Methods of Information in Medicine

M. A. Shwe, B. Middleton, D. E. Heckerman, M. Henrion, E. J. Horvitz, H. P. Lehmann, G. F. Cooper*

Probabilistic Diagnosis Using a Reformulation of the INTERNIST-1/QMR Knowledge Base

I. The Probabilistic Model and Inference Algorithms

Section on Medical Informatics, Stanford University, Stanford, CA,
* Section of Medical Informatics, University of Pittsburgh, Pittsburgh, PA, U.S.A.

Abstract: In Part I of this two-part series, we report the design of a probabilistic reformulation of the Quick Medical Reference (QMR) diagnostic decision-support tool. We describe a two-level multiply connected belief-network representation of the QMR knowledge base of internal medicine. In the belief-network representation of the QMR knowledge base, we use probabilities derived from the QMR disease profiles, from QMR imports of findings, and from National Center for Health Statistics hospital-discharge statistics.
We use a stochastic simulation algorithm for inference on the belief network. This algorithm computes estimates of the posterior marginal probabilities of diseases given a set of findings. In Part II of the series, we compare the performance of QMR to that of our probabilistic system on cases abstracted from continuing medical education materials from Scientific American Medicine. In addition, we analyze empirically several components of the probabilistic model and simulation algorithm.

Key-Words: Expert Systems, Computer-aided Diagnosis, Probabilistic Inference, Belief Networks

1. Introduction

The INTERNIST-1 system was developed at the University of Pittsburgh as an experimental decision-support tool in general internal medicine [1–4]. When presented with findings from a patient's history, physical examination, and laboratory-test results, INTERNIST-1 used a heuristic reasoning method with a quasi-probabilistic scoring scheme to suggest likely disease candidates and to guide the physician in the patient's workup.

In 1985, an extension of the INTERNIST-1 knowledge base (KB) was incorporated into a successor to INTERNIST-1, called Quick Medical Reference (QMR®) [3, 5]. QMR uses a new inference algorithm running on an IBM-compatible personal computer. QMR provides a clinician with many functions to access the QMR knowledge base. The developers of QMR classify the functions of the system into three levels [6]. First, a clinician can use QMR as an electronic textbook of medicine to display the findings associated with a disease, the diseases associated with a finding, or diseases related to a particular disease. Second, a clinician can use QMR as a diagnostic spreadsheet to show how particular groups of diseases and findings may co-occur. Third, a clinician can use QMR as an expert-consultant program, to analyze a diagnostic case. In this mode, QMR can provide diagnostic hypotheses containing multiple, pathophysiologically related diseases, or critique a diagnostic hypothesis that the clinician suggests.

In developing a decision-theoretic version of QMR, which we call *Quick Medical Reference, Decision Theoretic (QMR-DT)*, we are currently most interested in addressing the diagnostic functions contained in the expert-consultant mode of QMR. By a *decision theoretic* system, we mean a system that conforms to the principles of decision theory. Decision theory uses the axioms of probability theory and utility theory to provide a framework for choosing among alternative courses of action. Probability theory is a logic of degrees of belief, containing a set of axioms for expressing degrees of belief in propositions and for combining beliefs to derive measures of belief in related events. Utility theory comprises a set of axioms for ascribing

® QMR is a registered trademark of the University of Pittsburgh.

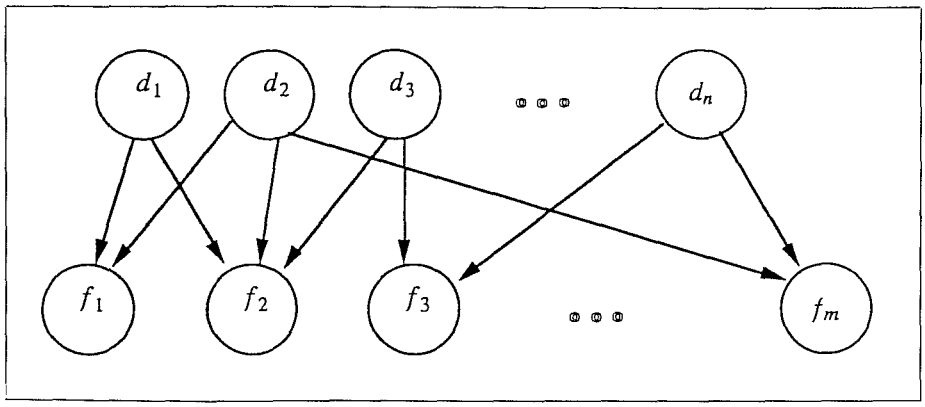

Fig. 1 The two-level belief-network representation of the current QMR-DT KB. The disease nodes are labeled $d_1,\ldots,d_n$ and the finding nodes are labeled $f_1,\ldots,f_m$. The probabilistic dependencies between diseases and findings are specified with directed arcs between nodes, where an arc points in the causal direction that we assume; that is, we assume that diseases cause findings.

numerical values to outcomes of events [7].

Our research to date has focused on building the probabilistic component of QMR-DT: a probabilistic model and method of inference for diagnosis in internal medicine. The initial goal of the project is to evaluate the performance of a probabilistic version of QMR, investigating the computational and representational tractability of a probabilistic approach. We are developing incrementally the QMR-DT system. More specifically, we build the first probabilistic model using as much of the QMR KB as possible, we test the accuracy of the inferential algorithms on this simple model, then we refine the model and algorithms, successively based on the performance of the system. Ultimately, we plan to add a utility model to the probabilistic component to make a complete decision-theoretic model.

We are currently using the INTERNIST-1 KB (circa 1986), rather than the more recent QMR KB. These two KBs are quite similar but the QMR KB contains newer disease profiles and updates of profiles from the INTERNIST-1 KB. The methods in this paper are applicable to transforming the QMR KB as well. For simplicity, where the distinction between the INTERNIST-1 KB and QMR KB is inconsequential, we will refer to the INTERNIST-1 KB as the QMR KB. The reader should note, however, that our probabilistic reformulation is based on the INTERNIST-1 KB.

We believe there are many reasons for seeking probabilistic systems in medicine. By using a probabilistic model, we make explicit our assumptions – those used in building both the KB and the algorithms for inference. Moreover, such a KB is built on the well-developed and widely understood language of probability, providing researchers with a theoretical basis for creating diagnostic systems and a common vocabulary for facilitating discussion and collaboration. The use of a common language also makes it possible to share independently developed probabilistic inference algorithms and KBs. We can, for example, incorporate statistics on the local prior probability of disease in various clinical settings. For those diseases about which we have sparse statistical data, we can rely on expert subjective estimates of prior probability. We can update these subjective probabilities incrementally as local clinical data are accumulated [8, 9].

The current output from QMR-DT is a differential of leading diagnoses with a posterior marginal probability associated with each disease. We believe that a probabilistic differential is a more meaningful measure of belief than is a differential with a heuristic score. Furthermore, with the development of a utility model, we can use these probabilities for expected-utility decision making, thus building a decision-theoretic system on top of our probabilistic one. We plan eventually to use the results of the expected-utility component of the system for cost-effective test ordering and therapeutic planning.

We chose QMR as the system to reformulate into a decision-theoretic version for several reasons. First, the QMR KB is the culmination of over 20 person – years of work; Version 10.729 (dated 6/14/89) of the QMR KB covers over 600 diseases and 4,000 findings [10]. Since the QMR KB is so large, inference on a probabilistic version of it presents a significant challenge. Second, the task of diagnosis in internal medicine can be a formidable one, since a patient often has more than one disease. We cannot make the simplifying assumption that only a single disease is present, as has been done in some other domains where probabilistic inference has been applied successfully [11, 12]. Third, the developers of INTERNIST-1 and QMR have demonstrated that the systems perform well on difficult cases: in a retrospective study, INTERNIST-1 performed comparably to hospital clinicians on cases from clinicopathological conferences (CPCs) [1], and in a prospective study, QMR performed better than the ward team on diagnostically challenging cases [13]. These results give us reason to believe that a probabilistic reformulation of QMR also might perform well as a diagnostic decision-support system, since they indicate that the relationships in the QMR knowledge base can support accurate diagnostic inference. Our research hypothesis is that a probabilistic reformulation of QMR will also provide accurate diagnostic performance relative to QMR.

In this paper, we build on the work of many researchers in probabilistic inference, probabilistic inference in medicine, and probabilistic inference in medicine using belief networks (see [14–16] for a review of probabilistic inference on belief networks). In Section 2, we describe how we built the QMR-DT model. Section 3 presents the algorithms used for inference on the model. In Section 4, we discuss limitations of the model and directions

for future research. The Appendix contains a summary of the notation used in this paper. In Part II of this series we discuss our laboratory evaluation of the probabilistic reformulation of QMR relative to QMR.

2. The QMR-DT Model

The QMR-DT model is built on a belief-network representation. A *belief network* is a graphical representation of probabilistic dependencies among variables [17]. Belief networks are also known as Bayesian belief networks and causal probabilistic networks. More specifically, a belief network is a directed acyclic graph in which each node represents a random variable or uncertain quantity [16]. The arcs in the graph often denote direct causal influences between variables, where the strength of the influence is specified by tables of conditional probabilities. Conversely, the absence of an arc between two nodes denotes an assertion of independence between the corresponding random variables. Many researchers have discussed the promise of using belief networks as an efficient, expressive knowledge representation for expert systems that reason under uncertainty [11, 15, 16, 18–20].

We have reformulated the associations between diseases and findings of the QMR disease profiles [1] into a belief-network representation. This reformulation is also described in [21, 22]. The QMR-DT KB consists of a two-level belief network of n diseases and m findings, as shown in Figure 1. Other work on inference using two-level networks appears in [23–25]. Each of the n diseases $\{d_1 \ldots d_n\}$ may be present or absent in a patient, and each of the m findings $\{f_1 \ldots f_m\}$ may be unobserved or observed to be present or absent. We refer to a disease hypothesis H as an assignment of presence or absence to each disease in $\{d_1 \ldots d_n\}$:

$$H = \{d_i = \text{present} \mid d_i \in H^+\} \cup \{d_i = \text{absent} \mid d_i \in H^-\}$$

where H^+ is the set of all diseases asserted to be present and H^- is the set of all diseases asserted to be absent. Note that $| H^+ | + | H^- | = n$.

