INTELLIGENT DECISION SUPPORT TOOLS FOR THE MANAGEMENT OF LABOUR: THE INFANT & DATA CARE EXPERT SYSTEMS

E C Ifeachor, J M Garibaldi and J Skinner

Introduction

In obstetrics, two major problems are:

1. how to use electronic fetal monitoring technology to accurately assess the condition of the fetus during labour to avoid unnecessary medical intervention (e.g. Caesarean and forceps deliveries) or a failure to intervene when necessary, and

2. how to assess the health of the baby immediately after birth in an objective and consistent way to guide immediate neonatal care and to provide feedback to obstetric clinicians to improve future management of labour.

Electronic fetal monitoring, introduced in the late 1960's, is based on the visual analysis of the cardiotocogram (CTG) which is a continuous trace of the fetal heart rate patterns and the uterine contractions (see Figure 1). Considerable expertise is required to interpret the complex changes seen on the CTG to accurately identify the fetus coping well with the stress of labour from the truly compromised fetus and to decide when to seek further information or intervene. Difficulties in the interpretation of the CTG during labour have led to incidents of unnecessary intervention as well as a failure to intervene when necessary. This may be an important factor in the number of infants who sustain potentially preventable brain damage or die because of events in labour. It is estimated that one third of Caesarean section deliveries are unnecessary (each costing about £1500). The cost of birth-related negligence claims is said to exceed £400 million each year in the UK alone.

![Figure 1 An example of the cardiotocogram (fetal heart rate pattern and the uterine activity).](image-url)
Existing methods of assessing the health of the newly-born infant are based on subjective measurements, such as the Apgar score, or on diagnosis of brain damage after birth. Neither of these distinguishes damage that occurred during labour from damage that occurred before or after labour. Analysis of the chemical acid-base status of blood taken from the umbilical cord of an infant, immediately after delivery, provides a measure of the response of the fetus to labour, and information on possible damage suffered by the infant due to lack of oxygen during labour. However, umbilical acid-base assessment is complex and error prone, and require significant expertise to carry out.

In both problems, the effectiveness of the available technology depends on the experience of the clinical staff responsible for routine care of patients and their ability to analyse, interpret and assess the significance of changes in the data produced by the equipment. Such expertise is not always available on labour wards day and night. In this paper, we describe two intelligent systems which have been developed to provide an objective interpretation of data, to match available technology to the needs of the user, and to assist in decision making. The INFANT (Intelligent Fetal Assessment) interfaces to electronic fetal monitors for objective analysis of CTG and provides advice to clinicians on the appropriate management of labour. The DataCare Expert system interfaces to a blood gas machine to validate and analyse the chemical acid-base status of blood taken from the umbilical cord immediately after delivery. An aim of the paper is to share our experience of the development and use of the two systems.

In both systems, our approach is to model the processes of judgement and intelligent interpretation of data used by human experts in clinical practice. The performance of such systems is likely to match that of an experienced clinician which should encourage widespread acceptance. This approach requires us to work closely with clinical experts to elicit knowledge which is then encapsulated within an automated framework.

The INFANT System

Previous computerised CTG analysis systems have gone some way in overcoming the inconsistencies associated with visual analysis, but they are limited in value because most rely on the CTG alone. The correct assessment of fetal condition does not depend on changes in heart rate alone. It also requires a knowledge of fetal physiology, fetal blood sampling, relevant patient medical history, the dynamics of the individual labour, as well as expert judgement, based on knowledge and experience. To be clinically useful, a practical decision-support tool for labour management should capture and encode such knowledge and expertise.

Over the past 8 years we have been developing the INFANT system to meet the above needs. The INFANT (see Figure 2) combines the strengths of various artificial intelligence techniques with conventional numerical analysis methods to ensure greater reliability and improved performance. Key features in the CTG (e.g. signal quality, baseline heart rate, heart rate variability, accelerations, the magnitude of deceleration and their timing in relation to contractions) are extracted and classified from the CTG using numerical algorithms and a small artificial neural network. Relevant clinical information, such as cervical dilatation, use of analgesia, fetal blood sampling and risk factors (intra-uterine growth retardation, placental abruption and thick mecanium) is then considered. The system interprets this combined data using a database of over 400 rules and then recommends action..
Following two successful in-house, retrospective studies to demonstrate the feasibility and validity of our approach, a more extensive retrospective study involving the comparison of the system's recommendation with those of 17 experienced clinicians from 16 leading centres in the UK, based on 50 cases was undertaken. All clinicians in the study have daily experience of labour management and were identified by their heads of departments as experts. The 50 cases represent a range of different outcomes and were reviewed by each expert separately twice, one month apart, under supervision of a research worker from our group. Two reviews of the cases were performed to establish each expert's consistency. This important study shows that genuine experts are consistent and can agree to a high degree, in the way they interpret the CTG and manage labour. The system's performance was the same as most of the experts and better than actual clinical practice.

