A Graphical Decision-Theoretic Model for Neonatal Jaundice

Article *in* Medical Decision Making · May 2007

DOI: 10.1177/0272989X07300605 · Source: PubMed

Manuel Gómez-Olmedo
University of Granada

Concha Bielza
Universidad Politécnica de Madrid

Juan A. Fernández del Pozo
Universidad Politécnica de Madrid

A Graphical Decision-Theoretic Model for Neonatal Jaundice

Manuel Gómez, Concha Bielza, Juan A. Fernández del Pozo and Sixto Ríos-Insua

Med Decis Making 2007; 27; 250
DOI: 10.1177/0272989X07300605

A Graphical Decision-Theoretic Model for Neonatal Jaundice

Manuel Gómez, PhD, Concha Bielza, PhD,
Juan A. Fernández del Pozo, MSc, Sixto Ríos-Insua, PhD

Background. *Neonatal jaundice is treated daily at all hospitals. However, the routine, urgency, and case load of most doctors stop them from carefully analyzing all the factors that they would like to (and should) take into account. This article develops a complex decision support system for neonatal jaundice management.* ***Methods.*** *The problem is represented by means of an influence diagram, including admission and treatment decisions. The corresponding uncertainty model is built with the aid of both historical data and subjective judgments. Parents and doctors were interviewed to elicit a multiattribute utility function. The decision analysis cycle is completed with sensitivity analyses and explanations of the results.* ***Results.*** *The construction and use of this decision support system for jaundice management have induced a profound change in daily medical practice, avoiding aggressive treatments—there have been no exchange transfusions in the past 3 years—and reducing the lengths of stay at the hospital. More information is now taken into account to decide on treatments. Interestingly, after embarking on this modeling effort, physicians came to view jaundice as a much more difficult problem than they had initially thought. Comparisons between real cases and system proposals revealed that treatments by nonexpert doctors tend to be longer than what expert doctors would administer.* ***Conclusion.*** *The system is especially designed to help neonatologists in situations in which their lack of experience may lead to unnecessary treatments. Different points of view from several expert doctors and, more interestingly, from parents are taken into account. This knowledge gives a broader picture of the medical problem—incorporating new action criteria, new agents to intervene, more uncertainty variables—to get an insight into the suitability of each therapeutic decision for each patient situation. The benefits gained and the usefulness perceived by neonatologists are worth the increased and time-consuming effort of developing this complex system. Although specially designed for a specific hospital and for neonatal jaundice management, it can be easily adapted to other hospitals and problems.* ***Key words:*** *neonatal jaundice; decision analysis; decision making; clinical decision support systems; influence diagrams; multiattribute utilities; socioeconomic factors. (****Med Decis Making 2007;27:250–265****)*

Neonatal jaundice is a very common disease in newborn babies. The breakdown of old and surplus red blood cells in the baby's system produces bilirubin that cannot be excreted into the intestines at a normal rate by a still immature liver. Therefore, most neonates have a higher bilirubin level than healthy adults do. The toxic effects of bilirubin on the central nervous system are well known, as is the likely irreversible brain damage,[1,2] leading to detrimental neurodevelopmental abnormalities, motor problems, cerebral palsy, deafness, and even death. Moreover, some pathologies promote an additional increase of bilirubin or emphasize its toxic effects. It is therefore a challenge to distinguish between so-called (benign) physiological jaundice and the more serious version, pathological jaundice or severe hyperbilirubinemia, related to the development of kernicterus (acute bilirubin encephalopathy). In this case, it is necessary to examine if the baby has risk factors.

DOI: 10.1177/0272989X07300605

During the 1950s, the exchange transfusion of the baby's blood with a blood bank, although risky (1% mortality), was the only treatment. During the 1960s, the introduction of phototherapy in jaundice management meant that lower bilirubin cases still causing (less severe) neurological disorders could be treated, which led to a drastic reduction of blood exchanges.[3] Obviously, this intensive exposure involves an extended hospital stay and keeps the baby away from the mother, hence disrupting breastfeeding.

Current jaundice guidelines are based on a few indications (mainly total serum bilirubin, birth weight, and age), and the type of treatment (just observation, phototherapy, or exchange transfusion) is independent of the cause of jaundice. The appraisal of several clinical tests to provide follow-up information sometimes turns out to be unnecessary, results have not always proved useful, and costs are considerable. Newman and Maisels[4] proposed a less aggressive approach that tries to balance the risks and costs of undertreatment and overtreatment. In general, guidelines rely on informal methods and notations, described as combining different formats as text, tables, and charts. They are continuously restated by the authorized pediatrics organizations (see some of the latest proposals for neonatal jaundice by the American Academy of Pediatrics [AAP][5–7]).

Even though guidelines are representative of more or less precise recommendations reached by consensus and based on the best empirical evidence and literature surveys, and although they can reduce the costs of care by up to 25%,[8] some problems come to light. First, there is no consensus on the circumstances under which each treatment should be administered. Also, when a treatment should be changed/started is not clearly stated, leaving the decisions to physicians' experience and skill. It is this experience and skill that should be modeled to set out the many uncertainties involved in the decision-making process and that physicians actually bear in mind when making such decisions. Second, there is a need to take into account different doctors' points of view and, more interestingly, parents' opinions. Parents represent their baby's (the patient's) position and may wish to have a say in some decisions. Thus, for moderate jaundices and depending on their cultural level, they may undertake to observe the baby at home rather than having him or her hospitalized. In fact, AAP recommendation 6.1 says "all hospitals should provide written and verbal information for parents at the time of discharge, which should include an explanation of jaundice, the need to monitor infants for jaundice, and advice on how monitoring should be done." This so-called shared decision making between health care providers and patients is in high demand in modern clinical practice,[9] and its importance has been emphasized in a rapidly growing literature.[10] Third, obviously, as the amount of money and resources that health care consumes continues to grow, there is a mounting incentive to cut costs. Hopefully, better decisions will decrease not only treatment risks but also the costs of diagnostics and therapeutics.

This article describes a decision-theoretic approach to neonatal jaundice management encompassing all the aforementioned points. The decision support system we built as a by-product was called IctNeo. It operates at the (public) Gregorio Marañón General Hospital in Madrid, Spain, one of the biggest hospitals in Europe, covering an area of 1142 km^2, serving approximately a population of 650,000, and having almost 800 doctors and 1800 beds. Newly refurbished, it is responsible for 1 of the 11 health areas of the autonomous region of Madrid. Over the past 4 years, the hospital attended an average of 3720 births and 522 admissions, 15% of which were due to jaundice. As explained in the Discussion section, our proposal is a novel idea in this field. We use decision analysis, an accepted paradigm that integrates many forms of medical information. The decision model is an influence diagram.[11]

Received 27 May 2005 from the Decision Analysis and Statistics Group, School of Computer Science, Technical University of Madrid, Spain. Financial support for this study was provided in part by grants PB97-0856 and TSI2004-06801-C04-04 from the Ministry of Education and Science, grant 97/0003-02 from the Ministry of Health, grants 07T/0009/1997 and 07T/0027/2000 from the Madrid Regional Government, and grant A9902 from the Technical University of Madrid. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. These results were presented in part at the following conferences: 14th International Conference on Multiple Criteria Decision Making, 8–12 June 1998, Charlottesville, Virginia; Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (AIMDM'99), 20–24 June 1999, Aalborg, Denmark; 2nd International Symposium on Medical Data Analysis (ISMDA 2001), 8–9 October 2001, Madrid, Spain; 8th Biennial Conference of the European Society for Medical Decision Making, 2–5 June 2002, Taormina, Italy; and INFORMS Annual Meeting, 17–20 November 2002, San José, California. Revision accepted for publication 30 October 2006.

Address correspondence to Concha Bielza, PhD, Decision Analysis and Statistics Group, School of Computer Science, Technical University of Madrid, Campus de Montegancedo, Boadilla del Monte, 28660 Madrid, Spain; e-mail: mcbielza@fi.upm.es.

