Converting a Rule-based Expert System into a Belief Network[*]

M. Korver & P.J.F. Lucas

Department of Computer Science, Utrecht University
P.O. Box 80.089
3508 TB Utrecht, The Netherlands
e-mail: lucas@cs.uu.nl

Abstract

The theory of belief networks offers a relatively new approach for dealing with uncertain information in knowledge-based (expert) systems. In contrast with the heuristic techniques for reasoning with uncertainty employed in many rule-based expert systems, the theory of belief networks is mathematically sound, based on techniques from probability theory. It therefore seems attractive to convert existing rule-based expert systems into belief networks. In this article, we discuss the design of a belief network reformulation of the diagnostic rule-based expert system HEPAR. For the purpose of this experiment, we have studied several typical pieces of medical knowledge represented in the HEPAR system. It turned out that, due to the differences in the type of knowledge represented and in the formalism used to represent uncertainty, much of the medical knowledge required for building the belief network concerned could not be extracted from HEPAR. As a consequence, significant additional knowledge acquisition was required. However, the objects and attributes defined in the HEPAR system, as well as the conditions in production rules mentioning these objects and attributes were useful for guiding the selection of the statistical variables for building the belief network. The mapping of objects and attributes in HEPAR to statistical variables is discussed in detail.

Keywords & Phrases: medical expert systems, belief networks, causal graphs, decision support systems.

1 Introduction

In heuristic, diagnostic expert systems, knowledge from a given domain is typically represented in the form of *production rules*, or rules for short. To express uncertainty in the domain, each conclusion of a rule is associated with a measure of confidence in its correctness. The exact meaning of such non-probabilistic measures of uncertainty usually is not clearly defined. An example of a method for handling uncertainty frequently applied in rule-based expert systems is the certainty-factor model developed by E.H. Shortliffe and B.G. Buchanan for the (E)MYCIN system [24, 1]. Certainty factors can be given a probabilistic interpretation, but

---

it turns out that in doing so usually an inconsistent specification of a probability distribution results [5, 3].

HEPAR is such a rule-based expert system; it aims at supporting the clinician in the initial assessment of the patient with a disorder of the liver or biliary tract [14]. The HEPAR system has been designed in such a way that it is capable of generating diagnostic explanations, based on clinical knowledge, with a similar amount of detail as employed by the clinician. The system incorporates the certainty-factor model to deal with uncertain medical knowledge. Although its diagnostic performance has been shown to be quite reasonable [11, 13], one of the problems with the HEPAR system is that its conclusions may be difficult to interpret, due to the unclear meaning of certainty factors.

In contrast with the heuristic models for reasoning with uncertainty, such as the certainty-factor model, the recent theory of belief networks is based on mathematically sound techniques, derived straight from probability theory [20]. A belief network permits the qualitative representation of dependencies and independencies among statistical variables. Since a causal relationship between two variables implies their statistical dependence, a belief network can be used to represent causal medical knowledge. The theory of belief networks then permits the use of this causal knowledge for diagnostic problem-solving. Such a causal model of a medical domain may be more easy to comprehend by medical students and unexperienced clinicians than a similar diagnostic rule-based system. Typically, in a rule-based expert system the heuristic knowledge encoded is too far removed from the underlying (patho)physiological processes to be understood by the non-expert physician. Furthermore, a belief network designed for diagnostic problem-solving can be applied to predict findings associated with (groups of) disorders. Although, in principle, the production rule formalism allows for the implementation of causal reasoning models, such a rule-based system cannot be used for diagnosis in a straightforward way. The sound mathematical basis and the expressiveness of the formalism of belief networks make it attractive to consider converting an existing rule-based expert system into a belief network.

In this article, we investigate the potentials of converting a diagnostic rule-based expert system into a belief network. This study is based on an actual experiment in which a representative part of the HEPAR system was converted into a belief network. Similar work has been carried out in the probabilistic reformulation of the INTERNIST-1/QMR knowledge base [25, 17]. However, the knowledge-representation and inference techniques used in this expert system differ considerably from those used in rule-based expert systems. We start with an introduction to the field of hepatology, and we will subsequently review the most important characteristics of the HEPAR system. The steps taken in converting the rule-based representation formalism applied in the HEPAR system into a belief network representation are discussed in Section 3. We conclude by summarizing the potentials and limitations of our approach to knowledge-base reformulation.

2 Overview of the HEPAR system

2.1 Clinical diagnosis in hepatology

Hepatology is a subspeciality of internal medicine which concerns the diagnosis and management of patients with disorders of the liver and biliary tract. Diagnosis of such disorders on purely clinical grounds is a difficult task, requiring much experience, as has been shown in several studies [15, 22, 4, 32, 33]. Clinicians with little experience in the field of hepatology

have shown to produce a correct specific diagnosis in patients with jaundice in less than 45% of the cases [31].

The main problems in the diagnosis of disorders of the liver and biliary tract are [34]:

- to distinguish between disorders of the liver (*hepatocellular disorders*) and of the biliary tract (*biliary obstructive disorders*),

- to differentiate between acute and chronic disorders, and

- to recognize whether a disorder is benign or malignant in nature.

These different disease groups require different diagnostic procedures, treatment plans and prognostic assessments. Especially the distinction between biliary obstructive and hepatocellular disorders is important, since the former typically require surgical treatment, whereas diseases from the latter group are managed conservatively. Both biliary obstructive and hepatocellular disorders may give rise to an impaired secretion of bile into the bile ducts, called *cholestasis*. This manifests itself by jaundice through the accumulation in the serum of the biliary substance bilirubin.
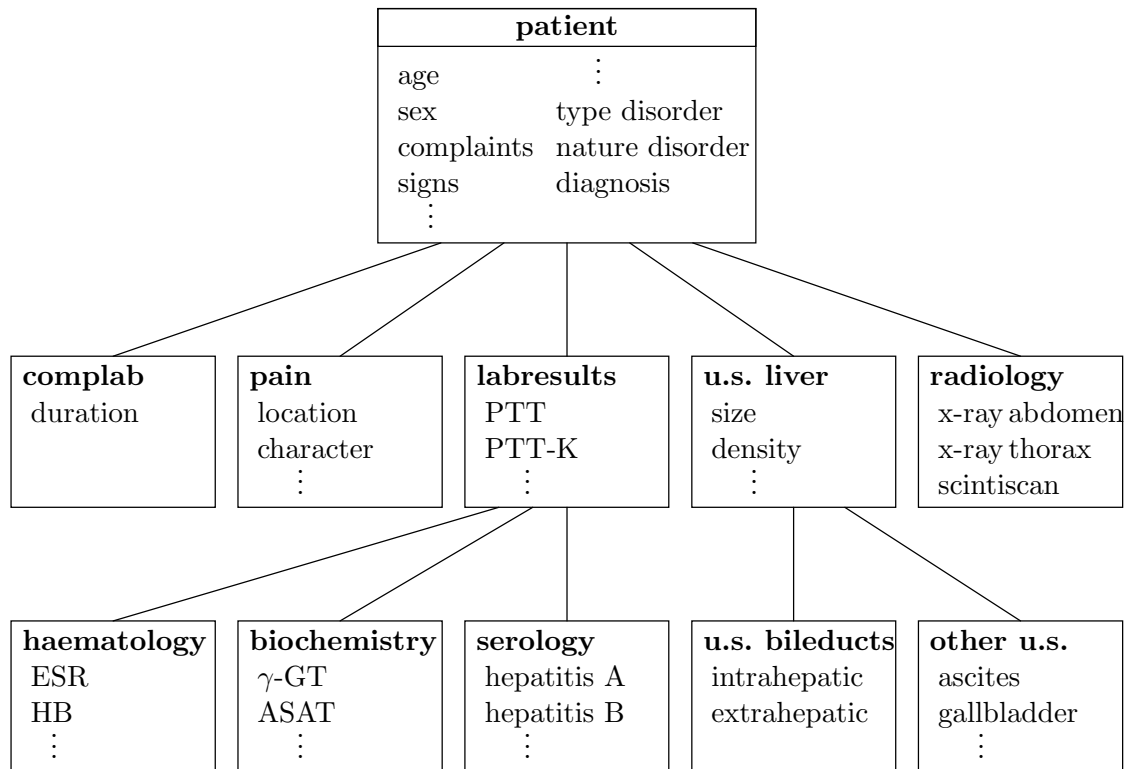
A large number of diagnostic methods is available to analyse patients with disorders of the liver and biliary tract. These include serological tests for several forms of viral hepatitis and autoimmune liver disease, and ultrasound imaging of the liver and biliary tract, which is of major importance in the diagnosis of biliary obstruction. In those cases where insufficient information is obtained from ultrasound, invasive techniques, such as percutaneous transhepatic cholangiography (PTC) and endoscopic retrograde cholangiopancreatography (ERCP), can offer help. In addition, liver biopsy can be employed in order to determine the aetiology and severity of chronic liver disease.

Of overriding importance in the diagnosis of hepatobiliary disease, however, is a detailed history and thorough physical examination. Together with a small number of routine laboratory tests, they frequently suffice for the formulation of a differential diagnosis. Supplementary diagnostic tests as mentioned above need then only be performed in selected cases, which minimizes not only expenses, but also diagnostic risk, since PTC, ERCP and liver biopsy are each associated with considerable morbidity and even a small mortality rate.

