# Computer-based Decision Support in the Management of Primary Gastric non-Hodgkin Lymphoma*

Peter Lucas
Department of Computer Science, Utrecht University
Utrecht, The Netherlands

Henk Boot & Babs Taal
Department of Gastroenterology & Medical Oncology,
Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital
Amsterdam, The Netherlands

## Abstract

Primary non-Hodgkin lymphoma of the stomach is a rare disorder for which clinical management has not yet been settled completely. Faced with the many uncertainties associated with the selection of a treatment for a patient with this disorder, it is difficult to determine the treatment that is optimal for the patient, as well as the prognosis to be expected. The development of a decision-theoretic model of non-Hodgkin lymphoma of the stomach is described. The model aims to assist the clinician in exploring various clinical questions, among others questions concerning prognosis and optimal treatment. Central to the model is a probabilistic network that offers an explicit representation of the uncertainties underlying the decision-making process. The model has been incorporated into a computer-based system, that can be used as a decision-support system. Preliminary evaluation results indicate that the performance of the model in its present form matches the performance of experienced clinicians.

*Keywords*: medical decision making, medical expert systems, probabilistic networks, influence diagrams, gastric lymphoma.

## 1 Introduction

Primary non-Hodgkin lymphoma (NHL) of the stomach is a relatively rare malignant disorder, accounting for about 5% of gastric tumours. Until recently, the aetiology of gastric NHL was unknown; it is now generally believed that the main factor in the pathogenesis of this disease is a chronic infection with the bacterium *Helicobacter pylori* [1, 2]. *H. pylori* has been shown to be an important causative factor in the development of mucosa-associated lymphoid tissue (MALT) in the stomach [3], which, by largely unknown mechanisms, may undergo malignant change [1, 4].

Various treatment modalities are in use for this disease, varying from chemotherapy, radiotherapy, surgery and, more recently, *H. pylori* eradication [5], i.e. elimination of the bacterium

---

from the stomach by means of antibiotic drugs, to particular combinations of these therapies. Due to the rare nature of the condition, reports on clinical experience with specific therapeutic regimes usually concern small numbers of patients. When the numbers of patients are larger, studies have been carried out over a long period of time, during which variation in diagnostic workup and treatment occurs (cf. [6]). Furthermore, most studies are retrospective in nature, without a predefined treatment regime, thus precluding a comparison between various therapeutic strategies. As a result, the prognostic impacts of particular patient features are still far from clear, even when only considering past experience at a single institution.

Several researchers have recognised the need for decision support in the clinical management of patients with NHL [7], NHL of the stomach included [8]. Therapy selection for gastric NHL is a complicated process, because only part of the patient findings necessary for therapy selection may be known at a particular stage of the disease, and knowledge of adverse reactions to particular treatments in patient groups may influence treatment selection significantly. Moreover, knowledge of outcome-specific clinical profiles, such as the typical clinical picture of a patient with microscopic evidence of tumour cell elimination after treatment, may help in clinical management and in the choice of appropriate therapy. Much work has been done in the identification of both pretreatment and treatment prognostic factors that help identifying patients at risk [9]. For example, the histological classification of NHL of the stomach in low-grade versus high-grade malignancy using the MALT concept has been shown to be such a prognostic factor [9, 10]. Recently, a number of centres have developed a prognostic model for agressive nodal NHL, called the International Prognostic Index (IPI), which includes five clinical features for predicting prognosis in patients [11]. This prognostic model has also been applied to NHL patients in general [12]. However, current prognostic models like the IPI are very restrictive. They can only be employed to predict survival; they have limited value in the selection of effective therapy, and provide no support in dealing with the other problems mentioned above.

In this paper, the technique of decision-theoretic networks is investigated with respect to its capability of generating patient-specific risk profiles, to predict prognosis, as well as to select optimal treatment for individual patients. The main goal of this study was to overcome the limitations of common prognostic models by means of a computer-based system that can be used by the clinician to answer many different clinical questions concerning gastric NHL. The development and possible clinical usage of this decision-theoretic model of gastric NHL is discussed in detail.

## 2    Methods

In this section, the basic techniques, and the development of a decision-theoretic model of primary gastric NHL are discussed.

### 2.1    Decision-theoretic networks

We briefly review the theory of decision-theoretic networks so far as is required for the reading of this paper. For a recent review of the subject, the reader is referred to [13, 14].

A distinction is made between two types of decision-theoretic network:

- probabilistic networks, and

- influence diagrams.
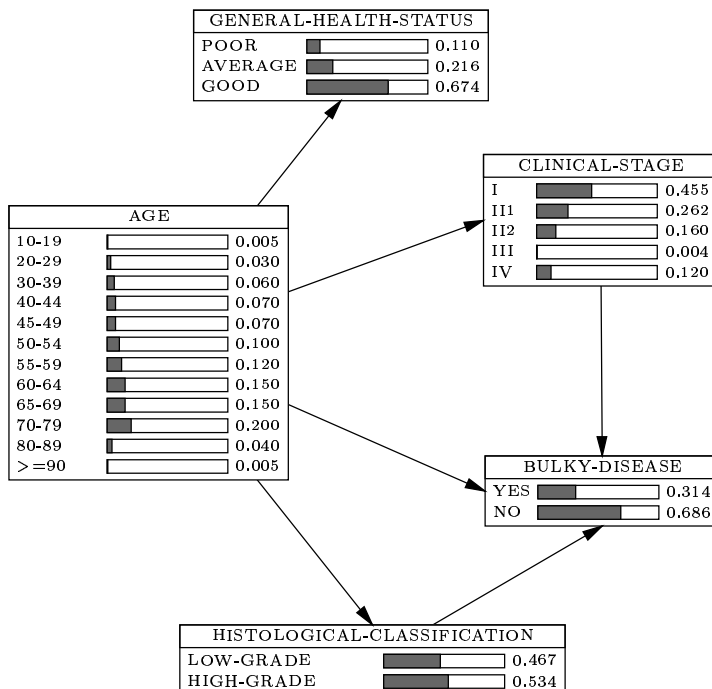
**Figure 1**: Simplified probabilistic network with gastric NHL pretreatment variables.

A *probabilistic network*, or *belief network*, is a directed acyclic graph $G = (V(G), A(G))$, consisting of a set of nodes $V(G) = \{V_1, \ldots, V_n\}$, called *probabilistic nodes*, representing discrete random variables, and a set of arcs $A(G) \subseteq V(G) \times V(G)$, representing causal relationships or correlations among random variables. Consider Figure 1, which shows a simplified version of a probabilistic network modelling some of the relevant variables in the management of gastric NHL. The presence of an arc between two nodes denotes the existence of a direct causal relationship or other influences; absence of an arc means that the variables do not influence each other directly. In Figure 1, the variable AGE is expressed to influence 'GENERAL-HEALTH-STATES', 'CLINICAL-STAGE', presence of 'BULKY-DISEASE', and 'HISTOLOGICAL-CLASSIFICATION'. Associated with a probabilistic network is a joint probability distribution Pr, defined in terms of conditional probability tables according to the structure of the graph. For example, for Figure 1, the conditional probability table

Pr(BULKY-DISEASE | CLINICAL-STAGE, AGE, HISTOLOGICAL-CLASSIFICATION)

has been assessed with repect to all possible values of the variables CLINICAL-STAGE, AGE, HISTOLOGICAL-CLASSIFICATION and BULKY-DISEASE. In general, the graph associated with a probabilistic network mirrors the stochastic (in)dependencies that are assumed to hold among variables in a domain. For a joint probability distribution defined in accordance with the structure of a probabilistic network, it, therefore, holds that:

$$\Pr(V_1, \ldots, V_n) = \prod_{i=1}^{n} \Pr(V_i \mid \pi(V_i))$$

where $V_i$ denotes a random variable associated with an identically named node, and $\pi(V_i)$ denotes the parents of that node. As a consequence, the amount of probabilistic information
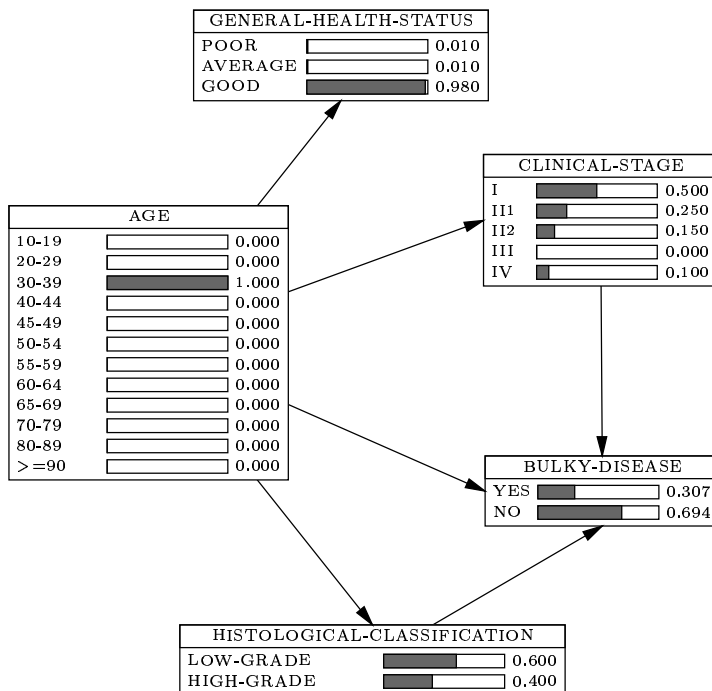
**Figure 2**: Updated probabilistic network after entering evidence concerning age.

that must be specified, exponential in the number of variables in general when ignoring the independencies represented in the graph, is greatly reduced. By means of special algorithms for probabilistic reasoning – well-known are the algorithms by Pearl [15] and by Lauritzen and Spiegelhalter [16] – a once constructed probabilistic network can be employed to enter and process data of a specific patient, i.e. specific values for certain variables, yielding, for example, prognostic information. Figure 2 shows the updated probabilistic network after entering evidence about a patient's age (30–39 years) in the network shown in Figure 1. Entering evidence in a network is also referred to as *instantiating* the network.