METHODS

Decision Model

After several interviews with 3 neonatologists acting as domain experts, we were able to model the structure of the problem by means of an influence diagram, identifying decisions to be made, the sequence in which they would arise, and the information available at each decision-making time, as well as all the uncertain events. It is well known that this initial structuring step is not automated yet and is almost an art in which most of the literature has taken little interest.[12] Only a few attempts to automatically construct decision models based on frames encoding medical domain knowledge and rules of correct model construction (decision trees or influence diagrams) have been made, such as the MIDAS system.[13] Although influence diagrams are conceptually simple, we will reveal here many difficulties that we encountered in our large diagram that can be extended for use in many other real problems.

The scope of the problem was delineated by considering infants born at the hospital and jaundice cases occurring in the early days of life, up to 4 days old, the critical period of time when the most harmful effects take place. Decision nodes involve the temporal axis of the model: first, to decide whether to admit the baby to the hospital and start treatment; second, if the baby is admitted, to decide the treatment (phototherapy, exchange transfusion, or simply observation) depending on some newborn factors and the results of any tests carried out. A medical test is useful not only because it reduces uncertainty but also because it permits the doctor to make better decisions, and its value must be evaluated in each particular case. Our proposal will take into account many more factors than usual (age, birth weight, bilirubin level). Decisions about treatments are repeated as many times as necessary until the problem is over, i.e., the infant is discharged or is given a treatment that is outside our specific problem (e.g., for hepatitis, Gilbert syndrome, etc.).

There are some constraints on the chain of treatment decisions, which makes the modeling process harder. For example, doctors cannot perform more than 2 exchanges per full treatment of a patient, 2 consecutive observations are never done, and each exchange must be followed and preceded by phototherapy, among others. If decision (identical) domains consist of the different treatment actions (each lasting 6 h) and hospital discharge, we would need up to 16 decision nodes to span 4 days. This would entail an intractable influence diagram and would not meet the constraints, containing sequences of treatments that are impossible. As a matter of fact, influence diagrams were not designed to model asymmetries such as these, although the recent literature suggests that this limitation could be addressed through the use of more sophisticated structures.[14–16]

We decided to consider domains based on combinations of the initial 6-h therapies, which, incidentally, the experts very often referred to in the clinical histories and during the interviews. Thus, it is common to perform a 12-h phototherapy in low-risk situations because this rules out having to retest bilirubin levels unnecessarily every 6 h. Moreover, there are even phototherapy sessions of up to 24 h. These combined treatments of different lengths became the alternatives at each decision point and satisfy most constraints, providing a diagram with only 5 decision nodes, whose domains are shown in Table 1. We dealt with the other asymmetries, not included in the new therapies domain, during the influence diagram evaluation (see the Implementation section). Note that the length of the full jaundice process varies; for example, some patients will need only 1 treatment stage. Therefore, subsequent decision domains are filled in with dummy therapeutic actions (do nothing or null), leading to a far more asymmetric diagram. Obviously, the system should not propose a treatment after patient discharge.

Before making the 1st decision on admission, the physician ascertains some factors about the neonate (and his or her mother) that inform about his or her state. These factors can be divided into 5 categories: mother's tests, mother's particulars, newborn particulars, newborn tests, and delivery data. The mother's tests are designed to determine her Rh factor, her blood group, and whether there is isoimmunization (an incompatibility between the mother's and the baby's bloods). The mother's particulars include 1) administrative data, such as age, race, whether she is a 1st-time mother, and whether she is ill; 2) social data, such as her cultural level and place of residence, included in a node called social cost, which determine whether some moderate jaundices could be discharged if parents—of perhaps a high social level and living near the hospital—undertake to observe the baby at home, thereby avoiding journeys to the hospital and neglecting other children or relatives; and 3) emotional data (emotional cost node) to assess emotional impact or anxiety because of the interruption of mother-baby bonding. Integrating social and emotional factors has been recently recognized as an important feature,[9] in this case for incorporating family criteria into jaundice treatment, as explained in the introduction, although it is not considered in the usual and current protocols.

Newborn particulars include birth weight, gestational age, and age in hours. Newborn tests include Coombs tests to determine blood group and Rh factor isoimmunization, tests to read bilirubin and hemoglobin concentrations, Apgar test and cord Ph test to study potential baby asphyxia, observation tests to measure how yellow the skin color is (a typical symptom of jaundice), and a test to evaluate the risk

Table 1 Domains of Decision Variables

1st Decision	2nd and 3rd Decision	4th Decision	5th Decision
No admission	Null	Null	Null
Admission + 6-phototherapy	Observation + discharge	Observation + discharge	Observation + discharge
Admission + 12-phototherapy	Observation	Observation + outside treatment	Observation + outside treatment
Admission + 24-phototherapy	Observation + outside treatment	6-phototherapy	
Admission + outside treatment	6-phototherapy	12-phototherapy	
	12-phototherapy	24-phototherapy	
	24-phototherapy		
	6-phototherapy + exchange transfusion + 6-phototherapy		
	12-phototherapy + exchange transfusion + 12-phototherapy		

of admission. Delivery data include whether instruments had to be used and whether the baby needed any kind of resuscitation. All this information is shown in the final influence diagram in Figure 1.

Apart from the variables known at the time of making the 1st decision, there are variables for monitoring the patient evolution after hospitalization, including both newborn data (age) and test results (bilirubin and hemoglobin serum concentrations). All of them, as well as the previous decision made, affect the next decision to be made, as represented by the arcs and configuring the different phases of the protocol. All the arcs in the chance nodes reflecting conditional probabilistic (in)dependencies were also elicited from the neonatologists. For example, bilirubin concentration depends on the baby's age and on the hemoglobin concentration (Figure 1).

Moreover, a number of pathologies may have an influence on hyperbilirubinemia: sepsis, congenital erythrocyte defect, inborn errors of metabolism, concealed hemorrhage, hypothyroidism, polycithemia, perinatal asphyxia, and isoimmunization (see the top of Figure 1, in which these pathologies are grouped inside an oval for the sake of clarity; i.e., an arc coming from that oval means that it comes from every node inside the oval).

Finally, the diamond-shaped node, called the value node, represents the utility function that models decision-maker preferences. Nodes with arcs leading to this node indicate the domain of the utility function, that is, the relevant variables for establishing the desirability of the consequences. Therefore, we constructed an objectives hierarchy with the help of the same physicians. The overall objective at the root of the hierarchy was to maximize patient well-being. By subdividing the objectives into more detailed lower level objectives, we clarify the intended meaning of the overall objective. Objectives were repeatedly tested for importance before inclusion in the hierarchy. The objectives tree was checked according to suitability criteria.[17] An objectives hierarchy for our jaundice problem is shown in Figure 2.

Note that we have the lowest level objectives aiming to minimize costs (X_1) and minimize injuries induced by administration of specific treatments (X_5) and by alteration of bilirubin levels (hyperbilirubinemia; X_6). Monetary costs are taken into account, although this was not so easy for physicians working in the Spanish National Health Service. Both secondary effects and bilirubin levels are objective (and conflicting), although the injuries they produce are subjective and must be assessed. Both X_5 and X_6 take into account the death risk as a possible extreme situation. We also have the minimization of the risk of being admitted to the hospital (X_4), a risk arising from infections, contagions, and so forth aimed at discouraging unnecessary admissions. Moreover, 2 objectives, social cost (X_2) and emotional cost (X_3), measure the preferences of parents as explained above. These 6 variables were added to the influence diagram (Figure 1). Therefore, we might state that the perspective being modeled is 3-fold: for the patient, for the hospital, and, all in all, for the society. Actual concerns involve the health of the patient, the family's quality of life, the cost of the treatment process, and the medical training.