An arc of probabilistic dependency between nodes representing a disease d and finding f exists in the QMR-DT KB if and only if there exists a link between d and f in the INTERNIST-1 disease profile of d. The QMR-DT KB thus contains $n = 534$ adult diseases and $m = 4{,}040$ findings, with 40,740 arcs depicting dependencies between diseases and findings. Recall the QMR-DT KB is based on the INTERNIST-1 KB, which also contains 534 adult diseases and 4,040 findings. We define the QMR-DT *model* as the combination of the two-level QMR-DT belief network depicted in Figure 1 and the assumptions that we describe in Section 2.1. Thus, we distinguish the QMR-DT *model* from the probabilistic inference *algorithms* that we apply to the model.

2.1 Assumptions in the Model

To reduce the representational and computational complexity of the current version of QMR-DT, we made several simplifying assumptions. In this section, we describe specific examples where these assumptions are not accurate. Although we know that the assumptions are inaccurate in such cases, we are taking an incremental approach in developing the QMR-DT model: we examine the performance of the system with these assumptions with the intention of eventually modifying those that appear most critical to accurate diagnostic performance.

Assumptions evident from Figure 1 include marginal independence of diseases and conditional independence of findings given any hypothesis of diseases. We model both findings and diseases as binary variables. Also, we model the influence of multiple diseases on a finding assuming causal independence. We discuss causal independence in detail in Section 2.1.4. Many of these assumptions have been used in previous probabilistic diagnostic systems [26]. The major assumption found in many of these systems, which we do not make in QMR-DT, is that the patient has at most one disease. In fact, we allow for the possibility that the patient may have any number of diseases. In the remainder of Section 2.1, we address the strengths and weaknesses of the assumptions that we make in the current QMR-DT model.

2.1.1 Marginal Independence of Diseases

The absence of arcs among disease nodes in the belief network of Figure 1 denotes the assumption that diseases are marginally independent. Under this assumption, we can compute the probability of a disease hypothesis H from the prior probabilities of the states of the diseases in H:

$$P(H) = \prod_{d_i \in H^+} P(d_i^+) \prod_{d_j \in H^-} P(d_j^-), \quad (1)$$

where d_j^+ is the event that disease d_i is present, and d_j^- is the event that disease d_j is absent. We have no reason to question the assumption of marginal independence between apparently unrelated diseases in the QMR-DT KB such as myocardial infarction and primary hyperparathyroidism. On the other hand, some disease combinations in the QMR-DT KB are clearly dependent. For example, the probability of congestive heart failure is greatly increased in a patient with significant aortic stenosis.

Although the QMR KB represents relationships among diseases, we have chosen initially to simplify QMR-DT by not including these relationships in the belief-network model. As we discuss in Section 3.2, however, we do use the relationships among diseases in one of our heuristic scoring functions to help improve the speed of inference.

2.1.2 Conditional Independence of Findings

The absence of arcs among finding nodes in Figure 1 denotes the assumption that findings are conditionally independent given any disease hypothesis. Let F be a set of findings that are observed for a particular patient,

where $m' = |F|$, F^+ is the set of findings observed to be present, and F^- is the set of findings observed to be absent. Note that many findings may be unobserved and thus appear in neither F^+ nor F^-.

The assumption of conditional independence of findings given any hypothesis allows us to compute the conditional probability of a set of findings F given a disease hypothesis H as follows:

$$P(F \mid H) = P(F^+ \mid H)\, P(F^- \mid H) = \prod_{f \in F^+} P(f^+ \mid H) \prod_{f \in F^-} P(f^- \mid H). \quad (2)$$

An example of a belief network in which this conditional-independence assumption is valid appears in Figure 2. This example depicts a conditional-independence assumption in the QMR-DT belief network. Suppose that we know that the only disease present in a particular patient is bronchial adenocarcinoma. Then, we have a certain belief (probability) that the patient has blood-streaked sputum. Suppose we are then told that the patient's endobronchial biopsy suggests adenocarcinoma. Since this new information does not change our belief that the patient has blood-streaked sputum[1], given that the disease state is bronchial adenocarcinoma only, we can compute P(BRONCHOSCOPY ENDOBRONCHIAL BIOPSY ADENOCARCINOMA, SPUTUM BLOOD STREAKED | only BRONCHIAL ADENOCARCINOMA) by multiplying the two component probabilities P(BRONCHOSCOPY ENDOBRONCHIAL BIOPSY ADENOCARCINOMA | only BRONCHIAL ADENOCARCINOMA) and P(SPUTUM BLOOD STREAKED | only BRONCHIAL ADENOCARCINOMA).

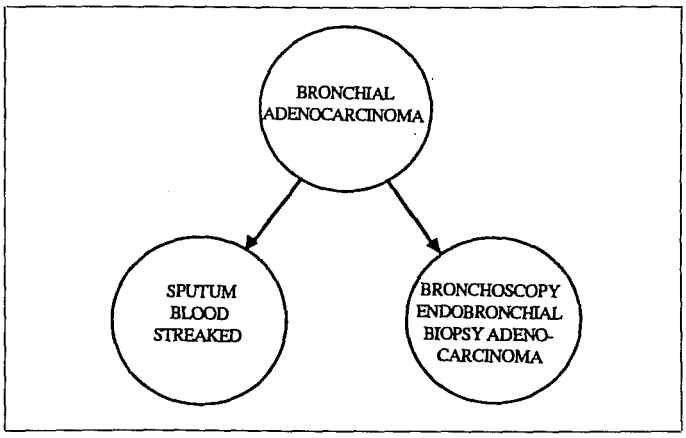

Fig. 2 A belief network representing conditional independence of the findings adenocarcinoma on endobronchial biopsy and blood-streaked sputum given that bronchial adenocarcinoma is present.

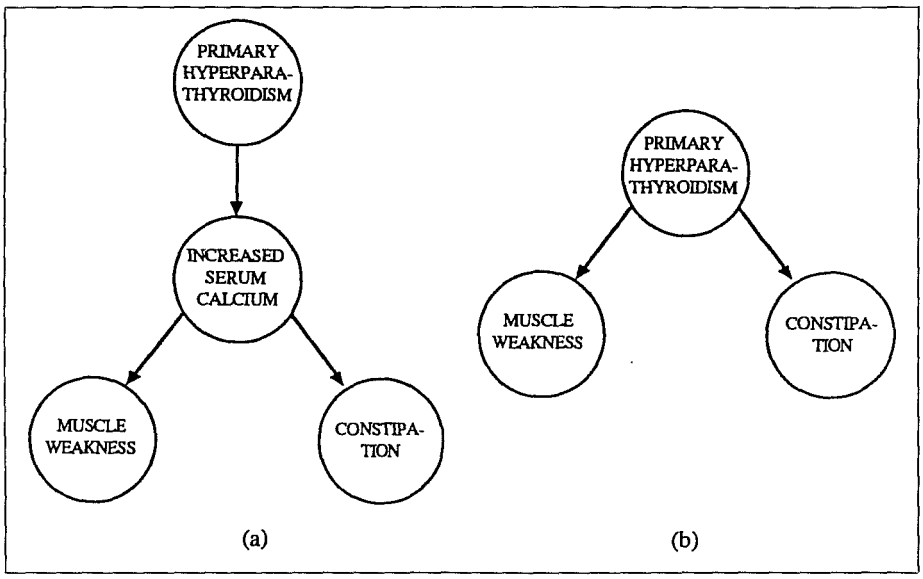

Fig. 3 Belief networks modeling conditional independence of findings given a disease. (a) The network displays the belief that muscle weakness and constipation are conditionally independent given the intermediate pathophysiologic state of increased serum calcium. (b) The network reflects the belief that muscle weakness and constipation are conditionally independent given primary hyperparathyroidism. Although the model in (a) is a more accurate depiction of the dependencies among variables, we currently use the model in (b) as an approximation.

Figure 3(a) shows a belief network representing an example in which there exists an intermediate pathophysiologic state between a disease and that disease's manifestations. Note that the assumption of conditional independence is accurate for muscle weakness and constipation, given increased serum calcium as depicted in Figure 3(a). On the other hand, muscle weakness and constipation are not conditionally independent given primary hyperparathyroidism, as depicted in Figure 3(b) and as represented currently in the QMR-DT KB. The independence relationship depicted in 3(b) is less accurate because of the existence of the intermediate state of increased serum calcium: Knowledge of constipation increases the belief in muscle weakness because it increases the belief in increased serum calcium, even when we are given that primary hyperparathyroidism is present. Although the model in Figure 3(a) is a more accurate representation of the dependencies among the variables, we currently use the model in 3(b) as an approximation.

2.1.3 Binary Diseases and Findings

We assume each finding and disease to be binary: diseases are either present or absent, and observed findings are either present or absent. This assumption simplifies the probabilistic transformation of the QMR KB, which likewise contains binary findings and diseases. Finer gradations in the representation of findings and diseases not only would be more intui-

[1] In this example, we ignore the event that the biopsy itself may cause blood in the sputum.

tive, but also would strengthen the correctness of the assumption of conditional independence of findings given any disease hypothesis. Consider the example from the QMR-DT belief network of conditional independence depicted in Figure 4(a). Suppose that we condition on (assume) the presence of Laënnec's cirrhosis – a disease with a wide spectrum of severity. The probability of decreased hepatic arterial vascularity is affected by the presence or absence of increased prothrombin time, and vice versa, even *given* that we know Laënnec's cirrhosis is present. That is, knowledge about one of the findings gives us information on the severity of the disease – more information than we gain by conditioning on simply the presence of the disease. Specifically, P(DECREASED HEPATIC ARTERIAL VASCULARITY | only LAENNEC'S CIRRHOSIS) is less than P(DECREASED HEPATIC ARTERIAL VASCULARITY | only LAENNEC'S CIRRHOSIS and INCREASED PROTHROMBIN TIME), since knowledge of increased prothrombin time shifts our belief to a more severe form of Laënnec's cirrhosis, increasing the probability of decreased hepatic arterial vascularity.