The probabilistic-network formalism provides only for probabilistic reasoning. For making therapeutic decisions, certain extensions to the probabilistic-network formalism are required, as offered, among others, by the influence-diagram formalism. Like a probabilistic network, an *influence diagram* is an acyclic directed graph $G = (V(G), A(G))$, except that in addition to probabilistic nodes $O(G) \subseteq V(G)$, two additional node types are distinguished: *decision nodes* $D(G) \subseteq V(G)$ and a *value node* $U \in V(G)$, where $O(G) \cap D(G) \cap \{U\} = \varnothing$. A probabilistic node is denoted by means of a circle, a decision node is indicated by means of a box, and a value node is indicated by a diamond.

As holds for probabilistic networks, arcs between probabilistic nodes denote direct causal relationships or correlations. An example of such an uncertain influential relationship in Figure 3 is the arc between 'AGE' and 'CLINICAL-STAGE'. Incoming arcs to a decision node indicate information that is required for the purpose of making a decision. The order of decision nodes in an influence diagram expresses the order in which the decisions are made in the domain; here, first a decision must be reached concerning surgery, next about a particular chemotherapy-radiotherapy (CT&RT) schedule (possibly none in both cases). Usually, it is assumed that this order is complete, although, in principle, a partial order can be employed
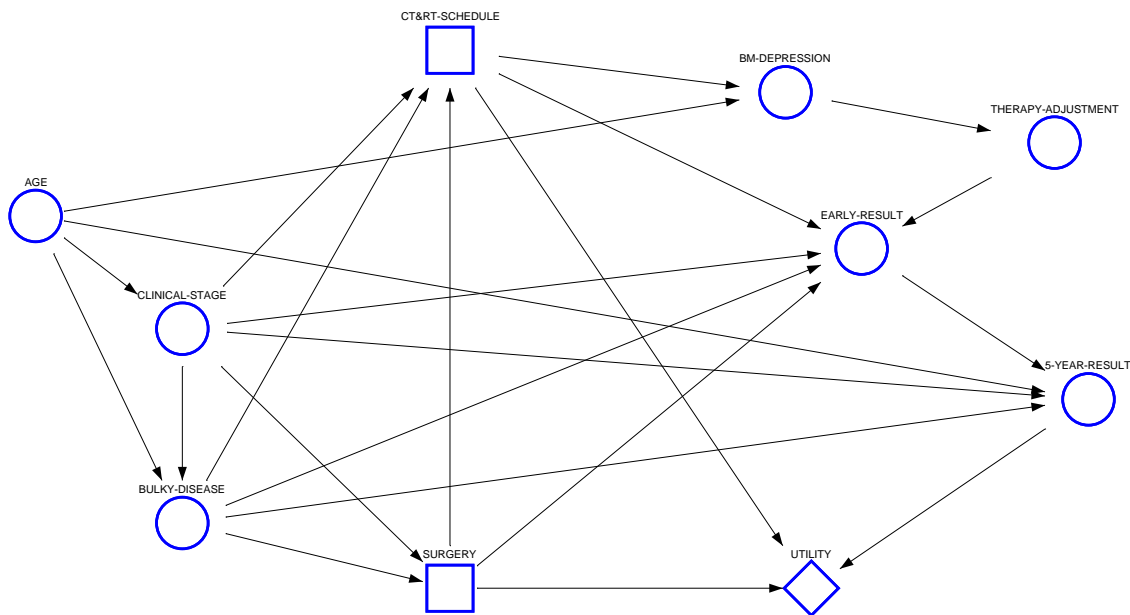
**Figure 3**: Simplified influence diagram for the treatment of gastric NHL.

to indicate alternative sequences of decisions. In order to select the most optimal sequence of decisions, assessments of these decisions with respect to other, partially uncertain, information must be available. These assessments, also called *utilities*, are represented by the value node. A utility in medicine is often a quantitative measure of the strength of a patient's preferences for particular therapeutic outcomes [17]. Incoming arcs of the value node indicate variables on which utilities must be defined, i.e. utilities can be expressed as a *utility function*:

$$u : C_{\pi(U)} \to \mathbb{R}$$

where $C_{\pi(U)}$ denotes the set of possible instantiations of parents of the value node $U$. In Figure 3 it is expressed that any decision concerning the best treatment for a patient will be based on the combined assessment of treatment alternatives (surgery in combination with a chemotherapy-radiotherapy schedule), and five-year survival (5-YEAR-RESULT). Since a value node represents the utilities of the variables corresponding to the nodes directly linked to the value node, it has no outgoing arcs.

As for probabilistic networks, the graph representation only denotes the qualitative relationships between variables; it must be supplemented with a quantitative representation to obtain an influence diagram. Various algorithms to evaluate influence diagrams are available, e.g. the algorithms proposed by Shachter [18] and Cooper [19]; the result of evaluation is an optimal sequence of decisions, a *decision strategy*, obtained by computation of the maximum expected utility (MEU) of every decision $D \in D(G)$ given particular evidence $E$:

$$\text{MEU}(D \mid E) = \max_{d \in D} \sum_{c \in C_{\pi(U)}} u(c) \Pr(c \mid d, E)$$

where possibly $D \in \pi(U)$, and where the evidence $E$ includes previous decisions. The reader is referred to [18] for a more detailed discussion on influence diagrams.

| Pretreatment Variables | Treatment Variables | Posttreatment Variables |
|---|---|---|
| age (AG) bulky disease (BD) clinical stage (CS) clinical presentation (CP) general health status (GH) histological classification (HC) | chemotherapy (CR) radiotherapy (CR) surgery (SU) combination therapy (CR) | bone-marrow depression (BM) early result (ER) gastric hemorrhage (HE) gastric perforation (PE) immediate survival (IS) post-CT&RT survival (PC) post-surgical survival (PS) 5-year result (5Y) therapy adjustment (TA) |

Table 1: Selected variables (abbreviated names).

## 2.2 Development of the gastric NHL decision model

The design of a structure for a decision-theoretic network model is usually such as to incorporate clinical experience as well as to reflect results from the medical literature. In principle, it is possible to construct such models automatically from a database with patient data. In practice, this is seldom done, due to the prohibitive amount of data and the huge computational resources required. Several researchers, however, have demonstrated that it is feasible to use subjective information, in addition to information drawn from the medical literature, as a basis for the design of decision-theoretic models (e.g. [20], [21], and [22]). Decision-support systems that include extensive domain models, at least partially based on subjective information, are also called expert systems [23]. Decison-support systems incorporating decision-theoretic models are therefore also known as *decison-theoretic expert systems* [21, 24].

In this section, the successive steps in the design of the gastric NHL model are reviewed. In the construction of the model, it has been assumed that the model will only be used for patients in whom the diagnosis primary gastric NHL has been confirmed histologically.

### 2.2.1 Identification of relevant variables

The selection of variables for a decision-theoretic network is influenced by the adopted view on the medical domain concerned. Often, knowledge about (patho)physiological mechanisms is used to guide the process of gathering relevant variables (cf. [25, 26]). For the gastric NHL model, however, we have chosen to incorporate only those variables into the model that are widely used by clinicians. The relevance of most of these variables is supported by literature on prognostic factors in primary gastric NHL [6, 9, 10].

First, the information used in the clinical management of primary gastric NHL was subdivided in pretreatment information, i.e. information that is required for treatment selection, treatment information, i.e. the various treatment alternatives, and posttreatment information, i.e. side effects, and early and long-term treatment results for the disease. The selected variables are presented in Table 1. The chosen subdivision of the variable 'age' into twelve intervals was based on the literature. It is said that a patient has 'bulky disease' if tumour size exceeds 10 cm in maximal diameter as observed endoscopically, or if there is invasive growth into surrounding tissues or organs. The variable 'clinical stage' is according to the Ann Arbor classification with Musshoff's modification for NHL, with as possible values I, $II_1$, $II_2$, III, and

IV [27]. This variable expresses severity of the disease, where the prognosis for patients in stage I is usually favourable and for patients in stage IV is rather grim. The variable 'clinical presentation' summarises the presence of particular gastric complications due to NHL at the time of presentation. Possible values are: 'hemorrhage', 'perforation', 'obstruction', and a variety of non-acute symptoms and signs, such as dyspepsia and weight loss, referred to by the value 'none'. Hemorrhage must be acute and be of sufficient significance to warrant blood transfusion. The variable 'general health status' represents the general health condition of the patient; possible values are: 'good' (WHO index 0 and 1), 'average' (WHO index 2) and 'poor' (WHO index 3 and 4). Histological classification is assumed to be based on the recent MALT concept of gastrointestinal NHL, with subdivision into low-grade and high-grade malignancy [28].

