Causal Networks for Modeling Health Technology Utilization in Intensive Care Units*

Max Chacón and Brenda Maureira

Informatic Engineering Department, University of Santiago de Chile,
Av. Ecuador 3659, PO Box 10233, Santiago, Chile
{mchacon,bmaureira}@diinf.usach.cl

Abstract. This study presents the application of Bayesian networks (Bn) to explain Neonatal Intensive Care Unit relationships. Information was compiled retrospectively from the medical records at two neonatal intensive care units of 523 neonates (63 deaths). A total of 31 variables were used for the model, eleven to characterize admission conditions and severity of illness as well as the 20 technologies. With mortality as the output variable, the K2 search algorithm and Geiger-Heckerman quality measures were used in the training that generated the Bn. Evidence propagation was used to assess the training, which yielded a sensitivity of 77.78% and a specificity of 91.30%, in the classification of mortality. Clinical criteria, correlations and logistical regression were used to analyse the relationships the model provided. The Bn found clinically coherent relationships as recognizable conditions that directly affect mortality such as congenital malformations are seen and it exposes the least effective technologies among those studied, bicarbonate treatment.

1 Introduction

Effectiveness in medical technology is essential for improving the quality of health-care [1]. When multiple medical technologies are used together, such as in premature births, Intensive Care Units (ICU), Acute Immune Deficiency Syndrome (AIDS) or multiple failure [2], it is necessary to establish the sequence (or causal order) in which the technologies are applied, to evaluate an individual technology's impact on the outcome. The limitations of using non-causal models in this area have become apparent [3-4].

Having a model that represents the causal way in which medical technologies are applied, delivers valuable clinical evidence concerning medical practices. In addition it offers the possibility of determining how changes to the order in which technologies are employed, impact the outcome, through the use of simulations.

Techniques such as decision analysis, Markov chains and path analysis have been used to describe the interaction of multiple technologies when few technologies are under scrutiny [4-5]. Such methods cannot cope with the surge in combinations when using more than five technologies, and do not have sufficiently efficient pruning mechanisms to be able to derive an easily interpretable model.

The current work explores the use of Bayesian networks (Bn) as an alternative to modelling the causal relationships in the utilization of multiple technologies. We have chosen a typical problem that requires the use of multiple technologies; Neonatal

---

Intensive Care Units (NICU). These units have proven their effectiveness at reducing neonatal mortality. Yet the continual rise in the introduction of new technologies in such units has not been adequately addressed. Little is known about how these technologies are used in daily medical practices, particularly in developing countries.

In medicine most Bayesian network applications centre on automatic diagnosis and in aiding treatment [6-7]. However, no Bayesian network applications are known of that assess health technologies. The closest study to the current work is that of Sierra et al [8], in which Bayesian networks are applied to intensive care units (ICU) aiming to improve the classification of patients within a unit, but the work does examine the use of ICU medical technologies.

In ICU there is no clear indication of the causal order in which the technologies are applied within the ICU; currently it is only possible to recognize the initial variables (admission conditions and the severity of the ailment) and the network's output variables (such as morbidity or mortality). Our proposal is to use structural learning for the Bn to determine just how medical technologies are used in the unit and how these relate to mortality.

2   Methods

2.1   Data Collection and Pre-processing

Data was collected retrospectively, from the medical records of the Fernandes Figueira Institute and the Pediatric Center of Lagoa, two neonatal intensive care units (NICU) in Rio de Janeiro, Brazil. A complete list of the variables collected is given in Table 1, which corresponds to health conditions at birth or admission, diagnostic hypotheses, and technology utilization. While the list is not exhaustive, the technologies selected, including both diagnostic and therapeutic means, were representative of the NICU armamentarium of the early 90s. Technologies used in nearly 100 % of the neonates, such as incubators and sedation, were not analysed. Cases involving internment over 45 days were also not included.

Initially, 78 different diagnostic hypotheses were found and these were grouped into the eight main diagnostic categories listed in Table 1. A consultant neonatologist who classified the neonate's condition as not affected, mild or severe for each diagnostic category reviewed each medical record. For the category congenital malformations, the number of different malformations was added up, assuming mild = 1 and severe = 2. These eight diagnostic variables, along with birth weight, gestational age, and the Apgar score (5th min), were used as a proxy for severity of illness.