A more accurate representation of the conditional independence appears in Figure 4(b), in which we model Laënnec's cirrhosis as being absent, mild, or severe. The assumption that P(DECREASED HEPATIC ARTERIAL VASCULARITY | only SEVERE LAENNEC'S CIRRHOSIS) is equal to P(DECREASED HEPATIC ARTERIAL VASCULARITY | only SEVERE LAENNEC'S CIRRHOSIS and INCREASED PROTHROMBIN TIME) is more accurate than the assumption that we make in the case of a binary disease modeling of Laënnec's cirrhosis. Because we have already conditioned on the severity of disease, knowledge of a finding such as increased prothrombin time tells us little more about the severity of Laënnec's cirrhosis, thus providing only a small update to our belief about decreased hepatic arterial vascularity. Although the model in Figure 4(b) is a more accurate representation of the conditional independence between the two findings, for simplicity, we currently model diseases and findings in QMR-DT as binary variables as in Figure 4(a).

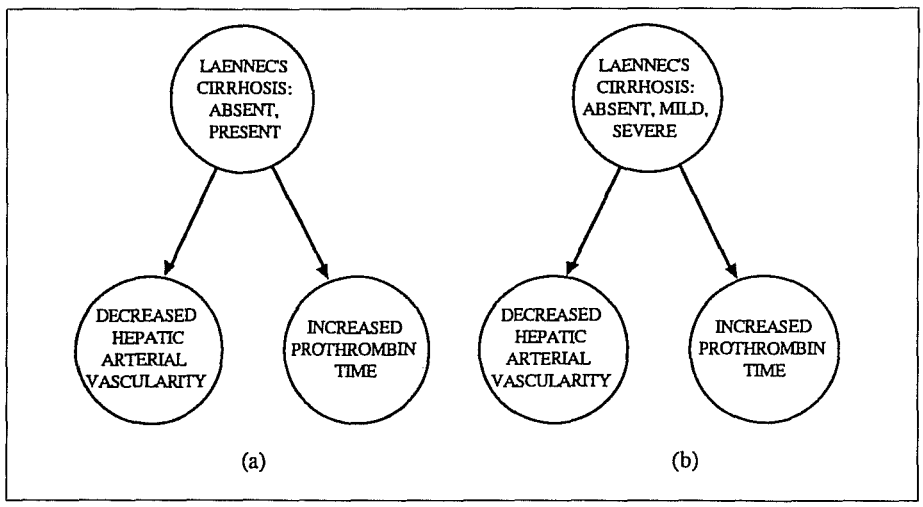

Fig. 4 Two belief networks that have different granularity of severity of disease. (a) Laënnec's cirrhosis is modeled as a binary disease: It can either be present or absent (b) Laënnec's cirrhosis is modeled with three states: absent, mild, or severe. In the QMR-DT KB, we currently model diseases and findings as binary variables, as in (a).

2.1.4 Causal Independence

We model the effects of multiple diseases on a single finding by assuming that the effects of the diseases on the finding occur independently. This assumption, called *causal independence,* has been described by a number of researchers, including Good [27]. We use a *noisy-OR gate* to model causal independence [16]. Several researchers have described the application of the noisy-OR gate to modeling the effects of diseases on manifestations [19, 21–23]. In addition, the developers of INTERNIST-1 implicitly assumed a noisy-OR gate interaction [28]. Under the assumption of a noisy-OR gate, we can avoid representing the full set of conditional probabilities of the state of a finding given each possible state of the finding's parents. Consider a belief net with binary finding f, where f has binary parents $d_1, d_2, \ldots, d_k$. To construct the complete conditional probability table associated with the arcs from $d_1, d_2, \ldots, d_k$ to f, we need to acquire 2^k conditional probabilities. If we assume causal independence, we need to acquire only k conditional probabilities of the form $P(f^+ \mid \text{only } d_i^+)$[2], for $1 \leq i \leq k$.

As its name implies, the causal-independence assumption maintains that the mechanisms by which diseases cause a finding operate independently of one another and independently of any other events that may cause the finding to occur, such as the influence of other findings. Figure 5 shows a belief network for a noisy-OR influence of two diseases d_1 and d_2 on a single finding f, as depicted by Heckerman in [21]. Assuming causal independence, we can model the influence of multiple diseases on a finding using the noisy-OR gate:

$$P(f^- \mid \text{only } d_1^+ \text{ and } d_2^+) = P(f^- \mid \text{only } d_1^+)\, P(f^- \mid \text{only } d_2^+). \quad (3)$$

The intuition behind (3) is that the probability of finding f not occurring (given a hypothesis H) is just the probability that, of the two mechanisms that can cause f to occur, neither succeeds in causing f to occur. Because we have modeled the findings as binary variables, we can rewrite (3) as:

$$P(f^+ \mid \text{only } d_1^+ \text{ and } d_2^+) = 1 - [1 - P(f^+ \mid \text{only } d_1^+)][1 - P(f^+ \mid \text{only } d_2^+)]. \quad (4)$$

[2] We distinguish $P(f^+ \mid \text{only } d_i^+)$ from $P(f^+ \mid d_i^+)$, where the former denotes the probability of the event that f occurs given that only disease d_i occurs, and, for all $j \neq i$, d_j are absent. By contrast, we use the notation $P(f^+ \mid d_i^+)$ to mean the probability of the event that f occurs given that d_i occurs and for all $j \neq i$, d_j occur based on their prior probabilities.

In the more general case of a disease hypothesis H, we can compute the probability of the presence of f given H as:

$$P(f^+ \mid H) = 1 - \prod_{d_i \in H^+ \cap \pi(f)} [1 - P(f^+ \mid \text{only} d_i^+)], \quad (5)$$

where $\pi(f)$ are the parents of f – that is, the diseases corresponding to those nodes in the QMR-DT belief network that have arcs directed to f. The probability of the absence of f is given simply by $1 - P(f^+ \mid H)$.

An example found in the QMR-DT KB where the causal-independence assumption is accurate, is pictured in Figure 6. In this example, diverticulitis may cause a severe mucosal inflammation in the diverticulum, which in turn causes mucosal breakdown and local hemorrhage, resulting in a positive guaiac test. Right colon cancer may cause a mucosal necrosis, which in turn causes mucosal breakdown and local hemorrhage, resulting in a positive guaiac test. Accordingly, in the example depicted in Figure 6, the variable DIVERTICULITIS CAUSES ⊕ GUAIAC TEST corresponds to the event that severe mucosal inflammation in the diverticulum is present or absent, and the variable RIGHT COLON CANCER CAUSES ⊕ GUAIAC TEST refers to the event that mucosal necrosis is present or absent. In general, the probability that mucosal necrosis is present in the diverticulum is largely unaffected by either the presence of diverticulitis or the presence of mucosal inflammation in the diverticulum. Also, the probability that severe mucosal inflammation is present in the diverticulum is largely unaffected by either the presence of right colon cancer or the presence of mucosal necrosis. Because the mechanisms by which diverticulitis and right colon cancer cause a positive guaiac test operate largely independently, we can justifiably apply the noisy-OR gate to model the influence of the two diseases on a positive guaiac test.

The causal independence assumption is less accurate in cases where diseases operate through a common pathway to cause a finding. Figure 7 shows a belief network modeling the effects of plasma-cell myeloma and primary hyperparathyroidism on muscle weakness. Note the common intermediate state of increased serum calcium. As a first approximation, the QMR-DT KB currently models this interaction with a noisy-OR-gate as in Figure 5, where d_1 = PLASMA CELL MYELOMA, d_2 = PRIMARY HYPERPARATHYROIDISM, and f = MUSCLE WEAKNESS.

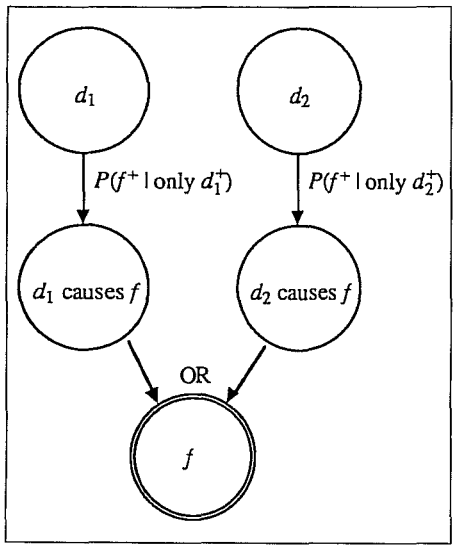

Fig. 5 A belief network depicting the causal-independence assumption of the noisy-OR gate. Consider a disease d_1 that causes finding f through some mechanism, and a disease d_2 that causes f through some other mechanism. This network depicts the assumption that d_1 and d_2 cause f through intermediate mechanisms that are independent of each other. If d_1 is present, it may or may not initiate a mechanism that causes f to be present. Our belief that d_1 initiates the mechanism "d_1 causes f" with probability $P(f^+ \mid \text{only } d_1^+)$ is represented by the arc from the node labeled "d_1" to the node labeled "d_1 causes f". The absence of an arc between the nodes labeled "d_1 causes f" and "d_2 causes f" represents an assumption of causal independence. That is, the probability that the mechanism "d_1 causes f" is active is not affected by whether the mechanism "d_2 causes f" is active. Also, the absence of the arc from the node "d_2" to the node "d_1 causes f" represents the causal independence assumption that the probability that the mechanism "d_1 causes f" is active is not affected by the presence or absence of d_2. The same causal independence assumptions apply to the mechanism by which d_2 causes f. The node with the double boundary is a deterministic node, which represents the belief that, if either of the two intermediate mechanisms occurs (that is, d_1 or d_2 succeeded in initiating the mechanisms by which it caused f), f will be present with certainty. (Adapted with permission from [21] Figure 2, page 165.)