The variables 'chemotherapy', 'radiotherapy', and 'combination therapy' were combined into one variable with name 'CT&RT-SCHEDULE' (abbreviated to CR) with possible values: chemotherapy (CT), radiotherapy (RT), chemotherapy followed by iceberg radiotherapy (CT-next-RT), and neither chemotherapy nor radiotherapy (none). Possible values for the variable 'surgery' are: 'curative', 'palliative' or 'none', where curative surgery means total or partial resection of the stomach with the complete removal of locoregional tumour mass.

The variable 'bone-marrow depression' denotes bone-marrow depression due to chemotherapy or radiotherapy, necessitating treatment adjustment. The variables 'gastric hemorrhage' and 'gastric perforation' represent the corresponding complications, solely due to therapeutic intervention. The variable 'early result' stands for the endoscopically verified result of the treatment, six to eight weeks after treatment. Possible outcomes of this variable are: complete remission, partial remission, no change or progressive disease according to standard WHO criteria. The variable '5-year result' represents the patient either or not surviving five years following treatment. The remaining variables in the table will be discussed in Section 2.2.2.

A number of clinical centres are presently investigating the effects of *H. pylori* eradication on gastric NHL. Recent clinical experience indicates that *low-grade* gastric NHL can completely regress after cure of *H. pylori* infection [5]. However, there is still insufficient clinical experience with combination therapies, and the period of follow-up of patients treated for *H. pylori* infection is too short to warrant inclusion of *H. pylori* treatment as a variable in the decision-theoretic model. As an indication of possible future extension, this decision variable was added to the final model, together with variables indicating the presence of *H. pylori* in the patient, and the result of treatment for *H. pylori*. This model is discussed further in the discussion.

### 2.2.2 Identification of relationships among variables

Next, the structure of the gastric NHL model was determined, using causality as the primary modelling concept, supplemented with knowledge of probabilistic dependency between variables. Information concerning the (in)dependencies among variables was derived from the literature and from clinical experience. For example, it is known that the clinical stage of gastric NHL depends on the patient's age; histological grading of obtained tumour tissue, on the other hand, does not influence the clinical stage directly. Directed graphs were used as visual aids in the representation of these relationships. During the design process, several nodes in the graph became very complicated, i.e. too many arcs entered these nodes. Since the number of probabilities that must be assessed is proportional to the product of the number of possi-
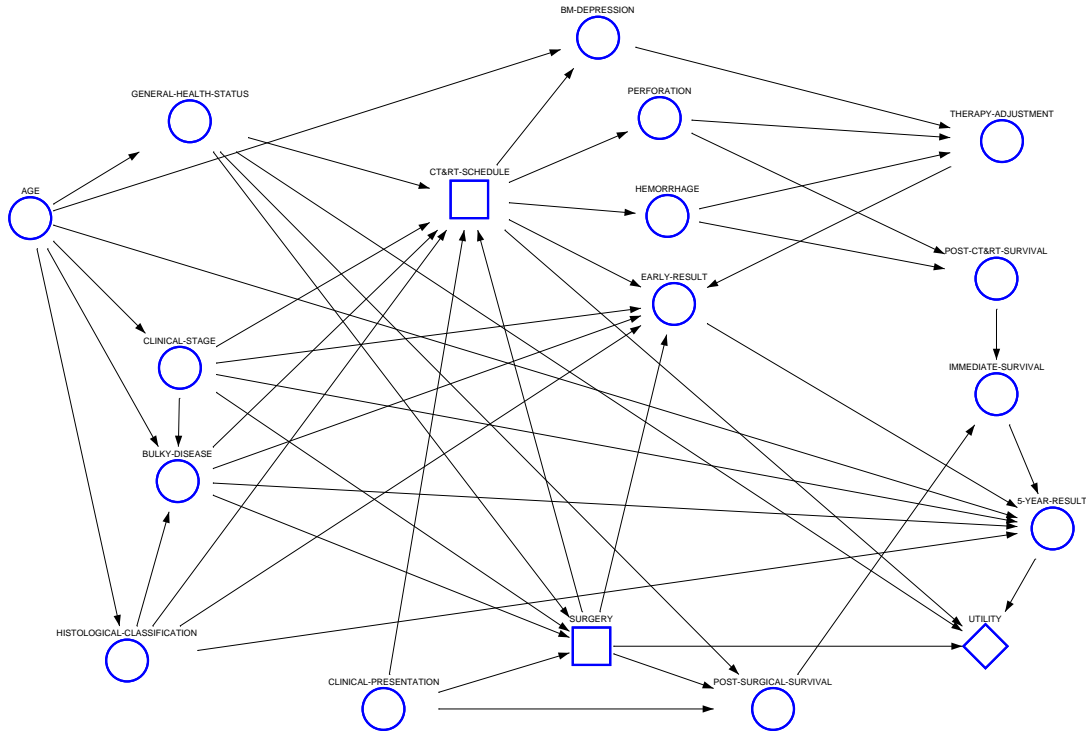
**Figure 4**: Complete influence diagram for the treatment of gastric NHL.

ble values for each of the variables involved, as determined by the number of incoming arcs of each node, probabilistic assessment can easily become unfeasible. To solve this potential problem, a small number of extra nodes was inserted into the graph in accordance with clinical practice, such as 'therapy adjustment', 'post-CT&RT survival', 'post-surgical survival' and 'immediate survival', summarising the effects of other variables. This technique is known as 'divorcing multiple parents' [29]. The variable 'post-CT&RT survival' represents not dying due to gastric hemorrhage or perforation, during the first eight weeks following treatment with chemotherapy or radiotherapy. The variable 'post-surgical survival' has a similar meaning with respect to surgery. The variable 'immediate survival' combines the possible effects of the two treatment decisions. These intermediate variables enabled a significant reduction in the number of incoming arcs in the graph, thus facilitating quantitative assessment.

The influence diagram resulting from this design process is shown in Figure 4. The value node present in the diagram was actually added in the last stage of the design process, to be discussed below.

### 2.2.3 Qualitative assessment of uncertainty and logical relationships

After the structure of the decision-theoretic model was established, it was necessary to gather probability tables and utility information, to obtain a probabilistic network and influence diagram, respectively. Although a database with data of a relatively small number of patients ($n = 137$) with gastric NHL was available from the Netherlands Cancer Institute to assess probabilistic information, it was decided to employ this database in the evaluation of the

| AGE | $\Pr(\text{GHS} \mid \text{AGE})$ |
|---|---|
| 10-69 | good > average > poor |
| 70-79 | average > good > poor |
| 80-89 | average > poor > good |
| $\geq 90$ | poor > average > good |

**Table 2**: Qualitative ordering of probabilities, used for assessment.

resulting network (see Section 2.2.6). The main reason for this was that therapeutic views have changed significantly during the past few years. For example, only a few years ago, radiotherapy was the treatment of choice for most patients with gastric NHL, where today, the role of chemotherapy has increased, and radiotherapy is frequently administered following chemotherapy. Hence, for many of the variables in the database it was known beforehand that they would be biased.

It is widely known that it is much easier to assess qualitative relationships between probabilities for particular values of variables, such as ordering relationships, than the probabilities themselves [30]. Such qualitative probabilistic information is often highly accurate. Therefore, the first step in the collection of probabilistic information has been the gathering of such qualitative information, using various techniques. Firstly, the logical relationships between variables were identified. Logical relationships yield probabilistic information with complete certainty, i.e. with probability 1 or 0; usually, logical information will not be a major source of inaccuracy in a model.

Secondly, we have collected information about qualitative probabilistic relationships between values of variables, represented by a number of tables. Table 2 shows the qualitative information that has been used in the assessment of the following conditional probability table:

$$\Pr(\text{GENERAL-HEALTH-STATUS} \mid \text{AGE})$$

For example, the following ordering: good > average > poor for patients with age between 10 to 69 years old, means:

$$\Pr(\text{GENERAL-HEALTH-STATUS} = \text{good} \mid \text{AGE} = \text{10-69}) >$$
$$\Pr(\text{GENERAL-HEALTH-STATUS} = \text{average} \mid \text{AGE} = \text{10-69}) >$$
$$\Pr(\text{GENERAL-HEALTH-STATUS} = \text{poor} \mid \text{AGE} = \text{10-69})$$

The tables with qualitative information proved extremely valuable in validating the numerical probabilities, which were assessed later on, because these numerical values could be checked agains this qualitative information.

### 2.2.4   Numerical probability assessment

After all qualitative information concerning the local probability distributions for the variables in the model was collected, the probabilities were assessed numerically. Various techniques were employed to render subjective, numerical assessment of the probabilities feasible.

Unconditional prior probabilities, such as for the variables AGE, and CLINICAL-PRESENT-ATION, were assessed in an iterative fashion. Probabilistic information so far as found in the literature on primary gastric NHL was collected. This probabilistic information was refined

further to reflect clinical experience at the Netherlands Cancer Institute. Note that these unconditional probabilities are likely to require reassessment for other clinical centres.

For the nodes in the network with incoming arcs, *conditional* probabilities had to be assessed. Unfortunately, these conditional probabilities could, almost without exception, not be found in the medical literature. To ease the assessment, a distinction was made between probabilistic nodes with a small number of incoming arcs (less than four) and the nodes EARLY-RESULT and 5-YEAR-RESULT, which each have six incoming arcs. The assessment of the conditional probabilities for the former non-complicated nodes, and the latter complicated ones, will be discussed separately.