In short, our custom-built influence diagram contains 5 decision nodes and 68 chance nodes. It took approximately 3 years to build, with 2-h interviews every 3 weeks. The diagram was continuously revised as new and refined knowledge was gained. This

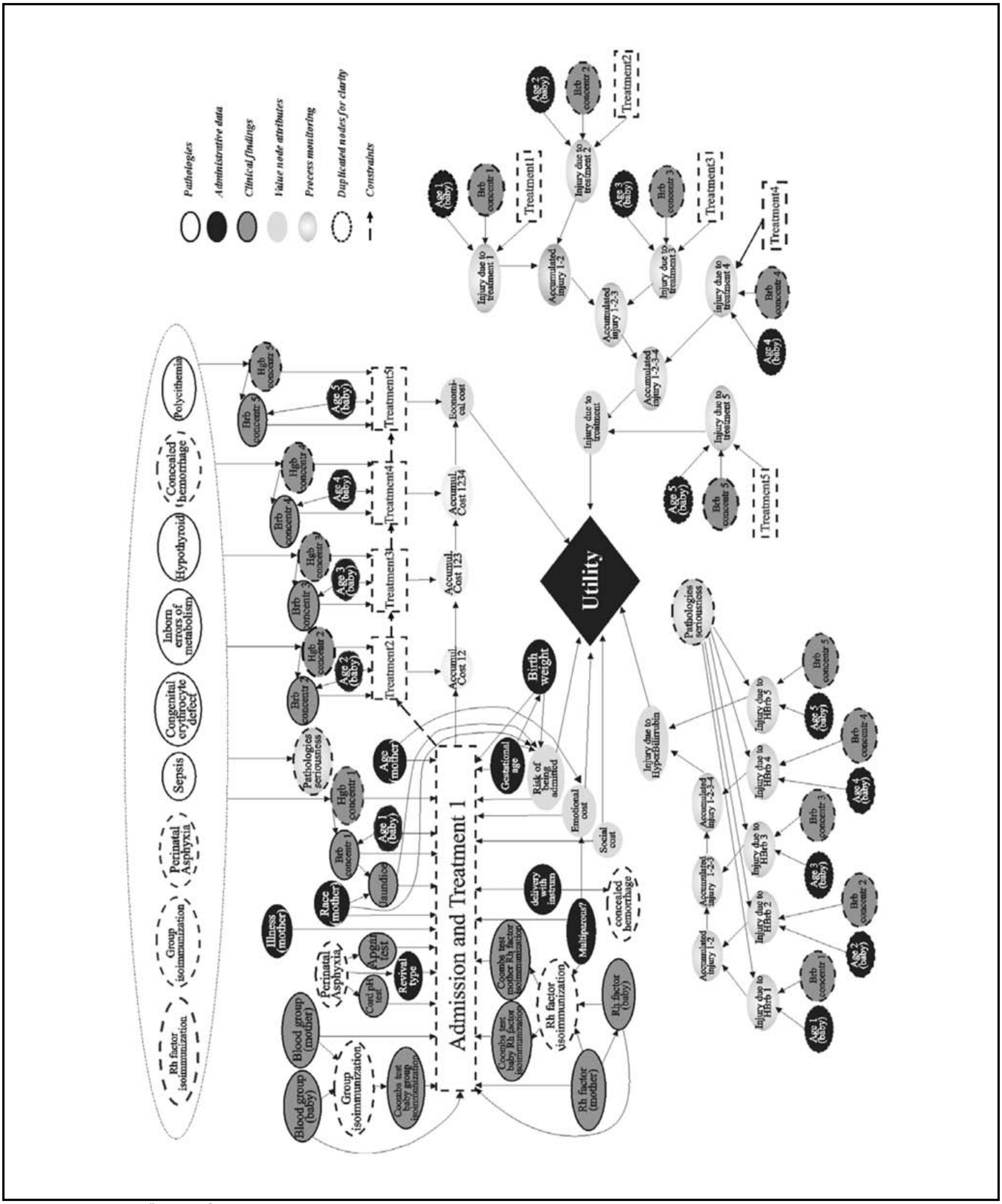

Figure 1 *IctNeo influence diagram.*

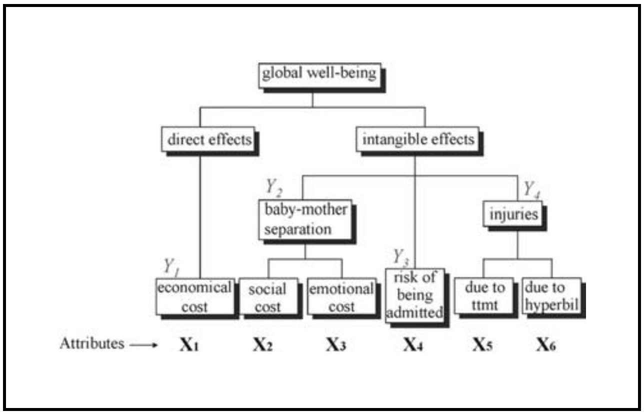

Figure 2 The objectives hierarchy for the jaundice management problem.

knowledge provides a broader picture of the medical problem—incorporating new action criteria, new agents to intervene, more uncertainty variables—to get an insight into the suitability of each therapeutic decision for each patient situation. Interestingly, after embarking on this modeling effort, physicians came to view the jaundice as a much more difficult problem than they had initially thought.

Modeling Uncertainty: Elicitation of Probabilities

The qualitative structure of the influence diagram is completed with the probabilistic information attached to chance nodes. For each node, we had to elicit a conditional probability distribution for that node (the random variable it represents) conditioned by its parents. The sample space or domain of each random variable was readily defined, even for continuous variables, because neonatologists decided to work with discretizations (e.g., standard intervals for weights for which they act similarly: 499–999 g, 1000–1499 g, 1500–2499 g, and more than 2500 g, or likewise for ages). Other more subjective variables related to medical appraisals were defined according to a scale close to the physicians' thinking. For instance, the scale for the pathology seriousness node comprises ordinal values (slight, serious, and very serious) commonly used by the experts and detailed enough to cover all the situations. All the pathologies are binary (absent/present), whereas bilirubin concentration ranges over normal, pathological, and very pathological, depending on the baby's age.

Having 68 random variables, we needed to elicit 13,521 probabilities. This task has been often cited as a major obstacle to building large probabilistic networks (see, e.g., van der Gaag and others[18] and the special issue of *IEEE Transactions on Knowledge and Data Engineering*, volume 12, number 4, 2000). From sources of probabilistic information, such as data from historical patient records and literature results, we were able to assess only 3.12% of the probabilities. The hospital maintains format-free patient histories, containing a lot of information that is, however, far from easy to automate in a computer and that does not match the variables considered.

Therefore, most probabilities were elicited from the knowledge and personal clinical experience of the 3 neonatologists, as subjective expert judgments. We followed the Stanford Research Institute encoding process and its extensions[19,20] as a formal protocol for probability elicitation. Many interviews with physicians were needed to overcome biases and poor calibration.[21] We assigned individual judgments from the 3 physicians, which we retained for sensitivity analysis. Group probabilities were obtained through behavioral aggregation, sharing knowledge by interaction among the doctors. In our case, there were no significant differences among individuals after their exchange of information.

Nevertheless, the main problem was how to get the probability table of a chance node with many parents from experts. The table contains as many entries as the product of the cardinalities of each parent domain and that of the chance node. Thus, we designed some mechanisms to dispense with so many elicitations. First, we added intermediate nodes to represent the partial and cumulative influence of the variables on a process step by step. For example, the economical cost variable that initially had the 5 decision nodes as its parents, requiring up to 5350 probability assessments, was turned into 4 progressive nodes requiring a total of only 700 assessments (see Figure 1). Other examples are injuries due to treatment and injuries due to hyperbilirubinemia. This could be considered a kind of "divorce the parents" strategy.[22]

Second, we identified logical constraints. For example, accumulated cost 1-2 (*C12*) is less than or equal to accumulated cost 1-2-3 (*C123*). Thus, if the economical cost has 5 possible outcomes (money intervals), probabilities $P(C123 = i \mid C12 = j$, treatment 3) do not have to be assigned for $i < j$, saving 10 assessments.

Third, we used generalized noisy-OR gates,[23] an extension of the noisy-OR gate.[24] This is a causal canonical model with some causes acting to produce an effect, with all the causes and the effect having values absent and present with various degrees of intensity. If some assumptions are held, the only assignments required to derive the others are those of the conditional probabilities of the effect given that all causes but 1 are absent, needing a number of assignments that is linear instead of exponential with respect to the number of causes, as in the general case. For example, see hemoglobin concentration in Figure 1 conditioned by all the pathologies. We obtained its conditional distribution by assessing only 18 probabilities rather than the required 1536.

For the whole jaundice problem, the 1st and 3rd mechanisms allowed us to reduce the number of parameters required to quantify uncertainty by 91.42%. The use of logical constraints reinforced this reduction further, reaching 97.83%. Therefore, only 3073 parameters (instead of 13,521) are used to define the probability distributions. The node requiring the largest number of assessments is the one related to the risk of being admitted (768 parameters). Obviously, several changes were made to the diagram structure to permit the use of these techniques. Gómez[25] showed all the details, which are available on request. These savings have a profound impact on the computational cost of the evaluation (see the Implementation section). For example, the 2nd mechanism dispenses with having to compute the expected utilities related to configurations marked as impossible by logical constraints.

