1997

INTELLIGENT TECHNIQUES FOR HANDLING UNCERTAINTY IN THE ASSESSMENT OF NEONATAL OUTCOME

GARIBALDI, JONATHAN MARK

INTELLIGENT TECHNIQUES FOR HANDLING UNCERTAINTY IN THE ASSESSMENT OF NEONATAL OUTCOME

by

JONATHAN MARK GARIBALDI

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Electronic, Communication and Electrical Engineering
Faculty of Technology

in collaboration with

Plymouth Postgraduate Medical School

November 7th, 1997

Intelligent Techniques for Handling Uncertainty in the Assessment of Neonatal Outcome

by

Jonathan Mark Garibaldi

Objective assessment of the neonatal outcome of labour is important, but it is a difficult and challenging problem. It is an invaluable source of information which can be used to provide feedback to clinicians, to audit a unit's overall performance, and can guide subsequent neonatal care. Current methods are inadequate as they fail to distinguish 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 information on any damage suffered by the infant due to lack of oxygen during labour. However, this process is complex and error prone, and requires expertise which is not always available on labour wards.

A model of clinical expertise required for the accurate interpretation of umbilical acid-base status was developed, and encapsulated in a rule-based expert system. This expert system checks results to ensure their consistency, identifies whether the results come from arterial or venous vessels, and then produces an interpretation of their meaning. This 'crisp' expert system was validated, verified and commercially released, and has since been installed at twenty two hospitals all around the United Kingdom.

The assessment of umbilical acid-base status is characterised by uncertainty in both the basic data and the knowledge required for its interpretation. Fuzzy logic provides a technique for representing both these forms of uncertainty in a single framework. A 'preliminary' fuzzy-logic based expert system to interpret error-free results was developed, based on the knowledge embedded in the crisp expert system. Its performance was compared against clinicians in a validation test, but initially its performance was found to be poor in comparison with the clinicians and inferior to the crisp expert system. An automatic tuning algorithm was developed to modify the behaviour of the fuzzy model utilised in the expert system. Sub-normal membership functions were used to weight terms in the fuzzy expert system in a novel manner. This resulted in an improvement in the performance of the fuzzy expert system to a level comparable to the clinicians, and superior to the crisp expert system.

Experimental work was carried out to evaluate the imprecision in umbilical cord acid-base parameters. This information, in conjunction with fresh knowledge elicitation sessions, allowed the creation of a more comprehensive fuzzy expert system, to validate and interpret all acid-base data. This 'integrated' fuzzy expert system was tuned using the comparison data obtained previously, and incorporated vessel identification rules and interpretation rules, with numeric and linguistic outputs for each. The performance of each of the outputs was evaluated in a rigorous validation study. This demonstrated excellent agreement with the experts for the numeric outputs, and agreement on a par with the experts for the linguistic outputs. The numeric interpretation produced by the fuzzy expert system is a novel single dimensional measure that accurately represents the severity of acid-base results.

The development of the crisp and fuzzy expert systems represents a major achievement and constitutes a significant contribution to the assessment of neonatal outcome.

To Emma with love;
to Anne and Anthony with pride.

Acknowledgements

I would like to acknowledge the support and encouragement of my supervisors:

Prof. Emmanuel Ifeachor for his engineering rigour, for his close supervision in times of stress, coupled with more relaxed supervision in times of progress, for his project management, words of wisdom and personal friendship, and for advice on the preparation of this thesis;

Prof. Keith Greene for his clinical vision.

I am indebted to my main experts, Dr Jenny Westgate and Prof. Karl Rosén, for their patience during knowledge elicitation for the crisp acid-base expert system, and their continuing collaborative efforts in the design and validation of the fuzzy expert system. Jenny has the great ability to verbalise her knowledge in a form that can be translated into expert system rules, and Korgi has the vast depth of knowledge that only comes from the familiarity gained through years of experience of acid-base analysis.

Thanks also to Maureen Harris, Mark Davies, Sarah Beckley, and Geoff Hughes for their clinical opinion during validation; to Mo and Mark for clinical assistance in experimental determination of acid-base errors; and particularly to Mark for his dedication in assisting me on occasions too numerous to list. Thanks also to Prof. James Low for his encouragement of this work, and for his help in the final validation study.

This work was supported and encouraged by **Chiron Diagnostics Ltd** by facilitating the supply of Model 238, Model 248 and 800 Series Blood Gas Analyzers complete with technical interface documentation. I would particularly like to thank Mike Brown for his contribution in getting others at Chiron interested in the concept of the Expert System, and in its initial marketing, and Bob Mayall for his technical and marketing support. Without the efforts of both Mike and Bob the Expert DataCare system would probably not have succeeded in achieving commercial release.

Special thanks to Emma for her love, support and encouragement, to Neil Rothwell Hughes for inspiration and technical discussions on Simulated Annealing. Also, thanks to Rob Keith for his initial engineering support, and for providing me the enthusiasm to begin.

Many thanks to Julian Tilbury for his massive assistance in the design, implementation and analysis of the final validation of the integrated fuzzy expert system.

Thanks to Dr Steve Shaw in the statistics department for discussions and general advice on statistics, and for checking the statistics for the evaluation of the crisp expert system.

Lastly, but not least, a special thank you to the consultants, doctors, midwives, but particularly the nursing auxiliaries at Derriford Hospital for their help in the implementation of the crisp expert system.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

This work was financed by *The Polytechnic Central Funding Council*, and *The Engineering and Physical Sciences Research Council*. The development of the crisp expert system was carried out in collaboration with the *Plymouth Postgraduate Medical School*, and *Chiron UK Ltd* (formerly Ciba-Corning Diagnostics UK Ltd), Colchester Road, Halstead, Essex.

Publications:

- J.A.Westgate, J.M.Garibaldi, K.R.Greene, "Umbilical cord blood gas analysis at delivery: a time for quality data", *British Journal of Obstetrics and Gynaecology*, Vol. 101, pp. 1054–1063, December 1994
- J.M.Garibaldi, J.A.Westgate, E.C.Ifeachor, K.R.Greene, "The development and implementation of an expert system for the analysis of umbilical cord blood", *Artificial Intelligence in Medicine*, Vol. 10, pp. 129–144, June 1997
- J.M.Garibaldi, E.C.Ifeachor, "Application of simulated annealing fuzzy model tuning to umbilical cord acid-base interpretation", submitted to *IEEE Transactions on Fuzzy Systems*
- J.M.Garibaldi, E.C.Ifeachor, K.R.Greene, "The clinical evaluation of an expert system for the analysis of umbilical cord blood", submitted to *Artificial Intelligence in Medicine*,
- J.M.Garibaldi, J.Tilbury, E.C.Ifeachor, "The validation of a fuzzy expert system for umbilical cord acid-base analysis", in preparation
- J.M.Garibaldi, N.Outram, E.C.Ifeachor, "Techniques for handling medical uncertainty with fuzzy logic", in preparation

Presentations and Conferences Attended:

- J.M.Garibaldi, J.A.Westgate, E.C.Ifeachor, K.R.Greene, "An expert system for umbilical cord blood gas analysis: the importance of quality data", presented at the *IEE Colloquium on Monitoring in Obstetrics and Gynaecology*, London, UK, December 1993
- J.M.Garibaldi, J.A.Westgate, E.C.Ifeachor, K.R.Greene, "The development of an expert system for the analysis of umbilical cord blood at delivery", presented at the *First International Conference on Neural Networks and Expert Systems in Medicine and Healthcare*, Plymouth, UK, August 1994
- J.M.Garibaldi, E.C.Ifeachor, K.R.Greene, "Interrelationships between validated umbilical cord acid-base parameters and Apgar scores", presented at the *Third International Symposium on Intrapartum Surveillance*, Dublin, Ireland, July 1995

- J.M.Garibaldi, E.C.Ifeachor, K.R.Greene, "An intelligent computer system for the analysis and interpretation of umbilical cord blood sampling", presented at the *27th British Congress of Obstetrics and Gynaecology*, Dublin, Ireland, July 1995

- J.M.Garibaldi, J.A.Westgate, E.C.Ifeachor, K.R.Greene, "The development of a crisp and fuzzy expert system for the analysis of umbilical cord blood at delivery", presented at the *Third World Congress on Expert Systems*, Seoul, Korea, February 1996

- J.M.Garibaldi, E.C.Ifeachor, "The comparison of a crisp and fuzzy expert system with practising and expert clinicians", presented at the *Second International Conference on Neural Networks and Expert Systems in Medicine and Healthcare*, Plymouth, UK, August 1996

- J.M.Garibaldi, E.C.Ifeachor, "The development of a fuzzy expert system for umbilical acid-base analysis", presented at the *Fifth Fuzzy Days Conference*, Dortmund, Germany, May 1997

- J.M.Garibaldi, E.C.Ifeachor, "A Novel Interpretation of Umbilical Cord Acid-Base Balance", presented at the *Fourth International Symposium on Intrapartum Surveillance*, Paris, France, June 1997

  Joint winner of the best poster prize.

Signed ..... JM Garibaldi .....

Date ..... 21 November 1997 .....

Contents

Appendices

List of Figures

List of Tables

Glossary of Terms

TERM	*MEANING*
acidemia[1]	low pH in the blood
acidosis	low pH in the tissues
aerobic	in the presence of oxygen
anaerobic	in the absense of oxygen
analgesia	pain relief
analgesic	pain relief drugs
antenatal	before birth
antepartum	before labour
Apgar	score a score of 0 (worse) to 10 (best) assigned to an infant
asphyxia	oxygen lack and acidosis with impaired organ function
base deficit	the amount of non-carbonic base removed from the buffering systems
birth asphyxia	loose term used in the past for any less than perfect outcome
crisp reasoning	using classical logic (all rules are either *true* or *false*)
encephalopathy	dysfunction of the brain
episiotomy	surgical incision to enlarge the vagina, usually during the second stage of labour
extra cellular fluid	fluid outside of the cells
fuzzy reasoning	using multi-valued, or fuzzy sets for approximate reasoning (rules can have intermediate truth values)
hypoxaemia	reduction of oxygen in the blood
hypoxia	reduction of oxygen in the tissues
hypoxic ischaemic encephalopathy	damage to the brain caused by low oxygen in the blood or no blood flow to the brain
intrapartum	during labour
intrapartum asphyxia	asphyxia that occurred during labour
metabolism	(loosely) the conversion of sugar (glucose) to energy
neonatal	newly born
neonatal encephalopathy	any brain damage found in a newborn infant
perinatal	around the time of birth (strictly between the seventh month of pregnancy and the first week of life)
plasma	fluid (noncellular) portion of the blood
postnatal	after birth
postpartum	after labour

[1] note that the international spelling of *acidemia* has been used throughout this thesis, rather than the british english spelling *acidaemia*

Chapter 1

Medical Background

1.1 Labour

1.1.1 The Normal Labour

Most labour is, by definition, normal. The process begins with the fertilisation of the egg, which proceeds to the uterus, where it attaches itself to the uterine wall. The fetus is a separate self-contained organism connected to the mother through the placenta via the umbilical cord. It receives all its oxygen and nourishment required for development and growth and removes the waste products of metabolism through the placenta.

The 'normal' labour occurs after the fetus has gestated for 40 weeks, during which time it has grown to approximately 3.5kg (7lb 12oz) and has prepared itself for the stressful process of childbirth. As labour begins, the maternal uterine contractions increase in intensity and frequency. During the *first stage* of labour the cervix (the opening of the uterus) dilates to 10cm diameter. This typically takes around 4 to 8 hours. When the cervix has reached its full dilatation, the *second stage* of labour begins. The contractions are now occurring at a rate of around one every two–three minutes, each lasting about one minute. The fetus begins its descent through the pelvis and the second stage culminates with the delivery of the infant after typically one to two hours.

The infant now starts to cry to expand its lungs and begin its independent existence. During the *third stage* of labour the uterus rapidly contracts and maternal blood supply to the placenta is terminated. Simultaneously, the infant's circulation to the placenta and umbilical cord gradually reduces and the now redundant organ eventually separates. In the western world the third stage of labour is usually managed by immediately clamping the umbilical cord with a pair of clamps near to the infant and the umbilical cord is cut to separate the infant. The infant is then briefly removed for cleaning and weighing whilst the third stage continues with the expellation of the rest of the umbilical cord and placenta. The mother and infant are quickly reunited and the happy family leave hospital.

1.1.2 Labour is Stressful

The process described above is a description of an imaginary ideal labour. In fact, labour is a critical time for the fetus: "*You are more likely to die on the first day of your life than on any other, except the last*". Throughout labour the maternal blood supply to the placenta is normally interrupted during each uterine contraction, so that oxygen levels in fetal blood fall during contractions and recover once placental blood flow resumes (Figure 1.1). Thus even during normal labour, the fetus is being regularly deprived of oxygen. If contractions are occurring every two minutes and lasting one minute each the fetus may be receiving oxygen for no more than half the time, whilst being subjected to the physical stress of childbirth.

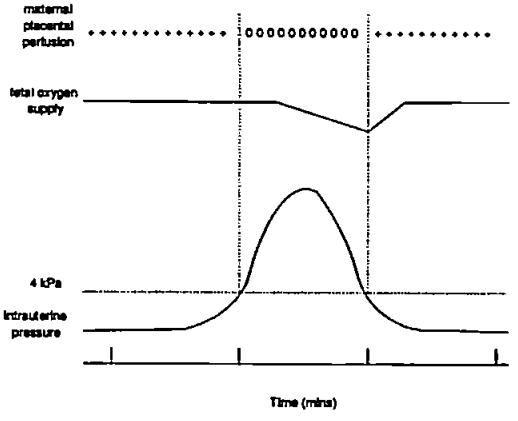


Figure 1.1: *oxygen supply to the fetus decreases during contractions because maternal perfusion of the placenta is interrupted as uterine pressure exceeds 4 kPa*

The fetus has developed a number of mechanisms to allow it to compensate for oxygen deficiency during labour. However although these compensatory mechanisms will be sufficient for the appropriately grown fetus at 40 weeks gestation to sustain a normal labour without suffering damage, they may not be adequate if the fetus is premature, growth retarded or the labour is unusually stressful. The important concept to bear in mind is that every fetus is being stressed during labour and its condition will be gradually deteriorating as its defences get used up. The art for the midwife, junior doctor or obstetric consultant in charge of childbirth is to achieve delivery *before* the infant's condition has deteriorated enough to cause the infant permanent damage.