2.2   HEPAR, a rule-based system for the diagnosis of hepatobiliary disease

As in many clinical specialities, in hepatology establishing an early working diagnosis is important. Firstly, because treatment must often be initiated as soon as possible. Secondly, because it determines which diagnostic tests are selected for further assessment of the patient. The application of simple data from the medical history, physical examination and laboratory tests has therefore been taken as the point of departure for the development of the HEPAR expert system. Using merely these data, the system is capable of providing an initial assessment of the patient with a disorder of the liver or biliary tract, determining:

- whether the patient is suffering from an acute, subacute or chronic disorder,

- whether the disorder is hepatocellular or biliary obstructive in nature, and

- whether benign or malignant features are present.

```
                          ┌─────────────────────────┐
                          │        patient          │
                          │ age            ⋮        │
                          │ sex         type disorder│
                          │ complaints  nature disorder│
                          │ signs       diagnosis   │
                          │ ⋮                        │
                          └─────────────────────────┘
```

complab	pain	labresults	u.s. liver	radiology
duration	location character ⋮	PTT PTT-K ⋮	size density ⋮	x-ray abdomen x-ray thorax scintiscan

haematology	biochemistry	serology	u.s. bileducts	other u.s.
ESR HB ⋮	γ-GT ASAT ⋮	hepatitis A hepatitis B ⋮	intrahepatic extrahepatic	ascites gallbladder ⋮

complab = complaints, clinical signs, lab abnormalities
u.s.   = ultrasound

Figure 1: The object tree of the HEPAR system.

When additional information derived from some more specific, but generally available tests like ultrasound is included, the HEPAR system is also capable of producing a differential diagnosis consisting of a subset of possible diagnoses out of a set of some 80 disorders, ordered by the amount of evidence for each diagnosis. No use is made of data obtained by PTC, ERCP and liver biopsy. Thus closely following the clinical approach to the hepatological patient, the HEPAR system may be a suitable tool to assist the doctor in the initial assessment of the patient.

The knowledge base of the HEPAR system consists of two separate components:

- a collection of *object definitions*, structured as a tree, where each object definition describes an entity from the field of hepatology by listing a collection of its features, and

- a *rule base*, containing the problem-solving knowledge concerning diagnosis in hepatology.

The features of an object are usually called *attributes*. For example, in the HEPAR object tree, which is depicted in Figure 1, the attributes 'age', 'sex', 'complaints', 'signs' and 'diagnosis' of the object named 'patient' describe identically named features of a patient. The object tree specifies for each attribute:

```
if
  lessthan(biochemistry,AP,60) and
  greaterthan(biochemistry,gamma-GT,100) and
  between(biochemistry,ASAT,[50,200]) and
  between(biochemistry,ALAT,[50,200]) and
  lessequal(biochemistry,total_bili,17)
then
  conclude(patient,type-disorder,biliary_obstructive)
    with CF = 0.25
  conclude(patient,type-disorder,hepatocellular)
    with CF = 0.60
fi
```

Figure 2: Example of a production rule from the HEPAR system.

- a *value type* (singlevalued, multivalued, yes/no),

- a *domain* (integer, real, boolean, textvalue), and

- a *trace-class.*

A singlevalued attribute can only take one value with absolute certainty at a time. The domain of singlevalued attributes may be a real or integer interval, boolean or textvalue (a set of strings). Contrary to singlevalued attributes, multivalued attributes can take more than one (text)value at the same time. For instance, a patient can be of only one age (singlevalued attribute), but can suffer from several complaints at the same time (multivalued attribute). A yes/no attribute is a special singlevalued attribute which can take only one of the two values 'yes' and 'no'.

The trace-class specifies for each attribute whether it should be asked from the user, denoted by the keyword *askfirst* (or *initialdata* if the attribute should be asked at the beginning of the consultation), should be inferred from the rule base, in which case it is declared as a *goal*, or neither of the two, in which case it is only considered by the inference engine when required for determining the values of some other attribute. All goal attributes in the system are contained in the 'patient' object. The main goal attributes are:

- the nature of the disorder (benign or malignant),

- the type of the disorder (hepatocellular or biliary obstructive), and

- the final diagnosis.

Furthermore, portal hypertension is a goal attribute, since it is a complication whose presence is very important to assess in view of the far-reaching consequences for therapy and prognosis. The goal attribute 'type of the disorder' is multivalued. In this way, it is expressed that hepatocellular and biliary obstructive features are not mutually exclusive alternatives, but can be present in the patient simultaneously. Similarly, several interrelated diseases may be present within the same patient, so the attribute 'diagnosis' is also multivalued.

The rule base of HEPAR consists of a collection of 533 production rules having 1032 conclusions. An example of a production rule selected from HEPAR is shown in Figure 2.

This production rule consists of five conditions and two conclusions which are drawn only if all conditions are satisfied. This depends on given data and derived facts concerning a specific case. With each conclusion, a certainty factor is associated. A certainty factor expresses on a scale from -1 to $+1$ the confidence in the correctness of the conclusion, provided that all conditions of the rule are satisfied with absolute certainty. Under the given conditions, the example rule concludes that a hepatocellular disorder is more likely (CF = 0.60) than a biliary obstructive disorder (CF = 0.25).

Most of the production rules concern data obtained from medical history, physical examination and routine blood-chemistry. A brief analysis according to the conclusions of the rules will give an impression of the complexity of the rule base. The duration of complaints, clinical signs and lab abnormalities (attribute 'duration') is deduced in only 3 rules (one for each of the values acute, subacute and chronic). A total of 58 rules conclude about the nature of the disorder, with 44 conclusions having malignant and 54 having benign as a value. Conclusions about the type of the disorder are drawn in 257 rules (249 conclude hepatocellular, 244 biliary obstructive); of these, 225 are similar to the one shown in Figure 2. About the diagnosis of the patient 339 conclusions are drawn in 224 rules.

During a consultation, the production rules are applied by the inference engine to derive values for the goal attributes. The algorithm used is known as *top-down inference* or *backward chaining*. The process of data collection is guided by the course of the inference, except as far as the initialdata attributes are concerned (see above). Part of the reasoning in the system concerns the manipulation of the certainty factors to propagate the uncertainty in problem-solving knowledge to the final conclusions concerning the goal attributes.

The representation of knowledge by means of production rules with object-attribute-value triples, as well as top-down inference and the certainty-factor model are described in more detail in [1, 12].

3 Design of the belief network

3.1 Basic notions of belief networks

The formalism of belief networks offers an intuitively appealing approach for expressing inexact causal relationships between domain concepts [7, 20]. A *belief network* consists of two components [3]:

- A *qualitative representation* of the variables and relationships between the variables discerned in the domain, expressed by means of a directed acyclic graph $G = (V(G), A(G))$, where $V(G) = \{V_1, V_2, \ldots, V_n\}$ is a set of vertices, taken as the variables, and $A(G)$ a set of arcs (V_i, V_j), where $V_i, V_j \in V(G)$, taken as the relationships between the variables.

- A *quantitative representation* of the 'strengths' of the relationships between the variables, expressed by means of *assessment functions*.

Note that the expressive power of the belief network formalism for representing causal knowledge is restricted by the acyclic nature of the directed graph; cyclic causal influences cannot be represented in the formalism. More about this restriction will be said in Section 3.3.4.

Each of the variables $V_i \in V(G)$ may take one value from the variable's outcome space. The presence of an arc (V_i, V_j) in $A(G)$ represents a direct influential or causal relationship from V_i to V_j; its absence denotes that the variables V_i and V_j are assumed not to influence

each other directly. Because of the causal meaning of the arcs, the qualitative representation of the domain will be called a *causal graph* in this article.

For each vertex V_i an assessment function, denoted by γ_{V_i}, is defined by the conditional probability distributions $P(V_i = v_i | c_{\pi(V_i)})$, where v_i denotes an arbitrary value of V_i and $c_{\pi(V_i)}$ a conjunction of arbitrary values for each of the variables in the set $\pi(V_i)$ of parents of V_i. In case the set of parents of the vertex V_i is empty, the probability function to be specified is just the prior probability distribution $P(V_i = v_i)$, for each value v_i. The assessment functions together define a probability distribution on the domain.

After a belief network has been specified, it can be used for reasoning about uncertain information in the domain concerned. For example, one may enter into a medical belief network information concerning certain symptoms and signs observed in a patient, after which an updated probability distribution will be computed. The process of computing the updated probability distribution after entering specific evidence into the belief network is called *evidence propagation*.

Currently, there are two frequently applied algorithms for evidence propagation in use in belief networks. The first, and oldest, scheme originates from J.H. Kim and J. Pearl [7]. This scheme is only applicable in belief networks in which the causal graph has the form of a singly connected graph, i.e. a directed acyclic graph in which at most one path exists between any two vertices. The second scheme for evidence propagation originates from the work of S.L. Lauritzen and D.J. Spiegelhalter [8]. This algorithm is applicable in arbitrary belief networks. However, several techniques are known for extending the method of Kim and Pearl to arbitrary belief networks [20]. For an in-depth account of belief networks, the interested reader is referred to [18, 20]. For an overview of the principles of belief networks, the reader is referred to [12].

There are now several expert system shells available for building belief networks. For the experiments described in this article, we used the IDEAL system [27], to which some small programs for the graphical display and printing of belief networks were added.