Clearly, 50 cases was insufficient to test the robustness of the system for clinical use, particularly the full range of abnormal outcome and interesting or unusual traces. The latest stage of development and validation involved the collection of 400 such cases, including 162 cases with extremely poor outcome. This database was supplemented with 100 normal cases. The aims were to extend the system's rule base and show that it could perform better than clinical practice and most clinical experts on a wide range of cases. System performance in the 162 poor outcome cases and 100 normal cases was examined in detail. Rule changes were made as appropriate to maximise the sensitivity and specificity of the system. When no more simple rule changes could be identified to improve system performance, all 500 cases were processed and the system's recommendations compared to the opinions of three external experts and clinical practice. Three of the best experts from the previous study prospectively managed all 500 cases using a computerised CTG review program. The system
recommended intervention in a similar number of cases as the experts and more than clinical practice, as would be expected with a database of largely abnormal cases. Two of the experts identified more cases of poor outcome than the system but had a correspondingly higher intervention rate in the normal cases. The expert who identified less poor outcome cases than the system had an extremely low unnecessary intervention rate. Clearly, there is a trade off between sensitivity and specificity of interventions both by the system and the experts. While the system and all the experts performed better than clinical practice, it remains a challenge to translate the system behaviour to a normal population where the incidence of fetal compromise is fortunately low.

The DataCare Expert System

Gas exchange (nutrients and wastes) between the fetus and mother takes place via the placenta and umbilical cord. The cord vein carries oxygenated blood from the placenta to the fetus and the two smaller cord arteries return deoxygenated blood from the fetus. Thus, the arterial blood reflects the fetal acid-base balance, whilst the venous blood reflects a combination of maternal acid-base balance and placental function. Analysis of the blood from both vessels immediately after birth can provide objective information on the condition of the fetus at the time of delivery.

Commercial blood gas analysers are available to measure various parameters of the blood including the pH, partial pressure of carbon dioxide (pCO2) and partial pressure of oxygen (pO2). The base deficit of the extracellular fluid (BDecf) can be derived from the pH and pCO2 to distinguish the cause of a low pH between the distinct physiological conditions of respiratory acidosis, due to a short-term accumulation of CO2, and a metabolic acidosis, due to a longer-term oxygen deficiency. Interpretation of the pH and BDecf parameters from both the artery and vein in combination can provide information concerning the occurrence, the type and length of a hypoxic event (lack of oxygen that occurred during labour). Such assessment of the acid-base status of umbilical cord blood has recently been recommended by the British Royal College of Obstetricians and Gynaecologists.

In practice, there are difficulties in obtaining the blood samples because the cord arteries are very small compared to the vein. Sometimes very little blood is available and occasionally the cord may be small, twisted or mis-shaped. Between 25% and 40% of data may contain errors preventing accurate interpretation, and many clinical staff lack the knowledge to interpret umbilical acid-base data.

The DataCare expert system for blood gas analysis was developed, in collaboration with a medical devices company (Chiron Diagnostics Ltd), to overcome the difficulties above and encapsulates the knowledge of leading obstetricians, neonatologists and physiologists. Figure 3 shows a block diagram of the conceptual model of the DataCare expert system. The expert system checks for errors in the results, alerting clinicians as necessary and passes the results to the rule module to provide an interpretation for the clinician. The system also archives the patient information together with the results and interpretation in a database for audit purposes. The system was successfully implemented for a year in Plymouth, six months in Exeter and now in several other hospitals.
Before the DataCare expert system was released into clinical use, it underwent an extensive process of evaluation. Evaluation of an expert system can be split into three main aspects:

1. **Verification** is the process of ensuring that the expert system is functioning according to its specification,
2. **Validation** is the process of ensuring that the knowledge embedded within the expert system is an accurate representation of the domain, and
3. **Assessment** is the process of determining the effect that the expert system has in the clinical setting.