The conditional probabilities for nodes with a small number of incoming arcs were assessed by distributing the probability mass among the various values of the variable corresponding to the node, where each time the values of the conditioning variables were kept fixed. The resulting probabilities were next checked against the tables with qualitative information previously collected. For example, the qualitative information concerning the variable CLINICAL-STAGE expressed that

$$\Pr(\text{CLINICAL-STAGE} = \text{I} \mid \text{AGE} = \text{10-29}) = \Pr(\text{CLINICAL-STAGE} = \text{II}_2 \mid \text{AGE} = \text{10-29})$$

which was used as a qualitative constraint for the assessment of four conditional probabilities.

The two more complicated nodes, EARLY-RESULT and 5-YEAR-RESULT, were assessed using a special additive model. First, the conditional probabilities were assessed by conditioning on four to five of the six variables for the value of the variable EARLY-RESULT, which would yield the largest probability in most cases. For the variable EARLY-RESULT, the value 'complete remission' (CR) was chosen as such. Next, the remaining probability mass was distributed among the remaining values for the variable EARLY-RESULT, again conditioning on the same variables. Finally, for both variables, the entire conditional probability table was computed by means of a computer program, that used the following additive model to compute probabilities $\Pr(A \mid C)$, based on partially specified information with respect to collections of variables with values in a given set $V$:

(1) for each $C \in V$: $\Pr(A \mid C) = \mathbb{P}(A \mid C)$;

(2) for each $C \in \wp(\Delta) \backslash V$, where $C$ contains a value for each variable in $\Delta$:

$$\Pr(A \mid C) = \sum_{\substack{C' \subset C,\, C' \in V \\ \forall C'' \in V,\, C'' \subset C\, :\, C'' \not\supset C'}} \Pr(A \mid C')$$

where $\wp(\Delta)$ denotes the power set of the set of all variables with their values, $\Delta$, and $\mathbb{P}$ denotes the partial specification of the probabilistic information. Note that according to this additive model, the most specific probabilistic information available is always used as a basis for calculating probabilities.

### 2.2.5 Refinement of the joint probability distribution

For the initial validation of the assessed joint probability distribution of the influence diagram, as discussed above, the probabilistic part of the influence diagram was extracted and examined. The two decision nodes, CT&RT-SCHEDULE and SURGERY, were converted to probabilistic nodes for which prior probability distributions were supplied, based in experience at

the Netherlands Cancer Institute. Using a probabilistic reasoning algorithm mentioned in Section 2.1, the prior marginal probabilities of this network were computed; the result is shown in Figure 5. A more compact screen version of the same probabilistic network, which will be used in the following, is shown in Figure 6. The accuracy of the assessments was investigated by comparing this prior probability distribution and the posterior distribution produced by the system by entering certain values for variables, with the clinical experience at the Netherlands Cancer Institute, and information from the literature. Data from a number of imaginary patients were processed by the system, and the resulting posterior, marginal probability distributions were again examined. A small number of adjustments in the probabilities were made, in all cases after reexamination of the literature.

### 2.2.6 Comparison with a clinical database

As mentioned in Section 2.2.3, a database with data from 137 patients with gastric NHL was available for further examination of the designed probabilistic network. Since the treatment policy for these patients has changed over the past few years, a major part of this database was considered unsuitable for constructing a reliable probabilistic network. The remaining part of the database contained data such as concerning age, which might have been used for probabilistic assessment. However, since it was our plan to integrate information from various literature sources with clinical experience in building the network, these data have not been used either. Although considered unsuitable as a basis for probabilistic assessment of the network, a comparison between a network with probabilistic information acquired from this database and the network described in the previous sections was expected to yield additional insight into the network's accuracy.

The comparison between the two networks was carried out in terms of the prior marginal probability distributions associated with individual nodes. It was expected that any significant difference between the two marginal probability distributions of identically named variables could be explained in terms of known biases in the database. For unbiased variables, approximate agreement between the associated marginal probability distributions was expected. Since the size of the database was relatively small, only part of the cases required for probabilistic assessment were included in the database. If none of the cases required for assessing the probabilities for a variable were available, a uniform (conditional) probability distribution was assumed. Experimental evidence from the literature suggests that a probabilistic network is relatively insensitive to probabilistic imprecision [31].

Let $p_i = \Pr(V = v_i)$ be the marginal probability for the value $v_i$ of variable $V$ in the network described above; similarly, let $p'_i = \Pr'(V = v_i)$ be the corresponding marginal probability for the network that was learnt from the database. Note that it holds that:

$$0 \leq \sum_{i=1}^{n} |p_i - p'_i| \leq \sum_{i=1}^{n} |p_i| + \sum_{i=1}^{n} |p'_i| = 2$$

where $|\cdot|$ denotes the absolute value, and $n$ is equal to the number of possible values of $V$ ($n \geq 2$). For the mean absolute difference in marginal probabilities $m$ for a variable $V$, it, therefore, holds that:

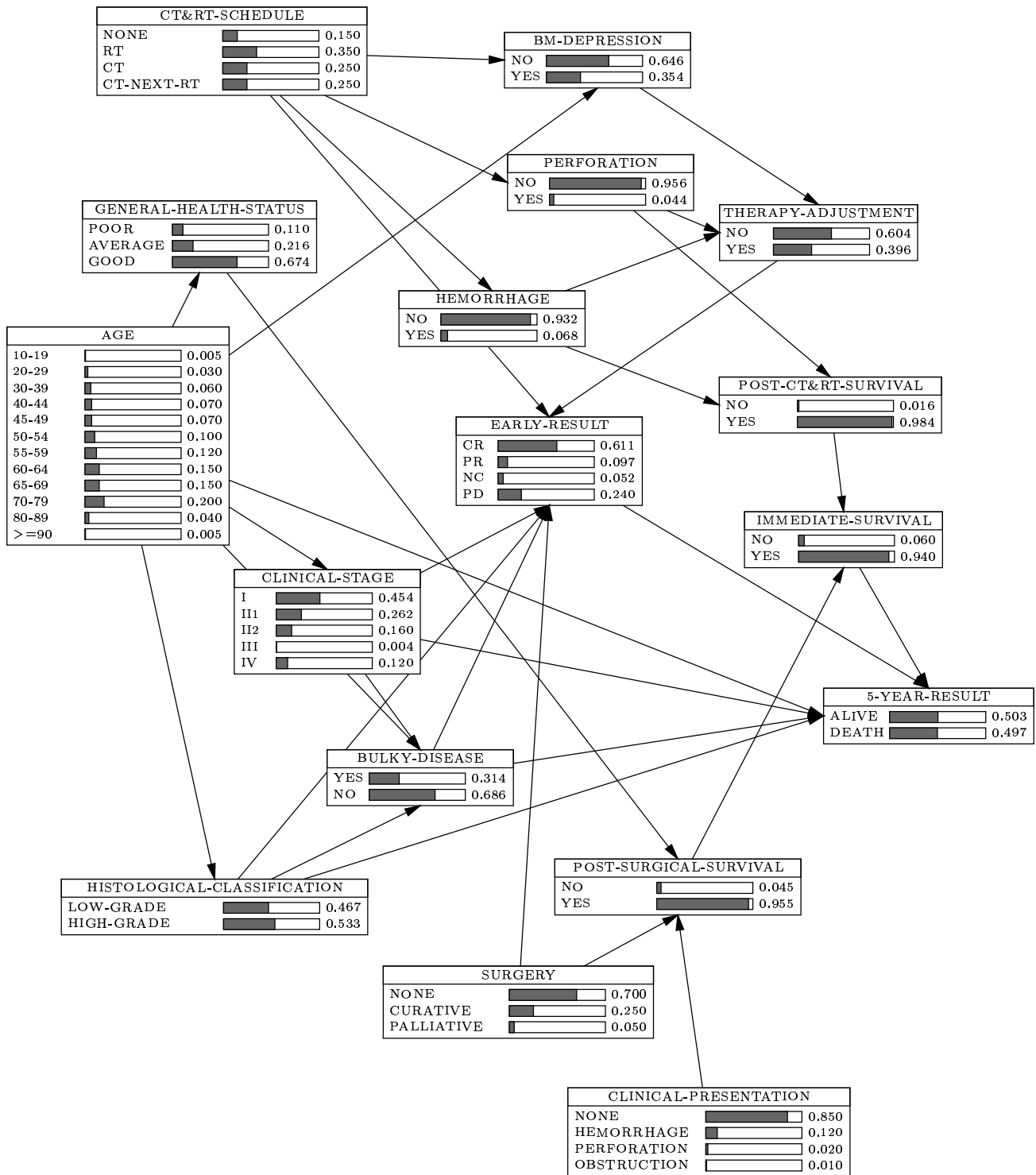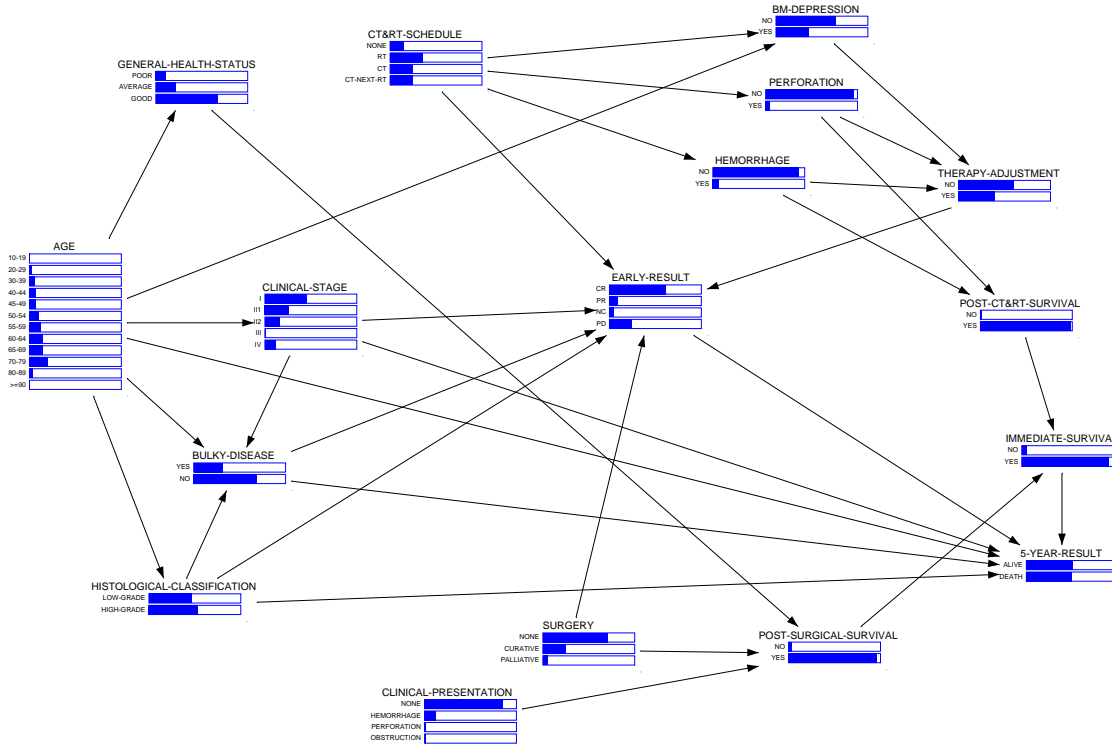$$0 \leq m = \frac{\sum_{i=1}^{n} |p_i - p'_i|}{n} \leq 1$$