2.1.5 Findings as Manifestations of Disease

There are various possible relationships between a disease and a finding. However, findings are all modeled as manifestations of disease in the current QMR-DT belief network. This assumption is generally not correct for historical findings. For example, neither a history of cigarette smoking nor a history of diabetes mellitus follows from acute myocardial infarction, as represented by the belief network

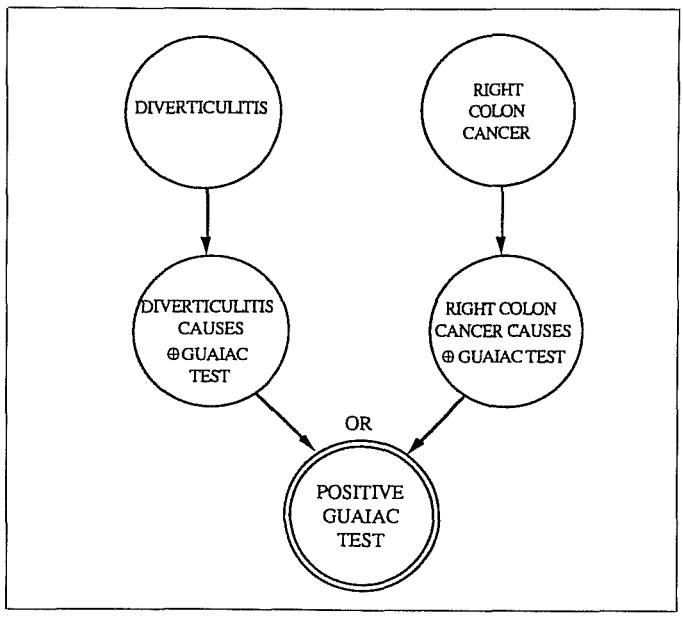

Fig. 6 A belief network representing the noisy-OR-gate interaction between diverticulitis, right colon cancer, and a positive guaiac test, as modeled in the current QMR-DT KB. This network models the belief that the mechanisms by which the two diseases may cause the finding of a positive guaiac test operate independently.

in Figure 8(a). A more accurate model of the influence of these two findings on acute myocardial infarction is pictured in Figure 8(b).

When we model historical findings as manifestations of disease, we introduce incorrect causal and conditional independences. For example, in such cases, causal independence is an inaccurate assumption, because we would not necessarily expect a historical finding to be absent when all the diseases associated with it fail to occur, since historical findings are not caused by disease. Conditional independence also is an inaccurate assumption, as demonstrated by the following example: Suppose we have a patient about which we originally have no data. We have a certain prior belief about whether this patient has a history of cigarette smoking. Suppose we are told that the only disease that the patient has is myocardial infarction. The information that the patient has myocardial infarction increases our belief that the patient has a history of cigarette smoking; let us suppose our probability of smoking changed from probability P_1 to probability P_2. Suppose we are next told that the patient has a history of diabetes mellitus. This new information decreases our belief that the patient has a history of cigarette smoking, P_2, since the history of diabetes mellitus tends to account for the myocardial infarction, and smoking is less needed as an explanation. In the first version of the QMR-DT KB, which we describe in this paper, we model historical findings as in Figure 8(a), although the relationship depicted in Figure 8(b) is more accurate. However, we model the historical findings age and gender not as in Figure 8(b), but rather as in 8(a) by conditioning the prior probabilities of diseases on age and gender. We describe the derivation of prior probabilities of diseases in Section 2.2.1.

Using the assumptions discussed in this Section, we can derive the connectivity of the QMR-DT belief network from the QMR KB. In Section 2.2, we discuss knowledge acquisition that we performed to add probabilities to the belief network.

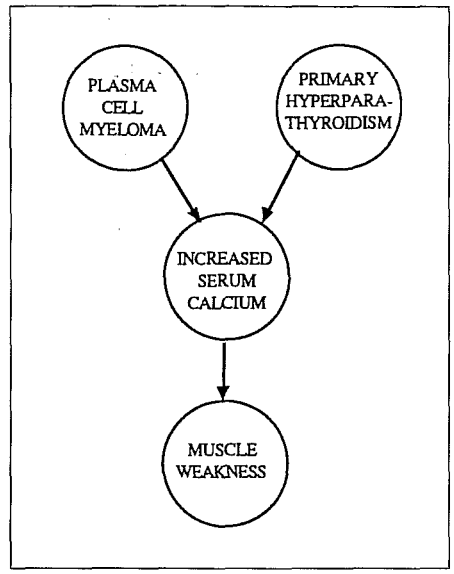

Fig. 7 A belief-network representation of two diseases (plasma cell myeloma and primary hyperparathyroidism) that operate through a common intermediate state (increased serum calcium) to cause a finding (muscle weakness).

2.2 Probabilities Used in the Model

Probabilities used in the QMR-DT model were derived both from the numbers contained in the QMR KB and from disease statistics. The necessary probabilities for our two-level belief network include the prior probabilities of diseases, and the conditional probabilities relating findings to disease. We describe the derivation of each of these in turn.

2.2.1 Prior Probabilities of Diseases

Our probabilistic model requires that prior probabilities on diseases be made explicit. We derived prior probabilities on diseases in the QMR-DT KB from data compiled by the National Center for Health Statistics (NCHS) on approximately 192,000 inpatients discharged from short-stay nonfederal hospitals in 1984 [29]. The diseases in the NCHS statistics are classified by the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) [30] coding system, but the INTERNIST-1 disease names do not always correspond directly to an ICD-9-CM name; therefore, we developed an approximate mapping between INTERNIST-1 names and ICD-9-CM codes. In about 60 percent of the cases, there was a close match between an INTERNIST-1 disease label and an ICD-9-CM disease label. When the INTERNIST-1 disease name encompassed more than one ICD-9-CM code, we summed the discharges for the subsumed codes. Conversely, when the ICD-9-CM code was more general than the INTERNIST-1 disease name, we denoted the discharges for the ICD-9-CM as an upper bound for the more specific INTERNIST-1 name. We reviewed the mapping for gross outliers and modified the prior probabilities using subjective estimates based on our clinical experience. The diseases with the

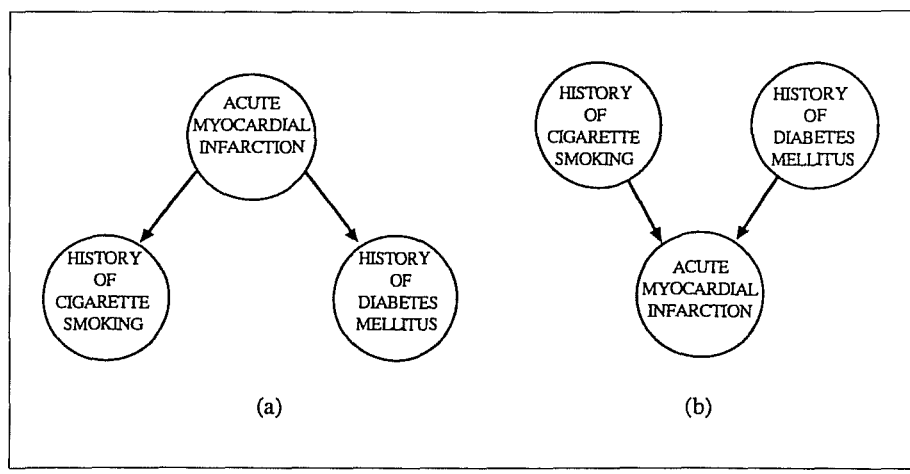

Fig. 8 Belief networks representing the relationship between diseases and historical findings (a) The current QMR-DT model assumes that all findings are manifestations of disease. (b) A more accurate model depicts a history of cigarette smoking and a history of diabetes mellitus as findings that may predispose a patient to acute myocardial infarction.

highest prior probabilities are bronchial asthma, angina pectoris, and diabetes mellitus, each with a prior probability of 0.02. The disease with the lowest prior probability is ectopic ACTH syndrome, with a prior probability of 2×10^{-5}.

The NCHS statistics provide discharge data categorized by age and gender. We chose to condition the prior probabilities of diseases on age and gender in the current QMR-DT model because age and gender are predisposing factors to diseases rather than manifestations of diseases. The prior probability in which we are interested for disease d_i is $P(d_i^+ \mid age, gender)$. The NCHS statistics do not provide us with these numbers, but rather provide data of the form $P(d_i^+ \mid age)$ and $P(d_i^+ \mid gender)$; therefore, we make the simplifying assumptions of both marginal independence of age and gender and conditional independence of age and gender given a disease. These assumptions are not essential to the QMR-DT model; we need only to acquire more specific statistics of the form $P(d_i^+ \mid age, gender)$. We recognize that the two assumptions are not mutually consistent [31]; however, we use them as first approximations. Assuming conditional independence of age and gender given a disease, we can write:

$$P(d_i^+, age, gender) = P(age \mid d_i^+) P(gender \mid d_i^+) P(d_i^+). \quad (6)$$

By the definition of conditional probability and the assumption of marginal independence of age and gender, we have:

$$P(d_i^+ \mid age, gender) = \frac{P(d_i^+, age, gender)}{P(age)\, P(gender)}. \quad (7)$$

Substituting (6) into (7), we derive an expression for the prior probability of d_i conditioned on both the age and the gender of the patient:

$$P(d_i^+ \mid age, gender) = \frac{P(age \mid d_i^+) P(gender \mid d_i^+) P(d_i^+)}{P(age)\, P(gender)}. \quad (8)$$

The NCHS data did not specify some of the age-specific or gender-specific discharges, when the number of discharges was deemed negligible. In these cases, we inferred the value by subtracting the other discharges from the general discharges for d_i and uniformly distributing the unaccounted discharges to the negligible categories. For example, the number of discharges of males with primary biliary cirrhosis is listed as negligible, whereas the number of females discharged with primary biliary cirrhosis is listed as 4,000. The total number of discharges of primary biliary cirrhosis is listed as 5,000. In this case, we calculated the value of discharges of males with the disease, 5,000 − 4,000 = 1,000. We also assumed a uniform distribution in each of the NCHS age categories, to distribute the age-specific discharges into the QMR age categories that did not correspond exactly to the NCHS categories. For example, we assumed a uniform distribution in the NCHS age category of 45 to 64 years to distribute the discharges for that category into the QMR age categories of *AGE 26 TO 55* and *AGE GTR THAN 55*. That is, we allocated (55 − 45)/64 − 45) = 0.53 of the discharges from the NCHS category of 45 to 64 years to the QMR category *AGE 26 TO 55* and 0.47 of the discharges to the category *AGE GTR THAN 55*.