Diagnostic methods and clinical procedures such as blood transfusions are examples of discrete technologies. The measure of the intensity of usage of these technologies was taken as the number of applications of the specific technology to each neonate. On the other hand, technologies that tend to be used on a continuous basis, such as drugs or mechanical ventilation, had their "dosage" measured by the total length of their application in days (Table 1). The day each treatment began was also recorded.

The Bayesian network adopted in this study requires the variables the model comprises to be represented by discrete values, which would later be transformed into binary values. For *birth weight* three ranges are used: low, normal and overweight; four states are defined for *gestational age* considering morphological criteria and foetal maturity. The 10 values for the *Apgar score* are grouped into four states. Diagnostic hypotheses are coded directly in three states; absent = 0, medium = 1 and se-

vere = 2, apart from congenital malformations for which six values are used, and correspond to the maximum value of the sum of the ailments found in the database. Medical technology usage (diagnostic and therapeutic) can be represented directly with binary values, taking 0 as not having used the technology and 1 as having used the technology, irrespective of how extensively it is employed. Finally, the output variable, mortality, considers 0 as surviving and 1 as dead.

Table 1. List of input variables and regression coefficients whit mortality.

Variable	Mean ± SD	Regress. Coeffi.	Variable	Mean ± SD	Regress. Coeffi.
Birth/admission weight (Weight)	2321 ± 865 [grams]	-11,81 *	Electrolytes (Elect)	2.20 ± 2.77 [events]†	15,76 *
Gestational age (GAge)	36 ± 3 [weeks]	-4,45	White blood cell count (WBCC)	2.42 ± 1.98 events†	-0,15
Apgar 5th min. (Apgar)	7.2 ± 2.2	-6,68 *	Antibiotics (Antib)	13.5 ± 13.4 [days]★	-6,70
Congenital malformations (ConMa)	0.19 ± 0.72 [sum]	17,27 *	Inotropic agents (InoAg)	0.97 ± 4.22 [days]★	10,54 *
Obstetrical conditions (ObsCo)	0.02 ± 0.17 [class]#	-2,44	Diuretics (Diure)	1.89 ± 7.54 days★	-2,03
Respiratory conditions (ResCo)	0.82 ± 0.68 [class]#	-0,29	Anticonvulsive therapy (Antic)	1.63 ± 6.95 [days]★	7,89
Asphyxia (Asphy)	0.36 ± 0.64 [class]#	1,22	Sodium bicarbonate (SodBi)	0.29 ± 1.12 [days]★	26,16 *
Cardiovascular conditions (CarCo)	0.03 ± 0.23 [class]#	-5,66 *	CPAP (CPAP)	0.71 ± 1.81 [days]★	-1,64
Haematologic conditions (HaeCo)	0.16 ± 0.46 [class]#	0,91	Oxygen therapy (oxihood) (Oxiho)	0.79 ± 1.61 [days]★	-24,58 *
Infections (Infec)	0.19 ± 0.59 [class]#	-4,82 *	Mechanical ventilation (MecVe)	0.91 ± 3.10 [days]★	-3,80
Metabolic conditions (MetCo)	0.07 ± 0.28 [class]#	-3,11	Phototherapy (Photo)	2.18 ± 2.94 [days]★	-3,38
Blood gases (BloGa)	4.49 ± 3.30 [events]†	4,44	Intravenous solutions (InSol)	4.85 ± 5.30 [days]★	4,55
X rays (XRays)	2.76 ± 8.40 [events]†	6,92	Parenteral nutrition (PaNut)	1.75 ± 4.54 [days]★	0,28
Ultrasound (UlSou)	0.79 ± 1.09 [events]†	-6,59	Packed red cells transfusion (PRCT)	0.52 ± 1.69 [events]†	14,39 *
Microhaematocrit (MiHem)	8.55 ± 8.72 [events]†	-18,77 *	Exchange transfusion (ExTra)	0.08 ± 0.38 [events]†	6,98
Blood culture (BloCu)	2.05 ± 1.71 [events]†	0,26			

