2000

The eBabies Project: Integrated Data Monitoring and Decision Making in Neonatal Intensive Care

Raymond Lister
University of Western Sydney, r.lister@uws.edu.au

George Bryan
University of Western Sydney, g.bryan@uws.edu.au

Mark Trac
Nepean Hospital, mtracy@ozemail.com.au

The e-babies Project; Integrated Data Monitoring and Decision Making in Neo-natal Intensive Care

Raymond Lister
Lecture
School Of Computing
and Information Technology
University of Western Sydne
PO Box 10, Kingswood,
NSW Australia
r.lister@uws.edu.au

George Bryan
Senior Lecturer
School Of Computing
and Information Technology
University of Western Sydne
PO Box 10, Kingswood, 2747
NSW Australia
g.bryan@uws.edu.au

Mark Trac
Consultant Neonatologist
Neonatal Intensive Care Unit
Nepean Hospital
Penrith
NSW Australia
mtracy@ozemail.com.au

Abstract- **Approximately 5% of newborn babies in Australia require intensive care admission after birth Many of these babies are very premature and suffer severe illness. Of those babies admitted 11% will die. Those who survive may have significant life long disability such as cerebral palsy, mental retardation and blindness. Nepean hospital located west of Sydney is one of eight tertiary neo-natal referral centers in the state of New South Wales. Tertiary neo-natal referral centers care of critically ill pre-term newborns. The Nepean center typically handles 700 babies annually. Several electronic instruments monitor baby's vital signs such as blood oxygen, blood pressure and heart rate. A major limitation in clinical management is that these monitors are not linked to provide an integrated picture of the baby's condition. This paper will describe the authors work on the "e-babies" project. The core objective of this project is to integrate this physiological data obtained from the monitors with a clinical database and provide tools for the analysis of this data.**

Introduction

In recent years, there has been an improvement in the mortality rates and absolute number of babies surviving in the Extremely Low Birth Weight (ELBW) categories. The reasons for this improvement include better obstetric care, transferring the birth and postnatal care of ELBW infants to centers of excellence (tertiary referral centers)[1,2], ne medications (e.g. surfactant replacement) and new technologies such as mechanical ventilators that offe synchronized and high frequency ventilation.

The Victorian Infant Collaborative group has studied a cohort of baby's born < 28 weeks gestational age in 1991-92. The survival rate of baby's born at 24 to 26 weeks gestation was 58%. At two years of age, the rate of blindness was 2.3%[3]. Studies in other countries show a similar finding [4]. The rates of severe sensorineural disability (cerebral palsy, mental retardation and blindness) was 6.8%[5,6]. The Victorian study group has followed the cohort to 5 years and found that all levels of sensorineural the incidence is 39.4%.

Given the cost of neonatal intensive care per patient and also the cost to society of severely disabled survivors, a number of scoring systems have been developed to identify independent variables that aid in the clinical management and predicted outcome of sick newborns. These scoring systems are known as The Clinical Risk Index for Babies (CRIB), and the Score for Neonatal Acute Physiology (SNAP) [7].

Both CRIB and SNAP suffer from the fault that the depend on the hand-annotated records of nursing staff, usually at 60 minute intervals, and 30 minute intervals at best. It is very common for critically ill babies to have significantl abnormal variation in the measured parameters minute by minute that are not recorded in the notes. Frequent transient falls in blood pressure and blood oxygen content, often with swings into the high range, may be of critical importance in survival and quality of survival free of significant disability. This issue has not been tested, but will be in this project.

Since the CRIB and SNAP score were developed other substantial advances in neonatal care have occurred. These include lung surfactant replacement treatment and micro sampling techniques for lactate measurements. This project will make use of these more recent advances.

The network used by the e-babies project, will enable clinicians to access real time data from all the existing separate systems. By capturing the electronic monitor data continuously, including new physiological variables such as dynamic compliance and blood lactate, and then integrating it with clinical data, we will be able to improve upon CRIB and SNAP. This project would not be possible without a sophisticated computer network to connect the different monitors, each monitor with separate data transmission protocols. The volume of raw data to be stored and analyzed is moderately large, with approximately 10 megabytes per baby per day

Methods:

The various physiological and clinical data elements will be integrated into a data server in real time. The researchers will enable the clinicians to data mine and analyze patterns with several methods. Primarily the research will utilise Bayesian analysis for the CRIB scoring and expansions of that score with more variables.