**Figure 5**: Probabilistic network with prior probability distributions for NHL of the stomach.

**Figure 6**: Probabilistic network with prior probability distributions: screen version.

Hence, the largest absolute difference between the marginal distributions of identically named variables in the two networks is at most equal to 2, and the mean absolute difference is at most equal to 1.

In Figure 7, the total absolute difference between the marginal probability distributions of the two networks is shown for each variable in the networks. Figure 8 shows the mean absolute difference between the marginal probability distributions, where in this case the number of possible values of the variables is also taken into account.

The probability distributions of the two treatment variables CT&RT-SCHEDULE (CR) and SURGERY (SU) exhibited a major difference between the two networks, as expected beforehand. It is a consequence of the evolution in treatment policy. The deviation of the variable BM-DEPRESSION (BM) appeared to be unacceptably large as well. The probabilistic information for this variable was derived from data concerning the presence of dose reduction in a patient, an indirect and not completely accurate measure of bone-marrow depression. The large differences between the marginal probability distributions for the variable THERAPY-ADJUSTMENT (TA) is mainly caused by the lack of patient cases whom experienced all complications of treatment. The variables POST-CT&RT-SURVIVAL (PC), POST-SURGICAL-SURVIVAL (PS), and IMMEDIATE-SURVIVAL (IS) participate in logical relations. The small differences in the probability distributions for these variables are, therefore, entirely due to rounding errors in floating point calculations. There were no patients included in the database with a poor general health, which explains why the mean difference in marginal probability for the variable GENERAL-HEALTH-STATUS (GH) exceeds 0.1. The conditional probability of the
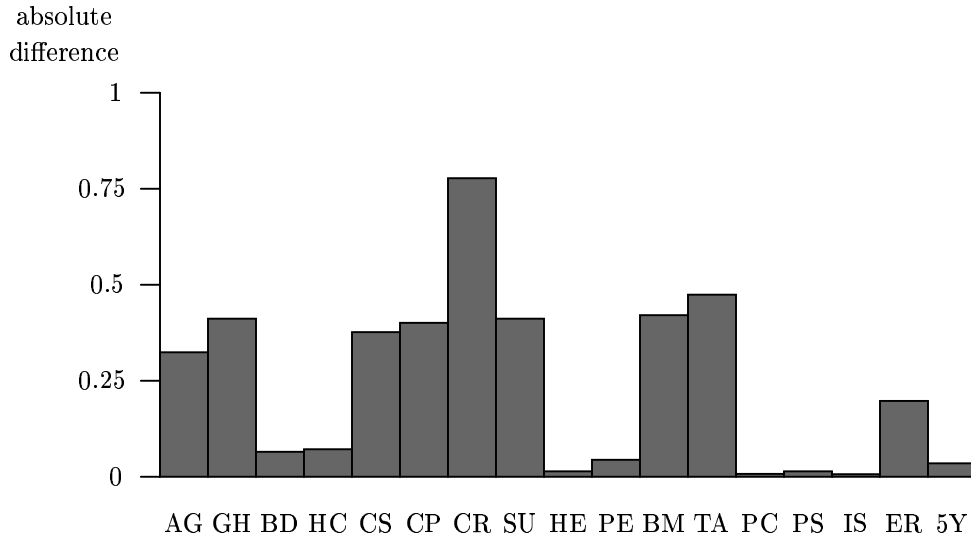
**Figure 7**: Total absolute difference between marginal probabilities of the learnt and estimated probabilistic networks. The meanings of the abbreviated variable names are given in Table 1.

value OBSTRUCTION for the variable CLINICAL-PRESENTATION (CP) was estimated from information concerning the presence or absence of vomiting in patients, which yields a significant overestimation. The differences in the marginal probability distributions of the remaining eight variables seem acceptable, however, given the distinct origins of the two probabilistic networks. In general, this comparative evaluation confirms that the probabilistic information incorporated in the network is fairly accurate.

### 2.2.7 Utility assessment

In order to obtain a complete influence diagram, a value node was added to the graph, and the assessment of probabilistic information discussed above was succeeded by the assessment of utility information for the value node. The following variables were marked out as being of major importance in deciding about optimal treatment for a patient with gastric NHL:

- general health status;

- CT&RT schedule;

- surgery;

- 5-year result.

The influence diagram, which resulted by choosing these variables as a basis for utility assessment, is shown in Figure 4. Utility assessment was carried out from the patient's perspective.

Two different methods were employed for utility assessment, thus facilitating a critical appraisal of the effects of alternative utility functions [17]:

- the direct scaling method, and

- the reference gamble method.

14
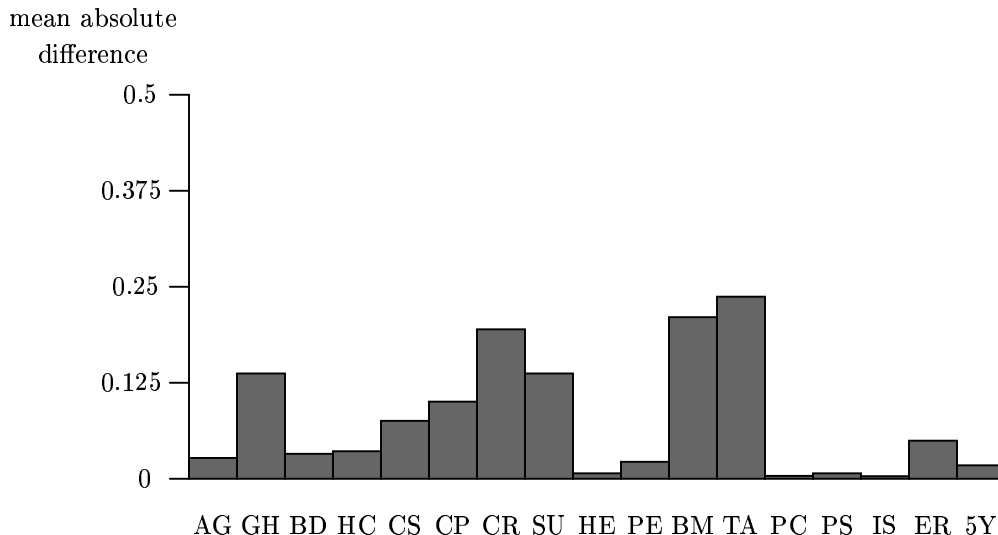
**Figure 8**: Mean absolute difference between marginal probabilities of the learnt and estimated probabilistic networks, taking the number of possible values of a variable into account. The meanings of the abbreviated variable names are given in Table 1.

In the *direct scaling method*, the possible clinical outcomes are assessed using a linear numerical scale, promoting the assessment of these outcomes relative to each other. This methods was applied twice. First, the position of the utilities of particular posttreatment outcomes, as described by the four variables mentioned above, was assessed on a numerical scale ranging from 0 (worst) to 1000 (best). The numerical values were next sorted; the resulting utility function is depicted in Figure 9. Note that many clinical outcomes were considered equivalent, as is reflected by their identical value according to the utility function. The sorted list of utilities with associated clinical outcomes was closely examined. Furthermore, the optimal treatment strategies generated by the system for a number of typical clinical patient cases were reviewed. In this way, it was discovered that the contribution of five-year survival in the numerical assessment had been underestimated. Part of the utilities were reassessed, again using the direct scaling method; the resulting influence diagram was again evaluated by examining its generated therapeutic advice for a number of typical patients. The resulting second utility function is also shown in Figure 9. Both utility functions display small incremental increases in their function values, showing a more or less even spread of utilities over intervals of the scale, indicating that further refinement may be required.