1.1.3 Changes in Obstetric Care

In the mid 1960's the cardiotocograph (CTG) was introduced, in which the fetal heart rate and maternal uterine contractions are recorded simultaneously. This electronic fetal monitoring was introduced with the implicit belief that it would improve care by providing the clinician in charge with additional information on the health of the fetus, thereby allowing the clinician to make more appropriate choices in the management of the labour. Other technological developments have included Fetal Blood Sampling (FBS), in which the pH of blood from the fetal scalp is measured, fetal pulse oximetry (pOx), in which the oxygen content of the fetal blood is continuously monitored in utero, or analysis of the fetal electrocardiograph (ECG), in which the shape of the fetal ECG changes are used to infer the fetal health.

In addition to the technological changes, there have been and continue to be a number of political changes in the delivery of healthcare. In the 1970's the trend was towards more hospitalisation and increased use of modern technology, but recently the emphasis has shifted once again. There is an increasing awareness among mothers of 'natural childbirth' in which modern technology is relegated to the background, while the Government is advocating a policy entitled *Changing Childbirth* [27], in which the standard hospital care is replaced by more personalised and often home-based care. There have also been much publicised changes in junior doctors hours, the introduction of independent NHS trusts and an increased

climate of audit with the introduction of Government league tables for many areas of public service.

Against these technological and political changes, there has been a massive increase in litigation. This litigation, which seems to occur whenever the result of labour is not a perfectly healthy infant, has caused clinicians to adopt a defensive style of practice, which has led to an increase in caesarean section rates as in the USA. Mothers expect higher standards of care, Governments promise higher standards of care and lawyers penalise failure to meet these ever higher standards. Meanwhile individual clinicians have difficulty in obtaining accurate feedback on whether their management for a particular case has been appropriate and whether their overall performance is up to standards.

1.1.4 Evaluating Care

How then should then effectiveness of care be measured? Crucially and more specifically, how should the contribution of *obstetric* care to the eventual neonatal outcome of the infant be measured? The perinatal mortality rate has fallen steadily to around 6–8 per thousand in the UK, USA and other developed countries [125]. Thus, in the developed world, perinatal mortality is simply not a sensitive enough variable to effectively compare obstetric practice. In contrast the cerebral palsy rate seems to have stabilised at around 2–5 per thousand [117].

There are many ways in which the outcome of labour can be measured. The health of the fetus is one obvious parameter, but it should not be forgotten that labour also entails risks for the mother. In the UK, the maternal mortality rate is currently around 8–10 per 100 000 [28] with many forms of maternal morbidity, such as caesarean section, also a possibility. Each caesarean section costs the health service an estimated £1 560 [21] and greatly increases the chance of the mother having a repeat caesarean section in all subsequent deliveries. There are forms of instrumental delivery such as forceps, and many other factors such as the type or amount of analgesia used, episiotomy rates or duration of labour. More ephemeral factors such as maternal satisfaction with the birth or comfort during delivery could also be considered as *outcomes* of labour.

In this thesis the term *outcome* is used specifically to refer to the health of the infant on delivery. However, were neonatal outcome to be considered the *only* important outcome measure then the optimal strategy would probably be to caesarean section each and every mother just before the onset of labour (whenever that might be). Even if this were to be done, there would still be an underlying rate of congenital abnormality and other problems that occur before the onset of labour. In other words, infants do not enter labour in a uniform condition. Hence, even given optimal care during labour, a small proportion of infants could still emerge with serious problems.

So, there are a heterogeneous population of fetuses entering labour at varying gestational ages, with varying body weight, that undergo labour and emerge in various states of health. The problem remains of how to measure the outcome of labour in terms of the obstetric contribution to the neonatal condition. Before attempting to address this, the process of labour and the physiology of the fetus will be discussed in more detail.

1.2 Fetal Physiology

1.2.1 Aerobic and Anaerobic Metabolism

In the primary mechanism of cell metabolism glucose is broken down in the presence of oxygen to release usable energy in the form of ATP (Adenosine Triphosphate), producing carbon dioxide and water as by-products. The exact chemical pathway is complicated and not fully understood, but the process, termed *aerobic metabolism* (metabolism in the presence of oxygen), can be summarised as:

$$1\,\text{Glucose} + 6O_2 \rightarrow 6CO_2 + 6H_2O + 36\ldots38\,\text{ATP} \tag{1.1}$$

If the oxygen supply to a cell is insufficient for it to maintain aerobic metabolism, a secondary mechanism of metabolism can take place, in which glucose is metabolised without oxygen to release energy and produce lactic acid. This process, termed *anaerobic metabol-*

ism (metabolism in the absence of oxygen), can be summarised as:

$$1\,\text{Glucose} \rightarrow 2\,\text{Lactic Acid} + 2\,\text{ATP} \tag{1.2}$$

As can be seen from the above equations, anaerobic metabolism produces only about 1/18 the amount of usable energy of aerobic metabolism, and it is therefore reserved as a short term emergency energy source or a longer term supplementary measure.

1.2.2 Oxygen Balance

The placenta is the organ of gas exchange for the fetus (Figure 1.2). Oxygen diffuses from maternal blood into fetal blood across the placental membrane and is carried to the fetus in the single large umbilical cord vein. Deoxygenated blood from the fetus returns to the placenta in two smaller arteries and the waste products of fetal metabolism, including carbon dioxide, are transferred or diffuse into maternal blood. The fetus has a number of special adaptive mechanisms which enable it to balance oxygen supply and energy demand to maintain aerobic metabolism (Figure 1.3).