Note that the probabilistic part of the model, a Bayesian network, can be used independently for diagnoses and to generate patient-specific risk profiles. However, our influence diagram goes further, as it is part of a decision support system that can answer many different clinical questions.

Modeling Preferences: The Multiattribute Utility Function Assessment Procedure

Following the decision analysis cycle, the acquisition of quantitative information for the influence diagram finishes after having elicited the utility function that represents decision-maker preferences for the consequences of decisions and outcomes. These were organized in the objectives hierarchy shown in Figure 2. For each of the 6 lowest level objectives, we then identified an attribute and a scale to indicate the extent to which objectives are achieved. For X_1, money was the natural proxy attribute. For the remaining objectives, we had to construct ad hoc scales with the help of neonatologists and even parents (for X_2 and X_3). For example, the values for X_4, the risk of being hospitalized, are $0 =$ *newborn not admitted to hospital, under observation*; $1 =$ *admitted for tests*; $2 =$ *admitted for medium care*; and $3 =$ *admitted for intensive care*.

The next task was to assess a utility function, $u(x_1, x_2, x_3, x_4, x_5, x_6)$, where x_i designates a specific level of X_i for 1 particular treatment strategy. A direct assessment of u had major practical shortcomings (we needed 5400 assessments!). Therefore, we investigated various sets of independence assumptions[17] about decision-maker attitudes to risk, to derive a consistent functional form of the multiattribute utility function.

To facilitate checking independence conditions and because of the homogeneity, on the one hand, of attributes X_2 and X_3 and, on the other hand, of X_5 and X_6, we assumed that such attributes could be structured temporarily, substituting X_2 and X_3 by only 1 attribute, Y_2, which represents "baby-mother separation," and X_5 and X_6 by Y_4, meaning "injuries." Hence, we should have for the time being 4 attributes denoted $Y_1 = X_1$, $Y_2 = (X_2, X_3)$, $Y_3 = X_4$, and $Y_4 = (X_5, X_6)$ see Figure 2. Our 1st aim then was to determine a utility function of the form $u(y_1, y_2, y_3, y_4) = f[u_1(y_1), u_2(y_2), u_3(y_3), u_4(y_4)]$.

To determine f, we verified the independence conditions that imply that the utility function u must be either additive,

$$u(y_1, y_2, y_3, y_4) = \sum_{i=1}^{4} k_i u_i(y_i), \qquad (1)$$

or multiplicative,

$$u(y_1, y_2, y_3, y_4) = \sum_{i=1}^{4} k_i u_i(y_i) + k \sum_{i=1, i>j}^{4} k_i k_j u_i(y_i) u_j(y_j) + k^2 \sum_{i=1, j>i, l>j}^{4} k_i k_j k_l u_i(y_i) u_j(y_j) u_l(y_l) + k^3 \prod_{i=1}^{4} k_i u_i(y_i), \qquad (2)$$

where k, k_i, $i = 1, 2, 3, 4$, are the scaling constants or weights. Specifically, these conditions were as follows: Y_2 was utility independent of its complement $\bar{Y}_2$, as was Y_4 with respect to its complement $\bar{Y}_4$, and $\{Y_4, Y_i\}$ $(i = 1, 2, 3)$ was preferential independent of its complement.

Then, we had to identify functions f_2 and f_4,

$$u_2(y_2) = f_2\lfloor u_2^x(x_2), u_3^x(x_3)\rfloor$$
$$u_4(y_4) = f_4\lfloor u_5^x(x_5), u_6^x(x_6)\rfloor,$$

where the u_i^xs are (single-attribute) utility functions over their respective domains. Fortunately, we found that f_2 and f_4 were additive functions. Neonatologists had to check whether X_2 and X_3 were conditionally additive independent given that Y_2 is fixed at any level and whether X_5 and X_6 were also conditionally additive independent given that Y_4 is fixed at any level. However, neonatologists did not find them easy to check because X_5 and X_6 varied jointly, and it was difficult to imagine situations in which one attribute was high and the other attribute was low. Therefore, we used the alternative test on Keeney and Raiffa's independence (see further details in Gómez and others[26]), getting an affirmative response. Hence, we obtained additive utility functions for Y_2 and Y_4 given by

$$\begin{aligned} u_2(x_2) &= k_2^x u_2^x(x_2) + k_3^x u_3^x(x_3) \\ u_4(x_4) &= k_5^x u_5^x(x_5) + k_6^x u_6^x(x_6). \end{aligned} \tag{3}$$

Moreover, to gain more confidence in the consequences, we decided to also apply an alternative test of additivity not involving lotteries,[27] with an interesting confirmation of the above results.

Next, we assigned the component utility functions u_i and u_i^x from the combination of 2 standard procedures, the probability equivalent (PE) method and the certainty equivalent (CE) method (see, e.g., Farquhar[28]), to mitigate the bias and inconsistencies of the elicitation process. We used the PE method known as extreme gambles, and the CE method known as the fractile method. Furthermore, instead of assessing only 1 number at each probability question as each method demands, we assessed a class of utility functions for each attribute.[29,30] This is less demanding because we asked physicians and parents to provide only incomplete preference statements by means of intervals, rather than point values. Figure 3A shows the ranges of 1 of these utility functions obtained using both methods, represented by the bounding utility functions u_1^L and u_1^U, where L (U) means lower (upper).

We compare the responses given by both inconsistency detection methods, in which inconsistencies are present if the intersection area obtained from both response types is empty in any range of the attribute. If inconsistencies were present, the preferences had to be reassessed until the expert provided a consistent range for the utility function. Thus, the intersection will be the range for the expert's utility functions. Figure 3B shows that range for utility u_1.

Because we need to rank the strategies, we built point-valued utility functions u_i. We fitted piecewise exponential functions $a + b^{-cx}$ using least squares

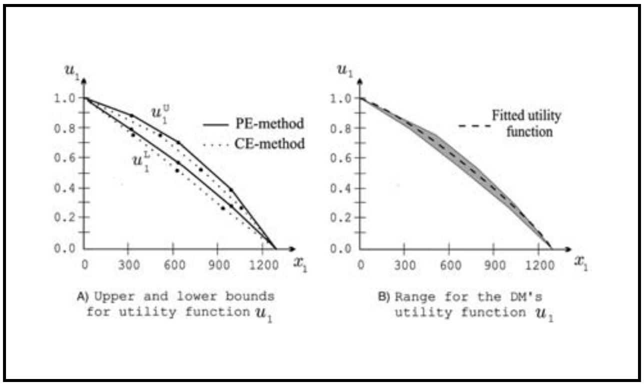

Figure 3 Utility functions for attribute X_1. PE = probability equivalent; CE = certainty equivalent.

with the midpoints of the utility intervals. This is illustrated by the dotted line in Figure 3B. All utility functions were assessed by physicians, except for the emotional cost attribute, which was also assessed with the help of parents. Table 2 shows the fitted component utility functions for all attributes.

As a utility function is defined for each attribute, the measures are comparable. The last step was to combine the individual utility functions into the overall function (1) or (2), with components (3). The procedure for establishing weights or scaling constants k_i was the tradeoff method, again providing imprecise assignments (i.e., an interval $[k_i^L, k_i^U]$). Therefore, for Y_1, the weight interval was [0.073, 0.145]; for Y_2, it was [0.019, 0.043]; for Y_3, it was [0.091, 0.271]; and for Y_4, it was [0.065, 0.152].

Now, because $\sum_{i=1}^4 k_i \in [\sum_{i=1}^4 k_i^L, \sum_{i=1}^4 K_i^U] = [0.248, 0.611] \not\supseteq \{1\}$ the multiplicative utility function (2) was always appropriated, and the additional constant k was found by observing that it is always $\sum_{i=1}^4 k_i < 1$, implying that $k \in (0, \infty)$. Therefore, we determined k as the solution to $1 + k = \Pi_{i=1}^4 (1 + kk_i)$.[17] Because we had weight intervals, we obtained a range [2.514, 19.163] for k, in which the extremes were obtained from the lower and upper values of those intervals, respectively. We used the same tradeoff-based procedure to assess the scaling constants k_i^x in the additive utility functions (3) but taking into account the consistency requirements $k_2^x + k_3^x = 1$ and $k_5^x + k_6^x = 1$. The resulting intervals were [0.262, 0.434], [0.247, 0.388], [0.374, 0.522], and [0.231, 0.467] for X_2, X_3, X_5, and X_6, respectively.

