NEONATE: Decision Support in the Neonatal Intensive Care Unit - A Preliminary Report

Jim Hunter [1], Gary Ewing [1], Yvonne Freer [3],
Robert Logie [2], Paul McCue [1], Neil McIntosh [3]

[1] Department of Computing Science, University of Aberdeen,
King's College, ABERDEEN, AB24 3UE, UK
[2] Department of Psychology, University of Aberdeen,
King's College, ABERDEEN, AB24 2UB, UK
[3] Department of Neonatology, University of Edinburgh,
EDINBURGH, UK

[1] {jhunter, gewing, pmccue}@csd.abdn.ac.uk
[2] r.logie@abdn.ac.uk
[3] yvonne.freer@luht.scot.nhs.uk, neil.mcintosh@ed.ac.uk

Abstract. The aim of the NEONATE project is to investigate sub-optimal decision making in the neonatal intensive care unit and to implement decision support tools which will draw the attention of nursing and clinical staff to situations where specific actions should be taken or avoided. We have collected over 400 patient-hours of data on 31 separate babies, including physiological parameters sampled every second, observations made by a research nurse of all the actions performed on the baby with an accuracy of a few seconds, occasional descriptions of the appearance, mobility, sleep patterns, etc of the baby. We describe our attempts to use this data to discover examples of sub-optimal behaviour.

1 Introduction

The original objectives of the NEONATE project (Hunter et al. 2003) were: (i) to identify situations in the Neonatal Intensive Care Unit (NICU) where sub-optimal performance might occur; (ii) to develop a number of data processing algorithms aimed at alerting the clinical staff to those situations; and (iii) to evaluate which approaches would be most effective in bringing about improvements in performance. It has been shown that simply displaying complex time series data does not automatically lead to improvements in patient care (Cunningham et al. 1998, McIntosh et al. 2000). The COGNATE project (Alberdi et al. 2000, 2001) concluded that some assistance in the form of additional data processing is necessary to support the decisions made by the clinical staff.

Identifying sub-optimal performance poses certain methodological problems. In complex domains such as medicine, judgments about what is sub-optimal can only be made by a recognized expert in the domain. However, if the expert is physically present on the ward, the normal (unobserved) behaviour of the junior decision-maker will almost certainly be changed by that presence; this would be in addition to the practical difficulty of obtaining long periods of an expert's time. The alternative that we have adopted is to capture in real time as much information as possible about the baby, and to present this to the expert at a later time. This paper describes the details of this approach:

- Through interviews with medical staff, we established a **lexicon of terms** used to describe a baby and the management actions that can be taken.
- We collected as **rich a data set as possible** for a number of babies through many hours of **on-ward observation**, noting the actions that were taken and descriptions of baby state ("descriptors"), as well as acquiring physiological and other data.
- We identified important **single actions** (e.g. **handbagging**) and attempted to acquire **protocols** from our expert clinician describing the circumstances under which these actions should be taken.

We have come to the (somewhat unexpected) **conclusion** that getting experts to comment on the appropriateness of individual actions is not practicable and we will discuss the consequences of this.

2 Developing a Lexicon of Observations and Actions

It is clear that medical staff acquire data about a patient through seeing, hearing and touching the baby as much as (or perhaps more than) by referring to physiological data acquired from instruments; we will refer to this information as "descriptors". In attempting to capture as complete a data set as possible, we considered it necessary to attempt to record these descriptors. A pre-requisite was to agree on a suitable lexicon (or ontology). We interviewed clinical staff at all levels asking them to say how they would describe the current state of a baby to a colleague. Thirty-two staff were interviewed and 552 descriptors were generated. Senior clinical staff subsequently reviewed these lists for consistency and to remove synonyms and singletons (words used by only one person), thus reducing the list to 166 terms. These terms were grouped under seven headings: Bowels (and urine), Crying (and facial expression), Feeding, Movement, Size (including shape and weight), Skin (including colour), Sleep (and demeanour). Examples of the 32 descriptors for Skin include: Pink, Good Capillary Refill, Blue, Jaundiced, Dry.