In Part II of the paper, we investigate the sensitivity of the QMR-DT belief-network model to prior probabilities of diseases. Specifically, we compare the performance of the system using the prior probabilities that we derived as discussed in this Section to the performance using uniform prior probabilities of diseases.

2.2.2 Mapping of QMR Frequencies to Probabilities

The developers of QMR define a *frequency* between a specific disease and finding as a measure of "how often patients with the disease have the finding" [1 (p. 489)]. For each QMR frequency value of 1 to 5, we assessed an interval on the value of $P(f^+ \mid \text{only } d_i^+)$ from R. Miller, one of the primary developers of INTERNIST-1 and QMR. We assumed a symmetrical distribution in each of the intervals to generate the mapping appearing in Table 1, where the values of $P(f^+ \mid \text{only } d_i^+)$ are the midpoints of the corresponding intervals. We use the probabilities in Table 1 in the noisy-OR gate described in Section 2.1.4 to compute $P(f^+ \mid H)$ as in (5).

2.2.3 Leak Probabilities

The developers of QMR estimate that they will have to model approximately 750 diseases within the QMR KB "to cover the most important problems of internal medicine" [4]. The version of the QMR KB that we are using models 534 adult diseases. It

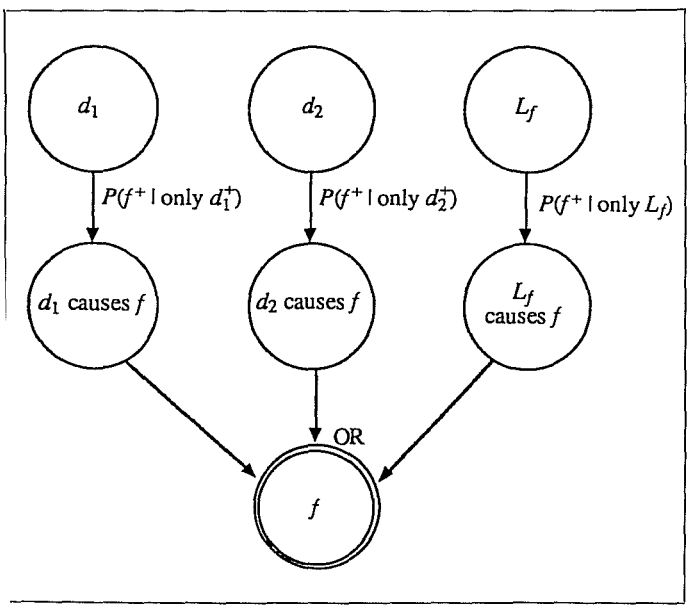

Fig. 9 A belief-network representation of the leak event L_f as an event that may cause f in a noisy-OR gate interaction with diseases d_1 and d_2. The prior probability that L_f occurs is 1 (not shown), and the probability that L_f triggers the sequence of events that leads L_f to cause f is $P(f^+ \mid \text{only } L_f)$. If any one of the three mechanisms are triggered (d_1 causes f, d_2 causes f, or L_f causes f), then f will be present.

is possible for a finding in the QMR-DT KB to occur as a manifestation of a disease that is not in the KB, or even to occur spontaneously in the absence of any disease whatsoever. (For example, some findings may be false-positive laboratory-test results.) We call either of these events *leak events.*

Consider a positive finding f^+. In the noisy-OR gate formulation of (5), $P(f^+ \mid H)$ becomes 0 when $\pi(f) \cap H^+ = \emptyset$. Since we compute $P(F^+ \mid H)$ according to (2), when there exists an $f \in F^+$ such that $P(f^+ \mid H)$ is 0, $P(F^+ \mid H)$ also becomes 0. Thus, the behavior of the noisy-OR gate as it appears in (5) is not practical with the QMR-DT belief network, since there would arise many occasions when $P(f^+ \mid H)$ would be 0. We augment the noisy-OR gate model with the leak event, providing for the possibility that f^+ may occur even if there does not exist a disease $d_i \in H^+$ that may cause it.

Let L_f be the event that the leak is active, regardless of whether or not other diseases may be causing f to occur. For each finding f, we model the leak in the QMR-DT belief network with a pseudo-disease node with prior probability $P(L_f) = 1$ and link weight $P(f^+ \mid \text{only } L_f)$, the probability that f occurs given that only the leak for f occurs. Figure 9 displays a belief network representing this model, in which the pseudo-disease node corresponding to finding f is labeled L_f. We assign a prior probability of 1 to this node to denote that the leak is present in every hypothesis H, so that $P(f^+ \mid H)$ is never 0 for any positive finding f. The leak L_f by itself causes f to occur with the probability $P(f^+ \mid \text{only } L_f)$. Alternatively, we say that the leak activates the leak mechanism "L_f causes f" with probability $P(f^+ \mid \text{only } L_f)$.

With the addition of the leak to the noisy-OR model, we derive a *leaky noisy-OR gate.* The probability that a finding f is absent, given a disease hypothesis H, is just the probability that all of the parent diseases of f fail to cause f and the leak fails to cause f as well. Adding the leak to (5), we have:

Table 1 A mapping between QMR frequencies and probabilities.

Frequency	$P(f^+ \mid \text{only } d_i^+)$
1	0.025
2	0.20
3	0.50
4	0.80
5	0.985

$$P(f^+ \mid H) = 1 - [1-P(f^+ \mid \text{only } L_f)] \prod_{d_i \in H^+ \cap \pi(f)} [1-P(f^+ \mid \text{only } d_i^+)]. \quad (9)$$

Thus, the leak is implicitly present in each disease hypothesis H[3].

We now explain how we derived the leak probabilities from the QMR KB. Recall that the import of a finding measures the "extent to which one is compelled to explain its presence in any patient" [1 (p. 469)]. The scale of import values ranges from 1 to 5, where an import of 1 denotes that the finding "is usually unimportant, occurs commonly in normal persons, and is easily disregarded", and an import of 5 denotes that the finding "must absolutely be explained by one of the final diagnoses" [10 (p. 463)].

We define a *significant disease* as one of the approximately 750 that the developers of QMR have modeled or intend to model in the QMR KB. Let D_f be the event that f is caused by a significant disease. Accordingly, let $P(D_f \mid f)$ be the probability that f is caused by a significant disease, given that f occurs. We wish to create a mapping between the import function and $P(D_f \mid f)$. To simplify this mapping, we assume that all significant diseases are contained in the QMR KB; the developers of QMR report that the import values for QMR findings were assessed originally assuming that all significant diseases will appear eventually in the QMR KB [28].

To derive a mapping between the import function and $P(D_f \mid f)$, we performed the following assessment. Approximately 10 findings of each import level were randomly selected to create a sample with a total of 44 findings.

[3] A proof that the leak term $P(f^+ \mid \text{only } L_f)$ is inconsequential for negative findings appears in [32].

For each finding, one of us (BM) performed the following steps:

1. List the most important significant diseases that can cause f,
2. List the most important nonsignificant diseases that can cause f,
3. Assess the false-positive rate of f,
4. Assess $P(D_f \mid f)$, the probability that f is caused by a significant disease, given that f occurs.

We used simple linear-regression techniques to determine the best straight-line relationship between the import values and assessed probabilities. The regression equation is able to explain a statistically significant amount of the variance in the assessed probabilities ($r^2 = 0.41$, $p = 0.001$). Table 2 shows, for each level of import, the fitted $P(D_f \mid f)$ value. In the QMR-DT KB, we currently use the mapping of imports to fitted values of $P(D_f \mid f)$ shown in Table 2.

In [32], we derive the following expression for $P(f^+ \mid \text{only } L_f)$:

$$P(f^+ \mid \text{only } L_f) = P(D_f) \frac{1-P(D_f \mid f)}{P(D_f \mid f)}, \quad (10)$$

where $P(D_f)$ is also derived in [32]:

$$P(D_f) = 1 - \prod_{d_i \in \pi(f)} [1-P(f^+ \mid \text{only } d_i^+)\, P(d_i^+)]. \quad (11)$$

We calculated the values of $P(f^+ \mid \text{only } L_f)$ for each of the 4,040 findings in the QMR-DT KB. To reduce the representational requirements for the leak probabilities, we did not condition the term for the prior probability of disease $P(d_i^+)$ in (11) on either age or gender. The largest calculated $P(f^+ \mid \text{only } L_f)$ value was for f = TACHYCARDIA, where $P(f^+ \mid \text{only } L_f) = 0.153$; the smallest value was $P(f^+ \mid \text{only } L_f) = 5.8 \times 10^{-8}$ for a number of findings, which included f = BLOOD CULTURE FRANCISELLA TULARENSIS.

In summary, the current QMR-DT model comprises a two-level belief network of findings and diseases. In this model, we assume marginal independence of diseases, conditional independence of findings given any hy-

pothesis of diseases, binary diseases and findings, and causal independence of diseases. We model a leak event as a pseudo-disease node, which may cause a finding with probability $P(f^+ \mid \text{only } L_f)$, which we derived using a mapping from QMR import values. We incorporate the influence of the leak event and other disease nodes on a finding using the assumption of causal independence in a leaky noisy-OR gate. We derived probabilities of the form $P(f^+ \mid \text{only } d_i^+)$ from a mapping of QMR frequencies. Finally, we used hospital-discharge statistics from the NCHS to compute prior probabilities of diseases. We refer collectively to the assumptions and probabilities summarized in this paragraph as the *QMR-DT model.* We distinguish the QMR-DT model from the algorithms that we use for inference on the model, which we discuss in the next section.

3. Algorithms for Inference

Given a set of positive and negative findings F and a model of the dependencies between diseases and findings in internal medicine, our goal is to compute $P(d_i^+ \mid F)$, the posterior marginal probability for each disease d_i: $1 \leq i \leq n$. Let μ be the assumption that diseases are mutually exclusive. We contrast $P(d_i^+ \mid F)$ with the posterior probability of a single-disease hypothesis, $P(\text{only } d_i^+ \mid F, \mu)$, where in the latter case we assume that the patient has exactly one disease. The assumption that diseases are mutually exclusive is clearly not applicable to the general problem of diagnosis in internal medicine, where patients often have several diseases simultaneously. The posterior marginal probability $P(d_i^+ \mid F)$, on the other hand, implicitly allows for the possibility that one or more diseases may be present in a patient. In Section 3.1, we discuss the complexity of calculating both $P(\text{only } d_i^+ \mid F, \mu)$ and $P(d_i^+ \mid F)$.

