# Analysing Time Series Medical Data-sets

Derek Sleeman, Zhi Luo, Gordon Christie, George Coghill
Computing Science Department, The University, ABERDEEN, AB24 3FX, UK
Telephone: +44 (0)1224 272296
Email: dsleeman@csd.abdn.ac.uk
Peter Andrews & Pat Jones.
Dept. of Clinical & Surgical Services (Anaesthetics)
Western General Hospital, EDINBURGH, EH4 2XU
Telephone: +44 (0)131 537 1666
Email: pandrews@ed.ac.uk

**ABSTRACT** At the 1999 AIME Conference [1] we reported a decision tree study on a subset of data, including time series physiological data, admissions characteristics, and outcome after traumatic head injury. That study analysed total duration for which patients had, for example, raised Intracranial Pressure, but it did not consider temporal relationships between various physiological and clinical events. In this study, we have addressed that issue in a number of ways. Firstly, by using a workbench, AAB, to display the Real Time data-set and asking clinicians to make predictions of expected outcome based on complete physiological and clinical data. Secondly, repeating the exercise with a reduced (more compact) representation for the physiological data. Thirdly, patterns were generated, including "adjacent" physiological parameters and clinicians were asked if they are likely/very unlikely to cause a particular major physiological event or outcome. Finally, we implemented a module to test patterns of the form:

IF X happens then Y will happen between T1-T2

against patient time-series data. Results of all these studies have so far not been conclusive [2]; it has been suggested that the brain is currently not very well understood physiologically, and that a similar set of analyses should be applied to a simpler organ. Given that significant amounts of data are now available for patients undergoing dialysis, we have chosen to do an analogous study in this area; also the physiology of the renal system is much better understood. We have outlined some additional studies we plan to undertake using data-mining, theory refinement and knowledge base refinement approaches.

## 1. INTRODUCTION

The structure of the note is as follows: the next section gives some background to the patients with Head Injuries, and discusses the data collected in the Edinburgh NICU, [3]. We subsequently discuss studies that we have undertaken on data which is available to the clinicians when they attend these patients; further we review a reduced data representation which we have introduced and a series of studies that we have undertaken with this representation.

We review the overall outcome of the head-injury studies, and as noted above concluded that at this point in time the cerebral system is too poorly understood, for much progress to be made. We conclude the paper by describing the series of experiments which we intend performing on the dialysis data-sets; both those undertaken in the head-injury domain and some additional approaches based on data-mining and theory revision.

## 2. DATA-SETS AVAILABLE FOR THE EDINBURGH HEAD INJURY PATIENTS

*Demographic data:* Demographic data describes patients' characteristics and their clinical state on admission to hospital. The following data were used in this analysis: Grade of injury (minor to severe), Glasgow Coma Score, or GCS (a measure of verbal, motor and eye responses), Age, Pupil response, Injury Severity Score (ISS), Sex, Cause of injury, Diagnosis (type of injury to the brain), where the patient was referred from, type of skull fracture (if any).

*Temporal data:* During their stay in the NICU, the bedside monitors for the patients in this study were connected to a data collection system. This recorded the values of various parameters once a minute. The degree of abnormality, that is the amount by which the parameter was out of range, was graded on a scale of 0 to 3. 0 being classed as normal and 3 being extremely abnormal. The data available is described below:

- *Parameters monitored and recorded:* ICP (Intracranial pressure), BP (invasive blood pressure), $SaO2_2$ (oxygen content in the arterial blood to the brain), $SvO2_2$ (oxygen content in the venous blood from the brain), $ETCO_2$ (end tidal carbon dioxide), T1 (temperature), HR (heart rate).

- *Insult types derived from recorded data:* Intracranial Pressure (pressure exerted on the skull by the brain), Hypotension (low blood pressure), Hypertension (high blood pressure), Cerebral Perfusion Pressure (CPP, cerebral blood flow), Hypoxia (low oxygen content in the arterial blood to the brain), Cerebral Oligaemia (low oxygen content in the venous blood leaving the brain), Cerebral Hyperaemia (high oxygen content in the venous blood from the brain), Hypocarbia (low expired carbon dioxide), Hypercarbia (high expired carbon dioxide), Pyrexia (high core temperature), Bradycardia (slow heart rate), Tachycardia (fast heart rate), Global cerebral hypoxaemia (low oxygen content in the blood in the brain), Global cerebral hyperaemia (high oxygen content in the blood in the brain).