Because we have to rank the strategies with (2), we need precise values for the scaling constants. Hence, we provided the average values for constants k_i, given by $k_i = (k_i^L + k_i^U)/2$, obtaining $k_1 = 0.109$,

Table 2 Single-Attribute Utility Functions

Attribute	u_i	Range
X_1	$u_1(x_1) = 1.604 - 0.604\exp(0.00077x_1)$	[0, 1260]
X_2	$u_2^x(x_2) = -0.1108 + 1.111\exp(-1.153x_2)$	[0, 2]
X_3	$u_3^x(x_3) = -0.225 + 1.225\exp(-0.8473x_3)$	[0, 2]
X_4	$u_4(x_4) = 1.277 - 0.2766\exp(0.5098x_4)$	[0, 3]
X_5	$u_5^x(x_5) = 1.361 - 0.361\exp(0.3316x_5)$	[0, 4]
X_6	$u_6^x(x_6) = 1.408 - 0.4083\exp(0.2476x_6)$	[0, 5]

$k_2 = 0.031$, $k_3 = 0.181$, and $k_4 = 0.109$. For the scaling constants in the additive functions, we provided the normalized average values, obtaining $k_2^x = 0.578$, $k_3^x = 0.422$, $k_5^x = 0.558$, and $k_6^x = 0.442$. From the k_i values, we found constant k. Note that this constant determines the type and degree of interaction between attributes. We found that $k = 6.329 > 0$ and $\sum_{i=1}^{4} k_i < 1$, so we have attributes that complement each other: Preferred values of attributes will yield high values for the overall utility.[12]

A detailed and technical explanation of this laborious process is given in Gómez and others.[26]

Implementation and Computational Issues

Reuse was the key motivation during the development of the IctNeo system. There are several reasons for this, the 2 most important being the following:

- The construction of a decision support system such as this requires continuous reconsideration of the model, adding and removing nodes and arcs (qualitative changes) and modifying probabilities and utilities (quantitative changes). It must be easy to make all these changes to the model. If this condition is not met, the system is not really useful.
- This system should be used as a learning and training project for dealing with new medical problems. Therefore, the core of the system must be completely separate from actual neonatal jaundice knowledge. The core should be concerned with specifying, solving, and analyzing models, whatever they are.

To achieve these purposes, we defined a 5-layer architecture (see Figure 4).

1. Specification layer. The 1st task was to define a language to express the knowledge about decision problems structured as influence diagrams. This language had to be general enough to cope with any decision problem. Obviously, it had to be able to express the variables (chance and decision), their states, the relationships between them (providing the parents for any variable), the probabilities and utilities, and qualitative knowledge about any particular domain (e.g., knowledge about the constraint on the application of more than 2 blood exchanges per patient). As a consequence of using this language, a compiler is required to check that the information about the influence diagram is properly written, to test that the probability distributions are well stated, to generate the utility function for a set of parameters, and finally, to create the data structures required to evaluate the influence diagram. That is, the tools of this layer are used to convert the knowledge coming from the experts into an automatically solvable influence diagram.
2. Analysis layer. It contains methods to analyze the influence diagram and to look for near-optimal evaluation strategies. The use of such methods is optional for small-size influence diagrams but mandatory for complex models. When solving complex models, the computations required may be substantially reduced by a good sequence of operations over the structure (arc reversals and node deletions).
3. Evaluation layer. This layer solves the influence diagram and presents the set of optimal policies to the user. It can work directly with the influence diagram, looking at itself for a sequence of operations, or it can receive this sequence from the analysis layer. The diagram is ready to receive evidence about the values of some variable(s) and to look for optimal policies, taking into account this knowledge. It should be noted that the evaluation of influence diagrams is NP-hard.[31] This module tries to overcome this problem, simplifying the computations as much as possible, delaying unnecessary computations until they are required, and using the qualitative knowledge about the problem. Detecting and using the qualitative knowledge about the problem domain is a fundamental task for reducing the computational complexity and ensuring the coherence of the system proposals. For example, the constraints on the decision variables make it that only 12.58% of their value combinations is allowed. These constraints can be directly specified through logical relationships, and thus, the system computations focus on this 12.58%, not computing the expected utility for the rest of them. This evaluation layer offers another important feature: During the evaluation of the model, it is possible to analyze the posterior distribution of the variables. We use this information to check model coherence, fixing values for critical variables and observing the changes in these probability distributions. Of course, all these results were presented to the experts so they could validate, step by step, structural and parametric changes in the influence diagram.
4. Knowledge base management layer, with 2 main functions: 1) to allow for an incremental evaluation of the influence diagram, adding knowledge for

particular subproblems and incorporating more and more knowledge as long as new computations are made, and 2) to structure the optimal policies to reveal the most important variables for the decision-making problem and using this information to get an optimal storage of the policies. This is essential in complex decision problems: Neither a complete evaluation of the model nor a direct storage of the decision tables is possible. The relevance of the variables can be used as an explanation for the system proposals too.

5. Query layer. It is responsible for receiving the queries from users about specific patients or more general questions, and it outputs the optimal policies for the treatment of those patients. The system proposals are justified by a report containing explanations or reasoning, which is very useful for physicians and makes the system an effective tool.[32] Computer-generated explanations are based on the values of the relevant variables included in the influence diagram, identified by a technique introduced in Fernández del Pozo and others.[33] Diagnosis outputs related to pathologies are also possible using their posterior distributions. Unlike the other layers, concerned with the generation of the final decision tables, this will be the query layer installed in the hospital computers, using the knowledge base produced by the previous layer. The knowledge base used by the query layer must be updated as long as more evaluations are carried out.

In Figure 4, the line connecting the query and the specification layers shows the iterative structure when constructing real decision support systems, making continuous reconsiderations as long as the system is working on new cases. The final influence diagram will actually be obtained after several cycles throughout the whole process.

To give an understanding of the complexity related to the development of a decision support system such as IctNeo, let us explain in further detail some computational difficulties faced when generating the optimal policies. These policies will consist of as many decision tables as decision nodes are present in the diagram. Each decision table will contain the known variables when the decision is to be made. To get an idea of problem complexity, see Table 3 for the number of relevant variables for every decision table (1st row), the number of states for every decision D_i (2nd row), and how many numbers must be computed to determine the optimal policies (3rd row).

Even for the 1st decision, the number of values for computation (and storage) is very high. But the situation is even worse for the other decisions. The

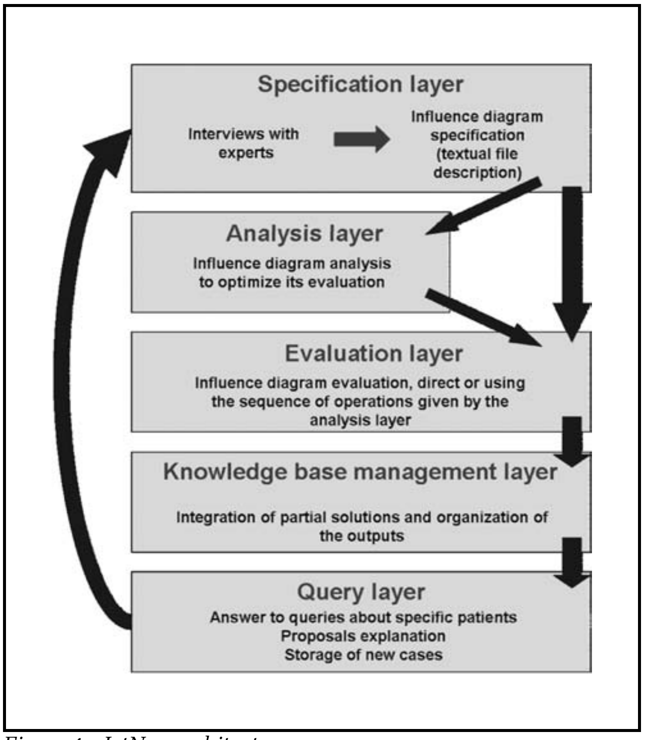

Figure 4 *IctNeo architecture.*

decision tables grow steadily until they exceed the storage capacity of any personal computer. Thus, computational techniques need to be used to evaluate such a model and overcome this problem. One of the most important techniques is called instantiation. By instantiating, the problem shifts to determine the optimal policy given the values of some variable(s), much in the spirit of a divide-and-conquer heuristic. Once these partial solutions have been found, the global solution to the problem should be built by putting together the partial solutions. For the IctNeo system, this is possible because the knowledge base obtained from the partial solutions is organized as a list-based compact representation introduced in Fernández del Pozo and others,[33] called the KBM2L list. Partial solutions are sequentially added to the KBM2L list by means of a learning mechanism that optimizes the list (minimizes its length) as new knowledge is entered.

