractions and approximations of our formal models for reasoning under uncertainty that can be fruitful. They should lead to models that are easier to understand because they correspond to important aspects of the ways in which people think about problems. They may also be easier to compute than the full-blown, often exponential models that currently arise in exact calculations of results.

5 Conclusion

We have reviewed the probabilistic and decision analytic approaches to medical reasoning under uncertainty, and examined some of the interesting directions in which related research is heading. The 1990's promises to be a decade of very rapid practical innovation as most hospitals and clinics rush toward computerization of their record systems and try to achieve a comprehensive, well-defined set of standards for the encoding, storage and dissemination of information about medical care. Prompted by government regulatory requirements, standards bodies, and cooperative agreements among insurers and health care providers, the popularity of outcomes studies and cost-benefit analyses will continue to grow. Many problems will be solved in *ad hoc* ways, because the pressures of short deadlines and practical needs will not allow the leisure of investigation that is often necessary to come to thoughtful solutions.

Nevertheless, these anticipated trends will create great opportunities for interesting research in medical reasoning under uncertainty. For the first time, we will move from problems of data scarcity to those of data overload, as the volume of recorded data explodes. Most of those data will be captured to satisfy regulatory and reimbursement needs, not primarily the needs of clinical care or of medical research. Therefore, tremendous opportunities will arise to develop insightful new approaches for taking advantage of this developing glut of data, helping to relate it to clinical and research needs and to the goals and preferences of tomorrow's patients.

Acknowledgments

The author's research is supported in part by grant number LM 05296 from the National Library of Medicine and by a grant from IBM.

References

1. Beck JR, Pauker SG. The Markov Process in Medical Prognosis. *Medical Decision Making* 1983;3:419-458.
2. Berzuini C, Bellazzi R, Spiegelhalter D. Bayesian Networks Applied to Therapy Monitoring. In: *Uncertainty in Artificial Intelligence '91*. Menlo Park, CA: Morgan-Kaufmann, 1991: 35–43.
3. Cooper GF. NESTOR: A Computer-based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge. Doctoral Thesis. Stanford University, 1984.
4. Cooper GF. The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks. *Artificial Intelligence* 1990;42(2-3):393-405.
5. Cooper GF, Herskovits E. A Bayesian Method for Constructing Bayesian Belief Networks from Databases. In: *Uncertainty in Artificial Intelligence '91.* Menlo Park, CA: Morgan Kaufmann, 1991:86–94
6. Duda RO, Hart PE, Nilsson NJ. Subjective Bayesian methods for rule-based inference systems. In: AFIPS Conference Proceedings. National Computer Conference, 1976: 1075-1082.
7. Elston R, Stewart A. A General Model for the Genetic Analysis of Pedigree Data. *Hum. Hered*. 1971;21:523-542.
8. Genesereth MR, Nilsson NJ. *Logical Foundations of Artificial Intelligence*. Menlo Park, CA: Morgan Kaufmann, 1987
9. Gorry GA, Barnett GO. Sequential Diagnosis by Computer. *Journal of the American Medical Association* 1968;205(12):849-854.
10. Gorry G, Kassirer J, Essig A, Schwartz W. Decision Analysis as the Basis for Computer-aided Management of Acute Renal Failure. *American Journal of Medicine* 1973;55:473-484.
11. Haimowitz I. Modeling All Dialogue System Participants to Generate Empathetic Responses. In: *Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care*. 1990: 51-57.
12. Halpern JY. An Analysis of First-Order Logics of Probability. *Artificial Intelligence* 1990;46(3):311-350.
13. Heckerman D. Probabilistic Interpretations for MYCIN's Certainty Factors. In: Kanal LN,