A. *Physiological Data to be Collected:*

Ventilators for set & delivered Peak Inspiratory Pressure, set & delivered mean airway Pressure, set & delivered Positive expiratory Pressure, Inspiratory time, rate, mode of ventilation eg trigger or high frequency ventilation (HFV), amplitude (HFV only), Hertz (HFV only), $Fi0_2$ fractional inspired oxygen concentration, tidal and minute volumes,

derived parameters of dynamic compliance and resistance, and-pressure/volume and pressure/flow loops

Hewlett Packard monitors for heart rate, transcutaneous oxygen saturation, invasive blood pressure, respiratory rate, and significant pauses in breathing (apnoeas) and temperatur

Dedicated pulse oxymetry monitors for transcutaneous oxygen saturation, heart rate and transcutaneous oxygen saturation

Dedicated pulmonary mechanics monitor (Ventrak) fo delivered Peak Inspiratory Pressure, delivered mean airwa Pressure, delivered Positive expiratory Pressure, Inspiratory time, rate, tidal and minute volumes, derived parameters of dynamic compliance and resistance, pressure/volume and pressure/flow loops

Arterial blood gas & Lactate measurements from blood gas machine integrate with clinical database

B. Clinical scoring factors:

Continuous electronic collection of the physiological variables stated above will be controlled for sex, gestational age, antenatal steroid use and degree of perinatal asphyxia. Various models combining these variables will be tested against the outcome measures.

The first model will combine the Oxygenation index (mean airway pressure $XFi0_2X100$/arterial 0_2 level) with blood lactate and dynamic compliance. Since the data is being collected continuously, the model can be tested at different times. eg 2 hrs, 4 hrs, 6 hrs, 12hrs etc.

Other variables will be added to the model a) hypotension (duration) b) hypoxia: percentage of time spent (duration) of discrete histogram analysis of transcutaneous oxygen saturation. and c) tidal volume corrected for birth weight.

C. Primary Clinical Outcome Measures:

1) Incidence of hospital death.

2) Chronic lung disease Y/N defined as 0_2 dependence at 28 days of age+abnormal CXR and as 0_2 dependence at 36 weeks corrected gestational age.

3) Intraventricular haemorrhage (severe forms highly likely to cause lasting severe brain damage eg cerebral palsy, mental retardation or blindness).

4) Periventricular leucomalacia (severe forms highly likely to cause lasting severe brain damage eg cerebral palsy, mental retardation or blindness).

5) Rentinopathy of prematurit . (mild to severe visual impairment “blindness”).

D. Statistical Analysis:

In this paper we will confine our discussion of statistical analysis tools to Bayesian Reasoning although it is envisaged our system will support other analytical tools. Bayesian Reasoning is a general statistical approach invented several hundred years ago by the Revered Thomas Bayes. It obtain an estimate of the posterior probabilit P(H) of an hypothesis H by combining the probabilities of one or more pieces of evidence e(i). "Naive" Bayesian Reasoning is the simplest form, where it is assumed that the e(i) are conditionally independent. Despite this assumption, the technique enjoys success in many fields, including medical applications [2, 3].

E. The Machine/Network Architecture:

The real-time data will be accrued at the rate of approximately 10 MB/Baby /Day. The collection and processing of this quantity of data will be addressed in three phases:

A Data Integrator (Figure 1) will collect the data from each of the monitoring instruments at a baby’s crib.

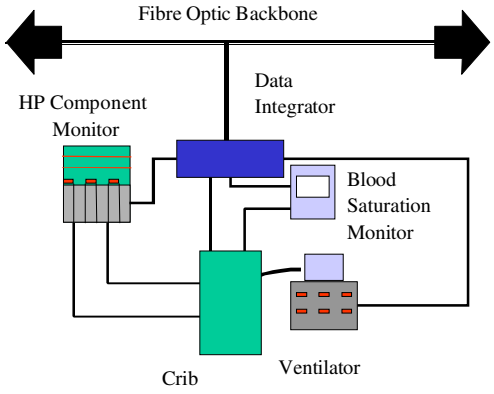

Figure 1, A Data Integrator

Each data integrator will forward its data onto the Local Data Collection PC. To provide the system with a degree of redundancy, the Data Integrator will be capable of storing a limited amount of data on its local disk unit. However, the principle data repository within the Neo-natal Intensive Care (NIC) Unit will be the Local Data Collection PC. This P will be capable of storing the physiological and clinical data on each of the babies in the unit.

The Data Collection PC will then forward its data onto a High Performance Data Server located at UWS, Nepean. Currently we have implemented a prototype to show the feasibility of accessing real time data from the existing separate physiological data collecting devices.