To decide the set of variables for instantiation, the following points are considered: It must be large enough to allow the evaluation of the associated subproblem, the variables for selection must be observable when the patient is admitted (therefore, they are directly observed when the patient is under examination), and their values are selected by

Table 3 Complexity of the Jaundice Problem

	D_1	D_2	D_3	D_4	D_5
Relevant variables	20	23	26	29	32
States for D_i	5	9	9	7	3
Numbers to be computed	1.9×10^9	6.1×10^{10}	1.1×10^{12}	1.5×10^{13}	1.2×10^{14}

analyzing cases previously examined at the hospital (as doctors pointed out). With this in mind, 18 variables were chosen for instantiation following the experts' indications, most of the resulting cases corresponding with frequent patients. Combining the values for these variables, we get 42,467,328 different configurations. Therefore, we would have the complete solution to the jaundice problem after carrying out this many evaluations.

Fortunately, it is possible to make some simplifications. For example, the mother race variable initially had 4 outcomes in its domain: Caucasian, gypsy, Asiatic, and black. But over the past 2 years, there was no Asiatic case and only 1 black case at the hospital. Therefore, the evaluation of the partial influence diagrams will begin by taking into account only 2 values: Caucasian and gypsy. A similar thing applies for the gestational age variable because preterm infants will be admitted whatever their bilirubin levels are. Another simplification proposed by the experts suggests considering some tests as error free. After these simplifications, the number of evaluations to be performed is reduced to 165,888. The results of an evaluation yield much smaller decision tables: The last one requires about 3.13×10^7 values (instead of 1.20×10^{14}; see Table 3).

The use of KBM2L lists[33] not only allows the aggregation of the partial results into the global solution to the problem but also a compact storage of policies, drastically reducing the memory space required to manage the complete decision tables. If 4 bytes are required to store a floating point number, the size of the decision table for the last decision is 119.4 MB, whereas the KBM2L list organizes the knowledge of all the decision tables in 18 MB. Moreover, this kind of knowledge organization can be interpreted as high-level explanations of the system proposals. Suppose a specific combination of values for several variables always leads to the same policy, regardless of the value of the remaining variables. The idea is that, in this case, the variables belonging to that combination explain the policy. The ability to explain the results is an enhancement that makes our system much more useful to clinicians. Of course, the explanations are also a means of validating the system. At present, the decision tables used at the hospital represent about 1% (although it seems a small part of the global solution, it includes about 4×10^{11} cases of the global final solution). More and more evaluations are added as long as they are required by the experts. Details about computational issues related to this decision support system are described in Bielza and others.[34]

Sensitivity Analysis

During the model construction and validation, sensitivity analysis techniques were applied to gain confidence about the model's behavior. Analysts may be unable to understand large complex models and explain the relationships between their components, and formal techniques may be required. Our analysis focused on several critical parameters (19) pointed out by experts, with whose assignments they felt more unhappy. Basically, these were the probabilities of some pathologies, the probabilities of the utility function attributes, and some weights used for the utility function. We first performed 1-way sensitivity analyses, mainly using tornado diagrams, to get starting guesses. The problem was especially sensitive to parameters related to the risk of being admitted to the hospital, perinatal asphyxia, and all the weights of the utility function. We then proceeded with multiway sensitivity analyses to explore the impact of several parameters at a time and find out possible interrelationships. Because of the complexity of our model, we implemented a simulation-based method based on the expected value of perfect information.[35] This indicator accounts for expected utility changes accompanying preferred decision changes. The jaundice problem was sensitive to its entire parameter set and to some joint variations of certain parameters, but it was insensitive to most individual parameters. Therefore, parameter interactions are important. The results of the sensitivity analysis led us to examine the parameters considered as relevant, as well as the interactions between them. Details of this analysis can be found in Bielza and others.[36]

RESULTS

Successful design and implementation of a decision support system depend not only on the aforementioned technical issues but also on how useful clinicians, as potential system users, perceive it to be. Thus, it is crucial to get their opinions about factors such as ease of use, no increase in workload, and precise information.[10] In general, clinicians were favorable to IctNeo, which fulfilled their requirements as follows. First, clinicians now have a way to centralize information about patients. Indeed, the query layer (see Figure 4) serves as a database for patient treatments containing structured information. The old format-free patient histories distributed across several databases have given way to automated and organized patient data entry and outputs and a useful repository of full clinical histories for statistical/clinical analysis. Second, IctNeo reports on comparisons between its recommended treatments and the treatments actually administered, looking for matches or discrepancies. It also reports on comparisons of diagnoses because a pathology's prior probabilities are updated via Bayes rule as long as the diagram is solved. It can also answer expert and/or incomplete questions. Third, the system is an aid that intern doctors can consult when they are in doubt or just for training.

Fourth, IctNeo is available on a common PC where the (partial) evaluation of the model is stored. The communication with IctNeo is simply via a user interface that launches a query via Microsoft Access at its internal database, obtaining the reply—optimal strategy and its explanation—at once.

In this sense, IctNeo plays the role of a doctor in a clinical session, the role of a teacher to help interns on the survey of a (real or virtual) case, and the role of an alarm device to help nurses track the patient. For each patient, the user may look at the system up to 5 times according to the 5 phases that the model considers.

Fifth, the workload was heaviest while the system was under construction. However, despite being a reference hospital receiving critical patients even from other hospitals and working under time pressure, this effort was worthwhile because it has been this introspection into the jaundice problem that has made such a profound impact on the daily clinical practice.

Specifically, an outstanding result is that the exchange transfusions performed in the past 3 years have decreased sharply to zero, instead of the usual 4 or 5 per year. Moreover, some routine and defensive practices by nonexpert doctors have been modified. They used to recommend nonrisky actions, for example, not discharging patients and keeping them at the hospital receiving phototherapy, especially on the weekends—when the expert doctors were not present—postponing the final decision until Monday after consultation with an expert. This practice was observed after analyzing 50 clinical histories used to validate the decision support system. Of course, these histories fit the 1% of cases already evaluated. When the decisions made by doctors were confronted with the IctNeo proposals, we observed that the system tended to recommend discharge before the patient was actually sent home. This happens in 78% of the cases. This leads to shorter stays at the hospital and combats a defensive medicine. Neonatologists were satisfied with the system recommendations, which are considered to be valuable training tools for nonexpert doctors.

The validation indicates that there is a good match between experts and system proposals. Only 6% of cases differ. These cases refer to low-weight babies with pathological levels of bilirubin and hemoglobin and with pathology-induced risk factors. For them, the system proposals tend to recommend medium-term phototherapies (about 12 h between controls), whereas experts think a shorter period of time between controls might be better. We think that this discrepancy can be attributed to the difference between thinking about an abstract patient (when the utility function is built, experts do not have a real patient in mind) and a real case. Therefore, although the utility function captures expert doctors' theoretical preferences, they are not always the preferences used for real patients.

Table 4 shows a comparison between the decision support system proposals and the experts' decisions. This is based on the set of 50 clinical histories mentioned before. From left to right, the 1st column shows the issue under consideration; the 2nd and 3rd columns contain the system's and experts' proposals, respectively; and the last column includes comments about the comparison.

