Using cognitive task analysis to facilitate the integration of decision support systems into the neonatal intensive care unit

Gordon D. Baxter[a*], Andrew F. Monk[a], Kenneth Tan[b], Peter R.F. Dear[b], and Simon J. Newell[b]

[a] *Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom*
[b] *Department of Paediatrics, St James' University Hospital, Leeds LS9 7TF, United Kingdom*

* Corresponding author:
Tel.: +44-1904-433170; fax: +44-1904-433181
E-mail: g.baxter@psych.york.ac.uk

Abstract

Objective: New medical systems may be rejected by staff because they do not integrate with local practice. An expert system, FLORENCE, is being developed to help staff in a neonatal intensive care unit (NICU) make decisions about ventilator settings when treating babies with Respiratory Distress Syndrome. For FLORENCE to succeed it must be clinically useful and acceptable to staff in the context of local work practices. The aim of this work was to identify those contextual factors that would affect FLORENCE's success.

Methods: A cognitive task analysis (CTA) of the NICU was performed. First, work context analysis was used to identify how work is performed in the NICU. Second, the critical decision method (CDM) was used to analyse how staff make decisions about changing the ventilator settings. Third, naturalistic observation of staff's use of the ventilator was performed.

Results: A. The work context analysis identified the NICU's hierarchical communication structure and the importance of numerous types of record in communication. B. It also identified important ergonomic and practical requirements for designing the displays and positioning the computer. C. The CDM interviews suggested instances where problems can arise if the data used by FLORENCE, which is automatically read, is not manually verified. D. Observation showed that most alarms cleared automatically. When FLORENCE raises an alarm staff will normally be required to intervene and make a clinical judgement, even if the ventilator settings are not subsequently changed.

Conclusions: FLORENCE must not undermine the NICU's hierarchical communication channels (A). The re-design of working practices to incorporate FLORENCE, reinforced through its user interface, must ensure that expert help is called on when appropriate (A). The procedures adopted with FLORENCE should ensure that the data the advice is based upon is valid (C). For example, FLORENCE could prompt staff to manually verify the data before implementing any suggested changes. FLORENCE's audible alarm should be clearly distinguishable from other NICU alarms (D); new procedures should be established to ensure that FLORENCE alarms receive attention (D), and false alarms from FLORENCE should be minimised (B, D). FLORENCE should always provide the data and reasoning underpinning its advice (A, C, D). The methods used in the CTA identified several contextual issues that could affect FLORENCE's acceptance. These issues, which extend beyond FLORENCE's capability to suggest changes to the ventilator settings, are being addressed in the design of the user interface and plans for FLORENCE's subsequent deployment.

Keywords: Decision making; Neonatal intensive care; Artificial respiration; Respiratory Distress Syndrome; Expert systems

1 Introduction

Medicine, like many other complex dynamic domains, is increasingly reliant on technology, with computer-based systems now routinely deployed in the management and delivery of patient care. This increased reliance means that the technology must also be dependable: staff need to know that the technology can be trusted to perform as expected, when required.

Technology is very much in evidence is the modern neonatal intensive care unit (NICU). In the NICU at St James' University Hospital (SJUH), Leeds, for example, premature babies may be connected to a ventilator as well as a collection of sensors measuring heart rate, blood pressure, oxygen saturation level, and blood gas parameters. The baby's cot incorporates a temperature monitor, and the ventilation air humidifier also has its own sensors. In addition, microprocessor-controlled intravenous devices are used to deliver drugs and food.

One of the fundamental problems faced by neonates is that key stages of development have to take place outside the womb. The brain, for example, develops faster during the final trimester of pregnancy than at any other time of life [1]. Clinical staff in the NICU are therefore concerned with ensuring that the baby's brain is adequately supplied with oxygen. This normally occurs through respiration. The lungs of premature babies are often underdeveloped, however, so they do not produce enough surfactant to adequately facilitate gas exchange between the lungs and the blood stream. This lack of surfactant production—respiratory distress syndrome, or RDS—is one of the commonest problems in the NICU. The symptoms of RDS appear at birth or very shortly afterwards, with the disease process usually peaking within 72 hours of birth [2]; the baby normally recovers within seven days. RDS is usually treated using mechanically-assisted ventilation, with a goal of maintaining pulmonary perfusion and normal gas exchange while minimising harm to the baby until the RDS recedes.

Controlling the ventilation of premature babies is a cognitively demanding task. Clinical staff have to diagnose the baby's condition before deciding how and when to act. The baby's respiratory processes are monitored using partial pressures of oxygen (PO_2) and carbon dioxide (PCO_2) in the arterial blood which need to be kept within strict limits during the first 72 hours.

The PO_2 is controlled by adjusting the inspired oxygen level (FiO_2) and the mean airway pressure (MAP). The latter is a derived quantity representing the average pressure applied over one respiratory cycle. MAP is controlled indirectly using the ventilator settings, and can be monitored via the flow graph displays. When MAP is optimally set, the FiO_2 is adjusted to adapt to the changes in PO_2.

The PCO_2 is controlled by adjusting the minute ventilation, which is the product of the tidal volume—the amount of air moved in and out in a single breath—and the breathing rate (including mechanically supported and spontaneous breaths). Minute ventilation is therefore controlled either by adjusting the difference between the inspiratory and expiratory pressure settings on the ventilator (and hence, the tidal volume) or by changing the ratio of the breath inspiratory time to expiratory time which will alter the breathing rate..

Neonatal lungs are very delicate, so great care has to be exercised when changing the ventilator settings. In particular, the pressure provided by the ventilator to maintain the inflation of the lungs needs to be carefully regulated to avoid possible damage.

A bespoke expert system for neonatal intensive care, ESNIC [3], was previously used at SJUH to assist staff in making decisions about changing the ventilator settings. ESNIC would suggest changes to the ventilator settings which the staff then made manually. For each use, however, staff had to manually copy data from the various devices into ESNIC, in addition to their other work. This extra burden contributed to ESNIC falling into disuse.

It is now possible to read data directly from some of the sensors into a PC for further processing, thanks to two recent advances in technology. First, the development of the Neotrend multi-parameter intra-arterial sensor [4], which allows continuous monitoring of a baby's blood gases [5]. Second, software developments have enabled the automatic reading of data from the Draeger Babylog 8000 Plus ventilators used in the NICU. It was therefore decided to exploit these developments in a new expert system, FLORENCE: Fuzzy LOgic REspiratory Neonatal Care Expert [6], that will support decision making in the NICU. The live data from the Neotrend and the ventilator will be used by FLORENCE to propose changes to the ventilator settings. FLORENCE is to remain open loop, in that the final decision about accepting its recommendations will still fall to the clinical staff who carry the ultimate burden of responsibility for patient care.