In a similar way, interviews elicited 191 terms to describe the actions that can be taken. This was reduced to 51 terms which were organised into a hierarchy; higher level nodes in this included intermediate abstractions such as: Care, Collect Data, Feeding, Respiration, Communication.

3 On-Ward Data Collection

A research nurse was employed for approximately four months to observe the activity at one or more cots and to make as accurate a record as possible. The information captured was:

- the equipment used to monitor, ventilate, etc.;
- the actions taken by the medical staff (see above);
- occasional descriptions of observable state (descriptors) (see above);
- the alarm limits in force on the monitors;
- the settings on the various items of equipment (including the ventilator);
- the results of blood gas analysis and other laboratory results;
- the drugs administered.

Data were entered with a timing accuracy of a few seconds on a laptop computer using a specially written program called 'BabyWatch' running under Windows. All data (with one or two exceptions) were entered by selecting from pre-compiled lists. In addition the research nurse could enter short free-text comments.

At the same time as data was being entered manually, the 'Badger' data collection system was automatically acquiring physiological data with a time resolution of one second. The actual parameters sampled depended on the monitoring in place but typically included heart rate, transcutaneous O_2 and CO_2, saturation O_2, core and peripheral temperatures, and blood pressures.

Before the observations began, a detailed protocol was established to set out how the study was to be conducted. This included guidelines for clock synchronisation, subject selection, descriptor recording and ethical considerations. Ewing et al. (2002a) describe this in more detail and Hunter (2002) describes the BabyWatch software.

4 Observational Results

Data collection started in mid October 2001 and finished in mid February 2002. We collected about 407 patient-hours of observations on 31 separate babies consisting of over 32,000 individual data records. Details of the data collected are available in Ewing et al. (2002a).

No experiment goes exactly as planned, and a certain amount of post processing was required. The BabyWatch and ‘Badger’ clocks had to be reconciled, obvious errors in data entry corrected; again details of the post-experimental processing are contained in Ewing et al. (2002a).

An existing tool, the Time Series Workbench (TSW), was adapted to allow us to present all of this data together. In addition to the usual presentation of time-series physiological data, it displayed:

- periods where the nurse was observing; periods where specific actions were taking place; the presence of observations entered by the nurse; the administration of medication, and the presence of blood gas and laboratory results; the existence of comments;

- the hierarchy of actions; the basic problem is that there are too many actions to display easily – our solution is to allow the user to select one (or a subset of) action(s) to be displayed by interacting with this hierarchy.
- the comments entered by the research nurse;

In addition, a tool was developed within the TSW which allowed us to view the data from the perspective of a particular type of action, observation, etc. and to collect overall statistics.

We believe that our database linking physiological measurements to simultaneous observations is one of the richest to have been collected.

5 Clinical Protocol Development

Recall that our initial objective was to identify sub-optimal decisions. From our perspective, a decision manifests itself as an observed action or the absence of such an action. We looked initially at the action of 'handbagging' - the manual ventilation of a baby. Overall we had 58 instances of this action with an average duration of 2 minutes. Handbagging often takes place more than once in a short period of time and we grouped related instances into 'episodes'; we had 29 such episodes. Because handbagging causes extended fluctuations in most physiological parameters, we only considered the first action in a given episode.

At first sight the methodology might appear obvious: given that we have almost complete data, get the expert to look at the episodes of handbagging and decide for each whether the action was performed optimally or not. However this is incomplete – the expert's attention is reasonably easily focused on the times when actions were taken (and perhaps should not have been), but we must also consider the occasions when an action should have been taken but wasn't. Without additional support, this would have required the expert to inspect all of the times when that particular action was not taken – many hours worth of data – and this is just not practical for a busy expert. To focus attention on possible candidates for such times, we asked the expert to define a simple protocol for when handbagging should be carried out:

((OX < 3 or SO < 60) for at least 10 seconds) and (HR < 100)

where OX is transcutaneous oxygen, SO is oxygen saturation and HR is heart rate.