Lemmer JF, ed. *Uncertainty in Artificial Intelligence*. North-Holland, 1986:167–196.
14. Heckerman DE. *Probabilistic Similarity Networks*. Cambridge, MA: MIT Press, 1991
15. Holtzman S. Intelligent Decision Systems.Addison-Wesley, 1989
16. Howard R, Matheson J. Influence diagrams. In R. Howard & J. Matheson (Eds.), *The Principles and Applications of Decision Analysis*. Menlo Park, CA: Strategic Decisions Group, 1984:719-762.
17. Jacquez JA, ed. *Computer Diagnosis and Diagnostic Methods*. Springfield: C. C. Thomas, 1972.
18. Jeffrey RC. *The Logic of Decision*. Chicago: University of Chicago Press, 1983.
19. Kjærulff U. Triangulation of Graphs—Algorithms Giving Small Total State Space. Institute for Electronic Systems, University of Aalborg, Denmark, 1990.
20. Lange A, Elston R. Extensions to Pedigree Analysis. *Hum. Hered*. 1975;25:95-105.
21. Leong T-Y. Dynamic Decision Modeling in Medicine: A Critique of Existing Formalisms. In: *Symposium on Computer Applications in Medical Care*. IEEE, 1993:478–484.
22. Long W. The Probability of Disease. In: *Symposium on Computer Applications in Medical Care*. IEEE, 1991:619–623.
23. Miller R, Pople HE, Myers J. INTERNIST-1, An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine. *New England Journal of Medicine* 1982;307:468-476.
24. Pauker S, Gorry GA, Kassirer J, Schwartz W. Towards the Simulation of Clinical Cognition: Taking a Present Illness by Computer. *American Journal of Medicine* 1976;60:981-996.
25. Pauker SG, Kassirer JP. Medical Progress: Decision Analysis. *New England Journal of Medicine* 1987;316:250-258.
26. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988
27. Pearl J, Verma TS. A Theory of Inferred Causation. In: *Principles of Knowledge Representation and Reasoning, KR '91.* Menlo Park, CA: Morgan Kaufmann, 1991:441–452.
28. Raiffa H. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. Reading, MA: Addison-Wesley, 1968.
29. Reggia JA, Nau DS, Wang PY. Diagnostic expert systems based on a set covering model. *International Journal of Man-Machine Studies* 1983;19:437-460.
30. Rieger C, Grinberg M. The declarative representation and procedural simulation of causality in physical mechanisms. In: *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*. 1977: 250-256.
31. Savage L. *The Foundations of Statistics*. New York: Dover Publications, 1972
32. Shachter RD. Evaluating Influence Diagrams. *Operations Research* 1986;34(6):871–882.
33. Shachter RD, Kenley CR. Gaussian Influence Diagrams. *Management Science* 1989;35(5):527–550.
34. Shachter RD, Peot MA. Simulation Approaches to General Probabilistic Inference on Belief Networks. In: Henrion M, Shachter RD, Kanal LN, Lemmer JF, ed. Uncertainty in Artificial Intelligence 5. Amsterdam: North-Holland, 1990: 221–231.
35. Shafer G. *A Mathematical Theory of Evidence*. Princeton University Press, 1976
36. Shortliffe E. *MYCIN: Computer-based Medical Consultations*. New York: American Elsevier, 1976.
37. Spiegelhalter DJ. Bayesian Analysis in Expert Systems. MRC Biostatistics Unit, Institute of Public Health, Cambridge, 1992.
38. Szolovits P, Pauker S. Categorical and Probabilistic Reasoning in Medicine Revisited. *Artificial Intelligence* 1993;59:167-180.

39. Wellman MP. Fundamental Concepts of Qualitative Probabilistic Networks. *Artificial Intelligence* 1990;44:257-304.
40. Wellman MP, Eckman MH, Fleming C, Marshall SL, Sonnenberg FA, Pauker SG. Automated Critiquing of Medical Decision Trees. *Medical Decision Making* 1989;9(4):272-284.
41. Wu T. A Problem Decomposition Method for Efficient Diagnosis and Interpretation of Multiple Disorders. *Computer Methods and Programs in Biomedicine* 1991;35:239-250.
42. Wu T. Domain Structure and the Complexity of Diagnostic Problem Solving. In: *Proceedings of AAAI-91*. 1991: 855-861.
43. Zadeh LA. Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems* 1978;1:3-28.