One of the main reasons many medical systems are never used is that they introduce ways of working that conflict with other aspects of the user's job or require changes to procedures that affect other people's responsibilities [e.g., 7]. It was therefore important to understand the broader work context of the NICU taking into account the roles, responsibilities and relationships of stakeholders.

The goal of this study was to understand the context of the NICU in order to identify how FLORENCE should be designed to fit in with local work practices. This mainly has to do with the way that information is presented by FLORENCE and the re-design of standard procedures to integrate it into the daily routines of the NICU. The study was not intended to elicit knowledge that would form the basis of the rule set. The functional requirements of FLORENCE and the use of Fuzzy Logic with this rule set are described in [8].

The next section presents a brief overview of the cognitive task analysis (CTA) used here. The application of each of the selected CTA methods is then described in turn, together with consideration of the implications arising from the results of applying that method. The paper concludes with a discussion of the utility of the methods in identifying issues that have implications for the design, acceptance and clinical use of FLORENCE.

2 Cognitive Task Analysis of the NICU

There are identifiable, agreed standards and guidelines for neonatal care [e.g., 9, 10]. The practice of neonatal care, however, varies between hospitals [e.g., see 11], often comprising a combination of locally-developed procedures and best practice. Evolution may be influenced by regulatory authorities and local factors.

For FLORENCE to be acceptable and clinically useful to the staff, it must support the ventilation of babies with RDS and, more generally, integrate with the practice of neonatal intensive care at SJUH. To understand these two broad requirements a CTA of the practice of neonatal intensive care at SJUH was carried out.

CTA extends more traditional task analysis techniques to facilitate the collection of information about the mental processes underlying observable task performance [12]. There are many CTA methods and tools available; using a CTA entails selecting and applying a combination of methods and tools appropriate to the task and domain being investigated. Since one of the main goals of the NICU study was to inform the development of FLORENCE, the Critical Decision Method (CDM) [13] was chosen to collect data about the decision making processes used by clinical staff when interacting with the ventilator.

Before applying the CDM it is important to become familiar with the domain [14]. To analyse the physical and social work contexts in which FLORENCE will be used, it was decided to use lightweight Rich Picture representations [15]. These provide a useful way of showing how the roles and responsibilities of the various system stakeholders' along with any concerns they may have about the work system. This method was supplemented by background reading, and informal visits and meetings to provide a more complete analysis of the work context.

The final part of the CTA was to analyse how people use the ventilator in situ. For this purpose it was decided to use naturalistic observation of work in the NICU.

3 Work context analysis

The analysis of the work context of the SJUH NICU included an initial period of domain and context familiarisation. This was followed by a rich pictures analysis of work in the NICU.

3.1 Domain and context familiarisation

3.1.1 Meetings and visits

Preliminary visits to the unit were used to gather background information on neonatal intensive care, on the SJUH NICU, and on FLORENCE. They were also partly used to gain acceptance by staff, given plans to subsequently carry out observation of behaviour on the unit.

The NICU can deal with up to eight babies at any one time. The babies' condition is monitored using a range of sensors, as noted above. Each of the sensors has an alarm setting, and one of the most noticeable features of the NICU is that there are alarms sounding fairly frequently. A very high percentage of these alarms are caused by artefacts rather than clinical events. If the baby moves, for example, the heart rate rises rapidly, which triggers an alarm, and then the heart rate quickly settles back within limits. Staff are trained to scan the displays and check the baby's condition to help them differentiate between artefacts and clinical events.

Ventilator alarms are allocated a higher priority by staff. These alarms do not all automatically require a change to the ventilator settings; some of them relate to problems with the ventilator circuit tubing, for example. Staff therefore use the mnemonic DOPE to remind them to check for: Displacement of the endo-tracheal tube (ETT); Obstruction of the ETT; Pneumothorax; and Equipment malfunction, before adjusting the ventilator settings.

The daily work on the unit is largely determined during the ward round, which marks the change over from the night shift to the day shift. In this meeting decisions are taken about treatment, including use of the ventilator. During one such ward round, a consensual decision was taken to ignore the anomalous blood pressure readings shown on the monitor for one baby because all the other available cues indicated that the baby's blood pressure was normal. Although FLORENCE does not use blood pressure when computing advice for setting the ventilator, this example does raise a general point about clinical decisions to ignore some reading as false that may not be accessible to an advisory system that takes data directly from monitoring equipment.

3.1.2 Implications for FLORENCE

The data in the NICU is inherently noisy which gives rise to a high level of false alarms [16], so it is important that FLORENCE should not exacerbate the situation by propagating existing alarms. FLORENCE will generate its own audible alarm when a ventilator intervention is required, so it is important that it should avoid generating false alarms. FLORENCE's alarm also needs to be easily distinguishable from the others, and prioritised so that staff respond appropriately.

Any automatic advisory system that takes data directly from monitoring equipment is particularly susceptible to problems of unreliable data. The interviews suggest two instances where such problems might arise: (i) the DOPE mnemonic was devised to test for possible false readings from the ventilator and other equipment; (ii) a ward round decision to ignore blood pressure readings. When data is manually entered into an expert system the operator will naturally check their validity; when data are read automatically this validation is less likely to happen. The procedures adopted with FLORENCE should ensure that the data the advice is based upon is valid. For example, FLORENCE could prompt the operator to carry out the DOPE checks before implementing any suggested changes to the ventilator settings. For the same reason, it is also good practice for advisory systems always to provide an explanation of why some particular advice is given along with the data on which it is based.

3.2 Lightweight Rich Pictures

Semi-structured informal interviews were carried out by the first author—an experienced interviewer—with eight members of staff. These ranged from trainee nurse to medical consultant from the clinical care side, and included a ward clerk from the administrative side. Interviewees were asked questions about their levels of experience, their roles and responsibilities in the working of the unit and any concerns that they had. The interviews, each of which lasted about an hour, were transcribed and then analysed to identify how the roles and responsibilities of the staff overlap and interleave to support work in the unit. The two central aspects of work that emerged from the analysis were the communication in and around the unit, and the way in which records are used. Rich pictures and accompanying textual descriptions were used to capture these aspects of work in the unit.

3.2.1 Communication

The communication links between the people on the unit are shown in Figure 1, which also gives a qualitative indication of their reported frequency of use. These frequencies can be used to provide a simple way of assessing the impact of FLORENCE on work in the unit. If the frequencies are significantly affected by FLORENCE, the qualitative nature of the changes may need to be investigated to identify whether the changes are adversely affecting work in the NICU.