To better validate the results of the decision support system, we conducted an analysis of the knowledge base containing the system proposals. Thanks to this analysis, we were able to produce a set of profiles for every decision, outlining the relevant variables for each one. Remember that the whole validation is not possible because of the enormous set of cases contained in the decision tables. An interesting result is that, unlike the old protocols, some factors, such as social and emotional costs, are

Table 4 IctNeo's Versus Experts' Proposals

Issue	IctNeo	Experts	Comments
Admissions	46 admissions, 4 not admitted	49 admissions, 1 not admitted	6% of difference
Outside treatments	4 cases, 1 at the 1st decision, 3 for the rest		100% match
Phototherapies	9 of short duration, 23 of average duration, 13 of long duration	8 of short duration, 21 of average duration, 16 of long duration	6% of difference; the 3 categories take into account the total phototherapy duration all over the stay; short duration (up to 24 h), average duration (24–48 h), long duration (more than 48 h)
Exchange transfusions	0		The experts took into consideration this alternative for only 1 patient
Total length of stay at hospital	35 with shorter stays, 10 with the same stay length		78% of the cases with shorter stays; this is measured by adding the actual duration of every patient treatment and computing the difference with respect to physician proposals

now considered. For example, the system proposals about the admission of healthy babies with low levels of bilirubin are highly dependent on social cost, which is consistent with the ideas outlined in the introduction. This attains another crucial objective that doctors pursued: to consider more variables apart from weight, age, and bilirubin levels.

These profiles describe general situations: patient is definitely not to be admitted, admission is not clear, admission for observation, and so forth. They were all validated one by one, analyzing cases belonging to each profile and testing the system proposals. The explanation of the profile provided through knowledge base analysis is very useful for this task. This explanation justifies a treatment according to the values of some variables (the most relevant ones for that profile). Following this procedure, the validation is conducted on complete sets of (related) cases and not case by case. Some authors[37] emphasize that although a decision support system is primarily a computerized decision-making tool, it is important to derive simple decision rules for situations of interest. We provide this added value.

As an example, in general, the most relevant variables for the 1st treatment are gestational age, bilirubin concentration, hemoglobin concentration, multiparity, Coombs test for baby's Rh isoimmunization, Coombs test for mother's Rh isoimmunization, Coombs test for baby's group isoimmunization, baby's age, and social cost. All of these variables can be used to produce an explanation of the system proposals, according to their values. For the other decisions, the most relevant variables are the ones related to previous decisions, plus the set of variables previously listed together with the birth weight, delivery with instruments, resuscitation type, and emotional cost. Details on their values to explain each specific treatment policy might be retrieved from the system (knowledge base management layer and query layer).

Evaluations have stopped at the 1% of evaluated cases. If, for example, there were new cases of black mothers at the hospital, it would be necessary to evaluate further and integrate this new knowledge into the previous databases. As yet, this has not been necessary. Today, the system is collecting data about new cases to perform a more in-depth validation.

DISCUSSION

To the best of our knowledge, our proposal is a novel idea in this field because the current scientific evidence in the medical literature amounts to no more than a differential diagnosis of jaundiced patients using decision trees[37] and expert systems[39] or formal logic-based approaches for the jaundice protocol.[40] Instead, we use decision analysis, an accepted paradigm that integrates many forms of medical information—professional expertise and patients' preferences included—as opposed to other methods such as evidence-based medicine, against which many criticisms have been raised.[41] The influence diagram model has many valuable capabilities for guiding the clinical decision-making process[42,43]: It determines optimal treatments according to the maximum expected utility principle; it yields diagnostic information about a specific patient given his or her particular characteristics and, possibly, therapeutic decisions; and, assuming that the final results of the treatment are known, it can be used to generate probabilistic profiles that fit these final results using backward reasoning.

Clinicians work under time pressure and competing obligations, leading them to be reluctant to use both a new technology (a computerized decision support system) and a new decision strategy that might modify their established routines. Furthermore, other factors such as the organization's attitude toward innovation and the lack of leadership support may hinder the acceptance of such systems. This did not apply to our doctors, who felt they needed to include more information when making decisions, who believed there were better decision strategies, and who were willing to perform additional tasks, such as eliciting and including patients' experiences and preferences into patient care, all of which are necessary conditions to bring about a change.[44] The tradeoff between a rich model yielding accurate results and the effort that goes into its construction is a concern conveyed in recent literature.[45] As demonstrated in this article, the benefits gained and the usefulness perceived by neonatologists, both interns and residents, which is indispensable,[10] were worth the added and time-consuming effort of developing this complex decision support system.

Several of the potential weaknesses for general decision support systems could be avoided with the involvement of experts and users during the development phase. Because of this, they did not perceive IctNeo to be an unpleasant tool, neither as a threat to clinical judgment nor as an inflexible oracle that cannot be put in doubt. Doctors do not see it as promoting the overreliance on software and limiting their freedom to think. Validating the results was a challenge for them more than a limitation. They pretended, from the beginning, to gain insight into the protocol for jaundice management (in fact, this is the reason for building IctNeo). Another weakness frequently outlined in articles about medical decision support systems is that they are time-consuming tools that lead to longer clinical encounters and create extra work. This is not the case here. IctNeo is a software that is very easy to use, and it is adopted as the tool for recording clinical data about patients. Using it is not a waste of time. The style, manner of advice presentation, recommendations, and justifications are very intuitive and easy to use and learn, reducing the need for training. Finally, there is a drawback that cannot be resolved (and is common to all decision support systems). It is very difficult and time-consuming to add new knowledge or guidelines into it, which may necessitate a major redesign of the influence diagram as well as a new evaluation period for getting the system ready for use again. However, this constitutes an objective for our future research, as explained below.

In the near future, there are several options for turning decision analysis into a common tool for real-world problems. It would be important to develop tools to automate the complex tasks of acquiring and structuring knowledge.[13] Also, knowledge verification tasks are still an open problem. In this sense, it would be very useful to check the coherence between new knowledge about the problem and the rest of the model, from both the qualitative and quantitative viewpoints. This should be integrated into graphical tools, making this task easy to do and check. Finally, the ongoing progress of powerful evaluation techniques such as simulation-based methods is bringing real-time decision support systems closer to reality. However, the efforts to compute large influence diagrams should be complemented with effective methods for building models if we want this work to be of widespread use and readily transferred to society. We think we have also contributed in this way, although we plan to share our knowledge further by turning our system into a distributed Web system.

CONCLUSIONS

Influence diagrams have proven to be powerful for communicating ideas in decision-making problems. Influence diagrams represent a new dimension for decision models.[38] As a special case of Bayesian networks, influence diagrams come from research into artificial intelligence, decision analysis, and statistics. The intersection of such disciplines has produced a useful instrument for representing and solving decision-making problems.

However, the application of influence diagram methodology in practice can be extremely involved for real large-scale problems. We had to tackle difficulties related to problem structuring (e.g., existence of constraints on the sequence of decisions), knowledge acquisition (probability and utility assignment), and computational limitations. As analysts, we gained experience in this and also a very general software platform (language and compiler) to be used in modeling other problems with influence diagrams. Presently, we are working on an extracorporeal membrane oxygenation project for the same hospital.

IctNeo is definitely an important aid for doctors. Doctors, both interns and residents, gained a custom-made tool for getting a better understanding of the jaundice problem and judging and debugging the routine protocol and the implications of any possible changes. The new insights into jaundice led doctors to change their view about this problem. Treatment aggressiveness is reduced. Over the past few years, there have been no new cases of exchange transfusions. The stays at the hospital are shorter, with the resulting economic (and emotional) impacts. For the 50 cases analyzed exhaustively, about 78% of the patients would have reduced their stays at the hospital. More variables are taken into account. The validation stage, with real cases, profile studies, and explanations, confirms that the clinical problem is faithfully represented.

Decision analysis has been very useful for clarifying the decision process of expert clinicians and to help to train new doctors. IctNeo plays a role in automating medical decision making by means of which the knowledge of 1st-class experts can be made available to the rest of the health care community. The hospital has gained an easy, centralized, and structured way to store knowledge about jaundice management cases, providing guidance on what kinds of information to collect about new cases.

Health care professionals are becoming increasingly aware of the important role that patient preferences ought to play in medical decision making. In this problem, the use of multiattribute utility analysis made it possible to combine the point of view of patients, doctors, and the hospital (the output decision tables show how social cost and emotional cost are relevant variables for the decision-making process). We allowed utilities to be expressed as ranges rather than point values, and we combined subjective estimates with objective evidence.