3.1 Exact Algorithms

Bayes' rule under the assumptions of mutually exclusive diseases and conditional independence of findings given any disease is sometimes referred to as *tabular Bayes'*[4] rule:

$$P(\text{only } d_i^+ \mid F, \mu) = \frac{P(F \mid \text{only } d_i^+)\, P(\text{only } d_i^+)}{\sum_{k=1}^{n} P(F \mid \text{only } d_k^+)\, P(\text{only } d_k^+)}, \quad (12)$$

where there are n diseases and $P(F \mid \text{only } d_i^+)$ is given by (2) when only d_i is present in H. Although the single-disease assumption is very restrictive, the tabular Bayes' formulation is appealing because of its low degree of computational time complexity – $\Theta(nm')$[5], where $m' = |F|$. In other words, the complexity is the product of the number of observed findings multiplied by the total number of diseases. Without a change in the order of computational time efficiency, we can augment the tabular Bayes' formulation by modeling the leak event, where we calculate $P(F \mid d_i^+)$ using the terms $P(f^+ \mid H)$ and $P(f^- \mid H)$ given by (9) when d_i is the only disease present in H. We will refer to the tabular Bayes' algorithm that includes the leak event as *TB*.

Straightforward application of Bayes' rule – without the assumption of mutually exclusive diseases – to the QMR-DT two-level belief network yields:

$$P(d_i^+ \mid F) = \frac{\sum_{H: d_i \in H^+} P(F \mid H)\, P(H)}{\sum_{H} P(F \mid H)\, P(H)}, \quad (13)$$

which has an inferential complexity of $O(m' n2^n)$[6]:

Suppose that we wish to compute $P(d_i^+ \mid F)$, given probabilities of the form $P(f^+ \mid \textit{only } d_i^+)$, $P(f^+ \mid \text{only } L_f)$, and $P(d_i^+)$. Assume that we compute the terms $P(H)$ and $P(F \mid H)$ using Equations (1), (2), and (9). To compute $P(F)$, the denominator of (13), we must compute 2^n terms of the form $P(F, H)$ in the summation, one for each possible hypothesis on n binary diseases. For each $P(F, H)$ term, we must compute $P(F \mid H)$ and $P(H)$. To compute $P(F \mid H)$, we need to compute $P(f^+ \mid H)$ for each $f \in F^+$ and $P(f^- \mid H)$ for each $f \in F^-$, which is a total of m' terms. For each of the terms $P(f^+ \mid H)$ or $P(f^- \mid H)$, we need to perform at most $n+1$ multiplications. To compute $P(H)$, we require n multiplications. Thus, to compute $P(F)$ we require $O([m'(n+1)+n]2^n) = O(m' n2^n)$ multiplications. Since the number of additions required to calculate $P(F)$ is bounded above by the number of multiplications required, we require $O(m' n2^n)$ *operations,* where an operation is either an addition or multiplication. We can compute the numerator of (13) for each disease d_i as we compute each term $P(F, H)$ in the denominator, adding $P(F, H)$ to a slot for each d_i, whenever

[4] The name *tabular Bayes'* rule is derived from the notion that we can compute $P(\text{only } d_i^+ \mid F, \mu)$ as in (12) from a $n \times m$ table of probabilities of the form $P(f_j \mid \text{only } d_i^+)$, where $1 \leq i \leq n$ and $\leq j \leq m$.

[5] We denote by $\Theta(n, m')$ the set of functions of the form $f(n, m')$ for which there exist positive constants c_1, c_2, n_0, m'_0, such that $c_1\, n\, m' \leq f(n, m') \leq c_2\, n\, m'$ for all $n \geq n_0$ and $m' \geq m'_0$. We say that the computational complexity of a procedure is $\Theta(n\, m')$ if the inference time for any problem of size (n, m') is equal to $f(n, m')$ for some $f(n, m') \in \Theta(n\, m')$.

[6] We denote by $O(m' n2^n)$ the set of functions of the form $f(m' n2^n)$ for which there exist positive constants c, n_0, m'_0, such that $0 \leq f(m' n2^n) \leq c\, m' n2^n$ *for all* $n \geq n_0$ and $m' \geq m'_0$.

[7] Suppose that we were to perform inference using (13) on machinery that could support 100 billion multiplications per second. By comparison, a Cray Y-MP/832 with eight processors has a limit of less than 3 billion floating-point operations per second [33]. Thus, our hypothetical machinery is over 30 times faster than a Cray Y-MP/832. Consider the time it would take to compute only the terms $P(H)$ in the summation of the denominator of (13). Each $P(H)$ term requires 534 multiplications. Thus, we would need more than $2^{534} \times 534 \approx 3 \times 10^{163}$ multiplications to compute the denominator of (13). On our hypothetical machinery, it would take more than 10^{144} years to complete the computation. Clearly, the brute-force application of Bayes' rule to inference on the QMR-DT belief network is intractable. Moreover, the problem of probabilistic inference on belief networks such as that of QMR-DT is known to be NP-hard [34].

$d_i \in H$. This calculation requires n additions in the worst case. Thus, we require $O([m' \; n+n]2^n) = O(m' \; n2^n)$ operations to compute $P(d_i^+ \mid F)$ for some d_i using an exhaustive application of Bayes' rule.[7]

We have sought to develop simulation algorithms to perform efficient inference on the QMR-DT belief network. Before describing simulation algorithms in Section 3.3, we briefly describe in Section 3.2 a heuristic algorithm that we have incorporated into a simulation algorithm.

3.2 A Heuristic Algorithm

We have developed a heuristic algorithm, which we call the *iterative tabular Bayes'* (ITB) algorithm. The ITB algorithm is similar to the INTERNIST-1 heuristic-reasoning scheme [1] in a number of ways, including ITB's use of *disease-to-disease* links to increase the score of diseases related to diseases that have been concluded previously. We created these links by mapping the frequencies of the QMR disease-to-disease links into probabilities of one disease given another, according to the scheme presented in Table 1 for disease-to-finding links.[8]

The heuristic ITB algorithm first uses TB to compute the posterior probability $P(\text{only } d_i^+ \mid F, \mu)$ for all diseases, where F is the set of all the findings in the diagnostic case. All diseases with a posterior probability greater than a threshold we have set empirically at 0.05 are concluded; that is, they are added to the final list of diagnoses from ITB. The prior probability of any disease that is associated (by a QMR disease-to-disease link) with any of the concluded diseases is then set to the probability associated with that link, if the current prior probability of the disease is less than the strength of the link. Each positive finding (excluding age and gender, which are used in setting the prior probabilities on diseases) that is associated with any of the concluded diseases is then removed from the set of findings. We call the steps of ITB just described a *stage*. We repeat ITB stages until there are no positive findings remaining with an import of greater than or equal to 3. At this point, we have completed a *round*. Note that after positive findings are removed upon the completion of a stage, the posterior probabilities computed by TB based on the remaining findings will generally lead to an entirely new set of diseases with a posterior probability greater than 0.05, which are thus concluded. These concluded diseases had posterior probabilities lower than 0.05 in previous stages, but because of the revised denominator of Bayes' rule in this particular stage, their posterior probabilities in this stage can exceed 0.05. Similarly, the algorithm may conclude different diseases in subsequent rounds since the prior probabilities of diseases are updated between rounds.

At the start of each round, the full set of positive and negative findings is restored. Also, the prior probabilities of all concluded diseases are set to 0 (for only the duration of the ITB algorithm). The algorithm repeats rounds until either it concludes a total of 20 or more diseases, or it completes 10 rounds. The ITB algorithm thus provides us with a set of heuristically selected diseases, where the cardinality of the set is typically about 20. We call this set the *heuristic-importance set*.

3.3 Simulation Algorithms

Because the general problem of probabilistic inference on belief networks is NP-hard, we have focused on developing simulation algorithms for inference on the QMR-DT belief-network model. In a stochastic simulation algorithm, the probability of an event of interest is estimated by the frequency with which the event occurs after a number of trials, where a trial consists of the instantiation of all nodes that have not been observed as evidence (unobserved nodes) to some value. Forward simulation is a type of stochastic simulation algorithm in which the order of the instantiation of the unobserved nodes is such that no node is instantiated before all of its parents are, and each node is instantiated to a value based on the probability of the different states of the node, given the states of its parents in the current trial [35]. The forward simulation algorithm that we use is called *likelihood-weighting*. This algorithm has been described previously by Fung and Chang [36] and Shachter and Peot [37]. The likelihood-weighting algorithm computes estimates of the posterior marginal probabilities of diseases that converge in the limit to the posterior marginal probabilities of diseases consistent with the QMR-DT model. (We will shall refer to the posterior marginal probabilities, given any evidence, consistent with a probabilistic model M, such as the QMR-DT model, as the posterior distribution *implied by* M.). In other words, the likelihood-weighting algorithm is guaranteed in the limit to compute the posterior distribution implied by the QMR-DT belief network, if allowed to run indefinitely. After a finite period of time, however, the estimates of the likelihood-weighting algorithm may deviate significantly from the probabilities implied by the model. (The reader is referred to Section 4 for discussion on the fidelity of the estimates of likelihood-weighting simulation.)

Consider a two-level belief network as in Figure 1, where $n = 3$ and $m = 4$. Suppose that of the four findings nodes in this sample network, only one node is observed, such that $f_1 =$ present. We wish to compute $P(d_i^+ \mid f_1)$ for $1 \leq i \leq 3$. We begin a *trial* of simulation by stochastically instantiating each of the three diseases in the network to present or absent according to the disease's prior probability $P(d_i^+)$. Suppose that the hypothesis instantiated is $H = \{d_1 =$ present, $d_2 =$ present, $d_3 =$ absent$\}$. We then compute the *sample score* $P(f_1^+ \mid H)$ and add this value to the variables for the present states of d_1 and d_2 and to the variable for the absent state of d_3. The first trial is thus completed. After any number of trials, we can estimate $P(d_i^+ \mid f_1)$ by dividing the value of the variable for

[8] Although we use heuristically the QMR disease-to-disease links in the ITB algorithm, we do not currently represent the disease-to-disease links as probabilistic dependencies in the QMR-DT belief-network model.

the present state of d_i^+ by the sum of the samples scores, computed for all of the trials completed thus far.