Full-time staff are explicitly shown in the figure, while those peripherally connected with the unit are shown as a single, separate group. Normally, staff communication follows an informal hierarchy: the first line of communication for a Nurse, for example, would be a Senior Nurse.

There is also a lot of communication through the use of records and simply noting what others are doing.

Figure 1 About Here

Decision making on the unit also normally follows a hierarchical structure based on the complexity of the decision and individual competence. So, for example, a Nurse may consult an SHO, then a Registrar, and ultimately a Consultant. Parents are also involved in decision making as the babies are moved out of the NICU.

The ward round is a major source of communication, and is the main source of in-service training and feedback provided by the consultants. Normally led by a consultant, it is attended by up to three registrars, the SHO from the night shift, the SHOs from the new day shift, one or two nurses, and often the unit's pharmacist too. Individual cases are discussed, and the results used to make clinical decisions and define care plans for the next 24 hours. These are recorded in the Doctor's Notes, and any deviations are noted on the Patient Charts.

In addition to the formalised handover encapsulated by the ward round, three other handovers also occur. The nurses, registrars and SHOs each have their own, more informal, handovers.

Staff also communicate through the medium of formal meetings. Sisters' meetings are held monthly, and unit meetings, which anyone who works on the unit can attend, are held infrequently. The latter deal with the everyday running of the unit.

3.2.2 Records

Many of the records kept by the unit serve as another mechanism for communication, because they provide a historical picture of a patient's status and treatment. Staff's use of these records is shown in Figure 2, which also gives a qualitative indication of the reported frequency of access to these records.

Figure 2 About Here

The Doctor's Notes are kept in folders updated by the SHO on the night shift. Events are added to the notes, such as details of long lines, arterial lines, and management decisions. Although the outcomes from the ward round are written in the Doctor's Notes, plans are only communicated verbally. These plans often have to be revised dynamically as staff react to events and re-prioritise tasks, to ensure that none get neglected indefinitely.

The Patient Charts contain comprehensive information, including details of vital signs, hourly observations, arterial blood gas levels, drugs, fluid input/output and blood sugar levels. In addition to the continuous monitoring of the babies' condition, both via the technological equipment and their physical appearance, clinical staff also normally make hourly observations; depending on the patient's status, may be less frequent (2, 3, or 4 hourly).

The doctors' Problem Sheets are held with the Doctor's Notes. These are used to record active problems. A similar scheme is being introduced for the nurses; these sheets will be kept with the Patient Charts.

Results of laboratory tests are available via the computer in the SHO's room. Details from the pathology database are transcribed into the Results Book, which gets included with the Doctor's Notes. The copying of test results is mainly performed by the SHOs.

In addition to the Doctor's Notes, there is also a Daily Update Book which is normally filled in by a Registrar during the ward round. It is basically a selection of tick boxes describing, for example, how the baby was fed, and what ventilation and antibiotics it received.

Whilst the SHOs maintain the Doctor's Notes and the Patient Charts, the nurses maintain a Cardex. This is usually written up retrospectively at the end of the shift. In certain circumstances, such as after a resuscitation, it is written directly following the event. In such cases the administered drug levels are called out and others will note the amounts involved, so that they can subsequently be recorded in the Cardex.

The Ward Clerks use the Daily Update Book to generate the Discharge Sheets when babies leave the unit. Much of the detail on the discharge sheets is written by the SHOs. These sheets are mainly read by clinicians, such as the baby's family doctor, although the registrars may consult them if the baby subsequently attends the weekly outpatient clinic at the unit. Patient documentation has to be kept for several years, due to legal requirements.

3.2.3 Implications for FLORENCE

Communication plays a central role in the operation of the NICU. It is therefore important that FLORENCE does not detract from existing lines and levels of communication.

Consideration needs to be given to how FLORENCE will integrate with the existing extensive paper records, and support non-verbal communication. Where FLORENCE is used, its decisions need to be recorded and incorporated: FLORENCE should supply a reason to accompany the suggested changes. If the user decides to override FLORENCE's decision, they should supply their own reason to support this. These reasons, which will be logged by FLORENCE, also need to be incorporated into the patient's record, since changes to the ventilator settings are currently logged manually on the patient chart. This also implies that it must be possible to print data from FLORENCE.

Anonymous reviewers of this paper have commented on the large number of records used in the NICU. Redundancy may be valuable as it increases access and makes possible the checks and balances included within daily procedures. However, if there are undetected inconsistencies in the data in parallel records it may lead to problems. It is important that the extra information that needs to be recorded from FLORENCE does not increase any problems that may already exist, indeed, its introduction may be a good opportunity to re-examine the need for all these records.

Changes proposed by FLORENCE have to performed manually. It is therefore important that these suggestions are displayed in a font that staff can easily read when standing in front of the ventilator. To avoid any ambiguities about the proposed changes, FLORENCE should show the size, direction and the new final value of the change, such as “PIP Up by 2 to 16”.

There is a need to provide accommodation for the PC that will be running FLORENCE. It needs to be located such that staff can see the display when standing next to the ventilator, without interfering with other staff or equipment. There will thus need to be space available for positioning FLORENCE next to the cot, with enough room to be able to use a mouse—consideration could also be given to using a touch screen display. There will also need to be an electrical power supply available. This implies that there could be problems with trailing power cables. Although these details appear somewhat trivial, they are often found to be critical to the acceptability of technology.

4 Critical Decision Method

The CDM has previously been successfully used in a neonatal setting to investigate decision making during the diagnosis of sepsis [17]. The purpose of the case study described here, however, was to identify issues that could inform the development and deployment of FLORENCE. The basic aim of FLORENCE is to empower the decision making of the clinical staff who provide the front line of care—SHOs and nurses—who interact with the ventilators. These people are collectively described below as front line carers rather than novices, because they have a higher level of knowledge and experience than is typically associated with novices.

Given the aims of FLORENCE, it was decided to interview staff with different levels of experience to identify differences in decision making when using the ventilator, and how FLORENCE could reduce these differences. Interviews were carried out with eight members of staff: four experts—two consultants, and two registrars—and four front-line carers—two senior nursing staff, and two SHOs. This was an opportunity sample, based on staff who volunteered to take part. The small sample sizes reflect the small parent populations (3-6 people in each category).

The interviewees' experience is summarised in Table 1. It should be noted that there is some gradation within the two categories. Among the experts, registrars will normally be called in before consultants—the consultants are more expert than the registrars. Among the front line carers, some of the senior nurses are very experienced in using the ventilator.