3.2   Design considerations

In developing the belief network, the HEPAR system was taken as the point of departure. However, as we shall see, using the HEPAR system as a source for statistical variables would result in a belief network consisting of over 300 vertices. We have therefore confined the scope of the experiment by focussing on three prototypical aspects of the field:

- cholestasis, which is an important pathophysiological concept used in the early assessment of the patient;

- cirrhosis, which is an important concept embodying a large number of chronic disorders of the liver and biliary tract;

- Wilson's disease, which is a specific chronic disorder affecting the liver. In this recessively inherited derangement of copper metabolism, cirrhosis may develop as a consequence of progressive copper accumulation in the liver; copper deposits in other organs may cause extrahepatic disease and peripheral cornea pigmentations, called Kayser-Fleischer rings.

In this paper, we confine ourselves to discussing the last two aspects of the field.

The development of the collection of causal graphs was undertaken in several steps:

1. The variables, with associated outcome spaces, playing a central role in the domain were discerned, yielding an initial set of vertices.

2. The possible set of states for each variable was determined.

3. The causal relationships between the various variables were established, resulting in a set of arcs and several additional intermediate vertices.

After this initial design of the causal graphs, which was based on medical considerations only and not on considerations with respect to the final implementation, cycles introduced in the graph were cut, and some vertices were removed or added to simplify probability assessment. Finally, the causal graph concerning Wilson's disease was transformed into a belief network by filling in the assessment functions.

In the following sections, the above-mentioned steps will be discussed one by one.

3.3 Design of the causal graphs

3.3.1 Choice of statistical variables

The object tree of HEPAR served as the point of departure for the choice of the statistical variables in the design of the belief network. The concept of attribute closely resembles the concept of variable, and the notion of the domain of an attribute is similar to the notion of the outcome space of a statistical variable. More precisely, a statistical variable may be viewed as a singlevalued attribute, because the values in the outcome space of a variable are mutually exclusive.

A detailed analysis of the HEPAR system revealed that its object tree contains 102 different attributes grouped by 11 objects. Subdivided by value type, there are 30 yes/no, 49 singlevalued and 23 multivalued attributes. Whereas translating yes/no and singlevalued attributes into statistical variables is straightforward, converting a multivalued attribute requires splitting up its domain into mutually exclusive subsets, thus obtaining a collection of variables with associated outcome spaces.

As an example of this translation process, consider the multivalued attribute 'blood count' which is contained in the object 'haematology'. Among others, the domain of this attribute contains the values 'normal', 'lymphocytosis', 'leukocytosis' and 'leukocytopenia'. In the translation process, related attribute values were combined as the possible values of a new statistical variable. To the outcome space of each of these variables we added the special value 'normal', thus obtaining a non-binary variable. In this case, we obtained the statistical variable 'leukocyte count' with outcome space {leukocytosis, normal, leukocytopenia}. Attribute values having no direct relationship to any other value were translated into binary statistical variables. For example, the value 'lymphocytosis' was translated into the variable 'lymphocytosis' with outcome space {yes, no}.

The number of values in the domain of the multivalued attributes in the HEPAR object tree varies from 2 (e.g. for the attribute 'type disorder') up to 87 for the attribute 'diagnosis' (of which 9 not yet concluded in production rules). Most multivalued attributes have domains in which neither of the values excludes any one of the others. The conversion of such a multivalued attribute with a domain consisting of n values will thus result in n binary variables. As a consequence, splitting up the 23 multivalued attributes in HEPAR will yield almost 250 variables. Hence, using only the HEPAR system as a source for variables to be discerned will already render a belief network consisting of over 300 vertices.

The theory underlying belief networks does not provide a means for grouping of variables; consequently, no analogy of the concept of object exists. By converting the object tree into a set of statistical variables, the structural coherence between closely related concepts provided in HEPAR is lost; variables with other than causal relations are dispersed over the network.

3.3.2 State assessment of variables

In the previous section, we have shown how the translation of the attributes of HEPAR into statistical variables was carried out. Now, if V is any statistical variable, we may want to express probabilistic statements of the form

$$P(V = v | c_{\pi(V)})$$

where V is an arbitrary variable and v a value of V, or

$$P(v_1 \leq V < v_2 | c_{\pi(V)})$$

for V a numerical (integer or real) variable. Expressions such as $V = v$ and $v_1 \leq V < v_2$ will be called *states* of a variable in this article.

One of the problems in distinguishing states for the statistical variables discerned was that the theory of belief networks is not capable of dealing with continuous probability distributions. As a practical solution to representing a continuously distributed variable in a belief network, the variable is simply made discrete by subdividing its outcome space into a finite number of intervals, and taking these intervals as states. In case of the HEPAR system, such a conversion was required for singlevalued attributes with a real or integer domain.

The assessment of variable states was guided by the conditions and conclusions in the production rules of HEPAR. For non-numerical attributes, the states distinguished for the associated variables corresponded in most cases to the values enumerated in the associated outcome space. Most of these values were referred to in the conditions and conclusions of the rules. Numerical attributes were transformed into discrete statistical variables with a finite number of interval states, where the assessment of the cut-off points of the intervals was based on the conditions in which the corresponding attributes appeared. If necessary, these intervals were further refined by using cut-off points mentioned in the hepatological literature [16]. Consider for example Figure 2. This rule provides us with one cut-off point for 'AP' (alkaline phosphatase), 'gamma-GT' and 'total_bili' (total bilirubin); two cut-off points can be derived for 'ASAT' and 'ALAT' from this rule.

In HEPAR, 20 of the 49 singlevalued attributes have a numerical domain. For 19 of these singlevalued attributes, from one up to nine cut-off points were distinguished. One attribute is only used in an arithmetical expression, so no cut-off points could be determined for this attribute. 'Age' was the attribute for which nine different cut-off points were discerned. Since many disorders are age-specific, it seems inevitable to distinguish so many states for this feature. In [2] for instance, its range is subdivided in seven categories.

The results of this translation process are summarized in Table 1. So far, the conversion of the HEPAR system into a belief network did not cause insurmountable problems. Only a small number of attributes could not be converted into discrete statistical variables without using information from other sources than the HEPAR system.

| Attributes in HEPAR | | Derived variables | | | | | | | | | | |
| | | No. of values | | | | | | | | | | |
Type	No.	No.	2	3	4	5	6	7	8	9	10	unknown
yes/no	30	30	30									
singlevalued	49											
textvalue	29	29	20	5	4							
numerical interval	20	20	8	3	2	1	2		2		1	1*
multivalued	23	> 246	242	4								−**

Table 1: Summary of the results of the conversion of HEPAR attributes into discrete statistical variables. *) No cut-off points could be derived from the HEPAR rule base. **) In the current implementation, no domain is specified for the multivalued attribute 'drug', nor is the attribute applied in the rules.

3.3.3 Detecting causal relations between variables

In rule-based systems, attributes are placed into relation with each other by means of production rules. In a diagnostic expert system like HEPAR, most of these relations concern associations between disease manifestations on the one hand and diagnostic conclusions on the other hand, which are largely heuristic and not causal in nature.

Although the notion of causality employed in belief networks is interpreted somewhat broadly as any relationship in which one variable influences another, the differences between heuristic and causal knowledge imply that rules cannot be translated into causal graphs in a straightforward way. The analysis of the rule base of the HEPAR system was thus guided by the question of how the associations had to be interpreted in terms of causality. It turned out that in broad outline three different possibilities can be distinguished:

1. The condition (e.g. age) is viewed as one of the '*causes*' of the conclusion (e.g. Wilson's disease).

2. The condition (e.g. Kayser-Fleischer rings) is viewed as one of the *effects* of the conclusion (e.g. Wilson's disease).

3. Both a condition (e.g. Kayser-Fleischer rings) and a conclusion (e.g. chronic duration of complaints) are viewed as the *effects* of a common *cause* (e.g. Wilson's disease).

In the last case, we see that a variable taking part in a causal relation is actually missing in the rule concerned. As we shall see in the following, in designing a belief network, taking a set of production rules as a point of departure will yield an incomplete belief network. So, it will be necessary to add several intermediate vertices in order to obtain a satisfactory description of the causal disease mechanisms involved. We have used additional knowledge from hepatological textbooks, in particular [34], in order to extend the heuristic associations in the rules to causal relations.

We shall illustrate this process with the help of the HEPAR rule shown in Figure 3. In this rule, findings from history and physical examination are associated with a conclusion stating that the patient's disorder is chronic. An analysis of the relations between the variables corresponding to the mentioned object-attribute-value triples in terms of causes and effects is summarized in the following paragraph. Variables occurring in the rule have been indicated in italic.

```
if
  greaterthan(patient,time_course,26) or
  same(patient,complaint,haematemesis) or
  same(patient,signs,xanthomata) or
  same(patient,signs,testicular_atrophy) or
  same(patient,signs,spider_angiomas) or
  same(patient,signs,palmar_erythema) or
  same(patient,signs,Kayser_Fleischer_rings) or
  same(patient,signs,gynaecomastia) or
  same(patient,signs,oesophageal_varices) or
  same(patient,signs,caput_medusae) or
  same(patient,signs,butterfly_erythema)
then
  conclude(complab,duration,chronic) with CF = 1
fi
```

Figure 3: Heuristic knowledge in the HEPAR system.