Naive Bayesian Reasoning Trial

The rest of this paper will discuss a trail of Naïve Bayesian Reasoning using comparing its performance with. The rest of this paper, will present a new method for developing risk scoring system for babies, based upon Naive Bayesian Reasoning. This approach is amenable to computing on a desktop PC, can be applied to both large and small datasets, and is more transparent than the current method Clinical Risk Index for Babies (CRIB). We present results for a small dataset which are superior in clinical accuracy to CRIB We used a dataset generated at the Nepean Hospital NIC, from babies treated between January 1995 and December 1997. It contains records of 287 ill preterm babies. Of these, 30 babies died.

F. Clinical Risk Index for Babies (CRIB)

The need for a performance measure is not confined to NSW. There is in fact a world wide need. An organization

known as the "International Neonatal Network" developed the CRIB system, for assigning a risk index to individual babies [10]. The CRIB scoring system assigns a positive number to a baby; the higher the score the more serious is the babies condition. CRIB assesses a baby across 6 parameters, based on routine data collected within 12 hours of birth. Tw of the parameters are birth weight, and weeks of gestation. A baby born after 24 weeks of gestation does not score any points under this criterion, but a baby born earlier incurs a single penalty point. (A baby born before 24 weeks is essentially a foetus, with very little chance of survival.) Birth weight in grams is broken into four categories, with penalty points as follows:

>1350 grams	scores 0 penalty points
851-1350 grams	scores 1 penalty points
701-850 grams	scores 4 penalty points
<= 700 grams	scores 7 penalty points

The point scoring system was arrived at by an elaborate statistical procedure that used logistic regression, and which requires sizable quantities of data. The CPU requirements fo this procedure are very high. Furthermore, a reasonabl sophisticated background in statistics is required to perform the process. These factors have inhibited further refinement of the CRIB system. Since t e CRIB system was developed, substantial advances in neonatal care have occurred. These include lung surfactant replacement treatment and micro sampling techniques for lactate measurements. These advances have not be integrated into CRIB. Even if a group, such as the International Neonatal Network, regularly updated CRIB, they could only incorporate parameters that have become standard. Individual NIC units, with their own augmentations to standard practice, would still be unable to apply CRIB to dividing babies into groups for clinical trials. Accuracy is obviously the paramount consideration when evaluating the utility of a predictive system, but individual clinicians require a scoring system that is also 1) easy to update, using a Personal Computer, and 2) applicable to the small data sets available in individual NIC units. The CRIB approach does not meet these extra two criteria. A further criterion is transparency. Clinicians lack the high level of statistical training required to perform the process by which the CRIB penalty numbers are derived.

G. Naive Bayesian Reasoning

Although the particular Naive Bayesian approach used in this study is based upon the technique in the textbook by Shinghal [11], the technique used in this study is substantially the same as that described in many textbooks. In this section, we illustrate the process of Naive Bayesian Reasoning by describing its application to a very simple example problem. In our example, there are three pieces of Boolean evidence. The relationship between the evidence and the hypothesis is defined by the following table:

e(1)	e(2)	e(3)	
false	false	false	false
false	false	true	false
false	true	false	false
false	true	true	true
true	false	false	false
true	false	true	true
true	true	false	true

Inspection of the table reveals that the hypothesis is true when two of the three pieces of evidence are true.

Before developing a model for this simple example, w need to introduce the concept of odds. While input and output is often expressed in terms of probabilities, calculations within Naive Bayesian Reasoning are performed in terms of odds. Any probability "p" has an equivalent odds "o" given by the following formula:

$o = p / (1-p)$ which rearranges to $p = o(1+o)$

A probability of zero maps to odds of 0, but a probability of 1 maps to odds of infinity.

A. Developing a Model

In our example, the prior probability of the hypothesis i P(prior)=3/7, as the hypothesis is true in three of the seven cases. The equivalent prior odds is

$O(prior) = (3/7) / (1 - 3/7) = (3/7) / (4/7) = 3/4$

The conditional probability of the hypothesis when e(1) is true is P(H|e(1)=true) = 2/3, since H is true in two of the three cases where e(1) is true. The equivalent conditional odds is:

$O(H|e(1)=true) = (2/3) / (1 - 2/3) = (2/3) / (1/3) = 2.$

The probability of the hypothesis when e(1) is false is

P(H|e(1)=false) = 1/4.