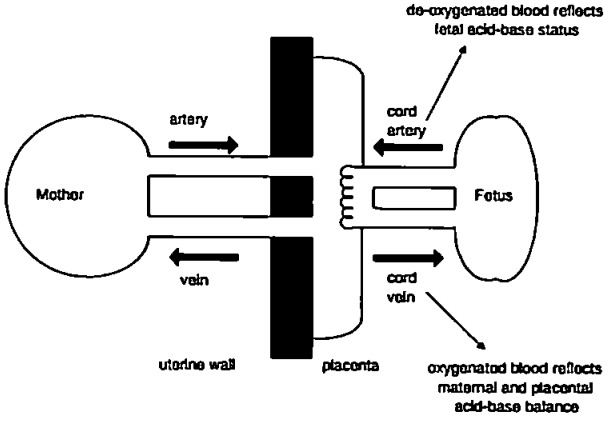


Figure 1.2: *diagrammatic representation of maternal and fetal circulations*

The amount of oxygen available to the fetus depends upon the blood flow, the partial pressure of oxygen in the blood (pO_2), the haemoglobin concentration, the type of haemoglobin and the oxygen saturation. The increased oxygen-carrying capacity of fetal blood is due partly to its higher haemoglobin concentration and partly to the greater affinity of fetal haemoglobin

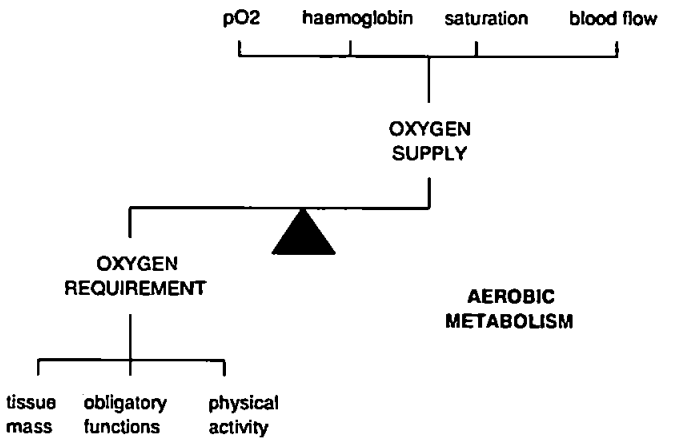


Figure 1.3: *oxygen balance is governed by oxygen supply and oxygen requirement*

for oxygen. This enables it to become saturated with oxygen at low partial pressures of oxygen. The amount of blood flow to an organ will also determine its oxygen supply. Fetal cardiac output is approximately four times higher than that of the adult per kilogram of body weight [97], with the majority directed to the placenta, so that the fetal organs are normally supplied with more oxygen than they require. This 'oxygen reserve' means that fetal oxygen extraction can be increased at the tissue if needed.

Oxygen requirement is determined by fetal size, fetal activity and essential fetal metabolic processes. Energy demand can be reduced by decreasing activity or growth. If oxygen supply and requirement are in balance the fetus has adequate oxygen to metabolise glucose aerobically to produce the energy required for organ function.

1.2.3 Impairment to Oxygen Supply

As described earlier, the exchange of oxygen via the placenta is greatly reduced or completely interrupted during uterine contractions when the pressure exceeds about 4kPa. This may occur for up to 60 seconds every 2–3 minutes. Additionally, as the fetus descends through the pelvis, the umbilical cord may often be compressed between the uterine body and fetal body, or may sometimes become tightly entangled around the fetal neck or limbs. A number of other mechanisms can act both before and during labour to result in a reduction of fetal oxygen supply, and these are summarised in Figure 1.4.

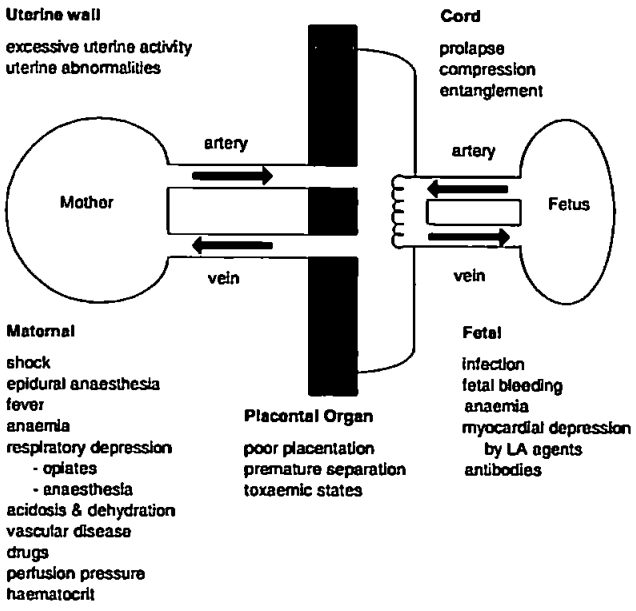


Figure 1.4: *a number of mechanisms can impair the oxygen supply*

The healthy fetus is able to adjust to episodes of diminished oxygen supply by a synchronised response, which involves behavioural and cardiovascular adjustments [44]. The most important of these is the centralisation of blood flow to the heart and adrenals with increased oxygen extraction at the placental bed and tissues. If these adjustments fail to maintain adequate oxygen supply to central organs, aerobic metabolism is supplemented by anaerobic metabolism of glucose to maintain cell and organ function. Glucose can have been stored by the fetus in the myocardium (heart muscle) and liver in the form of glycogen, a more complex carbohydrate. Glycogen stores can be mobilised with the onset of anaerobic metabolism to increase blood glucose levels, but these long term reserves are depleted in the process.

1.3 Acid-Base Balance

1.3.1 Oxygen Saturation and pO_2

It might be thought from the discussion above that the best indication of fetal condition would be obtained by measuring the oxygen content of the fetal blood, but in practice this is not the case. The *oxygen content* measures the quantity of oxygen in the blood, the *oxygen capacity*

is the maximum amount the blood can hold and the *oxygen saturation* is the percentage of content divided by capacity. The easiest parameter to measure is the partial pressure of oxygen dissolved in the blood (pO_2), but it is difficult to determine the oxygen content or saturation of the blood accurately from this. There is a complex relationship between the oxygen saturation and the pO_2 of blood due to the presence of haemoglobin, as shown in the sigmoid-shaped oxygen dissociation curve (Figure 1.5).

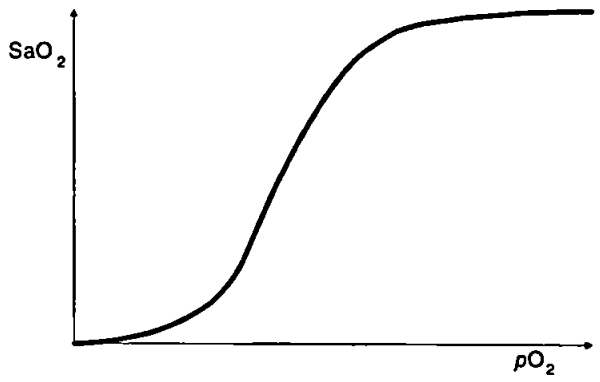


Figure 1.5: *oxygen dissociation curve*

Fetal haemoglobin is different from adult haemoglobin to allow the fetus to exist downstream of the maternal oxygen supply. The fetal oxygen dissociation curve lies to the left of the maternal curve to enhance the uptake of oxygen from the maternal blood. The fetal curve is also steeper so that the fetal blood loads and unloads large quantities of oxygen over a narrow pO_2 band. An effect known as the Bohr effect shifts the curve to the right with decreasing pH such that even with a precise and accurate pO_2 the oxygen saturation cannot be calculated without knowledge of the blood pH [15].