- *Outcome data:* Data are generally available concerning the outcome of each patient at 6, 12 and 24 months. Only outcome at 12 months was used in this study. Outcome is graded using the Glasgow Outcome Scale (GOS) [4], which classifies outcome into five categories: 1: Dead; 2: Persistent vegetative state; 3: Severely disabled; 4: Moderately disabled; and 5: Good recovery.

### 2.1 What Predictions Do Clinicians Wish to Make?

Clearly it is desirable to predict the actual GOS (Glasgow Outcome Score) for individual patients. However, given the quantity and the quality of the data available, this has not been possible so far. On the other hand, predicting survival has been achieved reliably. This, however, is not a very useful discrimination and so most studies have focussed their attention, for the moment, on predicting good or poor outcomes on the Glasgow scale. (Good outcome is usually defined as 4 & 5 on this scale and poor as 1, 2 & 3).

## 3. SOME STUDIES WITH THE INFORMATION PROVIDED BY EDINBURGH MONITOR-BROWSER© [9] SOFTWARE

The Edinburgh Monitor-Browser infers at least 10 secondary insult measurements from directly acquired physiological readings. As noted earlier, it also splits the insults generally into four levels: normal, moderate, high and extreme, that are displayed by BROWSER (off-line data processing) & MONITOR (on-line) systems as white, yellow, orange & red (CPP has a fifth category, green) respectively. These levels are also sometimes referred to as 0, 1, 2, & 3 ( and "a" for the mild CPP level). The clinicians use the pattern of insults to inform their decisions throughout patient care.

Below we give a diagram that shows how the current monitor looks to the clinician. Given the lack of space, we cite a display where only 2 channels are displayed instead of the usual 10:

| | | | | |
|---|---|---|---|---|
| P1 | 1 | 3 | | Major Event |
| P2 | 2 | | 1 | |
| time | 0 | 10 | 25 | 34 |

**Figure 1:** The format of information provided by the Edinburgh Monitor-Browser System. P = monitoring channel

The interpretation of this display is that at Time-0, the value of P1 is level-1, P2 is at level 2; at Time-10 P1's value moves to level 3; at Time-25 P2 reduces to a level-1 insult, and at Time-34 a major event occurs for this patient.

These types of displays are also available in an off-line system called the BROWSER that allows clinicians and analysts to review the data, which has been collected and stored in real-time by MONITOR. In experiment A, we reviewed 4 or 5 cases with several clinicians and asked them to:

i) Review the middle section of such traces for individual patients and asked each clinician to predict the likely outcome for a patient in terms of the GOS and/or major event that might happen to the patient. Generally the clinicians were unable to make such predictions.

ii) We then discussed the type of trace illustrated in Figure 1 where there had been a major event, e.g., a sudden increase in ICP, and asked the clinicians to review the events of the previous 1, 2 … hours, to see if they could predict the major event. Their response was that they could generate a number of explanations for each of these major events, but could not be sure what would happen in individual cases.

iii) We also asked them to discuss *types* of cases they had encountered; the clinicians tended to describe classes of patients in terms of the nature of the accident which led to head injury - such as young men on motorbikes or elderly people who have had bad falls. These patients were not classified in terms of their physiological parameters.

As a result of these experiments we decided to look at the nature of the information which the clinicians were using in both the MONITOR and BROWSER systems.

## 4. DEVISING A REDUCED REPRESENTATION and SUBSEQUENT EXPERIMENTS

One possible explanation for the inability of clinicians to interpret fully the data available to him from MONITOR (on-line) or BROWSER (off line) is that they are suffering from information overload (c.f., a pilot handling a plane in an emergency situation). So we hypothesized that a more user friendly display would allow the clinician to first **select** the channels which s/he wished to view and to only show the **transitions** between levels. These features have been incorporated into *Aberdeen Abbreviated Browser* (AAB) which displays the information inherent in figure 1 as:

| | | | | | |
|---|---|---|---|---|---|
| P1 | | 1 | 3 | 3 | Major Event |
| P2 | | 2 | 2 | 1 | |
| time | 0 | 10 | 25 | 34 | |

**Figure 2:** Showing the Reduced Representation used in ABB; this display summarises the identical information to Figure 1.

Experiment B: We repeated the same studies in Experiment A, using the reduced representation. The results for i) & ii) are essentially the same with the several clinicians interviewed being unable to make any clear predictions. However, in response to question iii), a clinician was able to characterize several classes of patients in terms of their physiological parameters:

One sub-group of patients have predominantly focal injuries (contusions) and exhibit a pattern of low blood pressure on days 1-3 of ICU admission and then high ICP and low CPP on days 4-10 of admission.