Note: 7 diagnostic categories have been coded as 14 binary variables thus increasing the total number of input variables to 31.

classified as absent = 0, mild = 1, or severe = 2.
† *number of applications of technology to each neonate.*
★ *total number of days of use in each neonate. *Significant coefficients at 5%.*

2.2   Bayesian Network Analysis

A Bayesian network is a directed acyclical graph whose nodes represent stochastic variables and the arc the conditioned dependencies between two related variables. The procedure for modelling a problem using Bn can be divided into two steps, learning and the propagation of evidence.

The learning process on which this work focuses involves two principal tasks: induction of a graphical model that best represents the problem at hand (structural learning), and the extraction of conditional probabilities that define a graphical model's dependencies (parametrical learning).

Learning, based on a database, is two-part process. First, a search algorithm tries to determine the possible parents of each node (conditional dependence of a precedent variable) and, second, a measure (or metric) of quality. The measure of quality calculates how successful the network is from the data. For its part, the search algorithm attempts to identify the structure offering the best measure of quality. As a rule, this is an NP-Complete problem and certain heuristic approximations are used to restrict the search space and so avoid the combinatorial explosion.

Structural learning is needed to find the order of technology application. We employed the most commonly used search algorithm at present for discrete variables, called K2 [8-10]. This algorithm searches for a network structure on the basis of discrete data, considering that the input variables are independent and that these are initially ordered. Wit this restriction there are $2^{n(n-1)/2}$ possible structures for representing a problem with n variables. The basic idea is to use a measure of quality to evaluate the quality of the network formed. In initial stages all the nodes lack parents, but in later stages those nodes that maximize the value of the measure of quality are added to the parental set. The process stops when addition of new nodes results in no improvement of quality or when the maximum number of parents u has been reached, this number being fixed for a particular problem. The complexity of the worst case for this algorithm is O(n4), when the maximum number of parents u is equal to the number of variables n.

Two different measures of quality were tried, the first was obtained from the Geiger-Heckerman measure (G-H) while the second is known as the Cooper-Herskovits measure (C-H) [11-12].

All of the algorithms used in this work were executed in release 12.1 of MATLAB software version 6.0 for the PC.

2.3   Initial Order and Evaluation

One of the K2 algorithm's requirements is that the variables that will comprise the Bayesian network are ordered first. Normally, expert opinion is sought when undertaking this initial ordering. For our purposes though, it was possible to employ the distribution of the first application of technologies during a patient's internment in the NICU as our criteria. To quantify these values we considered the first day of application of each technology, as the point when the technology had been applied in at least 80% of cases.

The dependability of network learning has to be determined when choosing the best network and to compare the measures of quality used. Thus, it was necessary to run a group of cases independently of the training group and to assess the results. Evidence propagation was performed using an exact method of general propagation,

known as "clustering" [10]. Training was carried out using 85% of the cases in the training set (445 cases) and 15%, selected at random, for the test set (78, 69 survivors and 9 deaths).

The general lack of work presenting the relationships of how neonatal intensive care units are used and the clinical knowledge required of such units makes quantitative means of comparison necessary. For such purposes, first the correlations of 23 real variables from the model are calculated (Admission weight, Gestational age, Apgar and the 20 technologies) followed by a logistical regression of the 31 variables shown in Table 1, using mortality as the output variable.

3   Results

Data were collected on 523 cases corresponding to 268 cases (48 deaths) from the Fernandes Figueira Institute and 255 (15 deaths) from the Paediatric Centre of Lagoa. The mean ± SD of the admission weight was 2320 ± 865 g (range 670-5680 g) and the gestational age was 36 ± 3 weeks (range 24-43 weeks). Table 1 gives the distribution of the other variables.

Training began with an initial ordering of the variables. The first 11 (admission conditions), used to approximate the severity of illness, are the primary candidates, as they are obtained at the point of admission to the unit. The technologies were then ordered, as described, by date of application. Table 2 shows the initial ordering, considering the day most of the technologies had been applied.