The equivalent conditional odds is:

$O(H|e(1)=false) = (1/4) / (1 - 1/4) = (1/4) / (3/4) = 1/3.$

The various conditional probabilities and odds fo e(2) and e(3) are the same as their analogous e(1) value.

Each of the conditional odds values is then used to calculate the effective influence measure M(i,v), for that parameter value, using the formula:

$$M(i,v) = O(H|e(i)=v) / O(prior)$$

and thus:

$$M(1,e(1)=true) = O(H|e(1)=true) / O(prior)$$
$$= 2 / (3/4)$$
$$= 8/3$$
$$M(1,e(1)=false) = O(H|e(1)=false) / O(prior)$$
$$= 1/3 / (3/4)$$
$$= 4/$$

The various influence measures fo e(2) and e(3) are the same as their analogous e(1) value.

Making a Prediction

We now show how the model developed above can be used to predict the probability of the hypothesis P(H) for a previously unseen combination of the evidence. Consider the case when all three pieces of evidence are true.The posterio odds is then:

O(H) = Odds(prior) * M(e1,true) * M(e2,true) * M(e3),true)

= 3/4 * 8/3 * 8/3 * 8/3

= 128/9

Which is equivalent to a posterior probability of 128/137. In general, if there are N pieces of evidence, where each piece of evidence has a value vi, then the posterior odds are:

O(H) = Odds(prior) * M(e1,v1) * M(e2,v2) * ... * M(en,vn) ... (1)

This simple example may give the impression that Naive Bayesian Reasoning can only be applied in cases where the evidence takes on Boolean values. That is not the case. The calculations in this section extend to evidence that takes on three or more discrete values. Numeric evidence can treated by categorizing it into several subranges, in the same way that CRIB breaks up birth weight and other numeric parameters.

H. Approach

Our aim was to perform the most direct comparison of CRIB and our Naïve Bayesian system. Consequently, our system used the same six parameters used in CRIB. Furthermore, our system broke the four numeric parameters, including birth weight, into the same subranges used b CRIB. We used the common leave-one-out methodology. That is, we set aside the record from one baby, build a Bayesian model, and then assess the performance of that model on that one baby. This process is repeated for each baby. Thus 287 Bayesian models were generated, and each model was assessed with a datum not used to develop the model. Leave-one-out is well suited to small datasets.

I. Implementation and results

The Naive Bayesian System was programmed in Java, was run in the slow JDK environment on a small UNIX machine, but produced the 287 Bayesian models in under 5 seconds. We compared the accuracy of our system to CRIB. We used the standard CRIB penalty points as given in [10]. The babies in the Nepean hospital dataset scored CRIB values between 0 and 17. The Bayesian posterior probabilities fo the same dataset varied, of course, from 0 to 1. Given some arbitrary threshold value, the numeric output value from either CRIB or the Bayesian system may be interpreted as supporting the hypothesis if the numeric output exceeds the threshold. If the threshold value is varied, then the interpretation of a given fixed numeric output may change. In clinical medicine, a Receiver Operating Characteristic (ROC) curve is widely used for assessing the performance of a predictive system, for predictive systems that produce a numeric measure of the likelihood of an hypothesis. (ROC was used in [10] to assess CRIB.) This technique generates a curve within the unit square of the Euclidean plane. The axes of the plane record the specificity and sensitivity of a predictive system, for a given threshold value. The curve of any predictive system runs between (0,0) and (1,1). Varying the threshold value generates it. A random choice ("coin toss") model generates a curve, which is the line between those two points. The area under that curve is 0.5. A perfect predictor would generate a curve that followed the left and topsides of the unit square (i.e. from (0,0) to (0,1) to (1,1)). The area under that curve is 1. A real predictor generates a curve somewhere between these two extremes. The area under the curve is frequently quoted as the measure of the quality of the predictor.

With the Nepea dataset, CRIB produced an area under th ROC curve of 0.85 and a 95% confidence interval of that area between 0.77 to 0.92. That figure is consistent with the performance of CRIB in [10]. Our Naïve Bayesian approach produced an area under the ROC curve of 0.96 with a 95% confidence interval of that area between 0.93 to 0.98. Th performance difference between CRIB and our Bayesian model is significant with p=0.006.

We have generated a scatter plot of CRIB score versus Bayesian Posterior probability, for all the babies used in the experiment. The CRIB system generates a distribution, which is relatively uniform: many babies score an ambiguous intermediate value. In contrast, Naive Bayesian Reasoning produces a distribution, which is relatively bimodal: the system gives a less ambiguous classification.