More formally, in the case of the two-level QMR-DT belief network described in Section 2, we estimate the marginal posterior probability of (13) by:

$$\hat{P}(d_j^+ \mid F) = \frac{\sum_{i=1}^{t} Z(H_i \mid F)\, U(d_j, H_i)}{\sum_{i=1}^{t} Z(H_i \mid F)}, \quad (14)$$

where H_i is the state of all the diseases as instantiated in the ith trial, and $U(d_j, H_i)$ is 1 if $d_j \in H_i^+$ and it is 0 otherwise. The sample score is given by:

$$Z(H_i \mid F) = \frac{P(F \mid H_i)\, P(H_i)}{P'(H_i)}, \quad (15)$$

where P' is the sampling distribution. In the simplest case (as in the example with three diseases), we instantiate the disease nodes based on their prior probabilities; that is, $P' = P$, and thus $Z(H_i \mid F) = P(F \mid H_i)$.

Alternatively, we can focus the likelihood-weighting simulation on certain instantiations of the network. Using this technique, called *importance sampling*, the algorithm instantiates the diseases not by their prior probabilities $P(d_i^+)$, but rather by any sampling distribution $P'(d_i^+)$ [37]. The only restriction on P' is that $P'(d_i^+) > 0$ whenever $P(d_i^+) > 0$. [Conversely, when $P(d_i^+) < 1$, it is necessary to make the restriction that $P'(d_i^+) < 1$.] The simulation's estimates of the posterior distribution will converge in the limit of infinity to the true posterior distribution as long as P' follows this restriction [38]. The estimates will converge most quickly when $P'(d_i^+)$ is equal to the true posterior distribution $P(d_i^+ \mid F)$ for each d_i, where $1 \leq i \leq n$ [39]. Of course, if we knew the true posterior distribution, then we would not have to perform the simulation. We can attempt to approximate the true posterior distribution using any method of our choosing to improve the convergence of the simulation. For instance, we can update $P'(d_i^+)$ at any point by setting it equal to simulation's current estimates of the posterior distribution $P'(d_i^+ \mid F)$. This technique is called *self-importance sampling* [37].

Table 2 A mapping between QMR imports and the probability that one or more significant diseases causes a finding f given that f is present.

Import	Fitted[a] $P(D_f \mid f)$	Std. Error $P(D_f \mid f)$
1	0.39	0.071
2	0.52	0.081
3	0.65	0.101
4	0.79	0.083
5	0.92	0.106

[a] The fitted $P(D_f \mid f)$ values were calculated by regressing the assessed values of $P(D_f \mid f)$ on the import values of the respective finding.

The self-importance updating function that we use is

$$P'_{\text{new}}(d_i^+) = \frac{P'_0(d_i^+) + g(t)\, \hat{P}_{\text{current}}(d_i^+ \mid N_F)}{g(t) + 1}, \quad (16)$$

where $g(t)$ is a linear function of the number of trials and P'_0 is the original sampling distribution (see next paragraph). We use P'_0 in the updating function so that very early in simulation the update will not converge to extreme probabilities (that is, close to 0 or 1).

We use the heuristic-importance set generated by ITB to derive the original importance distribution P'_0. We construct P'_0 such that the expected number of diseases instantiated from the heuristic-importance set is 1. That is, for all d_j in the heuristic-importance set, $P_0(d_j^+) = 1/N$, where N is the cardinality of the heuristic-importance set. We then set $P_0(d_j^+)$ for all d_j not in the heuristic-importance set to the greater of 10^{-3} or the prior probability on d_j. We use the threshold of 10^{-3} so that we can expect each disease to be instantiated during simulation a number of times before the sampling distribution is updated based on the simulation's probability estimates. We update the sampling distribution after the first 6,000 trials of simulation. Thus, we expect each disease not in the heuristic-importance set to be instantiated at least $6{,}000 \times 10^{-3} = 6$ times. The expected number of instantiated diseases not in the heuristic-importance set is approximately 1, based on the prior probabilities of disease and the threshold of 10^{-3} that we use. Thus, the expected total number of diseases instantiated per trial is about 2. The self-importance update to the sampling distribution is given by (16) where $g(t) = t/3{,}000$. Thus, during the first update of the sampling distribution P', we weight be estimates of the simulation $\hat{P}_{\text{current}}(d_j^+ \mid F)$ twice as much we weight the original sampling probability $P'_0(d_j^+)$. The simulation performs a total of 40,000 trials, updating the sampling distribution every 3,000 trials (after the first 6,000) with the same updating function.

4. Discussion

Although our approach to developing the QMR-DT model is to use as much of the QMR KB as possible, the current QMR-DT model does not use all of the information in the QMR KB. For instance, in the QMR-DT belief network, we do not use the QMR knowledge of evoking strengths, disease-to-disease relationships, and finding-to-finding relationships [10, 40]. We plan eventually to explore the possibility of incorporating these and other features of the QMR KB in the QMR-DT model. Other noteworthy issues in the development of the probabilistic model and inference algorithms include (1) the derivation of the prior probabilities, (2) the derivation of the leak probabilities, and (3) the estimates of the simulation algorithm. We address each of these issues in turn.

We encountered problems in deriving the prior probabilities of diseases to use in the model. We could not derive a one-to-one mapping between ICD-9-CM terms and QMR diseases. We attempted to remedy this problem by using subjective estimates for prior probabilities of diseases where the mapping was inexact. In addition, the NCHS hospital-discharge statistics are

only an approximation of the prior probabilities of diseases. There may be biases in the process of recording the discharge diagnoses. We believe, however, that the NCHS discharge statistics provide a reasonable first approximation for our initial investigation.

The current QMR-DT model uses the fitted leak probabilities from an assessment of findings of various types (history, physical, and laboratory tests) among imports of the range 1 to 5. We observe a substantial variation in the assessed $P(D_f \mid f)$ at certain import values (Table 2). We have yet to analyze this variation to determine whether it arises as a result of error in the assessment process, of variation in assessments across finding types, or for some other reason. It is possible that the leak probabilities will depend on the type of finding. Thus, in future research, we should perform a more detailed determination and sensitivity analysis of the leak probabilities used in the QMR-DT model.

Several researchers have successfully used belief networks to model medical domains for diagnostic inference. For example, researchers in the Pathfinder project use a belief network to model lymph-node pathology [11]. The primary difference between the QMR-DT belief-network model and the Pathfinder model is that the latter assumes that diseases are mutually exclusive. This assumption is accurate in lymph-node pathology, since diseases that occur simultaneously typically occur in different lymph nodes or in different regions of a single lymph node. In internal medicine, however, a patient often may present with more than one disease.

Researchers in the MUNIN project have built a belief network for modeling nerve disorders [41]. Like the QMR-DT belief-network model, the MUNIN model does not make the assumption of mutually exclusive disorders. Although the MUNIN model contains over 1,000 nodes with approximately 2,500 dependencies among them, the connectivity of the network is such that exact inference algorithms based on those developed by Lauritzen and Spiegalhalter [42] can be used on the MUNIN network. The QMR-DT belief network, however, is not amenable to these types of exact algorithms.

Since it is not practical to use exact algorithms to compute posterior probabilities of diseases from the QMR-DT belief network given an arbitrary set of findings, we use a simulation algorithm. The algorithm that we described in Section 3.3 produces estimates of the posterior marginal probabilities of diseases $\hat{P}(d_i^+ \mid F)$ that converge in the limit to the posterior marginal probabilities of diseases $P(d_i^+ \mid F)$ implied by the QMR-DT belief-network model. We attempt to decrease the time of convergence by incorporating two heuristics into the simulation: importance sampling and self-importance sampling. We derived empirically many of the constants that we use in the simulation. Using the constants we report in Section 3.3, we find that separate executions of the simulation produce similar posterior distributions on test cases of a total of 30 positive and negative findings.

Note that although we use a simulation algorithm, which itself uses heuristics, these heuristics are used to decrease the time of convergence, and they do not change the posterior probabilities that are derived. We use simulation algorithms solely to compute the posterior marginal probabilities of diseases implied by the QMR-DT belief-network model, rather than to compute heuristic scores according to some other model.

Ideally, to examine the convergence properties of the simulation that we are using, we would like to know the posterior distribution implied by the QMR-DT belief-network model – that is, a gold-standard distribution. For small test cases, we can use the quickscore algorithm developed by Heckerman [21] to compute the exact posterior marginal probabilities of diseases. This algorithm, which has a computational time complexity that increases exponentially with the number of positive manifestations in a diagnostic case, can compute the posterior marginal probabilities for a case of approximately nine findings in about one minute on a Macintosh IIci. Because of quickscore's computational requirements on cases with a large number of positive manifestations, however, we can not use the algorithm for inference on many diagnostic cases that are useful for testing the QMR-DT belief-network model.

On smaller test cases, we compared the distributions produced by our simulation to those produced by quickscore. We are confident that the distributions produced by the simulation are reasonably close to the distributions implied by the QMR-DT belief-network model because of (1) the similarity of the simulation's distributions to those of quickscore on smaller test cases, and (2) the reproducibility of the simulation's estimates on separate executions of the simulation on larger test cases. In Part II of this two-part series of papers, we analyze more formally the diagnostic performance of QMR-DT, using the simulation algorithm on several types of test cases. In addition, we analyze empirically the contribution of the importance sampling and self-importance sampling components of the algorithm.

We intend to augment the current QMR-DT belief network model with additional relationships among findings and among diseases. In addition, we intend to restructure the dependencies in the model to more accurately depict historical findings. We also plan to add intermediate pathophysiologic states, where necessary, to more accurately depict conditional independencies. Our belief is that as we more accurately model the relationships among disease and findings in internal medicine, the diagnostic performance of the QMR-DT system will continue to improve. All of the assumptions that we make in the current QMR-DT model are explicit. Based on empirical results, we can selectively address the particular assumptions that we identify as being most crucial to accurate diagnostic performance.