Oximetry techniques measure the oxygen saturation directly by passing light of visible red and infra-red wavelengths through tissue. Oxidised and deoxidised haemoglobin have different spectral absorption characteristics, such that oxidised haemoglobin absorbs more light in the visible red wave length than deoxidised haemoglobin. Comparison of the differential absorption of the visible red and infra-red allows the percentage of oxidised haemoglobin to be calculated. However, this is still inadequate, as it is impossible to determine whether the fetus has compensated for the overall lack of oxygen by, for example, redistributing its blood flow to essential organs only.

1.3.2 pH, pCO_2 and Base Deficit

The pH is a logarithmic function of the concentration of H^+ ions in a system, given by:

$$\mathrm{pH} = -\log\left[\mathrm{H}^+\right] \tag{1.3}$$

where $\left[\mathrm{H}^+\right]$ is expressed is $\mathrm{mol.l}^{-1}$. Hence, the pH decreases as the concentration of hydrogen ions increases. Some carbon dioxide (CO_2) produced by aerobic metabolism is held in the form of free CO_2, but some of the carbon dioxide and water (H_2O) combine to form carbonic acid (H_2CO_3), which in turn dissociates to form hydrogen ions (H^+) and bicarbonate ions (HCO_3^-):

$$\mathrm{CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons H^+ + HCO_3^-} \tag{1.4}$$

From the law of mass action, it can be shown that the pH of blood or tissue is directly related to the concentration of bicarbonate, $\left[\mathrm{HCO}_3^-\right]$, and inversely related to the concentration of carbonic acid, $[\mathrm{H_2CO_3}]$, as given by:

$$\mathrm{pH} = \mathrm{pK'} + \log\left(\frac{\left[\mathrm{HCO}_3^-\right]}{[\mathrm{H_2CO_3}]}\right) \tag{1.5}$$

where $\mathrm{pK'}$ is the negative logarithm of the dissociation constant of carbonic acid.

The dissolved CO_2, dCO_2, is the total of free CO_2 and hydrated CO_2: $dCO_2 = CO_2 + H_2CO_3$. The carbonic acid is also in equilibrium with free carbon dioxide (Equation 1.4), so that the concentration of free CO_2 can be expressed as a multiple of that of carbonic acid. Hence the concentration of dissolved CO_2 can be substituted for that of carbonic acid [15] to give:

$$\mathrm{pH} = \mathrm{pK''} + \log\left(\frac{\left[\mathrm{HCO}_3^-\right]}{[\mathrm{dCO_2}]}\right) \tag{1.6}$$

where $\mathrm{pK''}$ is approximately 6.1 for plasma.

The concentration of dissolved CO_2 is itself directly proportional to the partial pressure of CO_2 (pCO_2), by the solubility coefficient $\alpha \approx 0.0306$, giving the Henderson-Hasselbalch equation:

$$\mathrm{pH} = 6.1 + \log\left(\frac{\left[\mathrm{HCO}_3^-\right]}{0.0306\, pCO_2}\right) \tag{1.7}$$

A buffer is a solution that resists the changes in pH produced by adding acid or alkali (base). It is a mixture of a weak acid and its conjugate base, in which the equilibrium:

$$\text{acid}^z \rightleftharpoons H^+ + \text{base}^{z-1} \tag{1.8}$$

is neither fully displaced to the left-hand side or right-hand side.

As free hydrogen ions are potentially toxic to tissue the body has a number of buffering mechanisms that act to maintain a stable pH, by 'mopping-up' excess H^+ ions. The main buffers are bicarbonate, haemoglobin, and protein:

$$H_2CO_3 \rightleftharpoons H^+ + HCO_3^- \tag{1.9}$$

$$HHb \rightleftharpoons H^+ + Hb^- \tag{1.10}$$

$$HProt \rightleftharpoons H^+ + Prot^- \tag{1.11}$$

but there are many others, including organic and inorganic phosphates, and globulins.

As the concentration of H^+ ions increases the chemical gradient 'pushes' the buffer equations to the left hand side, to remove the free H^+ ions from circulation and prevent the pH from falling. As this happens the amount of buffers remaining is obviously used up. The *base deficit* is defined to be the amount of non-carbonic buffer base removed from the total normal buffering capacity of the body or, put another way, the excess of non-carbonic acid in the system. A *base excess* (or negative base deficit) indicates that the body temporarily has more than the normal amount of non-carbonic base available for buffering. The base deficit is not measured directly, but can be calculated from the pH and the HCO_3^-, which in turn can be calculated from the pH and pCO_2 by the Henderson-Hasselbalch Equation (1.7).

1.3.3 Respiratory Acidosis

The process of normal aerobic metabolism (or respiration) utilises O_2 and produces CO_2 and H_2O (Equation 1.1). This increase in concentration of H_2CO_3 pushes Equation 1.4 to the right hand side, resulting in a higher pCO_2 and lower pH. However, although buffering is involved to prevent the drop in pH that would occur in a system without buffers, there is no

increase in non-carbonic acids. Hence there is no reduction in the non-bicarbonate buffers and the base deficit remains the same. This is termed a *respiratory acidemia.*

The term *acidemia* refers to low pH in the blood whereas the term *acidosis* refers to low pH in tissues. The fetal arterial blood usually reflects the fetal tissue status, but this may not always be so, particularly if there is complete blockage of the umbilical cord preventing any blood flow. A respiratory acidemia is caused by retention of CO_2 in the blood, primarily as a result of insufficient placental gaseous exchange. It is primarily an acute event which has little or no significance to neonatal outcome [81]. A respiratory acidemia is usually corrected within minutes of birth with the establishment of adequate ventilation, as the excess CO_2 is blown off by the lungs.

1.3.4 Metabolic Acidosis

As discussed previously, when oxygen availability is insufficient for aerobic metabolism to fulfil requirements, it is supplemented by anaerobic metabolism. This produces lactic acid, which dissociates to form hydrogen ions and lactate:

$$\text{Lactic Acid} \rightleftharpoons H^+ + \text{Lactate}^- \tag{1.12}$$

The buffering systems attempt to remove the free H^+ ions, so that the bicarbonate levels fall and, in a closed system, the pCO_2 rises. Because non-carbonic (lactic) acid has been added to the system, the base deficit is increased — this is termed a *metabolic acidemia.* Although it is possible for other non-carbonic acids to enter the system, the base deficit is an indirect measure of the extent to which anaerobic metabolism has occurred. As the process is relatively inefficient in producing energy, H^+ ions quickly accumulate. There are two main reasons why a metabolic acidemia is more important than a respiratory acidemia [100]:

1. A metabolic acidemia indicates that a significant period of anaerobic metabolism has occurred as a result of oxygen insufficiency and therefore indicates either a longer duration of problem or a fetus that has insufficient defences remaining to maintain aerobic metabolism.

2. The circulating lactate diffuses across the placenta or lungs much slower than CO_2 and therefore the fetus or neonate will take longer to correct a metabolic acidemia.

In fact, as any cause of fetal oxygen restriction usually goes hand in hand with a restriction in CO_2 elimination, a fetal acidemia is rarely purely respiratory or purely metabolic. The term *mixed acidemia* is used to indicate both a respiratory and metabolic component.