The *time course* of the disorder is a direct result of its acute, subacute or chronic *duration*. Two sorts of chronic disorder of the liver are of importance in this respect: cirrhosis and chronic active hepatitis. The two major complications of cirrhosis are hepatocellular failure and portal hypertension. Portal hypertension may lead to the development of portasystemic collaterals such as *oesophageal varices* and *caput medusae* (dilated skin veins surrounding the navel); rupture of an oesophageal varix causes *haematemesis* (vomiting of blood). The pathogenesis of *testicular atrophy*, *gynaecomastia*, *spider angiomas* and *palmar erythema* in subjects with cirrhosis is not entirely clear, but seems related to abnormal sex hormone metabolism, among others a raised oestrogen concentration as a consequence of decreased liver function. (Oestrogens are metabolized largely by the liver.) *Kayser-Fleischer rings* are almost pathognomonic for Wilson's disease, but can also be found in primary biliary cirrhosis. In both disorders, the liver is usually cirrhotic. Primary biliary cirrhosis causes obstruction of the small intrahepatic biliary ducts, and since in chronic cholestasis very high plasma lipid levels can occur, *xanthomata* (deposits of lipids in the skin) may also appear in this disease. Chronic active hepatitis, in itself a type of chronic liver disease, is also a likely cause of cirrhosis. One of the disorders resulting in chronic active hepatitis is autoimmune chronic hepatitis, which is associated with *butterfly erythema*.

Part of the knowledge lacking in the production rule analysed above is represented in other HEPAR rules, namely the rules concerning portal hypertension, Wilson's disease, primary biliary cirrhosis and autoimmune chronic hepatitis. When represented as a causal graph Figure 4 is obtained. As can be seen, all conditions in the example rule from Figure 3, except the first, correspond to variables directly or indirectly influenced by a variable which also influences the variable corresponding to the rule's conclusion. The vertex 'cirrhosis' has been added to enable connecting the conclusion to the conditions 2, 4-6 and 8-10; the vertices labelled 'portal hypertension', 'portasystemic collaterals', 'liver function' and 'oestrogens' have been added here to obtain a more natural causal, pathophysiological description in correspondence with that provided in the textbooks. The disease causally related to condition 11 (butterfly erythema) is 'autoimmune chronic hepatitis', which has been added to the graph
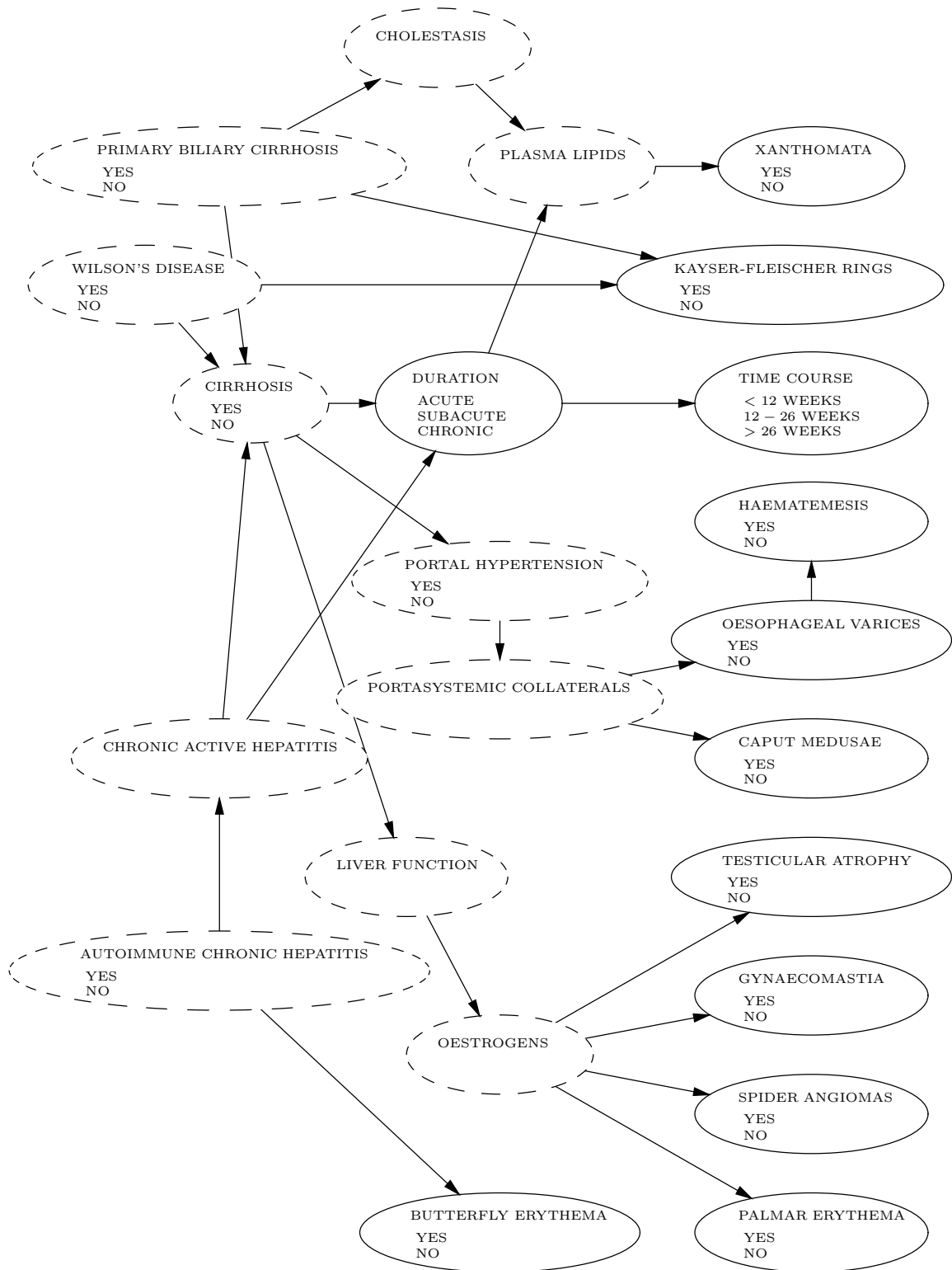
Figure 4: Causal graph corresponding to the Hepar rule in Figure 3; additional causal knowledge is derived from textbooks. Added variables are dashed.
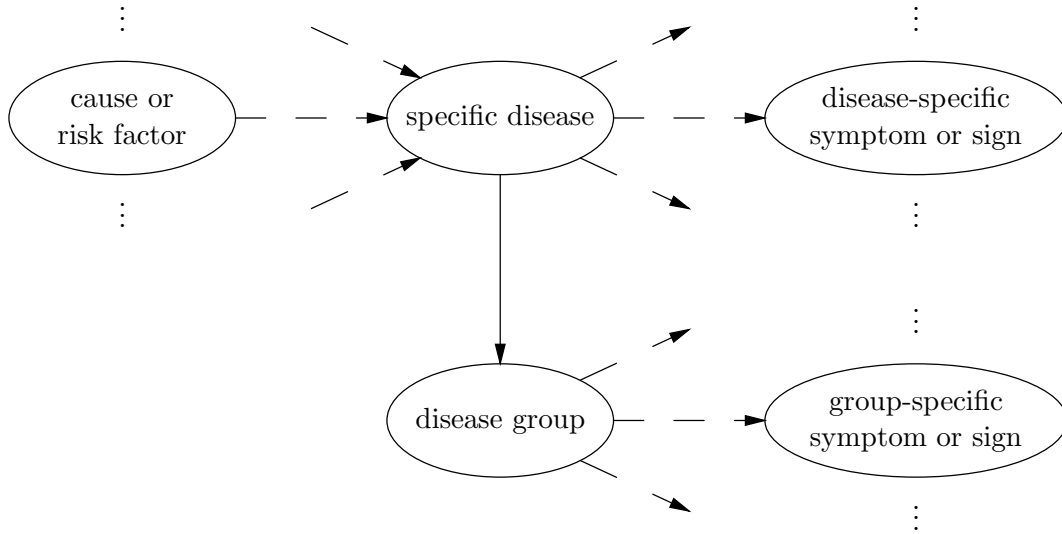
Figure 5: Global graph structure for a diagnostic belief network.

as a common cause of 'cirrhosis' (via 'chronic active hepatitis') and 'butterfly erythema'. In a similar way, condition 7 (Kayser-Fleischer rings) caused addition of the vertices 'primary biliary cirrhosis' and 'Wilson's disease', while condition 3 (xanthomata) has been explained by means of 'cholestasis' and 'plasma lipids'.

This example brings into the picture a third type of reasoning in addition to causal and heuristic reasoning: *taxonomic reasoning*. Taxonomic reasoning is based on hierarchical classification of concepts in groups and subgroups according to common features. The higher a concept is placed into a hierarchy, the more general it is. In HEPAR, for instance, this type of reasoning is applied in the rules concerning specific diseases. One of the questions that arose in representing taxonomic knowledge in a causal graph is whether the arcs between the vertices should be drawn from vertices representing specific concepts to vertices representing more general concepts, or the other way around. It is easy to see that, as far as taxonomies of disorders are concerned, we should follow the first approach, since disorders are categorized according to common features *resulting* from the disorders. The resulting method is depicted in Figure 5. Part of the graph in Figure 4 may thus also be interpreted as a hierarchy, with the vertex 'duration' representing the most general concept, 'cirrhosis' and 'chronic active hepatitis' as intermediate concepts, and 'primary biliary cirrhosis', 'Wilson's disease' and 'autoimmune chronic hepatitis' as the most specific concepts, being specific disorders.