The process of generating the Bayesian network took a maximum of u=10 parents for each node as the training parameter. Two networks were then obtained, one for each measure of quality with the training group (445 cases). Evidence propagation was then applied to the remaining 15% of cases (78 cases) to the two models. The results of the classification for the test group, expressed by sensitivity and specificity are presented in Table 3 for both measures of quality used.

While the two network structures are very similar, sensitivity and specificity analysis shown in Table 3 differ in that the G-H measure is more sensitive to the detection of deaths, and provides us the best tool for classifying the survival of a patient.

Table 2. Initial order of application of the technologies.

Technology	*0Day*	Technology	*1Day*
Intravenous solutions	2	Mechanical ventilation	4
Blood culture	2	Exchange transfusion	5
Blood gases	2	Diuretics	6
Microhaematocrit	2	CPAP	6
X rays	2	Parenteral nutrition	7
White blood cell count	2	Anticonvulsive therapy	7
Antibiotics	2	Inotropic agents	7
Electrolytes	3	Sodium bicarbonate	7
Phototherapy	4	Ultrasound	8
Oxygen therapy	4	Packed red cells transfusion	8

Table 3. Results of the classification when applying propagation of the evidence with 15% of the test group cases, for the two quality measures.

Measure	Sensitibity	Specifity
Cooper-Herskovits	55.56%	92.75%
Geiger-Heckerman	77.78%	91.30%

For the comparative analysis the correlations are doing for the 12-variable subgroup (from a total of 23 real variables). Results of the logistical regression on mortality are presented in Table 1.

Structural learning algorithms provided results ordering the parent nodes for each variable, using the probability of adding a new parent node and the probability once the node is added. For easier graphical interpretation we preferred to show a graph, simply showing the order in which each parent of the variable was chosen in the arc (between the parent and child). The graph for the network obtained from the G-H measures is shown as Figure 1.

4   Discussion

We will centre analysis of the results on comparison with basic existing neonatological relationships, correlations between the real variables of the studied data set, and the logistical regression of mortality (Table 1).

The first significant relationship provided by the chosen network (Figure 1) is that between *birth weight* and *gestational age*. This is well known in neonatology, since both variables are representative of immaturity and a high correlation exists between the two (r=0.75).

Asphyxia dependency on the *Apgar score* is accounted for, as the states of *asphyxia* are required as a component in calculating the *Apgar* points. The path that flows from *asphyxia* up to *anticonvulsive therapy* is also clearly explainable, as severe *asphyxia* induces convulsions.

The exchange transfusion dependence on haematologic conditions and birth weight is explained in the knowledge that this therapeutic technology is used in situations of incompatible blood with the mother, low birth weight and in cases of underdeveloped digestive systems that cause jaundice, which then requires a full blood transfusion.

One group of relationships that stand out include variables belonging to respiratory disorders.

Respiratory conditions as the parent of blood gases and x rays, which are the parents of mechanical ventilators and oxygen therapy, complete a classical cycle in medicine: problem (respiratory disorders), diagnosis (blood gases and x rays) and therapy (mechanical ventilators and oxygen therapy). All dependencies are confirmed by the correlations up 0.63. It is also possible to follow a route for cases in which respiratory infection is suspected that would require diagnosis using x rays, a dearer technology than blood gases, followed by a course of antibiotics to treat the infection. It is also worth noting that while CPAP ventilation therapy is applied after oxygen therapy, the former presents no relationship with mechanical ventilation.