In adults the base deficit is usually calculated from the blood compartment of the whole extracellular fluid (BD_{blood}). However, the fetus and neonate have a much larger extravascular fluid compartment, so that more of the buffering capacity is extravascular compared to the adult. This is particularly important when pCO_2 levels are high, because a high pCO_2 will unduly influence base deficit values calculated from the blood compartment alone. The fetus is more likely to experience high pCO_2 values than the adult as a build up in CO_2 occurs commonly during labour. Siggaard-Andersen showed that the influence of a high pCO_2 could be avoided if the base deficit was calculated from the whole extracellular fluid (BD_{ecf}) [114]:

$$BD_{ecf} = -(1 - 0.023\,Hb)\left(\left[HCO_3^-\right] - 24.1 + (2.30\,Hb + 7.70)(pH - 7.40)\right) \qquad (1.13)$$

where the $\left[HCO_3^-\right]$ is calculated from the Henderson-Hasselbalch equation (1.7), and:

$$Hb = 3.7 mmol.l^{-1} \qquad (1.14)$$

This calculation leads to values of BD_{ecf} of around 2–4 $mmol.l^{-1}$ (depending on pH and pCO_2) lower than BD_{blood}. Thus, BD_{blood} is not appropriate in the perinatal period and BD_{ecf} should therefore be used instead to prevent the over diagnosis of metabolic acidemia.

1.3.5 Arterial–Venous Differences

A fetus will be in a steady state when it has sufficient oxygen to maintain aerobic metabolism, there is normal blood flow through the umbilical cord and good placental gaseous exchange. The oxygenated venous blood to the fetus should have a higher pO_2, higher pH and lower pCO_2 than the deoxygenated arterial blood returning from the fetus with relatively

lower pO_2, lower pH and higher pCO_2. While CO_2 can diffuse rapidly across the placental membranes, H^+ and lactate ions take much longer to equilibrate. Thus, in a steady state situation there should be little or no difference in arterial–venous base deficit. In contrast, large arterial–venous BD_{ecf} differences (high arterial, low venous BD_{ecf}) reflect an acute onset of fetal metabolic acidosis. This can occur when normal placental function and gas exchange are interrupted by an acute reduction in fetal blood flow, for example severe cord compression or profound bradycardia (very slow heart-rate) [102].

In contrast, an acidemia in both artery and vein indicates that the hypoxia is not acute in onset. This is particularly so in the case of a metabolic acidemia. With the onset of metabolic acidemia, the H^+ and lactate ions diffuse from the fetal blood into the placental extracellular fluid compartment until equilibrium is reached. This gradual process delays the appearance of metabolic acidemia in the venous blood. As the fetus compensates for a reduced oxygen supply by increased blood flow to the vital organs and decreased blood flow to peripheral organs, a metabolic acidemia occurs later in the central organs than in the peripheral organs. A venous metabolic acidemia, indicating a non-acute event, therefore suggests that central organ metabolic acidosis is more likely to have occurred.

1.3.6 Mechanisms of Brain Damage

Precisely how the reduction in oxygen availability and consequent disturbances in acid-base balance lead to brain damage is not clearly understood. Within the brain, non-essential activity may be decreased to allow aerobic metabolism to be maintained. However, eventually the brain resorts to anaerobic metabolism and the breakdown of high energy cellular phosphate compounds. Likely candidates for immediate adverse effects (primary neuronal death) are high concentrations of intracellular H^+ ions and/or the lack of energy yielding substrate. There is now accumulating evidence that secondary neuronal death also occurs later as oxygen returns to the brain tissue. This may be caused by over-stimulation from the accumulation of excitatory amino acids, membrane destruction or the generation of oxygen free radicals [63].

It has long been known that the immature animal has a greater tolerance to suffocation than the mature animal. The anatomical and functional differences between a 26-week and 40-week fetus are considerable, and include lower metabolic rate, increased number of neurones and increased plasticity of the immature brain [62]. The complexity of these mechanisms means that, even if the acid-base assessment of the blood is known, it is still impossible to say exactly what levels of low pH or high base deficit will lead to brain damage. The precise levels will probably vary between individuals depending on factors such as prior reserves, tissue mass or gestational maturity. The only clear statement that can be made is that for any given individual as the pH gets lower (H^+ concentration increases), base deficit increases and duration increases, the more likely irreversible brain damage is to occur.

1.4 Evaluating Outcome

1.4.1 Birth Aspyhxia - An Incorrect Term

The term 'birth asphyxia' has haunted obstetrics for many years. In the past it has been used very loosely to indicate any neonate that appears to have neurological damage. The word 'asphyxia' means 'without pulse', but in the perinatal period asphyxia is strictly defined as the combination of oxygen lack and acidosis with impaired organ function. This only occurs after a sequence of events over time. These are:

1. Hypoxemia — a reduction in oxygen carried in the blood as a result of decreased pO_2 and decreased oxygen content. This may be compensated for by increased blood flow and increased oxygen extraction to maintain oxygen supplies to the tissues.

2. Hypoxia — oxygen supply is insufficient for tissue energy requirements and aerobic metabolism is supplemented by anaerobic metabolism to maintain energy balance.

3. Asphyxia — if hypoxia continues, the lactate and hydrogen ions produced by anaerobic metabolism accumulate, so that oxygen lack is accompanied by acidosis. Eventually, organ failure will occur with a risk of permanent tissue damage.

Asphyxia sufficient to cause damage may occur before labour (antepartum), during labour (intrapartum) or even after delivery (postpartum), but the term 'birth asphyxia' has implied that it occured around the time of birth. The confusion over the timing of the damage to the infant has meant that often obstetric practice gets unfairly blamed for subsequent neurological deficit. This contributes to defensive obstetric practice and confounds any attempt to evaluate obstetric care. The term 'intrapartum asphyxia' refers to an asphyxia as defined above that occured *only* during labour.

1.4.2 The Apgar Score

In 1953 Virginia Apgar proposed an 'objective' scoring system to provide 'a classification of newborn infants which can be used as a basis for discussion and comparison of results of obstetric practice' [8]. A score of 0, 1 or 2 is assigned to five factors that indicate the infant's health (Table 1.1) and totalled to give a score between 0 and 10. In Apgar's original paper these were assessed by an independent extra observer. In 1962 Apgar emphasised its use for classifying groups of infants and disclaimed its predictive value in individual infants [9]. Now in almost universal use, the Apgar score is commonly assigned at one minute after birth ($Apgar^1$) and five minutes after birth ($Apgar^5$), and often this is the only formal assessment of neonatal condition. In clinical practice the Apgar score has been applied subjectively and often retrospectively, with the Apgar scores being 'plucked out of thin air' sometime after the delivery and immediate neonatal care of the infant has been completed.