3.3.4   Cycles in graphs and pathophysiological knowledge

Representing causal knowledge as a graph as described in the previous subsection does not necessarily render a causal graph. As we have discussed in Section 3.1, a causal graph is by definition acyclic; in representing pathophysiological knowledge, however, cycles may be introduced. So, the restricted expressive power of belief networks in representing cyclic influences may hinder a natural representation of medical knowledge. In developing our causal graphs, we obtained directed cycles representing positive feedback influences, as well as cycles representing negative feedback influences. Although we do not provide a theoretical solution to dealing with directed cycles in the article, we show that by means of domain-specific ar-

guments, the cycles may be cut. Here, we only discuss the handling of negative feedback influences. The process is illustrated by a causal graph for cirrhosis which we developed in converting the HEPAR system.

The graph in Figure 6 represents the causal relations between cirrhosis, portal hypertension and ascites (fluid in the peritoneal cavity). As depicted, the distortion of the liver architecture in cirrhosis may lead to portal hypertension and, as a consequence of a partial block to hepatic venous outflow, to an increased exudation of hepatic lymph possibly exceeding the reabsorption capacity. In that case, the additional protein-rich lymph exudes from the liver surface into the peritoneal cavity. Another complication of cirrhosis is an impaired liver cell function, which may result in a reduced albumin synthesis. Both the albumin loss to the peritoneal cavity and the impaired synthesis of albumin affect the amount of serum albumin; the plasma volume determines its actual concentration. When portal hypertension is transmitted back to the splanchnic area, transudation of fluid into the peritoneal cavity will result, which is also favoured by a reduced colloid osmotic pressure consequent to reduced serum albumin. Because of this loss of fluid into the peritoneal compartment and of pooling of blood in the splanchnic circulation, the 'effective' blood volume is reduced. Renal perfusion therefore falls and the kidneys retain sodium (and water), so that the plasma volume increases. This renal regulatory mechanism is represented as the cycle numbered '1' in Figure 6. This *negative feedback* process, which is also effective under normal conditions, takes care of a stable plasma volume. Splanchnic pooling results through this cycle in a resetting of the plasma volume to a higher level. Because of this and since the renal regulatory mechanism is in fact beyond the scope of the domain, it seems justified to draw an arc between 'splanchnic pooling' and 'plasma volume', and to remove the remainder of the cycle from the graph. The second cycle depicted in the graph represents a pathological negative feedback process. The effect of splanchnic fluid transudation on the plasma volume, however, is undone by the above-mentioned renal regulatory mechanism. Therefore, cycle 2 may safely be cut by removing the arc from 'splanchnic fluid transudation' to 'plasma volume'.

3.3.5 Final adjustment of the causal graph

In addition to the problem of handling cyclic causal influences in a causal graph, we have to solve the problem of keeping the number of vertices as small as possible. The main reasons for this are:

- Since many of the corresponding variables are hardly accessible for observation, probabilities with regard to them will not be available in literature nor obtainable from experts.

- Inference speed will decrease considerably as the number of variables in the belief network increases.

We recognize two methods for reducing the number of vertices. The first one involves cutting off dead-end paths without any vertices representing diagnostic hypotheses or disease manifestations. Because paths not involving such vertices are not of interest in diagnostic problem-solving, they may be removed. In the second approach, we omit some of the intermediate vertices; the remaining ones are connected in an obvious way, as is illustrated in Figure 7 which has been obtained from Figure 6.

The gain expected from a reduction of the number of additional vertices should be weighted against the loss of information involved. Leaving out intermediate steps in causal mechanisms
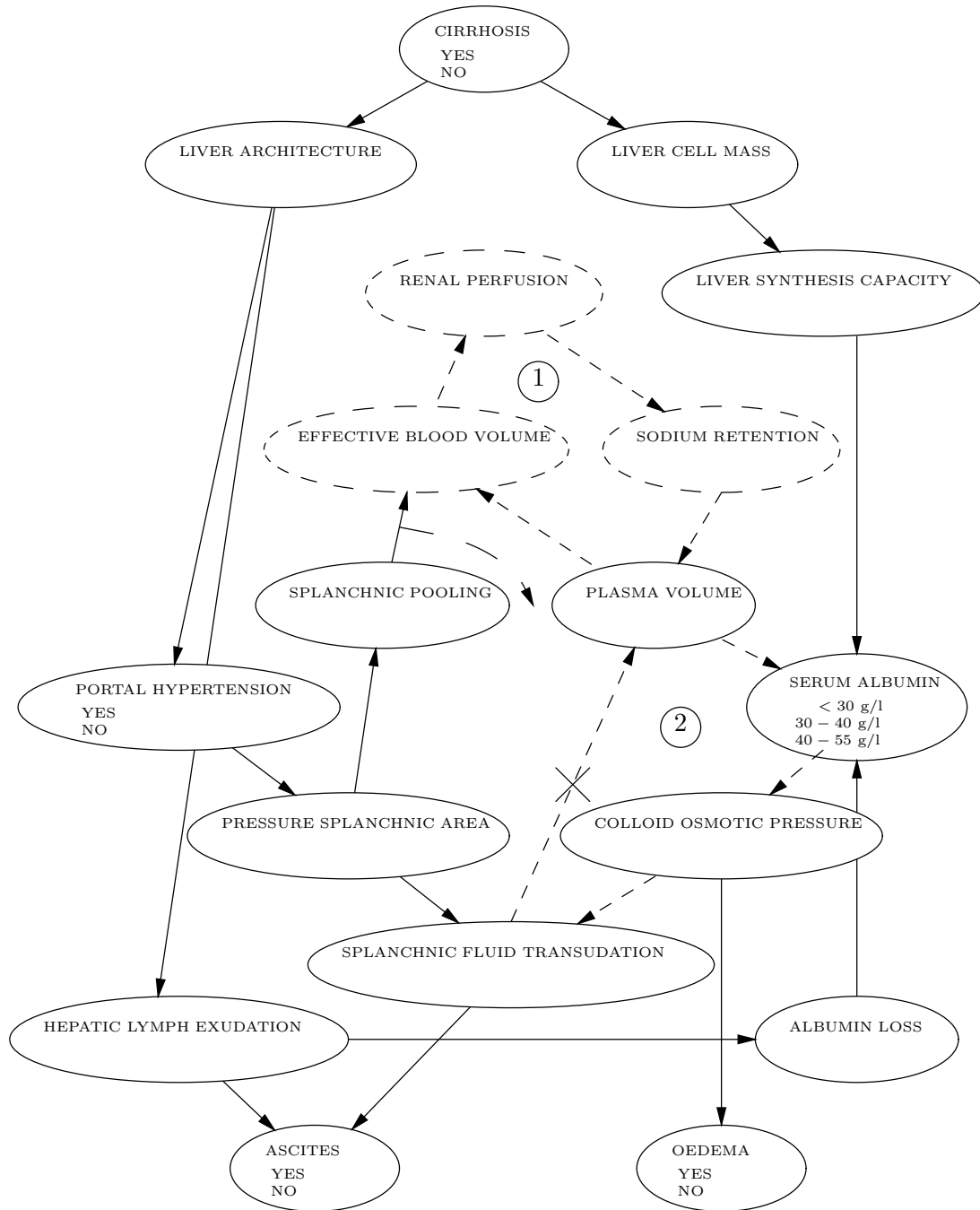
Figure 6: Cycle 1 represents a negative feedback process for keeping the plasma volume up to a constant level. The effect of splanchnic pooling on this process basically is an adjustment of the plasma volume to a higher steady state. It therefore seems justified to eliminate the dashed vertices and redirect the arc departing from 'splanchnic pooling' to 'plasma volume'. Since the effect of splanchnic fluid transudation on the plasma volume is undone by the above-mentioned regulatory mechanism, cycle 2 can harmlessly be interrupted by removing the crossed arc.
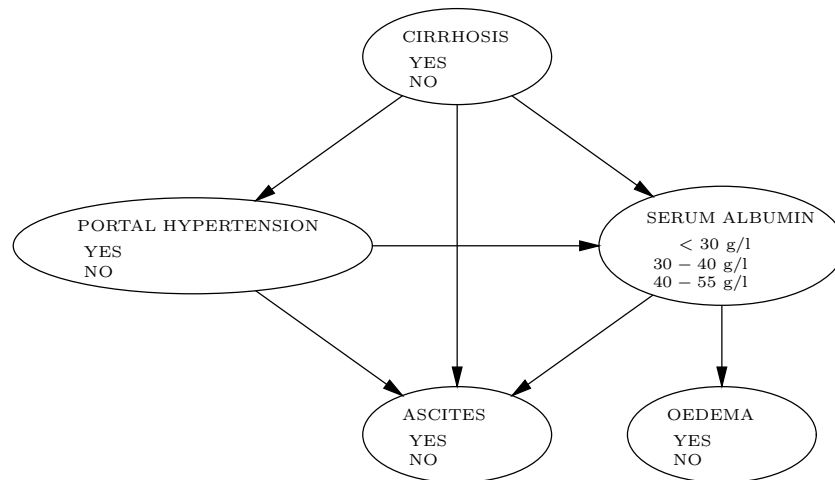
Figure 7: A strongly simplified version of the graph in Figure 6. Vertices not discerned in HEPAR have been removed. A pair of vertices has been connected by an arc iff in the original graph a path existed between them.

renders the network harder to interpret and harder to adjust. We have therefore been very reluctant in applying this method.