Another group of patients who have suffered diffuse injury (as shown by brain CT scans) and develop no ICP problems, do not make a good recovery. They are identified by the severity of their injury, CT scan, and admission GCS & ICP.

### 4.1 Generating possible patterns from the Reduced Representation

In the following discussion, notation P1(x) [ts1] denotes an insult P1 at level x during time-slot, ts1; on some occasions the time-slot will be given in detail as [0, 9] or [10,24]. The revised presentation provides us a basis for generating temporal patterns to explain major events. As this process could easily result in a very large number of patterns, we have made, for the moment, the *strong* assumption that an event only effects activities in the following time-slot. For instance, P2(2)[10,24] can cause P2(1)[25,33] or P1(3)[25,33], but it cannot cause a *major event*, which occurs at minute 34. The temporal relationship between two events is indicated by the phase "*followed by*"; for example, "P1(x)[i] followed by P1(y)[i+1]" implies the second event P1(y) occurs at time interval *i+1* and follows the event P1(x) at time *i*. Hence, the patterns generated are in fact sequences, and the temporal logic needed to map to Allen's interval calculus would be "BEFORE(*t1*, *t2*)".

In the example given above, the major event happened in minute 34. When looking at the previous 9 minutes (i.e. [25,33]), the major event might be caused by *P1(3)* solely or *P2(1)* solely or *P1(3)* and *P2(1)* jointly. However, if the window is enlarged to 24 minutes [10, 33], the following <u>additional</u> temporal patterns could be generated to explain the major event:

P1(3) followed by P1(3) or
P1(3) followed by P2(1) or
P1(3) followed by P1(3) & P2(1) or
P2(2) followed by P1(3) or
P2(2) followed by P2(1) or
P2(2) followed by P1(3) & P2(1) or
P1(3) & P2(2) followed by P1(3) or
P1(3) & P2(2) followed by P2(1)
P1(3) & P2(2) followed by P1(3) & P2(1)

The pattern generating process exhaustively enumerates all possible temporal combinations prior to the major event. It is understood that the patterns in a sense are generalizations as the actual time-slots are dropped. So the interpretation of the second pattern { P1(3) followed by P2(1) } is that any occurrence of a P1 at level 3 followed by a P2 at level 1, will cause the event, and that the lengths of these events are not significant. This we realise is a significant assumption which we may have to retract subsequently. This pattern generating process is similar to the one used in the Chemical structural generator of Stanford's DENDRAL project, [5].

<u>Experiment C</u>: The algorithm was used to generate patterns for several patient events, and we asked several clinicians to eliminate patterns (patterns of physiological events) which they thought would be impossible or very unlikely. If we are able to eliminate a sizable number of patterns then the remainder can be investigated in **detail** by clinicians. In the Head Injury domain, AAB produces patterns of the following form:

**Unit-01** <ICP: Normal> & <Hypertension: Grade 1> followed by
**Unit-02** <Raised ICP: Grade 1> & <Hypertension: Grade 1> Followed by
**Unit-03** <Raised ICP: Grade 1> & <Hypertension: Normal>

One of the clinicians did a preliminary classification of the patterns as very likely, pretty likely, somewhat unlikely, and very unlikely. An example of a **likely** pattern is: (Almost) any physiological derangement (eg, hypotension increase, hypertension decrease, hypoxia increase) after 1-15 minutes, will cause the ICP to increase.

### 4.2 An Alternative Approach: have the Experts specify Hypotheses and compare these against patient data-sets.

One of the Clinicians used AAB to test a particular hypothesis:

Does an increase in BP lead to a decrease in ICP?

He tried this for a number of patients. This led to the suggestions that we should provide a facility to test more general hypotheses of the form:

**IF X happens THEN Y will happen between T1 – T2**

In fact the clinicians have now provided us with 16 hypotheses for this domain (both positive and neutral). Below we give an example:

**IF HYPERTENSION increases THEN ICP will decrease within 1-15 minutes.**

The algorithm reports results for individual patients against each of the hypotheses. Further, we report the total number of X and X-Y-matches, and the percentage of matches, as it is important to know whether the 50% figure arises because of 1 X and 2 matches or because of 50 Xs and 100 matches (the later of course is a more interesting/reliable observation.)

Experiment D: reviewed the above results, identified a number of patients who satisfy particular hypotheses strongly, and then investigated their data-sets in greater detail. Not many of the patients complied with the positive hypotheses; but there were a considerable number of "hits" with the neutral hypotheses ie ones which suggest no change in the patient state. In particularly, we currently speculate that it will be necessary to include clinical interventions to get a fuller picture – more accurate hypotheses – for this domain. Although, this study was planned it has not yet been carried out due to lack of resources.