The relationship between microhematocrit and antibiotics (r=0.97), and between antibiotics and electrolytes (r = 0.88) present the highest correlations. It is possible to
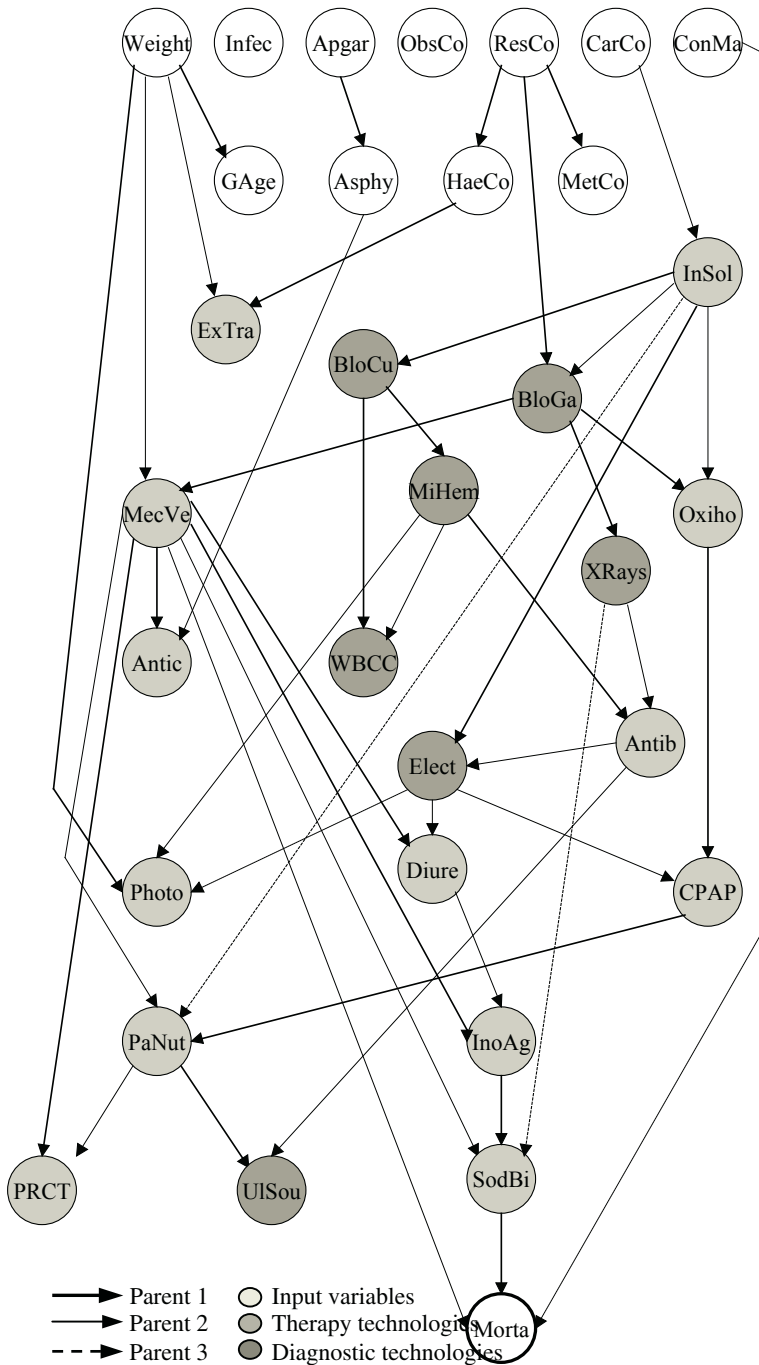
Fig. 1. Bayesian network with the Geiger-Heckerman measure.

explain this as both microhematocrits and electrolytes (r=0.8) are employed as antibiotic control technologies. Yet the Bayesian network affords further information on indicating that the microhematocrit is conducted before the use of antibiotics, while the electrolytes is applied afterwards. On examining the network however, it is possible to conclude that this relationship is, in reality, indirect, as it arises from the relationship caused by using antibiotics.

The only admission condition variables that are directly linked to mortality are congenital malformations, which is confirmed by high regression coefficient values (Table 1). The lack of technology intervention in this relationship indicates the effectiveness of technologies used in such disorders. In a significant number of cases of congenital malformations, and especially in more severe cases, medical intervention contributes little to patient recovery.

Various admission conditions, such as low birth weight, respiratory or cardiovascular conditions lead to the use of bicarbonate following the application of several diagnostic and therapeutic technologies. The relationship here with mortality is unequivocal and direct. Logistical regression sets bicarbonate as the variable most closely related with mortality. The clinical explanation for this is that this treatment technology is applied to compensate for acidosis of the blood arising from respiratory insufficiency. Use of bicarbonate treats the symptoms of the respiratory problem, but does not manage to revert the demise of the patient. The lower relationship between mechanical ventilators and mortality and the existence of multiple paths in the Bayesian network, reveal that this technology is more effective than bicarbonate at altering the initial morbid outlook.