For the dataset used in this study, Naive Bayesian Reasoning generated a model that is a better predictor of mortality than CRIB. The CPU requirements to generate a Naive Bayesian model are much less than CRIB. The CRI penalty numbers were derived from a dataset containing over 800 babies. Given the number of free variables in the logistic regression approach used to develop CRIB, it would be unwise to derive a set of CRIB penalty numbers, by logistic regression, on the much smaller dataset used in this study. In contrast Naive Bayesian Reasoning has a small number of free variables (the prior odds, and the influence measures), rendering it suitable for use on small datasets, which are the norm in clinical medicine. By taking the logarithm of both sides, equation (1) takes the following form:

$$\log O(H) = \log \text{Odds(Prior)} + \text{Sigma}^{N}_{i=1} (\log M(ei,vi)$$

This form demonstrates the similarity between CRIB and Naive Bayesian

Reasoning, if the CRIB penalty numbers are interpreted as the logarithm of likelihood measures: both approaches sum the logarithms of likelihood. Furthermore, this form highlights two statistical deficiencies of CRIB, compared to Naive Bayesian Reasoning. The first deficiency is that the CRIB approach does not account for the prior odds of death. The second deficiency is apparent after noting that all the penalty scores used by CRIB are positive. It follows (using the logarithm of likelihood interpretation) that CRIB onl treats evidence that increases the likelihood of the hypothesis.

Future Work

Following the success of this study, which only used data collected at Nepean hospital, we plan to intergrate a Bayesian Reasoning tool into our data analysis system.

We also plan to extend our Bayesian models to include the non-CRIB parameters that are of interest to Nepean hospital.

Progress in neonatal clinical practice is hampered by a paucity of data In an intensive care unit, nursing staff can only manually record data sporadically; perhaps every 15 minutes for each baby, often only once an hour. Clinicians agree that important information, such as peak heart rate, vary on time scales far shorter than the recording period.

We in our prototype we have shown the feasibility of automating this data collection process, by interfacing babies to a computer system, which will collect data on a second-by-second basis.

This will then able analyzed to be analized using a battery of tools including Naive Bayesian Reasoning to develop new ways to care for critically ill newborn babies.

We also plan to web enable our system the enabling centers to pool data to obtain larger datasets for analysis.

REFERENCES

[1] Anonymous. *The Vermont-Oxford Trials Network: very low birth weight outcomes for 1990. Investigators of the Vermont-Oxford Trials Network Database Project.* Pediatrics 1993; 91: 540-545.

[2] Field D, Hodges S, Mason E, Burton P. *Survival and place of treatment after premature delivery* [see comments]. Archives of Disease in Childhood 1991; 66: 408-10; discussion 410-1

[3] Anonymous. *Outcome at 2 years of children 23-27 weeks' gestation born in Victoria in 1991-92* . The Victorian Infant Collaborative Study Group. Journal of Paediatrics & Child Health 1997; 33: 161-165.

[4] Alberman E, Botting B. *Trends in prevalence and survival of very low birthweight infants, England and Wales: 1983-7.* Archives of Disease in Childhood 1991; 66: 1304-1308.

[5] Anonymous. *Improved outcome into the 1990s for infants weighing 500-999 g at birth*. The Victorian Infant Collaborative Study Group. Archives of Disease in Childhood Fetal & Neonatal Edition 1997; 77: F91-4.

[6] Doyle LW. *Outcome to five years of age of children born at 24-26 weeks' gestational age in Victoria* . The Victorian Infant Collaborative Study Group. Medical Journal of Australia 1995; 163: 11-14

[7] Richardson DK, GrayJE, McCormack, et al. *Score for neonatal acute physiology severity index for neonatal intensive care*. Pediatrics 1993; 91: 617-23

[8] Hanley JA, McNeil BJ. *A method of comparing the areas under the receiving operator characteristics curves derived from the same cases* . Radiology 1983; 148: 839-43

[9] Smith, I, Lister, R, and Ray, M, and Hawson, G., Naive Bayesian Prediction of Bleeding After Cardio-Pulmonary By-Pass Surgery, accepted for publication in, The Annals of Thoracic Surger

[10] International Neonatal Network, "*The CRIB (clinical risk index for babies) score: a tool for assessing potential neonatal risk and comparing performance of neonatal intensive care units* ", The Lancet, July 24, 1993.

[11] Shinghal R. *Formal Concepts in Artificial Intelligence.* 1st ed. London: Chapman & Hall, 1992: 347-89.