Factor	Score 0	Score 1	Score 2
Heart Rate	absent	heart rate below 100	heart rate above 100
Respiratory Effort	absent	slow or irregular	good strong breathing and crying
Reflex Irritability	no response	facial grimaces	cry (in response to stimulation)
Muscle Tone	limp	some flexion of extremities	active motion
Colour	blue or pale	pink body with blue extremities	pink all over

Table 1.1: *factors comprising the Apgar score*

Despite Apgar's clear guidelines, it is often incorrectly applied to individual prediction of subsequent neonatal development [89]. For many years the diagnosis of 'birth asphyxia' was made primarily on the basis of the Apgar score. While a baby who has been severely

hypoxic during labour can be expected to have low Apgars, asphyxia is *not* the only cause. Immaturity, infection, trauma, and congenital disorders are other possible causes. As a low Apgar score may have a congenital or antenatal cause it cannot be used to either define *intrapartum asphyxia* or fairly evaluate obstetric care.

1.4.3 Neonatal Encephalopathy

The condition of *neonatal encephalopathy* (neonatal brain dysfunction) was first defined in 1976 by Sarnat and Sarnat [107]. Three *stages* of encephalopathy were described, categorised by neonatal behaviour and measurement of the neonatal electroencephalogram (EEG) over a period of between one–three weeks. In this study, an attempt was made to exclude any cases in which the neonatal encephalopathy was **not** caused by intrapartum asphyxia. Having excluded other causes, the authors then referred to *hypoxia* (reduction of oxygen levels in tissues), *ischaemia* (reduction of quantity of blood flow) or *asphyxia* as having caused the encephalopathy. The study demonstrated that infants who did not enter stage 3 and who had signs of stage 2 for less than five days appeared normal at 12 months.

Although some of the concepts outlined in this study were adopted by others, the precise terminology and definitions varied [78, 74]. The terms *hypoxic-ischaemic encephalopathy* (HIE) [37] and *post-asphyxial encephalopathy* (PAE) [74] were introduced, with divisions of *mild*, *moderate* and *severe*, or *grade I*, *grade II* and *grade III*. Many studies followed in which the progress of infants in each of the three stages was investigated [37, 74, 78, 99]. It was found that the stage of 1 (I/mild), 2 (II/moderate) or 3 (III/severe) determined the prognosis for the infant fairly accurately: infants with stage 1 encephalopathy tend to have little or no long term damage, infants with stage 2 encephalopathy have a 20% chance of death or disability, those with stage 3 encephalopathy are certain to die or be disabled. Unfortunately, however, many of these follow-up studies failed to exclude non-asphyxial causes for the neonatal encephalopathy. Subsequently the terms HIE or PAE have been incorrectly applied to *any* infant exhibiting the symptoms of neonatal encephalopathy.

More recently, paediatricians have advocated that HIE or PAE should be the sole factor in

determining intrapartum asphyxia and for neonatal assessment [75]. However, as with the Apgar score, there are many causes of neonatal encephalopathy other than intrapartum asphyxia such as infection, hypoglycaemia, trauma and structural abnormalities [91]. It has been shown that 80–90% of neurological handicap is attributable to antepartum or postpartum events [90, 14]. Thus it is becoming accepted that the label HIE may be mistakenly attributing a pathology or cause (hypoxia-ischaemia) to symptoms (neonatal encephalopathy).

1.4.4 Long Term Follow-Up

The ultimate yardstick by which the neonatal condition at delivery may be measured is by the subsequent long term neurological development of the infant. Consequently, the performance or IQ of the child at various ages has been used as the 'gold-standard' [99]. Although this idea is superficially attractive it also presents a number of problems:

1. As the child develops there are many parental, cultural and environmental factors that may influence the child's neurological development, so that although the measurement of performance gets increasingly more accurate as the child gets older, the measurement reflects the process of birth less and less.

2. It is notoriously difficult and extremely costly to track and examine every child in the population at regular intervals. In the future it may be that each individual has lifelong, traceable, computerised central records which contain details of neurological development (e.g. school attainment tests), but currently this is a long way off, if ever.

3. There is necessarily a long time delay involved in obtaining a measurement, which makes it inherently impractical for evaluating care. i.e. if a new standard of care is introduced, it is unreasonable to wait 18 years before it is deemed unsatisfactory!

Any form of neonatal assessment must ultimately correlate to some degree with long-term neurological development, but for the first reason above a perfect correlation could never be expected or realised.

1.4.5 Umbilical Cord Acid-Base Balance

The physiology above has described the alterations to fetal acid-base balance that occur as a result of oxygen deficiency during labour. If the umbilical cord is 'double-clamped' with one pair of clamps near the fetus and one pair near the placenta, the isolated segment of cord may be removed for analysis. The arterial and venous blood contained in this isolated segment of umbilical cord provides a 'snap-shot' of the fetal acid-base status at delivery (the time of clamping).

Umbilical arterial blood reflects fetal condition and umbilical venous blood reflects a combination of maternal condition and placental function. Assessment of the pH and base deficit from both umbilical artery and vein provides information on the occurrence, timing and possible causes of oxygen deficiency. A cord arterial metabolic acidemia indicates a significant oxygen deficiency during labour; it does not, by itself, diagnose 'intrapartum asphyxia'.

In the past it has been suggested that analysis of the pH would be sufficient, sometimes without even specifying from which vessel the blood is taken. Cord arterial results are essential because the arterial blood most accurately represents the state of the fetal tissue conditions. Similarly, cord venous results are necessary for two reasons:

1. Venous results are necessary to ensure that arterial blood has definitely been obtained. The two arterial vessels are very small in comparison with the single large vein. If just arterial blood is being sampled, errors can lead to venous blood mistakenly being taken rather than arterial blood, resulting in incorrect diagnoses.

2. Venous results, and in particular the arterial–venous differences, can provide important information about the time course and possible causes of any oxygen deficiency. Both the duration and severity of oxygen deficiency are known to be important prognostic factors, so knowledge of both cord artery and vein status is valuable.

Some possible objections to umbilical acid-base assessment have been raised:

- *The fetus may encounter a period of severe hypoxia during labour, sufficient to cause brain damage, but then recover to be born with normal cord gases and possibly even normal Apgar scores.*

  Whilst it is possible that this may happen in the antepartum period, there is no clear evidence that it could happen during active labour. The recovery of responses and normalisation of acid-base balance would have to occur against the background of repeated episodes of hypoxaemia accompanying each uterine contraction.

- *A growth retarded fetus may be 'asphyxiated' without showing a metabolic acidemia as they do not have adequate stores of glucose or glycogen to metabolise anaerobically.*

  This is unlikely as there is considerable evidence from cordocentesis studies that growth retarded fetuses have higher lactate levels and lower pHs than appropriately grown controls. It has been demonstrated [80] that decreased weight for gestational age was the best predictor of metabolic acidemia and an increased incidence of arterial pH less than 7.10 in fetuses who were small for gestational age (10%) compared to appropriately grown fetuses (1%) has also been reported [41].

- *It is often said that pH does not correlate well with other diagnoses of 'birth asphyxia' or long-term outcome.*

  As stated above there are many other causes of symptoms such as low Apgar scores or neonatal encephalopathy and cord acid-base assessment provides additional objective information on fetal condition which can help in diagnosing the cause of these symptoms. Additionally, brain damage is the result of complex mechanisms such that a linear relationship of brain damage with pH should not be expected.

