The NEONATE Database

Jim Hunter [1], Gary Ewing [1], Lindsey Ferguson [2], Yvonne Freer [3], Robert Logie [4], Paul McCue [1], Neil McIntosh [2]

[1] Department of Computing Science, University of Aberdeen, King's College, ABERDEEN, AB24 3UE, UK

[2] Department of Neonatology, University of Edinburgh, EDINBURGH, UK

[3] School of Nursing Studies, University of Edinburgh, EDINBURGH, UK

[4] Department of Psychology, University of Aberdeen, King's College, ABERDEEN, AB24 2UB, UK

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

Abstract

The aim of the NEONATE project is to investigate decision support in the neonatal intensive care unit. This paper describes the considerable amount of observational and physiological data which was acquired and as well as the tools which have been developed to analyse them. Both the data and the tools are in the public domain. We believe that our database linking physiological measurements to simultaneous observations is one of the richest to have been collected.

1 Introduction

The NEONATE project (Hunter et al. 2003a) is concerned with the development of decision support tools for neonatal intensive care; a more general overview of the current state of the project is available in the main proceedings (Hunter et al. 2003b).

As part of the project we decided that it was necessary to collect as much data about neonates and their management as was possible. We employed a research nurse over a four month period to be present at the cot-side and to record all of the actions that were performed on the baby; she also made occasional observations of appearance and behaviour, and noted equipment settings. At the same time, the Badger neonatal system (Badger 2003) was automatically acquiring physiological data.

The paper describes: (i) the development of a suitable lexicon of actions and observations; (ii) the way in which the observational study was conducted, including a description of BabyWatch - the special-purpose software which allows rapid data entry at the cot-side; (iii) the post-collection processing of the data; (v) a summary of the data collected; (v) the software used to display and analyse the combined physiological and observational data – the Time Series Workbench (or TSW). The NEONATE database, consisting of about 407 hours of observational data, together with the BabyWatch and TSW software are publicly available.

2 Developing a Lexicon of Observations and Actions

It is clear that clinicians 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 through instrumentation; we will refer to this information as "observations". In attempting to capture as complete a data set as possible, we considered it necessary to attempt to record these. However 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 (eight each of junior nurses, senior nurses, junior doctors and senior doctors) were interviewed and 552 terms 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. In order to make it possible to design sensible input screens, the same senior staff allocated the terms to seven categories as shown in Table 1.

Table 1. Categories used for Observations

Bowels (and urine)	Size (including shape and weight)
Crying (and facial expression)	Skin (including colour)
Feeding	Sleep (and demeanour)
Movement	

By way of illustration, Table 2 shows the set of observations relating to Skin.

Table 2. Observations relating to Skin

Pink	Good Capillary Refill	Poor Capillary Refill	Washed Out
Poorly Perfused	Pale	Poor Colour	Mottled
Blue	Black	Bruised	Rash
Irritated	Blotchy	Dilated Veins	Plethoric
Jaundiced	Waxy	Appears Shutdown	Dry
Moist	Clammy	Gelatinous	Shiny
Thin	Transparent	Fragile Skin	Taut
Broken	Peeling Skin	Shabby	Puffy
Smooth	Loose Skin	Downy	Hairy

In a similar way, interviews elicited 191 terms to describe the actions that can be taken. Again, senior clinical staff reduced the number of terms to 51, and organised them into a hierarchy.

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;
- occasional descriptions of observable state (**observations**);
- the current **alarm limits** 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.

These 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 (Hunter 2002). All data (with one or two exceptions) were entered by selecting from pre-compiled lists derived from the lexicons described in Section 2. In addition the research nurse could enter short free-text comments.

Most actions are considered to take a finite time to perform, and are thus recorded as temporal intervals with start and end times. Additionally we noted who carried out the action: **N**urse, **D**octor or **P**arent. BabyWatch includes the facility to switch rapidly between different babies so that, in principle, the nurse could be observing more than one baby at a time; in practice this was little used.

The observer was a qualified nurse trained in neonatal intensive care and there was heavy reliance on her professional knowledge and judgment. A period of training was included at the start of the study and the data obtained during this time discarded.

Before the observational study began, a detailed protocol was established to set out how it was to be conducted. This included practical consideration of guidelines for:

- **Clock Synchronisation**: For the BabyWatch and Badger computer clocks.
- **Subject selection**: Initially the focus was on very high dependency babies. If there was more than one, observations were concentrated on the baby associated with the highest clinical activity at the time. Early in the data collection phase, once the clinical activity surrounding a baby under observation had settled down to a routine level, attention was moved to the next high-dependency baby, if one existed. However, by mid-phase, it was deemed appropriate to focus on the observation of individual babies for longer periods, in order to gain complete 'pictures' of activities.
- **Presence of the Observer**: The observer was to record when she was present at the cot-side in sufficient detail to indicate short absences, e.g. coffee breaks.
- **Observations of State**: At the start of a session, the state of the baby was to be recorded. Some aspects might not be observable at that time (e.g. feeding). If this was the case, this information was obtained from the nurse caring for the baby or from the patient notes. If there was any clinically significant change in the patient state (as judged by the observer), this was recorded. Whenever the baby was examined by a clinician (e.g. during ward-rounds), (s)he was asked to describe the baby's state.