3.4 Probability assessment

3.4.1 Sources of probabilistic information

After having carried out the steps discussed above, several causal graphs were obtained. In turning these into belief networks, we had to find a fill-in for the assessment functions by (conditional) probabilities.

On first thought, we hoped it might be possible to derive this probabilistic information from the certainty factors specified in the production rules in the HEPAR system. However, recall from the preceding sections that none of the production rules could directly be translated into a causal graph with the conclusion as a vertex having incoming arcs originating from vertices representing the conditions. So, the certainty factors specified in the production rules could not be applied as the basis of probability assessment. We have therefore consulted standard medical textbooks for collecting such probabilistic information [34, 21].

One of the difficulties encountered in obtaining assessment functions from medical literature is that usually in describing diseases a non-numerical notion of uncertainty is employed, featured by terms such as 'rare', 'common', 'almost all', etcetera. In case probabilistic information is specified, it is often incomplete. Also, available conditional probabilities may concern variables not causally related to each other, usually rendering these numbers useless for specifying the assessment functions in a belief network. Such information can, however, be applied for the *evaluation* of a belief network. Furthermore, probabilistic information from the literature is not always based on the population of the country concerned. Because factors such as hospital referral patterns and geographical variation in disease presentation play a role, probabilistic information available from studies carried out elsewhere may not be useful. Non-conditional probabilities seem to be more susceptible to population characteristics than conditional probabilities are.
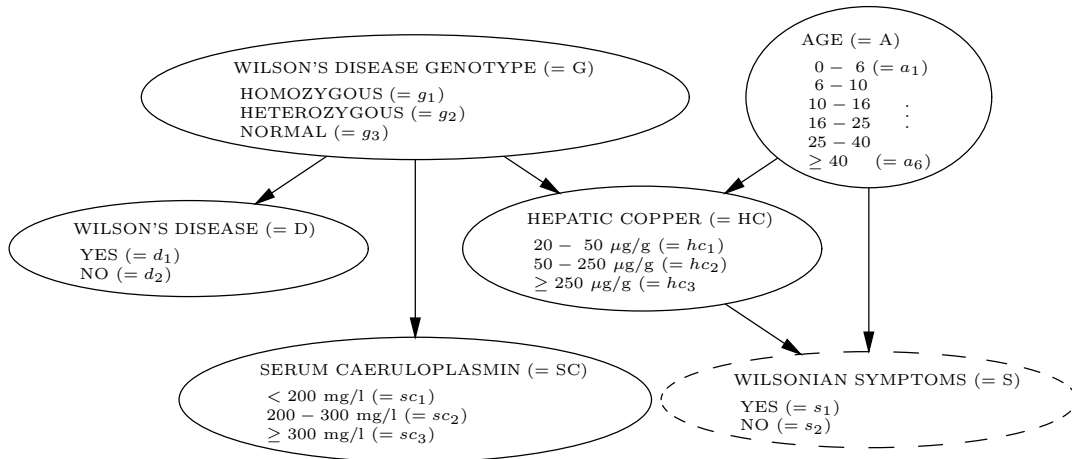
Figure 8: Causal graph representing some qualitative knowledge with regard to Wilson's disease.

One way to simplify obtaining probabilistic information is to reduce the number of parents of a given vertex in a causal graph. This may be done by introducing 'summarizing' intermediate vertices, a technique known as 'divorcing multiple parents' [19]. The main advantage of this technique is that conditional probabilities $P(V = v \mid v^1 \wedge \cdots \wedge v^n)$ where n is small, say less than 4, are much easier to estimate than for n large. For our application, the number of parents of vertices was always limited.

3.4.2 Estimating probabilities

We shall illustrate the process of probability assessment using information from medical literature, by discussing the belief network we obtained for Wilson's disease in some detail. This disorder is attended with low levels of serum caeruloplasmin and progressive copper accumulation in the liver. To be more precise, it is the genotype with regard to this disease which influences serum caeruloplasmin and hepatic copper levels. The capacity of the hepatocytes to store copper is eventually exceeded and release into blood and uptake in extrahepatic sites occurs, causing extrahepatic disease and Kayser-Fleischer rings. In Figure 8, the relationships among the above-mentioned concepts have been represented by means of a causal graph. Clinical manifestations, here summarized in the vertex 'Wilsonian symptoms', can be due to liver involvement as well as to involvement of other organs.

From the observation that Wilson's disease is a recessively inherited disorder all homozygous subjects will eventually have symptoms of, the values for the assessment function for the vertex 'Wilson's disease' follow immediately. Only subjects with a pair of 'Wilson's alleles' have the disease. Hence, the values of the assessment function γ_D are as depicted in the right half of Table 2.

According to Mendelian laws, we have

$$P(g_1) = P(g_1) \cdot P(g_1) + \frac{1}{2}P(g_1) \cdot P(g_2) + \frac{1}{4}P(g_2) \cdot P(g_2)$$

in non-consanguineous societies. Rewriting this equation results in

$$P(g_2) = -P(g_1) + \sqrt{P(g_1) \cdot (4 - 3P(g_1))}.$$

	$\gamma_G(g_i)$		$\gamma_D(d_j \mid g_i)$	Wilson's disease	
Wilson's disease genotype				yes	no
Homozygous	0.00003		Homozygous	1	0
Heterozygous	0.01092		Heterozygous	0	1
Normal	0.98905		Normal	0	1

Table 2: Assessment functions $\gamma_G(g_i)$ and $\gamma_D(d_j \mid g_i)$.

From the literature we have that the world-wide prevalence of Wilson's disease is about one in 30 000, we can now assess γ_G. Its values are given in the left part of Table 2.

Here, we have to keep in mind, however, that we are developing a belief network for use in hepatological patients, in whom the prevalence of Wilson's disease is much higher. With regard to 'age', a similar problem occurs: age distribution in the population of a specialized liver unit is very likely to differ from that in normal population. For serum caeruloplasmin, we found that 94.9% and 20.4% of homo- and heterozygotes for the Wilson's disease gene, respectively, have levels below 200 mg/l [30]. A table containing a summary of analytical data (means and standard deviations) in patients with Wilson's disease, heterozygous carriers, and control subjects [29], allowed us to estimate values for the assessment function γ_{SC} for caeruloplasmin levels, using statistical techniques assuming normal distribution.

		Hepatic copper	
		(μg/g dry weight)	
Group	No.	Range	Mean $\pm$ SD
Wilson's disease:			
Asymptomatic	36	152 – 1828	983.5 $\pm$ 368
Symptomatic	33	94 – 1360	588.3 $\pm$ 304
Heterozygous carriers	14	39 – 213	117.0 $\pm$ 51
Normal subjects	16	20 – 45	31.5 $\pm$ 6.8

Table 3: Summary of analytical data in patients with Wilson's disease, heterozygous carriers, and control subjects. Source: [29].

The same techniques were applied to the information with regard to hepatic copper levels, reproduced in Table 3. Contrary to serum caeruloplasmin, the hepatic copper concentration is not only influenced by the Wilson's disease genotype, but also by age. This applies only to homozygous patients, however. In normal subjects and heterozygous carriers, hepatic copper concentration does not vary with age, so that we have

$$P(hc_i \mid g_j \wedge a_k) = P(hc_i \mid g_j) \ \text{ if } j \neq 1.$$

These values are given in the lower half of Table 4, which has been derived from Table 3 as described above. With regard to homozygotes, conditioning on the values of S (Wilsonian symptoms) gives us

$$P(hc_i \mid g_1 \wedge a_k) = \sum_s P(hc_i \mid g_1 \wedge a_k \wedge s) \cdot P(s \mid g_1 \wedge a_k).$$

We next assumed that the influence of age on hepatic copper can be neglected once we know whether a Wilson's disease patient is symptomatic or not. The last equation can then be

Group	Hepatic copper (μg/g dry weight)		
	< 50	$50 - 250$	≥ 250
Wilson's disease:			
Asymptomatic	0.0056	0.0176	0.9768
	$0.0020 - 0.0139$	$0.0079 - 0.0356$	$0.9505 - 0.9901$
Symptomatic	0.0384	0.0951	0.8665
	$0.0166 - 0.0793$	$0.0542 - 0.1473$	$0.7734 - 0.9292$
Heterozygous carriers	0.0951	0.9004	0.0045
	$0.0294 - 0.2327$	$0.7666 - 0.9494$	$0.0007 - 0.0212$
Normal subjects	0.9967	0.0033	0
	$0.9857 - 0.9994$	$0.0006 - 0.0143$	

Table 4: Conditional probabilities of the values of hepatic copper concentration given Wilson's disease genotype and presence or absence of Wilsonian symptoms. The probabilities are derived from Table 3 by statistical methods assuming normal distribution. The probability intervals on the lower lines are meant as an indication of the accuracy of the corresponding probabilities on the upper lines.

simplified to:

$$P(hc_i \mid g_1 \wedge a_k) = \sum_s P(hc_i \mid g_1 \wedge s) \cdot P(s \mid g_1 \wedge a_k)$$

Note that eventually all untreated subjects with a pair of Wilson's disease genes will become symptomatic Wilsonian patients. Then, Figure 9 provides us with the required information for assessing the conditional probabilities $P(s \mid g_1 \wedge a_k)$. The conditional probabilities $P(hc \mid g_1 \wedge s)$ can be assessed from Table 4, so that we can compute the remaining values of γ_{HC}.