Table 1 About Here

At the start of the interviews, the CDM process was described to participants. The steps are summarised in Table 2. Interviews were recorded using a minidisc, and hand-written notes were simultaneously taken. One interview failed to record due to mechanical failure.

Table 2 About Here

The use of information cues by staff, and their situation assessment and decision making are discussed below. These features reflect the general structure of the iterative perceive-decide-act cycle that is typical of many control situations—here the controlled process is the baby's respiration. Initially, observations of the relevant information cues in the environment (such as current blood gas levels) are made. The current situation is assessed by comparing the perceived values with a set of expected values and deciding what action (if any) to take. If there is a difference between the perceived and expected values, some action is normally taken to stabilise the process. The selected action (which can be no action) is then performed.

The selected incidents represent a wide range of complexity, so it is not possible to always make direct detailed comparisons. Generally, the incidents described by front line carers were less complex, and occurred more frequently than those described by the experts.

4.1 Information cues

Experts are normally called in when front line carers encounter difficulties with a particular case. Although most of the reported incidents were described as typical, they were not usually frequent, with one expert noting that the selected incident was a particularly bad case. This highlights the fact that the same problem can present with different levels of severity; this is normally determined by consulting the appropriate cues.

For each incident a cue inventory was drawn up. The example in Table 3 shows some of the many sources of cues that are typically used in diagnosing the baby's condition. In addition to the equipment data (blood gases, ventilation, vital signs, and equipment), staff also checked the baby's external condition (skin colour, chest movements) and internal condition (X-rays, auscultation). They also checked the mechanics of the ventilator circuit to ensure that everything was working correctly. Experts additionally utilised verbal cues from other staff.

Table 3 About Here

In general, experts tended to use more cues than front line carers. This can be explained by the need for more information to diagnose more complex cases. Experts also utilised the same cues as front line carers. This appears to be partly as a check, and partly to ensure that the problem had not changed between the time they were initially called and when they arrived on the unit.

None of the front line carers mentioned checking the MAP. This is a relatively complex derived quantity represented by the area under the breathing wave form curve which can be shown graphically on the ventilator. Its adjustment is implicitly important in controlling the baby's PO_2. Some front line carers did talk about minute ventilation, however, which is also implicitly important in controlling the baby's PCO_2. Minute ventilation is a simpler derived quantity than MAP, in that it is just the product of the tidal volume and the breathing rate.

Staff often carried out cross-checks to validate the available evidence. The most obvious example was checking that equipment readings were consistent with the baby's physical appearance. In general, staff were very confident in the accuracy of the ventilator data, and only slightly less confident in the Neotrend data. The quality of the Neotrend data decays over time, so manual blood gas analyses are used to validate and recalibrate it. In addition, the position of the tip of the Neotrend can affect the readings, so it is sometimes checked using an X-ray.

Several staff reported talking out loud as they went through the decision making process. This has several functions. First, it provides an opportunity for other staff to comment. Second, it imparts knowledge about the decision making process to less experienced staff. Third, it provides a verbal audit trail, which could be useful in future reflections on the incident, because all staff who were present were kept informed of what was being done.

4.2 Situation Assessment and Decision Making

Whenever experts were called in, they reported that they always reviewed the situation on arrival. In other words, they always checked the data values shown by the equipment, the general condition of the baby, and took verbal reports on the progress of the problem. The process of assessing the situation and making decisions is an iterative one. It is summarised by a situation assessment record, an example of which is shown in Table 4. The initial situation assessment, labelled SA-1, consists of checking several cues. An expectation is then generated about what would happen if no changes were made. Goals are then set to identify what needs to be done, and a decision is taken about the actions required to achieve those goals (DP-1 in the table). The situation assessment can then be elaborated as the situation changes in response to any actions (SA-2 in the table). If some new important information comes to light which requires a reappraisal, a shift in situation assessment occurs (as illustrated by SA-3 in the table).

Table 4 About Here

In two of the incidents described by experts there was a shift in situation assessment. In each case this shift was caused by a new piece of information which was counter to the prevailing diagnosis of the baby's condition. In the first case, the new information was the sounding of the ventilator alarm. In the second, it was the results of the manual blood gas analysis which confirmed that the baby was being overventilated. In both cases the net effect was that further changes were made to the way that the ventilator was being used.

Although each incident only included a few critical decisions (usually 3 or 4), experts tended to be less cautious than front line carers when making an intervention. This finding, which is in line with research in other aspects of medical care [18, 19], can be partly explained by the fact that when experts are called in, several of the simpler possible diagnoses have been tried and (usually) ruled out. In general, arriving at the decision to change the ventilator settings involves a process of elimination. Most staff described using the DOPE mnemonic to remind them to check for alternative causes of the problem before adjusting the ventilator.

Once the decision to intervene using the ventilator has been made, a number of local procedures are followed. Generally, the nurses will only adjust the inspired oxygen and, possibly, the breathing rate. There is also a rule of thumb that suggests that if the pressure (PIP or PEEP) has to be increased, it is usually best to increase it by 2.

The MAP can be increased in several ways. Extended use of high pressures can damage the baby's lungs, so unless the ventilator pressures (PIP or PEEP) are noticeably low, it is usually preferable to adjust the other ventilator settings. In acute situations, however, the FiO2 and rate may be near their maximum acceptable limits by the time the experts are called in, leaving little room for manoeuvre. They therefore have to find alternative ways of achieving the desired effect, such as adjusting the inspiratory and expiratory times, or changing the pressure settings.

4.3 Implications for FLORENCE

FLORENCE is designed to empower the decision making of front line carers when dealing with RDS by reducing the length of time it takes them to learn to make more advanced decisions. It is not, however, intended to undermine the hierarchical channels of communication identified in the rich pictures. The re-design of working practices to incorporate FLORENCE, that will be reinforced through its user interface, must thus ensure that expert help is called on when appropriate. For example, there needs to be some way of discriminating between cases of RDS and more complex cases where FLORENCE's advice may not be appropriate, such as the case of persistent pulmonary hypertension identified in the expert's account above. In this way, FLORENCE can support the unit's existing hierarchy of decision making and communication, ensuring that appropriate cases still get referred to the relevant authority.

Both experts and front line carers reported having to take account of the current settings before deciding on changes to the ventilator. So, for example, when the normal solution required increasing the FiO_2, which was already close to its upper limit, an alternative set of changes that would bring about the same solution had to be used. FLORENCE therefore needs to take appropriate account of the current settings on the ventilator (which can be downloaded on request) to establish which settings can be changed. In such cases, FLORENCE should suggest an appropriate alternative course of action, which could include calling in a registrar or consultant.