- **Ethical considerations**: The observer was instructed not to interfere with the actions of, or prompt, the attending medical staff unless, in her professional opinion, not to do so would endanger the baby's life. No ethically controversial situation occurred.

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, O_2 saturation, core and peripheral temperatures, and blood pressures.

4 Post-Collection Processing

A certain amount of post processing was required; details are set out in Ewing et al. (2002), but included:

- **Clock synchronisation**: The offset between the BabyWatch and Badger clocks was estimated and the time stamps on the data collected using BabyWatch adjusted accordingly.
- **Presence of the Observer**: Occasionally we spotted periods where the 'observer present' data indicated the observer as being present, but where there was a complete lack of observations or actions. We concluded that she had probably forgotten to log out, and, acting conservatively, removed such periods from the 'observer present' data.
- **Durations of Intervals**: We wanted to establish whether there were any observational errors relating to the duration of action intervals. In particular, any actions that spanned more than one observational session were identified. Given that we cannot be sure that the action continued while the observer was not there, such intervals were broken into two intervals one with an 'undefined end' and the other with an 'undefined start'. In addition, we carried out a statistical analysis of each type of interval and identified those intervals with a duration of twice the standard deviation. If such an interval was clearly an outlier, it was replaced with an interval with an undefined end point.
- **Comments**: In reviewing the free-text comments, it became apparent that the observer had used these to report events that needed to be taken into account when post-processing the data - e.g. an incorrect action or observation had been made. Such corrections were made.
- **Validation of numerical input**: Numerical values were entered for: equipment settings, alarm limits, blood gas results, laboratory results, and medication dose. In entering such numerical values, it was possible for the observer to enter medically unrealistic values by mistake. BabyWatch has no checking built in for this (apart from checking that numerical values constituted legal numbers). The medical experts inspected the sets of values used. Any obvious misplacing of decimal points or confusion about units was corrected.
- **Overall medical validation**: Given the volume of data, it was not possible for the medical experts to examine all of the detail to check for entries that were medically inconsistent. However it is possible to print out the chronological sequence of observations for a given baby and any errors that were obvious here were investigated and corrected. Our medical expert did examine the routes of administration of drugs, and again, any obvious errors were corrected.

The Time Series Workbench (see Section 6) was augmented with editing features to simplify this processing.

5 Summary of Observational Results

Regular data collection started at the start of November 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; Table 3 shows how the records are distributed between the various categories.

Table 3. Distribution of records

Category	Number
Observer Present	403
Actions	19,610
Observations	1,831
Settings	4,512
Laboratory Results	1,343
Blood Gas	2,187
Medication	148
Comments	2,443
Total	**32,477**

The records have the fields shown in Table 4.

In general, time in the TSW is recorded as the number of days that have passed since 30 December 1899 plus the number of 'ticks' (a tick is 1/10 of a millisecond – 10^{-4} sec.) since midnight. In the published NEONATE database, to avoid compromising the anonymity of the babies, days are counted from the actual day of birth (which will not be released).

For events occurring at a single 'point' in time, the start and end times are the same.

These observations and the corresponding physiological data are formatted within the NEONATE project as tables in an 'Access' database and will be made publicly available in a number of formats.

6 Display and Analysis of Results

An existing tool, the Time Series Workbench (TSW), was adapted to allow us to present both physiological and observational data. In addition to the usual presentation of time-series physiological data, it displays:

- 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 (or observations); 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.

It is also possible to analyse all similar records and to accumulate statistics such as average duration, distribution of durations, etc.

7 Conclusion

We believe that the NEONATE observational database constitutes a unique resource which ought to be of interest to the community concerned with the interpretation of ICU data. For this reason we are placing it, its documentation and the supporting software in the public domain.

References

Badger P, "The Badger System", *http://www.clevermed.com*, 2003

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*, 2002.

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/*, 2003a.

Hunter J, Ewing G, Freer Y, Logie R, McCue P, and McIntosh N, 'NEONATE: Decision Support in the Neonatal Intensive care Unit – A preliminary Report', *AIME-03*, 2003b.

Table 4: Record structure.

Name of Field	Type	Content
ContextName	Text	Always set to 'Badger'.
SourceName	Text	Anonymous internal identifier for the baby.
ChannelName	Text	The nature of the data e.g. 'ACTION', 'OBSERVATION', 'SETTING', 'OBSERVER PRESENT'.
Full Descriptor	Text	The specific action, observation, etc.
VValue	Number	For those data items which have an associated numerical value (e.g. ventilator FiO_2 setting), this is stored here.
StartDateTime Days	Number	The date that the interval starts.
StartDateTime Ticks	Number	The time that the interval starts.
EndDateTime Days/Ticks	Numbers	The date/time that the interval ends.
CreatedDateTime Day/Ticks	Numbers	The date/time that the interval was entered into the database.
OriginatorType	Text	Always set to 'Observer'.
OriginatorName	Text	Always set to the initials of the observer.
Context Dependent Data	Text	Structured text; used here to indicate the degree of success in synchronizing the clocks.