A number of specific evaluation tasks were carried out to address each of these aspects as summarised in Table 1. These were:

1. **Subsystem validation**: this involved extensive ‘destruction testing’ of the software in which every line of code was examined to ensure that its behaviour was well determined.
2. **Face validation**: the complete rule set was given to a number of experienced clinicians who were asked to highlight any interpretations that they would disagree with, and all non-normal results obtained during the field trials were reviewed by the resident experts.
3. **Hazard analysis**: each potential hazard or external event that might cause software failure was identified and the appropriate software behaviour anticipated. The event was then instigated or simulated (as far as possible) and the software behaviour observed.
4. **Sensitivity analysis**: almost 1000 test samples were generated to systematically check each classification rule boundary to ensure that the expected result was obtained.
5. **Economic assessment**: a cost benefit analysis was carried out to ensure that the expert system could be expected to be of economic benefit to the hospitals.

### Table 1: how each evaluation task relates to aspects of overall evaluation

<table>
<thead>
<tr>
<th>Task</th>
<th>Verification</th>
<th>Validation</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subsystem validation</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face validation</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Hazard analysis</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Economic assessment</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>Field tests in Plymouth</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Field tests in Exeter</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: conceptual block diagram of the DataCare expert system.
6. Field tests in Plymouth: the system was placed on trial at the local hospital for six months and the views of the clinical users were regularly checked.

7. Field tests in Exeter: the system was placed in a nearby hospital that had previously not carried out any umbilical cord blood analysis to obtain the views of novel, independent users of the system.

The system has now been installed in twenty five hospitals in the UK, and as yet there have been no cases of software problems or maintenance required. The expert system has since been updated to connect to two new models of blood gas analyser, and again this modification was incorporated seamlessly and without problem. The lack of bugs and software maintenance requirements is a reflection of the time and effort put into the original software testing and evaluation process. In terms of project time, at least as much time was occupied by the evaluation process as for the rest of the project (from specification through knowledge elicitation to coding) combined. Although the evaluation requirements of the industrial partner initially seemed onerous and overly restrictive, in fact the effects of quality assurance on the project have been beneficial.

Conclusions

Two practical artificially intelligent systems for obstetrics have been described, one of which is already in use in clinical practice and the other is going through key stages of clinical evaluation and development. These are a testimony that intelligent medical systems technology has matured. However, we are aware that few intelligent medical systems are in routine clinical use. This is due in part to the long time scale that it takes to carry out the underpinning research, design, fully develop and transfer the new technology into clinical practice - typically between 4 and 10 years depending on the clinical area. In this safety critical area, there are no short-cuts. There are other factors that have limited the transfer of the intelligent systems technology into clinical practice. These include technical and socio-political problems (e.g. validation issues, difficulty with performance metrics/gold standards, lack of genuine experts, problems with genuine, multi-disciplinary collaboration, lack of sufficient number of “good quality” data - including clinical information - safety/legal issues, human resistance in its various forms, and funding issues).

The research to underpin the next generation of the systems described in this paper is already in progress. The first key research question, namely how to manage the imprecision and uncertainty inherent in obstetric data and knowledge, has already been addressed using fuzzy logic. The next phase of the work is to integrate the analysis of a variety of data in a manner that would provide a fuller picture of fetal well-being during labour and immediately after delivery.

Acknowledgements:

The authors wish to acknowledge the contributions of many colleagues in the development of the INFANT and DataCare expert systems, especially those of Professor Karl Rosen (Sweden), Professor Keith Greene (UK), Dr Westgate (New Zealand), Dr Robert Keith (USA), Professor James Low (Canada) and the experts who took part in system evaluation. We are grateful to Chiron Diagnostics Ltd who supported the development of the DataCare Expert systems (and now markets it), the EPSRC, MRC and South West Regional Health Authority for financial support.
References


8 Garibaldi, JM. and Ifeachor, EC: “Application of Simulated Annealing Fuzzy Model Tuning to Umbilical Cord Acid-Base Interpretation”, IEEE Transactions Fuzzy Systems, Accepted for publication.