    The resulting belief network for Wilson's disease is shown in Figure 10. The bar charts in this figure show the prior probability distributions for the variables. We have added vertices for the inheritance of Wilson's disease (Wilson's disease genotype of mother, father and sibling). Wilsonians symptoms have been futher specified as Kayser-Fleischer rings, cirrhosis, psychiatric, neurological and renal disease; the corresponding assessment functions have been obtained from [34].

3.4.3 Preliminary evaluation of the belief network

Evaluation of the belief network for Wilson's disease has been carried out by using data from patients with Wilson's disease, and by comparing the probabilities concerning symptoms and signs, when certain evidence was entered into the graph, with statistical information concerning Wilson's disease in the literature. For purpose of the evaluation, we have tried to simulate the situation of an internal medicine clinic. In that case, the prevalence of Wilson's disease is much higher than in the general population. We assumed that Wilson's disease will be encountered in about one in 200 patients. As an example, using data of a patient that has been used in the evaluation of the HEPAR system, resulted in a probability of Wilson's disease of 0.774; the HEPAR system was also capable of concluding the patient to suffer from Wilson's disease with a certainty factor equal to 0.69. Similar results were obtain for five patients reported in the medical literature [6].
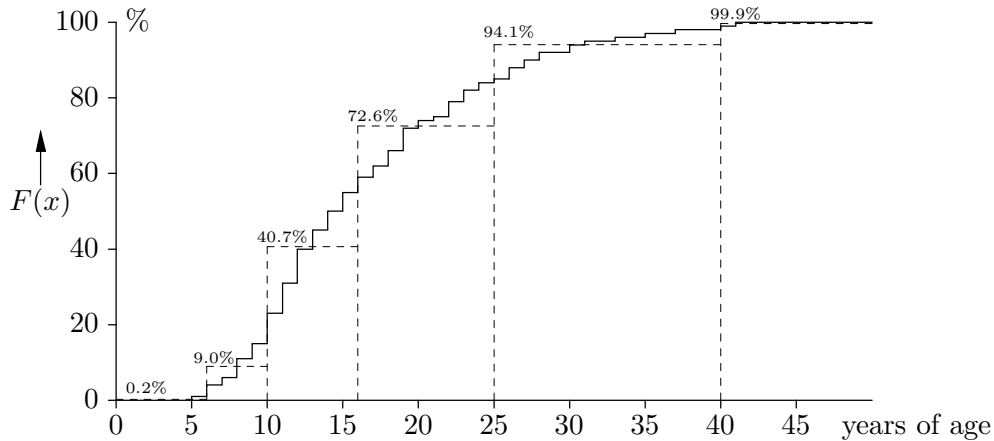
Figure 9: Cumulative percentage of 121 patients, $F(x)$, with onset of overt symptoms of Wilson's disease before age x. Source: [29].

Since eventually all untreated homozygotes for the Wilson's disease gene will become symptomatic Wilsonian patients, the dashed lines and corresponding percentages give us the probabilities of symptomatic Wilson's disease in homozygous subjects from the distinct age groups.

The belief network was next evaluation in a test in which, after entering certain combinations of evidence mentioned in the literature, computed probabilities were compared with qualitative and quantitative statements concerning Wilson's disease that have appeared in literature. Because most of these statements were qualitative in nature, it was not possible to obtain precise evaluation results in this study. We had the impression that the probabilities computed using the belief network, fitted the descriptions in the literature quite closely.

4   Discussion

Diagnostic problem-solving in medicine can be considered from many different perspectives. The same is true for computer-based decision-support systems, such as medical expert systems. On the one hand, assistance in medical diagnosis can be offered by expert systems employing heuristic reasoning models, based on the experience of clinicians in the management of their patients. In such systems, decisions as to which diagnostic procedures ought be untertaken for a given patient, are implicitly embodied in the knowledge encoded and is applied by means of some diagnostic strategy [14]. On the other hand, decision theory, being a combination of probability theory and utility theory, makes it possible to explicitly reason about such decisions, allowing the analysis of cost/benefit considerations of any action undertaken by the clinician, an application not readily accessible by the heuristic models. However, the development of such decision-theoretical systems for areas of the size that can be handled by current expert system technology, is a major undertaking. It therefore seems attractive to consider starting the development of such decision-theoretical models by taking existing heuristic expert systems as a point of departure. Belief networks may be taken as a starting point for the development of decision-theoretical models.

In the present article, we have described an experiment on converting a rule-based expert system into a belief network, looking at the notion of belief network primarily as a means for representing causal medical knowledge. Rule-based systems for medical diagnosis, however,
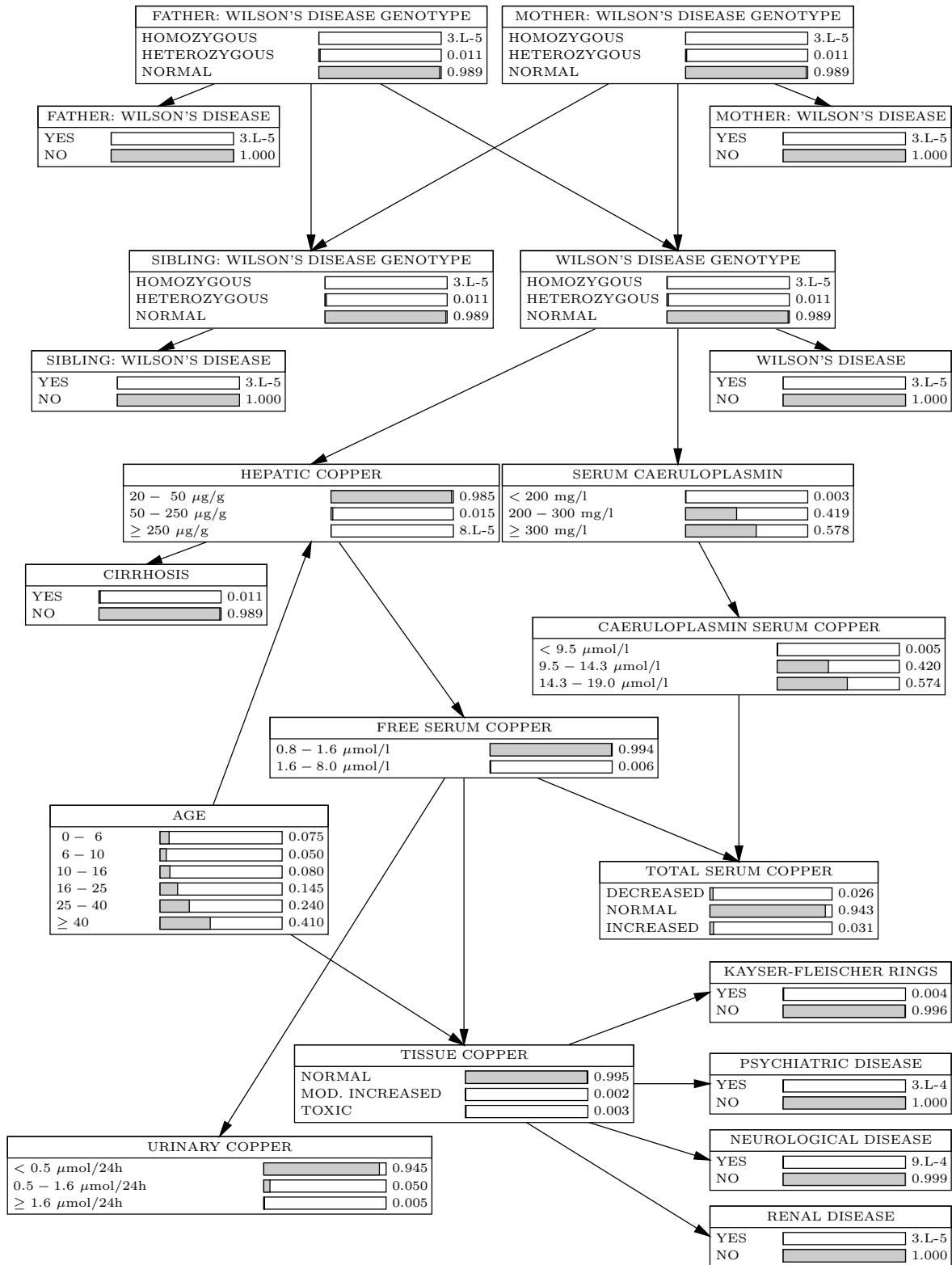
Figure 10: Belief network for Wilson's disease.

mainly represent heuristic knowledge; they lack knowledge about causal directions in relationships between variables. As a consequence, it is generally not possible to design a mapping that permits the automatic conversion of a rule-based expert system into a causal graph. Moreover, our study shows that for estimating a probability distribution for a belief network, the measures of uncertainty associated with the rules in the corresponding rule-based expert system are of little value. Due to these differences in the type of knowledge represented and in the formalism used to represent uncertainties, much knowledge required in a belief network cannot be derived from a rule-based system. Hence, converting a rule-based system into a belief network requires (re)consulting medical literature and renewed knowledge elicitation from experts. However, our experiment shows that the knowledge represented in a rule-based system may be well suited as a *foundation* for building a belief network upon. Major variables as well as their outcome spaces can rather easily be derived from it. Moreover, although the relationships expressed in the rules do not represent causal direction, they still can serve as an aid in searching for causal relations.