All levels of staff reported using data from the ventilator and the Neotrend. FLORENCE combines (and can display) data from the ventilator and the Neotrend, which means that there is some potential data redundancy. The ventilator and Neotrend displays will both still be available, so guidance needs to be given to staff about when to consult the different displays.

If FLORENCE is to be effective in teaching front line carers to take more advanced decisions, they must be able to understand the reasoning behind the advice to change ventilator settings. This is another reason for requiring FLORENCE to display explanations along with the data used to make the decision.

5 Observation

When a baby arrives on the unit with RDS, it is normally connected to a ventilator for between five and seven days. Since RDS normally peaks within 72 hours of birth, it was decided to focus the observation on the first few days after birth, in an attempt to maximise the likelihood of observing front line carers interacting with the ventilator.

A preliminary observation period was used to scope the requirements for observation, and to provide further insights into the domain of neonatal intensive care. It was decided to record the observations manually, because this is the least intrusive method: the NICU is a highly stressful place for parents and babies, particularly immediately after birth.

5.1 Case studies

Two babies were observed, starting on the second day after their admission to the unit. Each observation session lasted approximately two hours. All the sessions were carried out at the same time of day, during the afternoon. The first baby was observed for only one day; the observation was stopped when the Neotrend was removed by the doctors. The results are included here because they still offer valid insights into use of the ventilator and the Neotrend that can help to inform the development of FLORENCE. The second baby was observed for a total of three days.

Observations of alarm related events where there was no interaction by staff are summarised in Table 5. Most of the alarms were intermittent (43 out of 47), in that they only rang briefly (in some cases only once), and then cleared without any intervention from staff. The occurrence of an alarm, and its spontaneous clearance were recorded as separate events.

Table 5 About Here

Table 6 shows how many times staff interacted with the various pieces of equipment and the number of other actions or events that occurred, such as checking the baby's condition. The interactions include reading data as well as physically adjusting the controls. Although many interactions occur in response to alarms, in some cases staff tried to anticipate and prevent alarms. So, for example, when a front line carer was going to suction a baby's ETT, they knew that this often causes a ventilator alarm, so they adjusted the ventilator settings beforehand, increased the inspired oxygen level, and muted the audible alarm. Depending on how long the procedure took, they also had to mute the alarm again.

Table 6 About Here

Taken together, the tables show that approximately two-thirds of the observed actions and events are related to the equipment used in caring for the baby. If anything, the numbers are likely to be underreported. In some instances where an audible alarm sounds, the allocated nurse may be dealing with another baby, but may glance at the cardiac monitor[1], for example, from across the room. It is not always possible to capture such events, particularly when there may already be another member of the clinical staff attending to the baby.

Staff interacted with the ventilator, the cardiac monitor and the Neotrend in all of the sessions. This is to be expected, since the hourly observations involve checking and recording the current status of these devices. In every instance except one, however, the number of interactions was higher than the number of interactions observed as part of the routine recording of hourly observations by the staff. The only exception is for the Neotrend on day 3 of case 2.

5.2 Implications for FLORENCE

Most of the alarms observed cleared automatically, without any intervention from staff. However, the way FLORENCE is intended to be used is rather different. When FLORENCE raises an alarm staff will normally be required to intervene and make a clinical judgement, even if the ventilator setting are not subsequently changed. For this reason: (i) the audible alarm used for FLORENCE should be clearly distinguishable from the others; (ii) new procedures should be established to ensure that FLORENCE alarms receive attention; (iii) false alarms from FLORENCE should be minimised. One way of reinforcing (ii) would be to ensure that FLORENCE alarms never self cancel.

6 Conclusions

The particular combination of methods selected for the CTA used here proved to be successful in meeting both of the goals of the study. The implications for the development and use of FLORENCE are summarised in Table 7, which also shows the particular method that was used

[1] The cardiac monitor shows the baby's heart rate, blood pressure and oxygen saturation levels.

in generating the implication. Note that each implication is only shown once, even where it was identified by more than one of the methods.

Table 7 About Here

The methods clearly identified the fundamental importance of communication (both verbal and non-verbal) in the functioning of the NICU. Many of the implications in the table are related to maintaining the existing levels of communication by ensuring that FLORENCE does not unnecessarily burden staff with extra false alarms, for example.

The implications cover a wide range of issues that affect the clinical utility of FLORENCE, and its acceptability. The former can be addressed as part of the development of FLORENCE; the latter relate to its deployment and how it can be integrated into the work of the NICU without adversely affecting the practice of neonatal intensive care.

The value of applying a variety of methods is clearly demonstrated in Table 7. The chosen methods have identified a set of implications that go beyond the functional capability of FLORENCE to calculate changes to the ventilator settings. The fact that several of the implications were identified as the results of applying more than one of the methods serves to confirm the importance of those implications.

Rich picture analysis is not normally mentioned in texts on CTA. On the basis of its usefulness here, in identifying the underlying work structures in the practice of neonatal care and generating implications for the acceptability of FLORENCE, it should be considered more often when carrying out a CTA. In particular it has helped to identify the areas where FLORENCE can provide support for current practice (such as prompting staff to follow the DOPE mnemonic), where it could change existing practice (such as the need for the data records produced by FLORENCE to be incorporated into the patient records, and highlighting possible data redundancy issues), and more traditional ergonomic issues (such as the need for space to accommodate FLORENCE, and the need for clear, legible displays that can be unambiguously read when staff are standing at the ventilator).

The implications identified here are to be incorporated into the development and the ensuing deployment of FLORENCE. Some initial work has already been carried out to compare the decision making abilities of FLORENCE with a group of neonatal care specialists [6]. When the implementation of FLORENCE is complete, the findings described above will be evaluated during the testing of FLORENCE, and its early stages of deployment.

Acknowledgements

The authors would like to thank the staff at SJUH who voluntarily participated in the study. Steve Smye provided extensive comments on an earlier draft of the paper, and Jennifer Wills transcribed the interviews.

This study is part of the Dependability Interdisciplinary Research Collaboration (DIRC) funded by EPSRC grant number GR/N13999. Development of FLORENCE was funded by the Sports Aiding Medical Research for Kids (SPARKS) charity.

7 References

[1] Reid T, Freer Y. Developmentally focused nursing care. In: Boxwell G, editor. *Neonatal intensive care nursing*. London: Routledge, 2000:14-42.

[2] Levene MI, Tudehope D. *Essentials of neonatal medicine*. 2nd ed. London: Blackwell Scientific, 1993.

[3] Snowden S, Brownlee KG, Dear PR. An expert system to assist neonatal intensive care. *Journal of medical engineering technology* 1997;21(2):67-73.