Our automated tool covers the following aspects: representation and refinement of the influence diagram, consistency tests, data acquisition, influence diagram evaluation, presentation of system proposals and explanations, and clinical data storage and use. Unlike the heuristic judgment processes that adopt strategies that minimize the effort when confronted with complex situations, our significant effort has been worthwhile and is demonstrable through the positive results and opinions.

ACKNOWLEDGMENTS

We are grateful to Drs. M. Sánchez Luna, S. Caballero, and D. Blanco, the experts involved in the IctNeo adventure. We also wish to thank the Decision Analysis Society of the Institute for Operations Research and Management Science (INFORMS) for encouraging us to publish this work after voting us a finalist for the 2002 Decision Analysis Society Practice Award.

REFERENCES

1. Gartner LM. Neonatal jaundice. Pediatr Rev. 1995;16:22–31.

2. Mollison PL, Cutbush M. A method of measuring the severity of a series of cases of hemolytic disease of the newborn. Blood. 1951;6:777–88.

3. Rubaltelli FF, Zanardo V, Granati B. Effect of various phototherapy regimens on bilirubin decrement. Pediatrics. 1978;61: 838–41.

4. Newman TB, Maisels MJ. Evaluation and treatment of jaundice in the term infant: a kinder, gentler approach. Pediatrics. 1992;89: 809–30.

5. American Academy of Pediatrics, Subcommittee on Hyperbilirubinemia. Clinical practice guideline: management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. Pediatrics. 2004;114(1):297–316.

6. Worsley JB. Hyperbilirubinemia guidelines and unintended harms. Pediatrics. 2004;114(4):1134–5.

7. Alkalay AL, Simmons CF. Hyperbilirubinemia guidelines in newborn infants. Pediatrics. 2005;115(3):824–5.

8. Clayton P, Hripsak G. Decision support in healthcare. Int J Biomed Comput. 1995;39:59–66.

9. Ball MJ, Lillis J. E-health: transforming the physician/patient relationship. Int J Med Inform. 2001;61:1–10.

10. Ruland CM. A survey about the usefulness of computerized systems to support illness management in clinical practice. Int J Med Inform. 2004;73:797–805.

11. Howard RA, Matheson JE. Influence diagrams. In: Howard RA, Matheson JE, eds. The Principles and Applications of Decision Analysis. Vol. 2. Menlo Park (CA): Strategic Decisions Group; 1984.

12. Clemen RT. Making Hard Decisions: An Introduction to Decision Analysis. 2nd ed. Pacific Grove (CA): Duxbury; 1996.

13. Sonnenberg FA, Hagerty CG, Kulikowski CA. An architecture for knowledge-based construction of decision models. Med Decis Making. 1994;14:27–39.

14. Bielza C, Shenoy PP. A comparison of graphical techniques for asymmetric decision problems. Manage Sci. 1999;45(11):1552–69.

15. Nielsen TD, Jensen FV. Representing and solving asymmetric decision problems. Int J Inform Technol Decis Making. 2003;2(2): 217–63.

16. Demirer R, Shenoy PP. Sequential valuation networks for asymmetric decision problems. Eur J Oper Res. In press.

17. Keeney RL, Raiffa H. Decisions with Multiple Objectives. Preferences and Value Tradeoffs. 2nd ed. Cambridge (UK): Cambridge University Press; 1993.

18. van der Gaag LC, Renooij S, Witteman CLM, Aleman BMP, Taal BG. Probabilities for a probabilistic network: a case-study in oesophageal cancer. Artif Intell Med. 2002;25(2):123–48.

19. Merkhofer MW. Quantifying judgmental uncertainty: methodology, experiences, and insights. IEEE Trans Syst Man Cyber. 1987;17:741–52.

20. Shephard GG, Kirkwood CW. Managing the judgmental probability elicitation process: a case study of analyst/manager interaction. IEEE Trans Eng Manage. 1994;41:414–25.

21. Morgan MG, Henrion M. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge (UK): Cambridge University Press; 1990.

22. Jensen FV. Bayesian Networks and Decision Graphs. Berlin: Springer; 2001.

23. Díez FJ. Parameter adjustment in Bayes networks: the generalized noisy-OR gate. In: Heckerman D, Mamdani A, eds. Uncertainty in Artificial Intelligence. San Mateo (CA): Morgan Kaufmann; 1993. p 99–105.

24. Pearl J. Probabilistic Reasoning in Intelligent Systems. San Mateo (CA): Morgan Kaufmann; 1988.

25. Gómez M. IctNeo: A decision support system for newborn jaundice management [dissertation, in Spanish]. Madrid (Spain): Department of Artificial Intelligence, Technical University of Madrid; 2002.

26. Gómez M, Ríos-Insua S, Bielza C, Fernández del Pozo JA. Multiattribute utility analysis in the IctNeo system. In: Haimes YV, Steuer RE, eds. Research and Practice in Multiple Criteria Decision Making. Lecture Notes in Economics and Mathematical Systems 487. Berlin: Springer; 2000.

27. Delquié P, Luo M. A simple trade-off condition for additive multiattribute utility. Multicriteria Decis Analysis. 1997;6:248–52.

28. Farquhar PH. Utility assessment methods. Manage Sci. 1984; 30:1283–300.

29. Von Nitzsch R, Weber M. Utility function assessment on a micro-computer: an interactive procedure. Ann Oper Res. 1988; 16:149–60.

30. Ríos S, Ríos-Insua S, Ríos Insua D, Pachón JG. Experiments in robust decision making. In: Ríos S, ed. Decision Theory and Analysis: Trends and Challenges. Boston: Kluwer; 1994. p 231–42.

31. Cooper GF. The computational complexity of probabilistic inference using Bayesian belief networks. Artif Intell. 1990;42: 393–405.

32. Quaglini S, Bellazzi R, Locatelli F, Stefanelli M, Salvaneschi C. An influence diagram for assessing GVHD prophylaxis after bone marrow transplantation in children. Med Decis Making. 1994;14:223–35.

33. Fernández del Pozo JA, Bielza C, Gómez M. A list-based compact representation for large decision tables management. Eur J Oper Res. 2005;160(3):638–62.

34. Bielza C, Gómez M, Ríos-Insua S, Fernández del Pozo JA. Structural, elicitation and computational issues faced when solving large influence diagrams. Comput Oper Res. 2000;27(7–8):725–40.

35. Felli JC, Hazen GB. Sensitivity analysis and the expected value of perfect information. Med Decis Making. 1998;18:95–109.

36. Bielza C, Ríos-Insua S, Gómez M, Fernández del Pozo JA. Sensitivity analysis in IctNeo. In: Ríos Insua D, Ruggeri F, eds. Robust Bayesian Analysis, Lecture Notes in Statistics 152. Berlin: Springer; 2000.

37. Peña D. A decision analysis model for a serious medical problem. Manage Sci. 1980;26(7):707–18.

38. Helfand M, Pauker SG. Influence diagrams: a new dimension for decision models [editorial]. Med Decis Making. 1997;17(3):351–2.

39. Molino G, Marzuoli M, Molino F, et al. Validation of Icterus, a knowledge-based expert system for jaundice diagnosis. Methods Inform Med. 2000;39(4–5):311–8.

40. Marcos M, Balser M, ten Teije A, van Harmelen F. From informal knowledge to formal logic: a realistic case study in medical protocols. In: Gämez-Pírez A, Benjamins RV, eds. Lecture Notes in Artificial Intelligence 2473. Berlin: Springer; 2002.

41. Cohen AM, Stavri PZ, Hersh WR. A categorization and analysis of the criticisms of evidence-based medicine. Int J Med Inform. 2004;73:35–43.

42. Owens DK, Shachter RD, Nease RF. Representation and analysis of medical decision problems with influence diagrams. Med Decis Making. 1997;17:241–62.

43. Nease RF, Owens DK. Use of influence diagrams to structure medical decisions. Med Decis Making. 1997;17:263–75.

44. Payne JW, Bettman JR, Johnson EJ. The Adaptive Decision Maker. Cambridge (UK): Cambridge University Press; 1993.

45. Chenoweth T, Dowling KL, St. Louis RD. Convincing DSS users that complex models are worth the effort. Decis Support Syst. 2004;37:71–82.