A general comparison between the logistical regression model and Bayesian network of Figure 1, clearly shows a regression model's limitations at representing such problems, since the regressions solely manage to identify variables that are directly and only linked to the outcome (mortality), but do not manage to represent the series of causal relationships that exist in the NICU before the results are known.

5   Conclusions

The network shown in Figure 1 enables people to obtain and to interpret the relationships clearly between admission conditions, usage of technology and outcome, in forming a better idea than is possible with conventional statistical methods of a complex usage of technologies in medical establishments such as in ICUs. It is also possible to trace out an idea of the effectiveness of certain technologies with regard to others, on comparing their direct or indirect association with the results.

A method is surely needed though to quantify the impact of the technologies upon results; a direct measure of effectiveness. This would require designing a method that enables the simulation of amendments to the application of a technology in the Bayesian network, and to establish the impact of such modifications on the results. A simple way of generating such a method would be to eliminate arcs or nodes in the trained Bayesian network and then use a propagation of evidence that would allow the propagation of cases in networks with different structures to those for which they were trained. This would make it possible to quantify the simulation of changes representing the non-deployment or different ways of employing a particular technology to a group of patients.

References

1. Evans, RW Health care technology and the inevitability of resource allocation and rationing decisions, JAMA. 249 (1983) 2047-2053.
2. Slonim, AD; Patel, KM; Ruttimann, UE & Pollack, MM The impact of prematurity: A perspective of pediatric intensive care units, Crit. Care Med. 28 (2000) 848-853.
3. Almeida, RT; Panerai, RB; Carvalho, M & Lopes, JMA. Analysis of multiple technologies in neonatal care, Int. J. Techol. Assess. Health Care. 7 (1991) 22-29.
4. Panerai, RB; Chacón-Pacheco, ML & Almeida, RT Path Analysis in Health Technology Assessment, International Society of Technology Assessment in Health Care, Seventh Ann. Meet., Helsinki, Finland, (1991).
5. Panerai, RB; Almeida, RT; Portela, MC; Carvalho, M; Coura-Filho, M & Costa,TP Estimating the effectiveness of perinatal care technologies by expert opinion, Int. J. Techol. Assess. Health Care. 7 (1991) 367-378.
6. Kahn, CE Jr; Roberts, LM; Shaffer, KA & Haddawy, P. Construction of Bayesian Network for mammographic diagnosis of breast cancer. Comput. Biol. Med. 27(1) (1997) 19-29.
7. Sierra, B & Larranaga, P. Predicting survival in malignant skin melanoma using Bayesian networks automatically induced by genetic algorithms. An empirical comparison between different approaches. Artif. Intell. Med. 14(1-2) (1998) 215-230.
8. Sierra. B; Serrano, N; Larranaga. P; Plasencia, EJ; Inza, I; Jimenez, JJ; Revuelta, P & Mora, ML. Using Bayesian networks in the construction of a bi-level multi-classifier. A case study using intensive care unit patients data. Artif. Intell. Med. 22(3) (2001) 233-248.
9. Heckerman, D Bayesian networks for knowledge discovery, In: *Advances in Knowledge Discovery and Data Mining*, Fayyad, UM.; Piatesky-Shapiro, G; Padhraic, S. and Uthurusamy, R. (Eds.), Mit Press, Melo Park, CA. 273-305, (1996).
10. Castillo E; Gutiérrez, JM & Hadi, AS. Expert systems and probabilistic network models. Springer-Verlag, New York, (1997).
11. Gaiger, D & Heckerman, D. A characterization of the Dirichlet distribution with application to learning Bayesian networks. *In: proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Pub., San Francisco, CA, (1995) 196-207.
12. Cooper GF & y E. Herskovits E A Bayesian Method for the induction of probabilistic networks from data. Machine Learning. 9 (1992) 309-347.