[4] Philips. Trendcare continuous blood gas monitoring system: Technical Overview. Böblingen, Germany: Philips Medical Systems, 2001.

[5] Morgan C, Newell SJ, Ducker DA, Hodgkinson J, White DK, Morley CJ, et al. Continuous neonatal blood gas monitoring using a multiparameter intra-arterial sensor. *Fetal & Neonatal* 1999;80(2):F93-F98.

[6] Tan K, Baxter G, Brownlee KG, Newell SJ, Dear PRF, Smye S. Fuzzy logic expert system for ventilation of the newborn infant. *Archives of Disease in Childhood* 2003;88(Supplement 1):A47.

[7] Berg M. Implementing information systems in health care organizations: myths and challenges. *International Journal of Medical Informatics* 2001;64:143-156.

[8] Tan K, Snowden S, Evans C, Baxter G, Brownlee KG, Newell SJ, et al. Fuzzy logic expert system for neonatal ventilation. In: Jervis BW, editor. *Proceedings of the Fifth International Conference on Neural Networks and Expert Systems in Medicine and Healthcare (NNESMED 2003) and 1st International Conference on Computational Intelligence in Medicine and*

Healthcare (CIMED 2003). Sheffield, UK: School of Engineering, Sheffield Hallam University, 2003:61-4.

[9] British Association of Perinatal Medicine. Guidelines for good practice in the management of neonatal respiratory distress syndrome: Report of the second working group of the BAPM. London: BAPM, 1998.

[10] British Association of Perinatal Medicine. Standards for hospitals providing neonatal intensive and high dependency care (Second edition) and Categories of babies requiring neonatal care. London: BAPM, 2001.

[11] Redshaw ME, Harris A, Ingram JC. *Delivering neonatal care*. Norwich, UK: HMSO, 1996.

[12] Chipman SF, Schraagen JM, Shalin VL. Introduction to cognitive task analysis. In: Schraagen JM, Chipman SF, Shalin VL, editors. *Cognitive task analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000:3-23.

[13] Klein G, Calderwood R, MacGregor D. Critical decision method for eliciting knowledge. *IEEE Transactions on System, Man, and Cybernetics* 1989;19:462-472.

[14] Hoffman RR, Shadbolt NR, Burton AM, Klein G. Eliciting knowledge from experts: A methodological analysis. *Organizational Behavior and Human Decision Processes* 1995;62(2):129-158.

[15] Monk AF. Lightweight techniques to encourage innovative user interface design. In: Wood L, editor. *User interface design: bridging the gap between user requirements and design*. Boca Raton: CRC Press, 1998:109-129.

[16] Miksch S, Seyfang A, Horn W, Popow C. Abstracting steady qualitative descriptions over time from noisy, high-frequency data. In: Horn W, Shuhar Y, Lindberg G, Andreassen S, Wyatt J, editors. *Artificial Intelligence in Medicine. Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making, AIMDM'99, Aalborg, Denmark, June 1999, Proceedings. Lecture Notes in Computer Science 1620.* Berlin: Springer, 1999:281-290.

[17] Crandall B, Getchell-Reiter K. Critical decision method: A technique for eliciting concrete assessment indicators from the intuitions of NICU nurses. *Advances in Nursing Science* 1993;16:42-51.

[18] Boreham NC, Mawer GE, Foster RW. Medical diagnosis from circumstantial evidence. *Le Travail Humain* 1996;59:69-85.

[19] Boreham NC, Foster RW, Mawer GE. Strategies and knowledge in the control of the symptoms of a chronic illness. *Le Travail Humain* 1992;55:15-34.

8 Appendix A: Critical Decision Method Probe Questions

The basic probe questions that were used during the application of the critical decision method are listed below. The critical decision method is a semi-structured interview technique. Where appropriate, interviewees were asked further questions to provide more detail on particular aspects of the described incident.

The questions are divided into two lists. The first are general questions about the situation that the interviewee described; the second are the more detailed questions which were asked for each of the decision points identified during the described incident.

General questions
- Data types, variables etc.
  - What were you seeing?
  - What were you hearing?
- Analogies to past cases
  - Were you reminded of previous similar experiences?
  - Which of them is most similar?
- Standard scenarios
  - Is this case a typical scenario?
  - Is it the sort of scenario you were trained to deal with?
  - How many incidents like this have you been involved with?

Decision point questions
- Available options:
  - What other options were available?
  - What other options were considered?
- Rationale for selection of options:
  - How was the option selected/rejected?
  - What rule were you following?
  - Were you conscious of making a decision?
- Shifts/changes in situational awareness:
  - If you had to describe the situation to someone taking over on the next shift how would you summarise it?
- Information requirements at given points:
  - What information did you use in making this decision?
  - How was it obtained?
  - What training or experience was necessary or helpful in making this decision?
  - How confident were you in the information?
- Missing information:
  - If the decision was not the optimal one what knowledge or information that was not available might have helped?
- Speculation:
  - If a particular feature (identified in the incident) had been different, how would it have affected your decision?
- Goals:
  - What were your goals at this point?
  - What were your command objectives?
- Imagery/mental simulation:
  - Did you imagine the consequences of the action?

Did you imagine the events that would transpire given the situation and information available?

Errors:

Did you acknowledge if your situation assessment or option selection were correct?
How might a novice have behaved differently?
What mistakes are likely at this point?

Potential aiding:

If decisions were not optimal, what knowledge, or information was missing which could have aided decision making?

Decision making time/effort:

How much time pressure was involved in making this decision?
How long did it actually take to make this decision?