## 5. CONCLUSION TO HEAD INJURY STUDY

Use various Statistical analyses to detect events in patient physiological data which appear to be correlated. If such correlations exist, this would enable AAB's generative algorithm to use a smaller number of "building blocks", and hence to produce a more tractable search space.

The general conclusions to this study is that no, or few, correlations were formed as the physiology of the brain is not sufficiently understood. For this reason we have decided to repeat (and extend) the experiments in another area of medicine – namely renal dialysis as outlined in the following section.

## 6. PLANNED STUDY OF DIALYSIS PATIENTS

Patients in the end stages of renal failure are often given dialysis treatment several times a week. Modern dialysis machines are able to record a number of parameters each minute. Again blood pressure and heart rate are among the parameters recorded; in this domain one of the major events which can occur is hypotension which can lead to a heart attack.

Initially we plan to repeat, with the dialysis data-sets, the four experiments performed above; we are optimistic that the clinicians in this domain will be able to provide some detailed predictions given that the physiology of the renal system is so much better understood. Additionally, we plan using data-mining techniques to address the following questions:

- Can major events (like hypotension) be predicted (& hence avoided)?
- Are there distinct groups of patients and how does this effect their dialysis regime?
- What are the major differences between patients who are successfully dialysed and those that are not?
- Is it possible to create optimum dialysis strategies for most / all patients?

Further, we plan to develop theory revision approaches which use background knowledge (eg equations & process information) to see if individual patient data-sets and the above background knowledge can be made consistent. In this sub-project we will draw on our experience of knowledge base refinement, [6] , [7] and model learning [8].

Thus far the original AAB workbench (written in PASCAL) has been completely rewritten in JAVA in order to extend its ability to interact with data from different sources on and off the internet. This tool has successfully analysed a range of data-sets.

## 7. REFERENCES


1. A. McQuatt, P. J. D. Andrews, D. Sleeman, V. Corruble & P. A. Jones (1999). The analyses of Head Injury data using Decision Tree techniques. *Artificial Intelligence in Medicine* W Horn et al (eds) (Proceedings of AIMDM'99 Conference, Aalborg, Denmark, June 1999) Springer-Verlag, pp336-345.
2. Derek Sleeman, Zhi Luo, Peter Andrews & Pat Jones (2001). What insights can be obtained from Temporal Physiological Data-sets collected during Intensive Care for Head-Injury patients? Workshop on Computers in Anaesthesia & Intensive Care: Knowledge-Based Information Management at AIME-01 (Cascais, Portugal).
3. P.A. Jones, P.J.D. Andrews, S. Midgley, S.I. Andrerson, I.R. Piper, J.L. Tocher, A.M. Housley, J.A. Corrie, J.Slattery, N.M. Dearden, & J.D. Miller (1994). Measuring the burden of secondary Insults in Head-Injured Patients during Intensive Care, *Journal of Neurosurgical Anaesthesiology, Vol. 6, No. 1,* pp4-14, 1994.
4. B. Jennett, & M. Bond (1975) Assessment of outcome after severe drain damage. A practical scale. The Lancet, vol 1, pp 480-484.
5. A. Barr, & E.A. Feigenbaum (1981) *The Handbook of Artificial Intelligence, Vol 1*. Menlo Park, CA: William Kaufmann, pp 106-123.
6. S. Craw & D. Sleeman (1990) Automating the Refinement of Knowledge-Based Systems. In Proceedings of ECCAI-90. Luigia Aiello (Ed). London: Pitman, pp 167-172.
7. A Aiken, & D Sleeman (2003) Refiner++: A Knowledge Acquisition and Refinement Tool, presented at KCAP-03 Workshop on “Capturing knowledge from domain experts: Progress & Prospects” Sanibel Island, Florida, 26 October 2003.
8. Coghill G. M., Garrett S. M. & King R. D., (2002) “Learning Qualitative Models in the Presence of Noise”, Proceedings of the 16 th International Workshop on Qualitative Reasoning: QR'02, Sitges, Spain, June 2002, pp 27 – 36.
9. TP Howells. *Edinburgh Monitor-Browser* ©Software, [v1.0.0] Computer Program 1994. tim.howells@nc.uas.lul.se



## ACKNOWLEDGEMENTS

A studentship from the University of Aberdeen allowed Zhi Luo to undertake this work; more recent support has been provided by the EPSRC through the AKT IRC. We acknowledge discussions on this topic with Dr. Ric Bellazzi (Pavia), and Dr. Nick Fluck (Aberdeen Royal Infirmary).