With the *standard reference gamble method*, every possible outcome is assessed in probabilistic terms by determining the risk neutral equivalent of two uncertain outcomes, assuming that the outcome being assessed occurs with absolute certainty. Hence, utility assessment is not carried out in an incremental fashion. A utility function was designed, using the standard reference gamble method; this third utility function is also depicted in Figure 9. Notice that this last utility function reveals a bit more variation in its function values as compared to the other two utility functions.

Results obtained by the reference gamble method are generally considered more accurate than those resulting from the direct scaling method; the variation in the values of the last utility function is due to the more subtle expression of differences among the various outcomes.
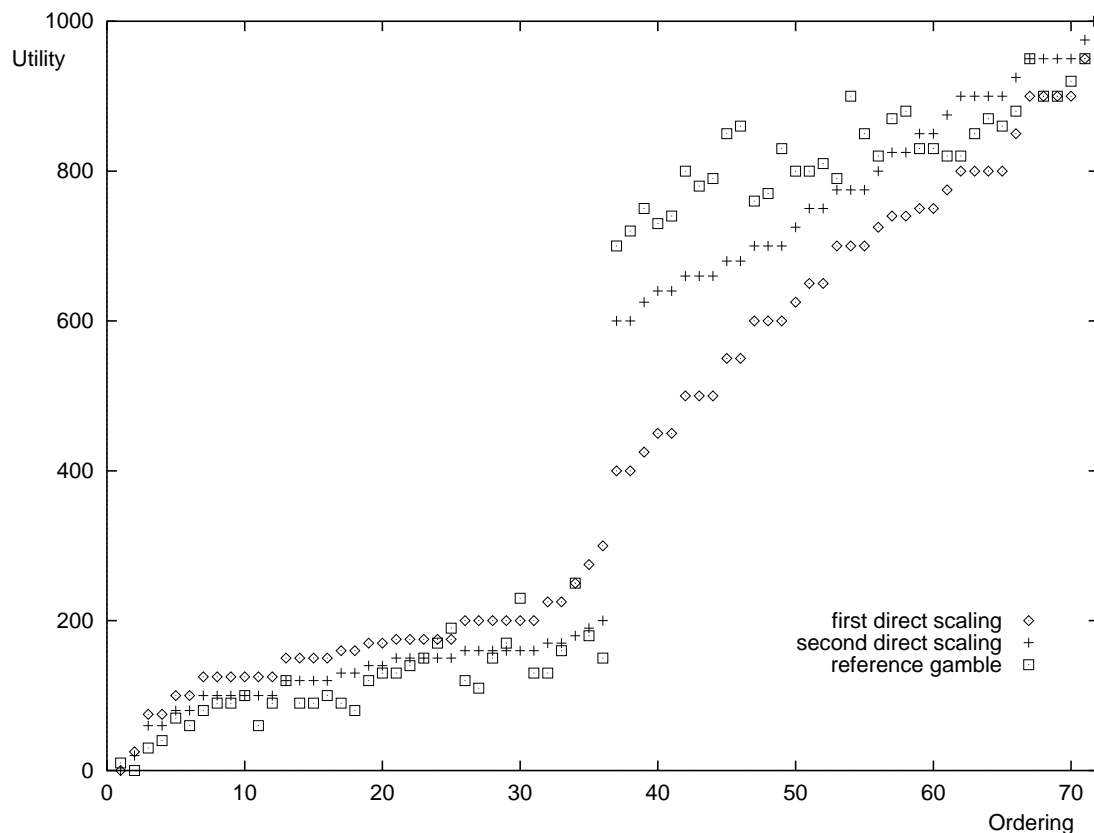
**Figure 9**: Graphical plot of three utility functions.

The successive application of these two assessment methods promoted the convergence of the utility-assessment process to a clinically faithful utility function. Finally, a fourth utility function was defined, where only the five-year survival following treatment was adopted as a basis for assessment. As the first utility function, this utility model was later discarded because its generated therapeutic advice was systematically too drastic. Two utility functions remained; of these, the function obtained by the reference gamble method was thought to reflect patient's preferences best.

The influence diagram and associated probabilistic network were eventually implemented using an improved version of the *Ideal* probabilistic expert system shell [32]. Hence, the decision-theoretic model of gastric NHL actually consists of two complementary, and closely related, representations. The Ideal system offers a collection of reasoning algorithms and a graphical user interface to display results of probabilistic reasoning.

# 3    Results

In this section, the clinical usefulness of the decision-theoretic model of gastric NHL is examened, and the results of a preliminary evaluation of the model are discussed.
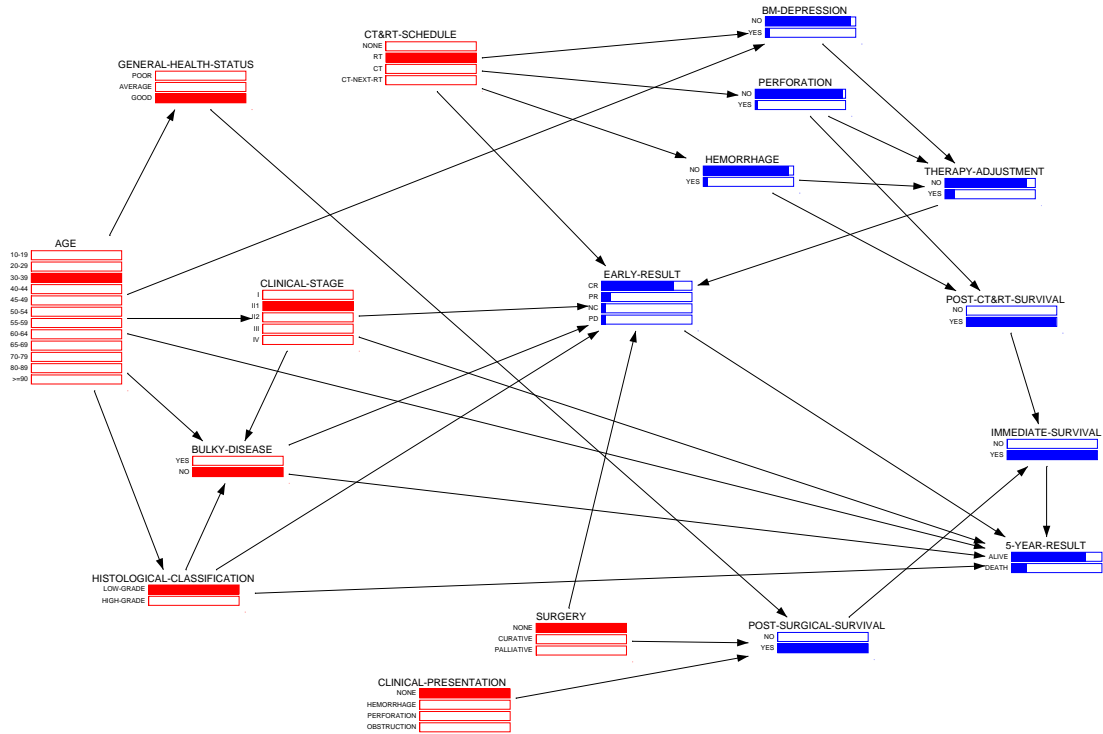
**Figure 10**: Prognostic prediction, assuming that the patient has been treated with radiotherapy.

## 3.1 Clinical usefulness

The decision-theoretic model of gastric NHL (influence diagram and associated probabilistic network) can be employed to answer a number of different clinical questions. Here, some possible clinical uses of the model will be discussed.

### 3.1.1 Prognostic prediction

The probabilistic network for gastric NHL can be applied to the data of a specific patient, yielding prognostic information about the patient, given particular patient features and therapeutic decisions. For example, in Figure 10 all pretreatment variables have been entered. If we assume that this patient will be treated with radiotherapy only, the probabilistic network can be employed to indicate the future likelihoods of the various possible consequences (side effects, remission, death, etc.) of the chosen therapy, given the clinical characteristics of the patient. Others clinical questions can be explored along similar lines. Note that it is not necessary that all patient data are available. It is even possible to leave the therapy choice open, and to examine consequences of the 'average' therapy.

### 3.1.2 Selection of optimal treatment

By reverting to the original influence diagram in Figure 4, it is possible, using Shachter's reasoning algorithm described in Section 2.1, to determine which treatment is optimal for

the patient introduced above. The treatment advice generated by the system depends on the utility function reflecting patient preferences. Taking the two utility functions discussed in Section 2.2.7, where the former was obtained by the direct scaling method and the latter by the standard reference gamble method, yielded the same optimal treatment: radiotherapy (20 Gy on the stomach followed by 20 Gy on the abdomen) with expected utility of 833 and 825 (on a scale from 0 to 1000), respectively. The small difference in expected utility can be explained by the fact that for the utility function assessed by the standard reference gamble, inconvenience to the patient from radiotherapy was rated higher than for the utility function obtained by the direct scaling method.

Adoption of the utility function in which only the expected five-year survival following treatment was taken into account yielded a more drastic treatment: chemotherapy followed by radiotherapy with expected utility equal to 831. A utility equal to 1000 is associated with a survival of five years or longer, and a utility of 0 with surviving less than five years, which is interpreted as treatment failure.