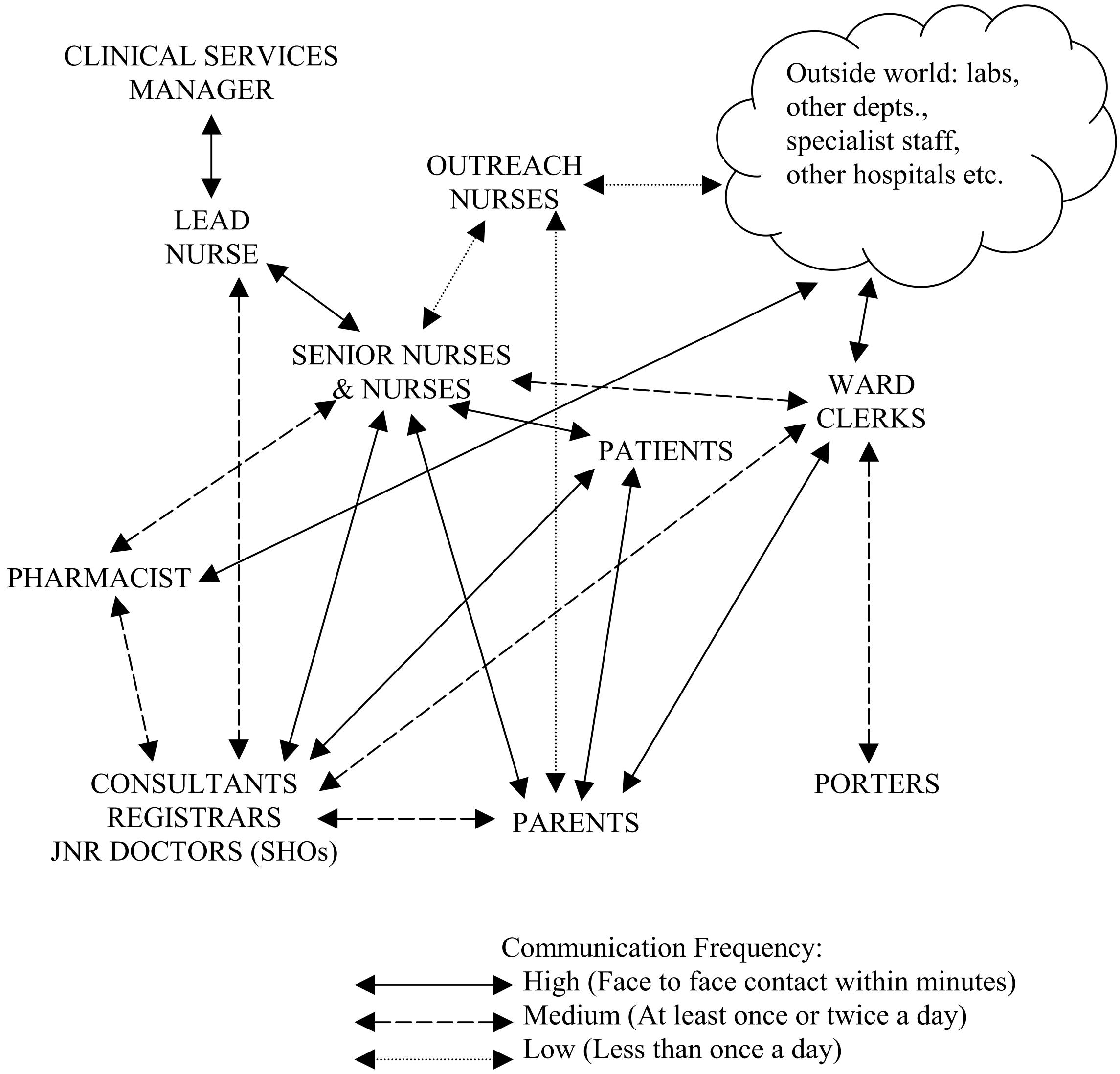

Figure 1. Who talks to whom. Rich picture showing communication links and their frequency of usage between the various staff working in or with the neonatal unit.

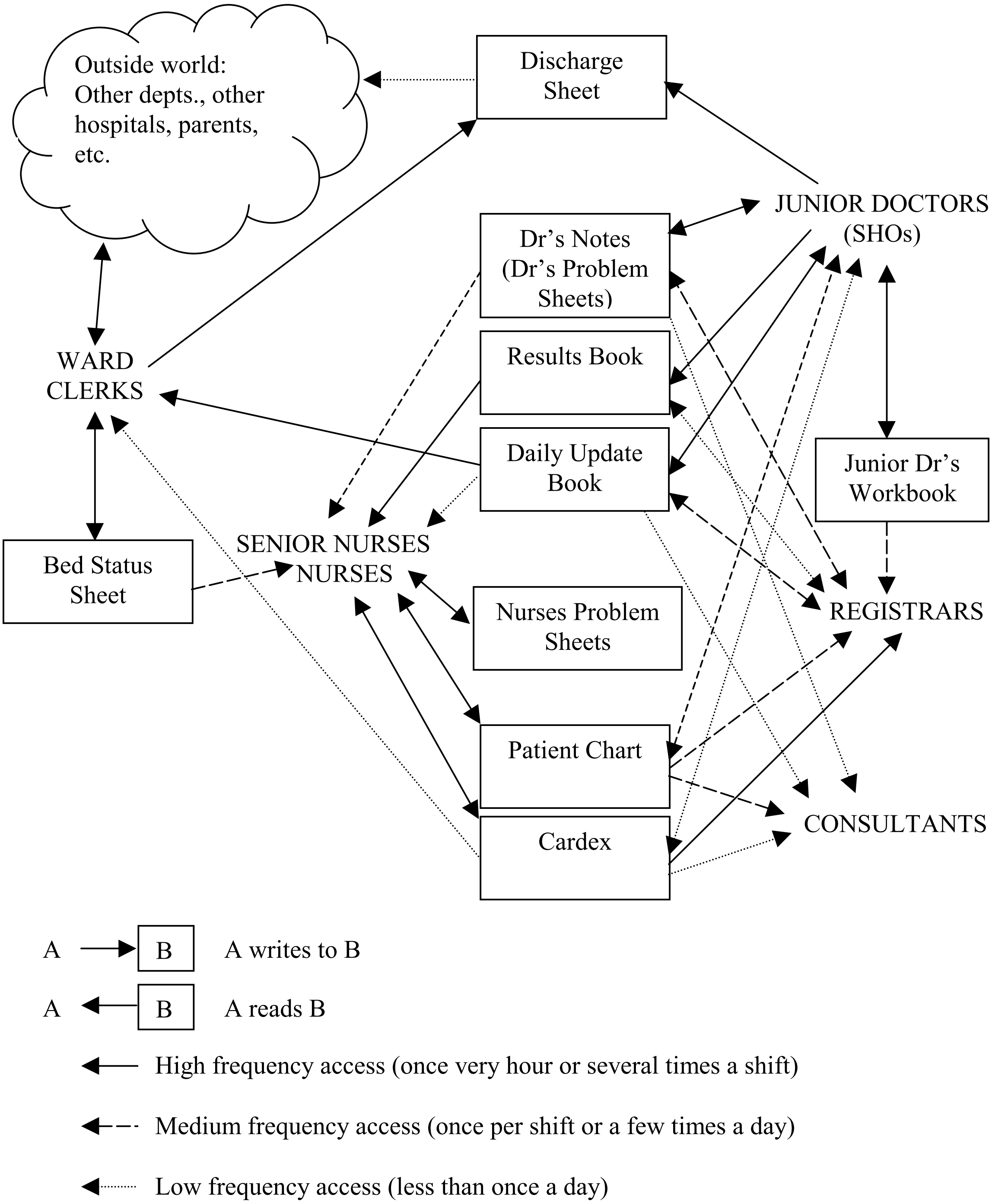

Figure 2. The use of written records. Rich picture showing which staff members access which data records, and how often. (Doctor's problem sheets are bracketed with the doctor's notes, because they are stored together in the unit.)