Acknowledgements

We are grateful to Randolph Miller for providing us with the INTERNIST-1 KB and for collaborating with us on the QMR-DT project. Ross Shachter and Mark Peot provided in-

sight on the likelihood-weighting algorithm. Lyn Dupré provided valuable comments on drafts of this paper.

This work was supported by the National Science Foundation under Grant IRI-8703710, the U. S. Army Research Office under Grant P-25514-EL, the National Library of Medicine under Grant RO1LM04529, and the National Center for Health Services Research Grant T2HS00028. Computing facilities were provided by the SUMEX-AIM Resource under National Library of Medicine Grant LM05208.

Appendix: Notation and Abbreviations

1. Algorithms

TB Tabular Bayes' algorithm. A heuristic algorithm that uses Bayes' rule under the assumptions of 1) mutually exclusive disease hypotheses and 2) conditional independence of findings given any disease

ITB Iterative tabular Bayes' algorithm. A heuristic algorithm that applies TB successively to various subsets of the set of findings.

2. Knowledge base (KB)

d_i A disease in the KB

f_j A finding in the KB

m The number of diseases in the QMR-DT KB

n The number of findings in the QMR-DT KB

F A set of findings that are observed

$|F|$ *The number of elements of F.* Also, $m' = |F|$

F^+ A set of positive findings that are observed

F^- A set of negative findings that are observed

H A hypothesis of diseases, in which each disease is assigned a value of *present or absent*

$\pi(f)$ The set of diseases that are parents of finding f (that is, those diseases with an arc leading to f).

REFERENCES

1. Miller RA, Pople HEJ, Myers JD. Internist-1: An experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med 1982; 307: 468–76.
2. Miller RA. INTERNIST-1/CADUCEUS: Problems facing expert consultant programs. Meth Inform Med 1984; 23: 9–14.
3. Miller RA, McNeil MA, Challinor SM, Masarie FEJ, Myers JD. The INTERNIST-1/QUICK MEDICAL REFERENCE project – Status report. West J Med 1986; 145: 816–22.
4. Miller R, Masarie FE, Myers JD. Quick medical reference (QMR) for diagnostic assistance. MD Computing 1986; 3: 34–48.
5. First MB, Soffer LJ, Miller RA. QUICK (Quick Index to Caduceus Knowledge): Using the INTERNIST-1/CADUCEUS knowledge base as an electronic textbook of medicine. Comp Biomed Res 1985; 18: 137–65.
6. Miller RA, E. MF. Use of the Quick Medical Reference program as a tool for medical education. Meth Inform Med 1989; 28: 340–5.
7. Von Neumann J, Morgenstern O. Theory of Games and Economic Behavior, Princeton. New York: Princeton Univ Press, 1947.
8. Spiegelhalter DJ, Franklin RCG, Bull K. Assessment, criticism and improvement of imprecise subjective probabilities for a medical expert system. In: Machine Intelligence and Pattern Recognition: Uncertainty in Artificial Intelligence 5. Henrion M, Shachter R, Kanal LN, Lemmer JF, eds. Amsterdam: North-Holland Publ Comp 1990: 285–94.
9. Harris N, Spiegelhalter DJ, Bull K, Franklin RC. Criticizing conditional probabilities in belief networks. In: Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care. Miller RA, ed. Los Alamitos, CA: IEEE Comp Soc Press, 1990: 805–9.
10. Masarie FE Jr, Miller RA, Myers JD. INTERNIST-1 properties: Representing common sense and good medical practice in a computerized medical knowledge base. Comp Biomed Res 1985; 18: 458–79.
11. Heckerman DE, Horvitz EJ, Nathwani BN. Update on the Pathfinder project. In: Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care. Kingsland LC, ed. Los Alamitos, CA: IEEE Comp Soc Press, 1989: 203–7.
12. deDombal FT, Leaper DJ, Horrocks JC, Staniland JR, McCain AP. Human and computer-aided diagnosis of abdominal pain: Further report with emphasis on performance. Brit Med J 1974; 1: 376–80.
13. Bankowitz RA, McNeil MA, Challinor SM, Parker RC, Kapoor WN, Miller RA. A computer-assisted medical diagnostic consultation service: Implementation and prospective evaluation of a prototype. Ann Inter Med 1989; 110: 824–32.
14. Cooper GF. Current research directions in the development of expert systems based on belief networks. Appl Stoch Models Data Anal 1989; 5: 39–52.
15. Horvitz EJ, Breese JS, Henrion M. Decision theory in expert systems and artificial intelligence. J Approx Reas 1988; 2: 247–302.
16. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, CA: Morgan Kaufman, 1988.
17. Howard RA, Matheson JE. Influence diagrams. In: Readings on the Principles and Applications of Decision Analysis. Howard RA, Matheson JE, eds. Menlo Park, CA: Strategic Decisions Group, 1981: 721–62.
18. Chavez RM, Cooper GF. Hypermedia and randomized algorithms for medical expert systems. Comp Meth Progr Biomed 1990; 32: 5–16.
19. Cooper GF. NESTOR: A Computer-Based Medical Diagnostic Aid That Integrates Causal and Probabilistic Knowledge. (Ph. D. Diss). Medical Information Sciences. Stanford, CA: Stanford University, 1984.
20. Andersen SK, Olesen KG, Jensen FV, Jensen F. HUGIN – A shell for building Bayesian belief universes for expert systems. In: Eleventh International Joint Conference on Artificial Intelligence. Sridharan NS, ed. San Mateo, CA: Morgan Kaufmann Publ 1989: 1080–5.
21. Heckerman DE. A tractable inference algorithm for diagnosing multiple diseases. In: Machine Intelligence and Pattern Recognition: Uncertainty in Artificial Intelligence 5. Henrion M, Shachter R, Kanal LN, Lemmer JF, eds. Amsterdam: North-Holland Publ Comp, 1990: 163–72.
22. Henrion M. Towards efficient probabilistic diagnosis in multiply connected networks. In: Influence Diagrams, Belief Nets and Decision Analysis. Oliver RM, Smith JQ, eds. Chichester: Wiley, 1990: 385–407.
23. Peng Y, Reggia JA. A probabilistic causal model for diagnostic problem solving – part I: Integrating symbolic causal inference with numeric probabilistic inference. IEEE Trans Syst Man, and Cybern 1987; SMC-17: 146–62.
24. Ganote DP. A Bayesian Set Covering Model Applied to Diagnosis in Blood Typing. (M. Sc. Thesis). Department of Computer Science. University of Ohio, 1989.
25. Habbema JDF. Models for diagnosis and detection of combinations of diseases. In: Decision Making and Medical Care. De Dombal FT, Grémy F, eds. New York: North-Holland Publ Comp 1976: 399–411.
26. Miller MCI, Westphal MC, Reigart JR, Barner C. Medical Diagnostic Models: A Bibliography. Ann Arbor, Mich.: University Microfilms International, 1977.
27. Good IJ. Good Thinking: The Foundations of Probability and Its Applications. Minneapolis, MN: Univ of Minnesota Press, 1983.
28. Miller RA. Personal communication, 1987.
29. Lawrence L. Detailed diagnosis and surgical procedures for patients discharged from short-stay hospitals. Vital & Helath Statistics 1986; 13.
30. Health Care Financing Administration. International Classification of Diseases (Ninth Revision) – Clinical Modification. Washing-

ton, D. C.: U. S. Department of Health and Human Services, Public Health Service, 1980.
31. Johnson RW. Independence and Bayesian updating methods. Artif Intell 1986; 29: 217–22.
32. Shwe MA, Middleton B, Heckerman DE, Henrion M, Horvitz EJ, Lehmann HP, Cooper GF. Probabilistic Diagnosis Using a Reformulation of the INTERNIST-1/QMR Knowledge Base – Part 1: The Probabilistic Model and Inference Algorithms. Knowledge Systems Laboratory Memo no. KSL-90-09. Stanford, CA: Stanford University, 1990.
33. Dongarra JJ. Performance of Various Computers Using Standard Linear Equations Software in a FORTRAN Environment. Mathematics and Computer Science Division Technical Memorandum no. 23. Argonne IL: Argonne National Laboratory, 1988.
34. Cooper GF. The computational complexity of probabilistic inference using Bayesian belief networks. Artif Intell 1990; 42: 393–405.
35. Henrion M. An introduction to algorithms for inference in belief nets. In: Machine Intelligence and Pattern Recognition: Uncertainty in Artificial Intelligence 5. Henrion M, Shachter R, Kanal LN, Lemmer JF, eds. Amsterdam: North-Holland Publ Comp, 1990: 129–38.
36. Fung R, Chang KC. Weighting and integrating evidence for stochastic simulation in Bayesian networks. In: Machine Intelligence and Pattern Recognition: Uncertainty in Artificial Intelligence 5. Henrion M, Shachter R, Kanal LN, Lemmer JF, eds. Amsterdam: North-Holland Publ Comp, 1990: 209–20.
37. Shachter RD, Peot M. Simulation approaches to general probabilistic inference on belief networks. In: Machine Intelligence and Pattern Recognition: Uncertainty in Artificial Intelligence 5. Henrion M, Shachter R, Kanal LN, Lemmer JF, eds. Amsterdam: North-Holland Publ Comp, 1990: 221–31.
38. Rubinstein RY. Simulation and the Monte Carlo Method. New York: John Wiley & Sons, 1981.
39. Geweke J. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 1989; 57: 1317–39.
40. Heckerman DE, Miller RA. Towards a better understanding of the INTERNIST-1 knowledge base. In: MEDINFO 86. Salamon R, Blum B, Jørgensen M, eds. Amsterdam: North-Holland Publ Comp, 1986: 22–6.
41. Olesen KG,. Kjaerulff U, Jensen F, Jensen FV, Falck B, Andreassen S, Andersen SK. A MUNIN network for the median nerve – a case study on loops. Appl Artif Intell 1989; 3: 384–404.
42. Lauritzen SL, Spiegelhalter DJ. Local computations with probabilities on graphical structures and their applications to expert systems (with discussion). J Royal Statist Soc 1988; B50: 157–224.

Address of the authors:
Blackford Middleton, M.D.,
Section on Medical Informatics,
Stanford University,
Medical School Office Building, X 215,
Stanford Ca 94305-5479,
U.S.A.