Note that in this way alternative treatment strategies associated with various utility function can be examined. These alternatives are a result of differences in patient preferences. Using the corresponding probabilistic network shown in Figure 5, the predicted prognosis for a particular treatment for a specific patient can be examined, and used as a further guide in clinical decision making.
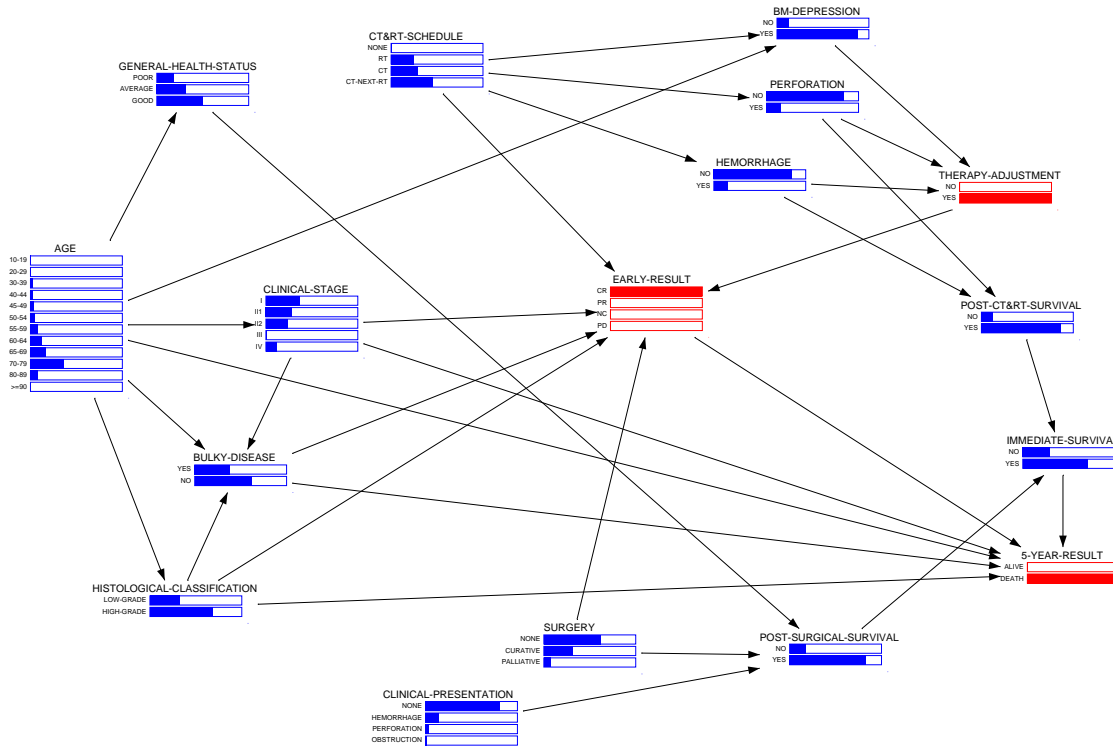
### 3.1.3 Risk profiles

Above, we have essentially used the decision-theoretic model of gastric NHL for prognostic assessment, where in the last case we have also determined the optimal treatment for a specific patient. The same model, however, can also be used to reason in the reverse direction: it is possible to start with the final results of treatment, and to generate probabilistic profiles that fit these final results. For this purpose, we use again the purely probabilistic version of the model. Consider for example the situation where we have been able to reach complete remission after treatment, but were required to adjust treatment, resulting in death within five years. The resulting profile is depicted in Figure 11.

### 3.2 Evaluation

There are various aspects of a computer-based system that may be evaluated, but the most important is, of course, the accuracy of the model. It is well-known among researchers that a rigorous evaluation of a decision-support system is difficult and time-consuming [33, 34, 35], but nevertheless required to obtain a system that can be used in daily clinical practice. In this section, a preliminary evaluation of the system is described. Evaluation of the system was hampered by the lack of a large clinical database with data of patients, who have received consensus treatment. As discussed in Section 1, no such database currently exists.

The developed influence diagram was evaluated by means of a blinded cross-validation study. A database with data of 28 imaginary patients was compiled such as to reflect a wide range of clinical pictures. The optimal treatment according to the influence diagram was first computed for these patients using the two utility functions discussed in Section 2.2.7. There was a complete agreement about optimal treatment for the influence diagram employing the two different utility functions in 79% of cases, and disagreement in the remaining 21%.

The same pretreatment data of patients as entered into the system were also presented

**Figure 11**: Risk profile for patients with therapy adjustment, complete remission, but death within five years following therapy.

to both clinicians from the Netherlands Cancer Institute (NCI) involved in the development of the system. Their therapeutic advice was recorded. In 59% and 54%, respectively, of the patient cases their advice agreed with the advice generated by the system when supplied with the utility functions obtained by the direct scaling and reference gamble methods, respectively. This clinical advice, together with the advice generated by the influence diagram supplied with the two different utility functions, was next reviewed by two experienced clinicians of the Leiden University Hospital (Leiden UH) (one haematologist, JCK-N, and one radiation oncologist, EMN). It was impossible for these clinicians to determine whether a particular recorded treatment advice had been produced by the system or by their colleagues. They were requested to select the treatment advice with which they fully agreed, if any. Treatment decisions considered acceptable but suboptimal, could also be modified to reflect own clinical practice and experience. If none of the recorded possiblities was considered acceptable, their judgment was recorded as being in 'disagreement'. The results of the evaluation are summarised in Table 3.

As can be seen, the two specialists from Leiden University Hospital slightly favoured the system supplied with the utility function obtained by the standard reference gamble over the recorded clinical advice, although the difference is not statistically significant ($P > 0.05$). The advice by the system that had been supplied with the utility function obtained by the direct scaling method was judged to be least acceptable. Interestingly, the two medical specialists from Leiden University were more positive about the system's performance when

| Clinicians of Leiden UH | Direct Scaling $n$ (%) | Reference Gamble $n$ (%) | Clinicians of NCI $n$ (%) |
|---|---|---|---|
| complete agreement | 15 (53) | 19 (68) | 17 (61) |
| agreement after | | | |
|   slight modification | 5 (18) | 2 (7) | 5 (18) |
| disagreement | 8 (29) | 7 (25) | 6 (21) |
| Total (patients) | 28 (100) | 28 (100) | 28 (100) |

**Table 3**: Performance evaluation results: 'direct scaling' represents the influence diagram with utility function obtained by the direct scaling method; similarly, 'reference gamble' represents the influence diagram with utility function obtained by the standard reference gamble method (see text).

it was supplied with the utility function obtained by the reference gamble method than the two NCI clinicians. However, again this difference is not statistically significant.

## 4  Discussion

A number of prognostic models of NHL, and of gastric NHL in particular, have been developed in the past. Examples of such models are those developed by Valicenti et al. [8], by Radaszkiewicz et al. [9], and by Azab and colleagues [6]. These models are based on univariate and multivariate analysis of data from patients with gastric NHL. Although these statistical techniques are useful for identifying relevant prognostic factors, the resulting models are of limited value from a clinician's point of view. As Wyatt and Altman have noted, many medical prognostic models have never reached the stage of actual clinical use [36]. This is unfortunate, because most of these models have been developed with the explicit goal of supporting the clinician in dealing with difficult clinical problems. An important reason why these prognostic models are not widely used in a clinical setting may be their restrictive nature: they cannot be used in the process of selecting optimal treatment.

In contrast to other researchers, we have chosen to develop a computer-based expert system, using the techniques of decision-theoretic networks, i.e. probabilistic networks and influence diagrams, as a basis for a model of gastric NHL. The design of a medical decision-theoretic network model is guided by the clinical experience of medical specialists, who are capable of indicating the random variables that are of importance in medical decision making, as well as the relationships, causally or otherwise, that hold between these variables. This information, supplemented with information from the medical literature and from clinical databases, offers a suitable foundation for building reliable decision-theoretic network models. A decision-theoretic network yields an attractive medical model, because the contributions of the various findings in dealing with a particular clinical problem are explicitly shown, a feature not shared by conventional statistical methods. Another advantage is that a decision-theoretic expert system can be employed by the clinician to explore many different clinical questions, as illustrated by the various examples above. Although the technique has been introduced fairly recently, a number of successful medical systems have already been developed, using this technique (cf. [20, 21, 24]).

The present basis of the system discussed in this paper is clinical experience, supplemented
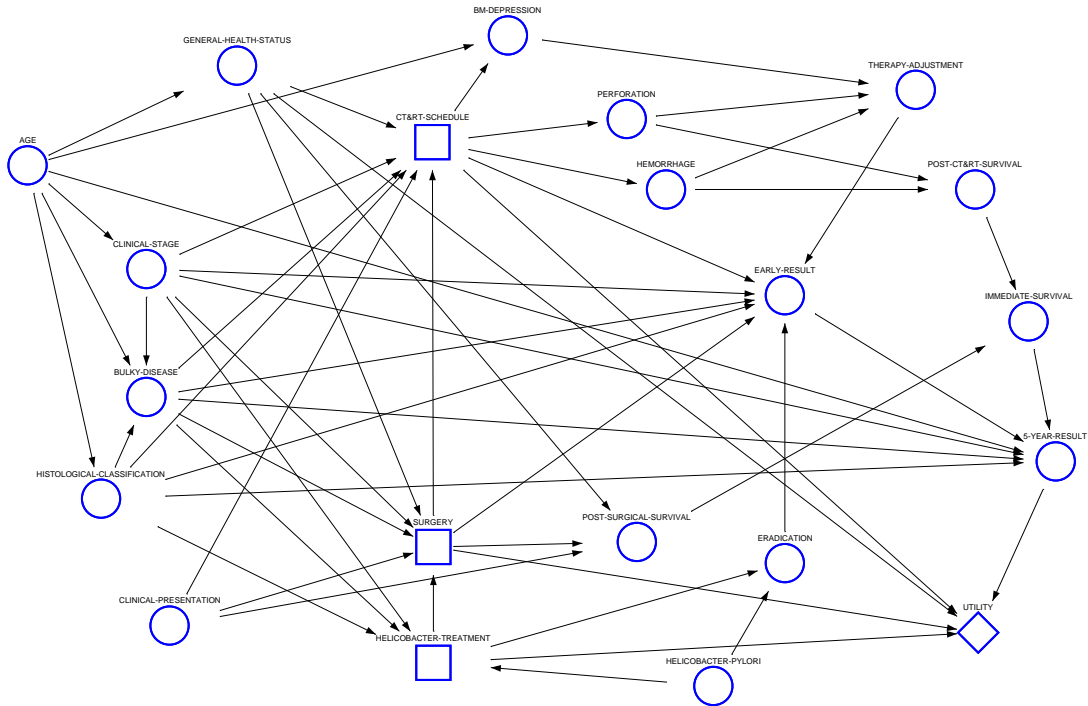
**Figure 12**: Influence diagram for the treatment of gastric NHL, after adding treatment for *H. pylori* infection.