Table 1. Summary of experience of interviewees. Front Line Carers are those staff who routinely deal with the babies in the first instance (nurses and SHOs).

Interviewee	Relevant experience
Expert 1	> 20 years
Expert 2	> 20 years
Expert 3	> 3 years
Expert 4	> 3 years
Front Line Carer 1	> 20 years
Front Line Carer 2	< 1 year
Front Line Carer 3	> 20 years
Front Line Carer 4	< 1 year

Table 2. Summary of the phases of the CDM

Phase	Description
Instruction	The rationale of the study was explained. Participants were given a general introduction to knowledge elicitation, including an overview of the ideas underpinning the CDM.
Incident identification and selection	Participants were guided to identify cases (incidents) where they were involved, where time was a factor, and they had to interact with the ventilator.
Event recall	Participants were asked to describe the selected incident in detail. As they described the incident, the interviewer drafted out a timeline for the incident to identify important events, decisions and actions.
Timeline verification Decision point identification	When participants had finished recalling the incident, the interviewer used the timeline from the Event Recall phase to describe the incident to the participant. It was simultaneously used to identify (and verify) the locations of the decision points that provided the basis for the Event recall/Probe questioning phase.
Event recall/Probe questioning	A semi-structured interview was conducted where the interviewer tried to identify in detail what was happening around the decision points. Participants were guided back through the incident using the timeline and a set of questions designed to probe for more information. (The basic set of probe questions appears in Appendix A.)
Data analysis and verification	The data that was collected was analysed and transformed to generate a cue inventory, a situation assessment record, and a list of temporal issues.

Table 3: Critical Cue Inventory for an incident reported by Expert2

Cue Category	Description
Blood gases	PO2; PCO2
Ventilation	FiO2; MAP, Ti, Te
Vital signs	Blood pressure (if it is too low, then oxygen levels are not of primary importance, because the oxygen is not reaching the brain at a timely rate anyway)
Tube	Tube size (leakage occurs if too small)
Skin	Colour
Lungs	Lung volume
Chest	X-ray (to check position of Neotrend probe)
Equipment	Neotrend alarms (PO2, PCO2); Audible alarms silenced; Calibration
Verbal reports	From other staff

Table 4: Situation Assessment Record (SAR) for critical incident reported by Expert2

<u>SA-1</u>

Cues/knowledge:	Verbal reports from other staff; neotrend display; neotrend alarms (low PO2; high PCO2); neotrend probe position checked via X-ray
Expectations:	Baby's condition would remain unstable and worsen
Goals:	To increase tidal ventilation; to increase MAP
<u>DP-1</u> :	Increase PIP by more than normal

<u>SA-2</u> (Elaboration)

Cues/knowledge:	Flow displays on the ventilator; blood pressure; ventilator alarms (low PO2, high PCO2)
Expectations:	Baby would take too long to stabilise
Goals:	To increase oxygenation
<u>DP-2</u> :	Increase TI

<u>SA-3</u> (Shift)

Cues/knowledge:	Ventilator alarm indicates leakage in ETT circuit; No leak in the tubes or connections was found
Expectations:	Leak would persist; Baby's condition would not improve at desired rate
Goals:	To eliminate the leak in the ventilator circuit and to ensure that the oxygen was reaching the lungs.
<u>DP-3</u> :	Replace ETT with tube of a larger diameter

Table 5. Summary of observed alarms within two hour observation period. The cardiac monitor alarms encompass ECG, BP and Oxygen saturation alarms. The first figure in each column shows the number of alarms observed. The second figure, in brackets, is the number of alarms that cleared without any staff intervention. The total row shows the sum of the alarms and the spontaneously cleared alarms.

Alarm Events	Case 1: Day 1	Case 2: Day 1	Case 2: Day 2	Case 2: Day 3
Ventilator alarm	2 (0)	15 (13)	4 (4)	3 (3)
Cardiac monitor alarm	7 (7)	0 (0)	1 (1)	3 (3)
Neotrend alarm	2 (0)	3 (3)	0 (0)	3 (3)
Humidifier alarm	0 (0)	0 (0)	0 (0)	2 (2)
Platform alarm	0 (0)	0 (0)	0 (0)	4 (4)
Total alarm related events	18 (7)	34 (16)	10 (5)	30 (15)

Table 6. Summary of observed actions taken by staff excluding dealing with alarms within two hour observation period.

Actions/Events	Case 1: Day 1	Case 2: Day 1	Case 2: Day 2	Case 2: Day 3
Ventilator	3	21	10	13
Cardiac monitor	7	5	6	4
Neotrend	3	5	6	3
Humidifier	0	0	0	1
Platform	1	1	2	1
Total interactions	14	32	24	22
Other actions/events	15	26	22	27

Table 7. Summary of the implications for the development and deployment of FLORENCE identified by the different CTA methods.

CTA Method	Implications
Domain and context familiarisation (meetings, visits and rich picture analysis)	Staff need to be aware of any contingencies (such as ignoring anomalous data) before implementing FLORENCE's suggestions.
	FLORENCE should prompt staff to follow the DOPE mnemonic.
	The FLORENCE audible alarm needs to be distinctive.
	Staff need to be trained how to respond to a FLORENCE alarm.
	The size of the text used to display FLORENCE's advice needs to legible when staff are stood at the ventilator.
	The wording of FLORENCE's suggestions needs to be clear and unambiguous (e.g., PIP Up by 2 to 16)
	Space needs to be made available for the PC running FLORENCE, and a mouse (unless a touch screen is used).
	Consideration needs to be given to what data from FLORENCE needs to be included with existing paper records.
	It must be possible to print data from FLORENCE for inclusion with the other patient records.
	There needs to be a power socket available for the PC running FLORENCE.
Critical Decision Method	The limitations of FLORENCE need to be made explicit to staff.
	FLORENCE should check the current ventilator settings before deciding what changes are required.
	Staff need to be made aware of the potential data redundancy problem, because FLORENCE will display trends of data that are available on the ventilator and Neotrend.
	FLORENCE should be able to explain its decisions on request.
	Staff should be able to override FLORENCE's suggestions, as long as they can supply a reason for doing so.
Observation	FLORENCE should attempt to avoid generating intermittent alarms.