As with all knowledge acquisition, our expectation was that the first attempt as formalisation would be inaccurate and incomplete. To begin with we intended to take as our gold standard the actions that were actually taken by the clinical staff – i.e. we would assume that they always made the right decision. We expected that the protocol as implemented would generate false positives and negatives (with respect to the decisions actually made to act or not to act) – as well as true positives and negatives. Once the protocol had been refined over several iterations, we anticipated that a change of emphasis would occur, in that our expert's attention would be focused more and more on the 'false' positives and negatives and we anticipated that (s)he would start to query whether they really were 'false'. In other words the assumption that the correct decision was always made would become increasingly subject to question and we might decide in some cases that the protocol was correct and that the decision made was in some sense sub-optimal.

6 (Somewhat Unexpected) Conclusion

We discovered that our experts were reluctant to comment on the appropriateness or otherwise of a specific single handbagging action (whether recommended or actual) without reviewing the way in which the respiratory function of the patient had been managed over a considerable period of time (including the ventilator settings, drugs, X-rays, suction, repositioning, etc). It was clear that to put pressure on them to come to a view based on purely local (in a temporal sense) information would be counter-productive. We now consider that we need to identify and formalise the protocol for respiratory management taken as a whole. Such a protocol will be much more complex; however we believe that languages such as ASBRU (Shahar et al. 1998) are sufficiently rich to express it. Ultimately we are convinced that in the highly complex environment of the ICU, protocols must represent complete management strategies if our medical experts are going to be willing to devote time to developing them, and if the end users are going to see their advice as appropriate.

References

Alberdi E, Gilhooly K, Hunter J, Logie R, Lyon A, McIntosh N and Reiss J, 'Computerisation and Decision Making in Neonatal Intensive Care: A Cognitive Engineering Investigation', *Journal of Clinical Monitoring and Computing*, Vol 16, No 2, pp 85-94, 2000.

Alberdi E, Becher J-C, Gilhooly K, Hunter J, Logie R, Lyon A, McIntosh N and Reiss J, 'Expertise and the Interpretation of Computerized Physiological Data: Implications for the Design of Computerized Monitoring in Neonatal Intensive Care', *International Journal of Human Computer Studies*, Vol 55, No 3, pp 191-216, 2001.

Cunningham S, Deere S, Symon A, Elton RA and McIntosh N, 'A Randomized, Controlled Trial of Computerized Physiologic Trend Monitoring in an Intensive Care Unit', *Crit.Care Med*, Vol 26, pp 2053-2060, 1998.

Ewing G, Ferguson L, Freer Y, Hunter J and McIntosh N, 'Observational Data Acquired on a Neonatal Intensive Care Unit', *University of Aberdeen Computing Science Departmental Technical Report: TR 0205*, 2002a.

Hunter J, 'BabyWatch User Manual', *University of Aberdeen Computing Science Departmental Technical Report: TR 0206*, 2002.

Hunter J, Logie R, McIntosh N, Ewing G and Freer Y, 'NEONATE: Effective Decision Support in the Intensive Care Unit', *http://www.csd.abdn.ac.uk/~gewing/neonate/*, 2003.

McIntosh N, Becher JC, Stenson BJ, Laing IA, Lyon AJ, and Badger P, 'The clinical diagnosis of pneumothorax is late: use of trend data and decision support might allow preclinical detection', *Pediatric Research*, Vol 48, pp 408-415, 2000.

Shahar Y, Miksch S, and Johnson P, 'The Asgaard Project: A Task-Specific Framework for the Application and Critiquing of Time-Oriented Clinical Guidelines', Artificial Intelligence in Medicine, Vol 14, pp 29-51, 1998.