with information gathered from the medical literature. Although it has been previously shown that such information may be a sufficient basis for the development of a reliable decision-support system [21], further improvement may be expected from adapting the underlying decision-theoretic network model using data from a large, reliable clinical database. We are currently planning to use data from the Netherlands Cancer Registry for improving the system, as well as for evaluation of the underlying decision-theoretic network model. This database, however, is inadequate for obtaining information about patients that have been treated for infection with *H. pylori*. The clinical experience with this treatment, in particular when combined with other treatment modalities, is currently insufficient for being included in the final model. The existing clinical experience, however, could easily be added to the model, by inserting three nodes into the influence diagram from Figure 4, yielding the influence diagram shown in Figure 12.

Although the development of a system as described in this paper is time-consuming, it has repeatedly been shown that computer-based decision support systems can improve health care [37]. We believe that a computer-based system that assists clinicians in exploring difficult clinical problems in a number of different ways, like the one presented in this paper, may aid to further improvement of clinical care.

# References

[1] Eidt S, Stolte M, Fishcer R. *Helicobacter pylori* gastritis and primary gastric non-Hodgkin's lymphoma. *J Clin Pathol* 1994; **47**: 436–439.

[2] Wortherspoon AC, Doglioni C, Diss TC, Pan L, Moschini A, de Boni M, Isaacson PG. Regression of primary low-grade B-cell gastric lymphoma of mucosa-associated lmphoid tissue type after eradication of *Helicobater pylori*. *Lancet* 1991; **342**: 575–577,

[3] Wyatt JI, Rathbone BJ. Immune response of the gastric mucosa to *Campylobactor pylori*. *Scan J Gastroenterol* 1988; **23** (suppl 142): 44–49.

[4] Hussell T, Isaacson PG, Crabtree JE, Spenser J. The response of cells from low-grade B-cell gastric lymphomas of mucosa-associated lyphoid tissue to *Helicobacter pylori*. *Lancet* 1993; **342**: 571–574.

[5] Bayerdörffer E, Neubauer A, Rudolph B et al. Regression of primary gastric lymphoma of mucosa-associated lymphoid tissue type after cure of *Helicobacter pylori* infection. *Lancet* 1995; **345**: 1591–1594.

[6] Azab MB, Henry-Amar M, Rougier P, Bognel C et al. Prognostic factors in primary gastrointestinal non-Hodgkin's lymphoma. *Cancer* 1989; **64**: 1208–1217.

[7] Shipp MA. Prognostic factors in aggressive non-Hodgkin's lymphoma: who has "high-risk disease"? *Blood* 1994; **83**(5): 1165–1173.

[8] Valicenti RK, Wasserman TH, Kucik NA. Analysis of prognostic factors in localized gastric lymphoma: the importance of bulk of disease. *Int J Radiation Oncology Biol Phys* 1992; **27**: 591–598.

[9] Radaszkiewicz T, Dragosics B, Bauer P. Gastrointestinal malignant lymphomas of the mucosa-associated lymphoid tissue: factors relevant for prognosis. *Gastroenterology* 1992; **102**: 1628–1638.

[10] Taal BG, Boot H, Van Heerde P, De Jong D, Hart GAM. Primary non-Hodgkin of the stomach: endoscopic pattern and prognosis in low versus high grade malignancy in relation to the MALT concept. *Gut* 1996; **39**: 556–561.

[11] Shipp MA, Harrington DP, Anderson JR, Armitage JO, Bonadonna G, Brittinger G, Cabanillas F, Canellos GP, Coiffier B, Connors JM, Cowan RA, Crowther D, Dahlberg S, Engelhard M, Fisher RI, Gisselbrecht C, Horning SJ, Lepage E, Lister A, Meerwaldt JH, Montserrat E, Nissen NI, Oken MM, Perterson BA, Tondini C, Velasquez WS, Yeap BY. A predictive model for aggressive NHL: The International Non-Hodgkin's Lymphoma Prognostic Factors Project. *N Engl J Med* 1993; **329**; 987.

[12] Hermans J, Krol AD, van Groningen K, et al. International Prognostic Index for agressive non-Hodgkin's lymphoma is valid for all malignancy grades. *Blood* 1995; **86**: 1460–1463.

[13] Gaag, LC van der. Bayesian belief networks: odds and ends. *The Computer Journal* 1996; 39: 79–113.

[14] Jensen, FV. *An Introduction to Bayesian Networks*. London: UCL Press, 1996.

[15] Pearl J. *Probabilistic Reasoning in Intelligent Systems*. San Mateo, California: Morgan Kaufman, 1988.

[16] Lauritzen SL, Spiegelhalter DJ. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society (Series B)* 1987; **50**: 157–224.

[17] Sox HC, Blatt MA, Higgins MC, Marton KI. *Medical Decision Making*. Boston: Butterworths, 1988.

[18] Shachter RD. Evaluating influence diagrams. *Operation Research* 1986; **34**(6): 871–882.

[19] Cooper GF. A method for using belief networks as influence diagrams. In: *Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence*, 1988: 55–63.

[20] Andreassen S, Woldbye M, Falck B, Andersen SK. MUNIN — A causal probabilistic network for interpretation of electromyographic findings. In: McDermott, J, ed. *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, Los Altos, CA: Morgan Kaufmann, 1987: 366–372.

[21] Heckerman DE, Nathwani BN. Towards normative expert systems: part II – probability-based representations for efficient knowledge acquisition and inference. *Methods Inf Med* 1992; **31**: 106–116.

[22] Spiegelhalter DJ, Franklin RCG, Bull K. Assessment, criticism and improvement of imprecise subjective probabilities for a medical expert system. In: Henrion M, Shachter RD, Kanal LN, Lemmer JF, eds. *Uncertainty in Artificial Intelligence 5*. Amsterdam: North-Holland, 1990: 285–294.

[23] Lucas PJF, Gaag LC van der. *Principles of Expert Systems*. Wokingham: Addison-Wesley, 1991.

[24] Heckerman DE, Horvitz EJ, Nathwani BN. Towards normative expert systems: part I – The Pathfinder project. *Methods Inf Med* 1992; **31**: 90–105.

[25] Korver M, Lucas PJF. Converting a rule-based expert system into a belief network. *Med Inf* 1993; **18**(3): 219–241.

[26] Lucas PJF. Knowledge acquisition for decision-theoretic expert systems. *AISB Quaterly* 1996; **94**: 23–33.

[27] Musshoff K. Klinische Stadieneinteilung der Nicht-Hodgkin-lymphome. *Strahlentherapie* 1977; 153: 218–221.

[28] Isaacson PG. Gastrointestinal lymphoma. *Hum Pathol* 1994; **25**: 1020–1164.

[29] Olesen KG, Kjaerulff U, Jensen F, Jensen FV, Falck B, Andreassen S and Andersen SK. A MUNIN network for the median nerve — A case study on loops. *Applied Artificial Intelligence* 1989; 3: 301–319.

[30] Wellman, MP. Fundamental concepts of qualitative probabilistic networks. *Artif Intell* 1990; **33**(3); 257–303.

[31] Pradhan M, Henrion M, Provan G, Del Favero B, Huang K. *The Sensitivity of Belief Networks to Imprecise Probabilities: an Experimental Investigation*. Report SMI-95-0595, Stanford CA: Stanford University, 1995.

[32] Srinivas S, Breese J. *IDEAL: Influence Diagram Evaluation and Analysis in Lisp Documentation and Users Guide*. Palo Alto: Rockwell International Science Center, 1993.

[33] Hilden J, Habbema JDF. Evaluation of clinical decision aids – more to think about. *Med Inf* 1990; **15**: 275–284.

[34] Miller PL. The evaluation of artificial intelligence systems in medicine. *Comput Programs Biomed* 1986; 22: 5–11.

[35] Wyatt JC, Spiegelhalter DJ. Evaluating medical expert systems: what to test for and how? *Med Inf* 1990; **15**(3): 205–217.

[36] Wyatt JC, Altman DG. Commentary: prognostic models: clinically useful or quickly forgotten. *BMJ* 1995; **311**: 1539–1541.

[37] De Dombal FT, Dallos V, McAdam WA. Can computer-aided teaching packages improve clinical care in patients with acute abdominal pain? *BMJ* 1991; **302**: 1495–1497.