The main bottleneck in the conversion of a rule-based system into a belief network is the phase in which assessment functions are to be estimated. In our case, it turned out to be mandatory to consult other sources than the original rule base, in particular the medical literature. Unfortunately, the medical literature usually contains little suitable probabilistic information and uncertainty is often described in qualitative terms. When probabilistic information is provided, the probability distribution is often incomplete or conditioned on the 'wrong' variables. Furthermore, probability assessment is essentially population specific; the patients visiting a specialized medical centre form a population different from the normal population, which is likely to affect especially unconditional probabilities. Variables such as age and disease prevalence often are associated with unconditional probabilities, whereas these very probabilities are usually mentioned in literature for normal population. In for instance [28, 9, 10, 23], similar problems have been recognized in transferring diagnostic systems to other populations.

The problem of probability assessment can partly be solved by keeping the causal graph as simple as possible. A great deal of thought should therefore be paid to the construction of the causal graph, in which the availability of probabilistic knowledge should be taken into account. Using subjective probabilities may also be a solution. In a study concerning the assessment of simple conditional probabilities (viz. of certain findings given certain diseases), Spiegelhalter et al. showed that reliable probability assessments can be obtained from experts [26]. Moreover, they propose a practical procedure for improving such subjective probabilities by combining them with observed data. However, since we have tried to construct a belief network that provides a clear representation of the causal, pathophysiological knowledge involved, subjective probability assessment becomes much more complicated. In order to obtain a reliable belief network in such cases, the availability of probabilities computed from a large series of real-life patients therefore seems inevitable.

References

[1] B.G. Buchanan and E.H. Shortliffe (Eds.), *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison-Wesley Publishing Company, Reading, Massachusetts, 1984).

[2] F. Burbank, A computer diagnostic system for the diagnosis of prolonged undifferentiating liver disease, *American Journal of Medicine* 46 (1969) 401–415.

[3] L.C. van der Gaag, Probability-based models for plausible reasoning, Ph.D. Thesis, University of Amsterdam, Amsterdam, 1990.

[4] A. Haubek, J.H. Pedersen, F. Burcharth, J. Gammelgaard, S. Hancke and L. Willumsen, Dynamic sonography in the evaluation of jaundice, *American Journal of Radiology* 136 (1981) 1071–1074.

[5] D.E. Heckerman, Probabilistic interpretations for MYCIN's certainty factors, in: L.N. Kanal and J.F. Lemmer (Eds.), *Uncertainty in Artificial Intelligence 3* (North-Holland, Amsterdam, 1986) 167–196.

[6] T.U. Hoogenraad, C.J.C. van den Hamer, J. van Hattum, Effective treatment of Wilson's disease with oral zinc sulphate: two case reports, *British Medical Journal* 289 (1984) 273–276.

[7] J.H. Kim and J. Pearl, A computational model for causal and diagnostic reasoning in inference systems, in: *Proceedings of the 8th International Joint Conference on Artificial Intelligence* (Karlsruhe, West Germany, 1983) 190–193.

[8] S.L. Lauritzen and D.J. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society (Series B)* 50 (1988) 157–224.

[9] G. Lindberg, Studies on diagnostic decision making in jaundice, M.D. Thesis, Karolinska Institute, Stockholm, 1982.

[10] G. Lindberg, C. Thomsen, A. Malchow-Møller, P. Matzen and J. Hilden, Differential diagnosis of jaundice: Applicability of the Copenhagen Pocket Chart proved in Stockholm patients, *Liver* 7 (1987) 43–49.

[11] P.J.F. Lucas, R.W. Segaar and A.R. Janssens, HEPAR, an expert system for the diagnosis of disorders of the liver and biliary tract, *Liver* 9 (1989) 266–275.

[12] P.J.F. Lucas and L.C. van der Gaag, *Principles of Expert Systems* (Addison-Wesley Publishing Company, Wokingham, England, 1991).

[13] P.J.F. Lucas and A.R. Janssens, Second validation of the HEPAR system, an expert system for the diagnosis of disorders of the liver and biliary tract, *Liver* 11 (1991) 340–364.

[14] P.J.F. Lucas and A.R. Janssens, Development and validation of HEPAR, an expert system for the diagnosis of disorders of the liver and biliary tract, *Journal of Medical Informatics* 16 (1991) 259–270.

[15] W.B. Martin, P.C. Apostolakos and H. Roazen, Clinical versus actuarial prediction in the differential diagnosis of jaundice, *American Journal of Medical Sciences* 240 (1960) 571–578.

[16] P. Matzen, A. Malchow-Møller, J. Hilden, C. Thomsen, L.B. Svendsen, J. Gammelgaard, E. Juhl and the Copenhagen Computer Icterus Group, Differential diagnosis of jaundice: a pocket diagnostic chart, *Liver* 4 (1984) 360–371.

[17] B. Middleton, M.A. Shwe, D.E. Heckerman, M. Henrion, E.J. Horvitz, H.P. Lehmann and G.F. Cooper, Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base, II. Evaluation of diagnostic performance, *Methods of Information in Medicine* 30 (1991) 256–267.

[18] R.E. Neapolitan, *Probabilistic Reasoning in Expert Systems — Theory and Algorithms* (John Wiley & Sons, New York, 1990).

[19] K.G. Olesen, U. Kjaerulff, F. Jensen, F.V. Jensen, B. Falck, S. Andreassen and S.K. Andersen, A MUNIN network for the median nerve — A case study on loops, *Applied Artificial Intelligence* 3 (1989) 301–319.

[20] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann Publishers, San Mateo, California, 1988).

[21] R.G. Petersdorf, R.D. Adams, E. Braunwald, K.J. Isselbacher, J.B. Martin and J.D. Wilson (Eds.), *Harrison's Principles of Internal Medicine*, tenth edition (McGraw-Hill International Book Company, Auckland, 1983).

[22] S. Schenker, J. Balint, L. Schiff, Differential diagnosis of jaundice: Report of a prospective study of 61 proved cases, *American Journal of Digestive Diseases* 7 (1962) 449–463.

[23] R.W. Segaar, J.H.P. Wilson, J.D.F. Habbema, A. Malchow-Møller, J. Hilden and P.J. van der Maas, Transferring a diagnostic decision aid for jaundice, *Netherlands Journal of Medicine* 33 (1988) 5–15.

[24] E.H. Shortliffe, *Computer-Based Medical Consultations: MYCIN* (Elsevier, New York, 1976).

[25] M.A. Shwe, B. Middleton, D.E. Heckerman, M. Henrion, E.J. Horvitz, H.P. Lehmann and G.F. Cooper, Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base, I. The probabilistic model and inference algorithms, *Methods of Information in Medicine* 30 (1991) 241–255.

[26] D.J. Spiegelhalter, R.C.G. Franklin and K. Bull, Assessment, criticism and improvement of imprecise subjective probabilities for a medical expert system, in: M. Henrion, R.D. Shachter, L.N. Kanal and J.F. Lemmer (Eds.), *Uncertainty in Artificial Intelligence 5* (North-Holland, Amsterdam, 1990) 285–294.

[27] S. Srinivas and J. Breese, IDEAL: Influence diagram evaluation and analysis in Lisp; documentation and users guide, Technical Memorandum No. 23, Rockwell International Science Center, Palo Alto Laboratory, Palo Alto, 1990.

[28] R.B. Stern, R.P. Knill-Jones and R. Williams, Use of computer program for diagnosing jaundice in district hospitals and specialized liver unit, *British Medical Journal* 2 (1975) 659–662.

[29] I. Sternlieb and I.H. Scheinberg, Prevention of Wilson's disease in asymptomatic patients, *New England Journal of Medicine* 278 (1968) 352–359.

[30] I. Sternlieb and I.H. Scheinberg, Wilson's disease', in: [34] 949–961.

[31] A. Theodossi, An assessment of the value of diagnostic techniques in hepatobiliary disease, M.D. Thesis, University of London, London, 1986.

[32] A. Theodossi, R.P. Knill-Jones, A. Skene, G. Lindberg, B. Bjerregaard, J. Holst-Christensen and R. Williams, Inter-observer variation of symptoms and signs in jaundice, *Liver* 1 (1981) 21–32.

[33] A. Theodossi, D. Spiegelhalter, B. Portmann, A.L.W.F. Eddleston and R. Williams, The value of clinical, biochemical, ultrasound and liver biopsy data in assessing patients with liver disease, *Liver* 3 (1983) 315–326.

[34] R. Wright, G.H. Millward-Sadler, K.G.M.M. Alberti and S. Karran (Eds.), *Liver and Biliary Disease*, second edition (W.B. Saunders Company, London, 1985).