1.5 The Problem

1.5.1 An Integrated Approach to Assessment of Outcome

The informal use of the term 'birth asphyxia' for various conditions ranging from simply a low Apgar score to any neonatal encephalopathy has fuelled the medicolegal debate and hampered the evaluation of obstetric care. It is vitally important to distinguish intrapartum events from antepartum or postpartum events, each of which can affect Apgar scores and neonatal development. The use of any single factor is likely to result in a heterogeneous population of neonates, only some of whom will have experienced significant intrapartum hypoxia. It is increasingly accepted [7] that accurate diagnosis of intrapartum asphyxia requires the presence of metabolic acidosis, low Apgar scores *and* neonatal complications (Figure 1.6).

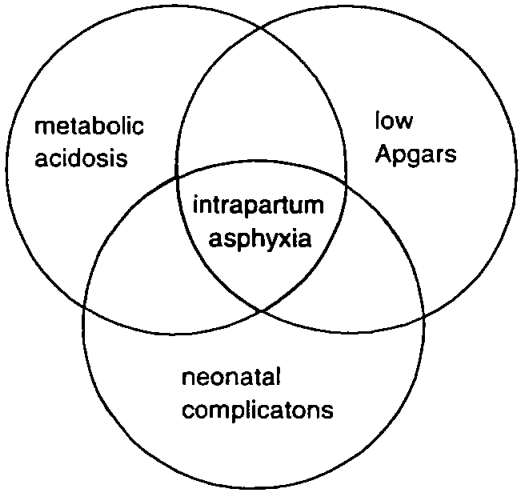


Figure 1.6: *the combination of low Apgar scores, metabolic acidosis and neonatal encephalopathy defines true intrapartum asphyxia*

Additional background information on the maternal past history, antenatal development of the fetus and knowledge of other labour events may all contribute to enhancing the accuracy of a diagnosis of 'intrapartum asphyxia' and of determining the neonatal condition. Conceptually, there are two related but distinct conditions that need to be evaluated:

1. Neonatal condition. This represents the accurate and precise evaluation of the condition of the infant at the moment of delivery.

2. Intrapartum damage. This represents the amount of additional damage that the infant has suffered as a *direct consequence* of labour. It would be derived from subtracting the neonatal condition on delivery from the antenatal fetal condition on entry to labour.

These might be determined numerically, for example on an arbitrary scale of 0 (dead) to 100 (perfect), so that:

antenatal condition:	86
neonatal condition:	75
intrapartum damage:	11

Alternatively the output may be displayed linguistically, for example:

antenatal condition:	the antenatal condition was satisfactory
neonatal condition:	but the neonate had very low Apgars
	and severe metabolic acidemia in both vessels
	and is now having fits with abnormal EEG
intrapartum damage:	the infant has suffered severe intrapartum asphyxia

If such a diagnosis of intrapartum events could be determined the result would be a major contribution to obstetric care, and would be of national and international significance.

1.5.2 The Need For Uncertainty Handling

An integrated approach, combining a number of different factors as outlined above, might appear relatively straightforward. However, there is a significant problem inherent in all aspects of medical care, and labour is no exception. As the biological systems (mother, fetus, process of childbirth) are all so immensely complex, it is impossible to describe and characterise them precisely and accurately. The fetus is relatively inaccessible both before and during labour, so that clinicians and mothers are naturally reluctant to resort to intervention unless forced to. The clinicians have responded to this problem by often categorising knowledge with linguistic labels rather than attempting to evaluate terms numerically.

The Apgar score, for example, is comprised of a number of factors having vague linguistic labels in their scoring definition. To determine *skin colour*, the clinician must decide whether the infant has *pink extremities* or is *pink all over*. Similarly, in scoring *muscle tone*, an arbitrary dividing line must be drawn between *some flexion of extremities* and *active motion*. It has been shown that paediatricians differ from other obstetricians in their assignment of Apgar scores [20]. For grading severity of neonatal encephalopathy between grade 2 and 3, the distinction between *common seizures* and *uncommon seizures* must be made. Although the gestational age of the fetus can now be determined fairly accurately using ultrasound, clinicians usually refer to *pre-term*, *term* or *post-term* fetuses, and doubt can still occur if ultrasound is not performed.

In acid-base assessment, where there are essentially four numeric parameters (arterial pH, arterial base deficit, venous pH and venous base deficit) to consider, there are still elements of uncertainty. Blood gas machines can only measure the pH and pCO_2 parameters to a certain accuracy, and it is not uncommon for example for two venous blood samples or one venous sample and one mixed arterial-venous sample to be taken accidentally. Even when the pH and base deficit are known, at what levels of base deficit does a low pH indicate a *respiratory*, *mixed* or *metabolic* acidemia? What is a *low* pH?

1.5.3 The Expert System Solution

Once each of the input variables has been assessed, labelled and/or numerically determined, the knowledge must be combined and processed to produce the required diagnosis. Whilst Apgar scores and neonatal complications are relatively easy to interpret in isolation, umbilical cord acid-base status is a difficult task in itself. Acid-base assessment requires considerable experience and expertise, which resides in relatively few clinicians. The ability to interpret cord acid-base status in conjunction with Apgar scores, neonatal progress and other clinical information will not be available on every labour ward, 24 hours per day. A computerised expert system, in which the knowledge of experts is captured, represented and processed would be a suitable solution. Such an expert system could process information

consistently, reliably and could be available continuously. The expert system must eventually incorporate uncertainty handling, so the uncertainty inherent in the domain must be characterised and modelled.

The ultimate goal of this work would be to produce a complete working expert system for the assessment of neonatal outcome of labour. A conceptual model of the final envisaged system is shown in Figure 1.7. The system would comprise five main modules for interpreting

1. umbilical cord acid-base parameters (arterial and venous pH, pCO_2 and pO_2),
2. other umbilical cord blood parameters, for example co-oximetry parameters, electrolytes such as sodium (Na^+), potassium (K^+), calcium (Ca^{++}) and chloride (Cl^-) ions, and metabolites such as glucose and lactate,
3. Apgar scores,
4. neonatal encephalopathy, and
5. other labour information.

Each of these modules would utilise uncertainty handling, with most of the primary input parameters having explicit uncertainty information. A further module would combine the outputs from each of these to form a comprehensive overall assessment of neonatal outcome. The final output(s) of the combination module would have to be presented in a form that was acceptable to the clinicians. This may well be a combination of numeric and linguistic outputs, with information on the derived uncertainty in each.

The *Apgar* module would take the five basic components of the Apgar score (see Table 1.1), and produce an output representing an assessment of the health of the infant. This work has in fact already been carried out by others [111, 110], who have published details of a *fuzzy-logic* based expert system (see Section 2.4 for details of fuzzy-logic). It is possible that this could be incorporated directly as the *Apgar* inference module. Unfortunately, however, it appears that this fuzzy expert system only utilised the one minute Apgar score, and thus missed out important information contained in the five minute score. Another problem is that

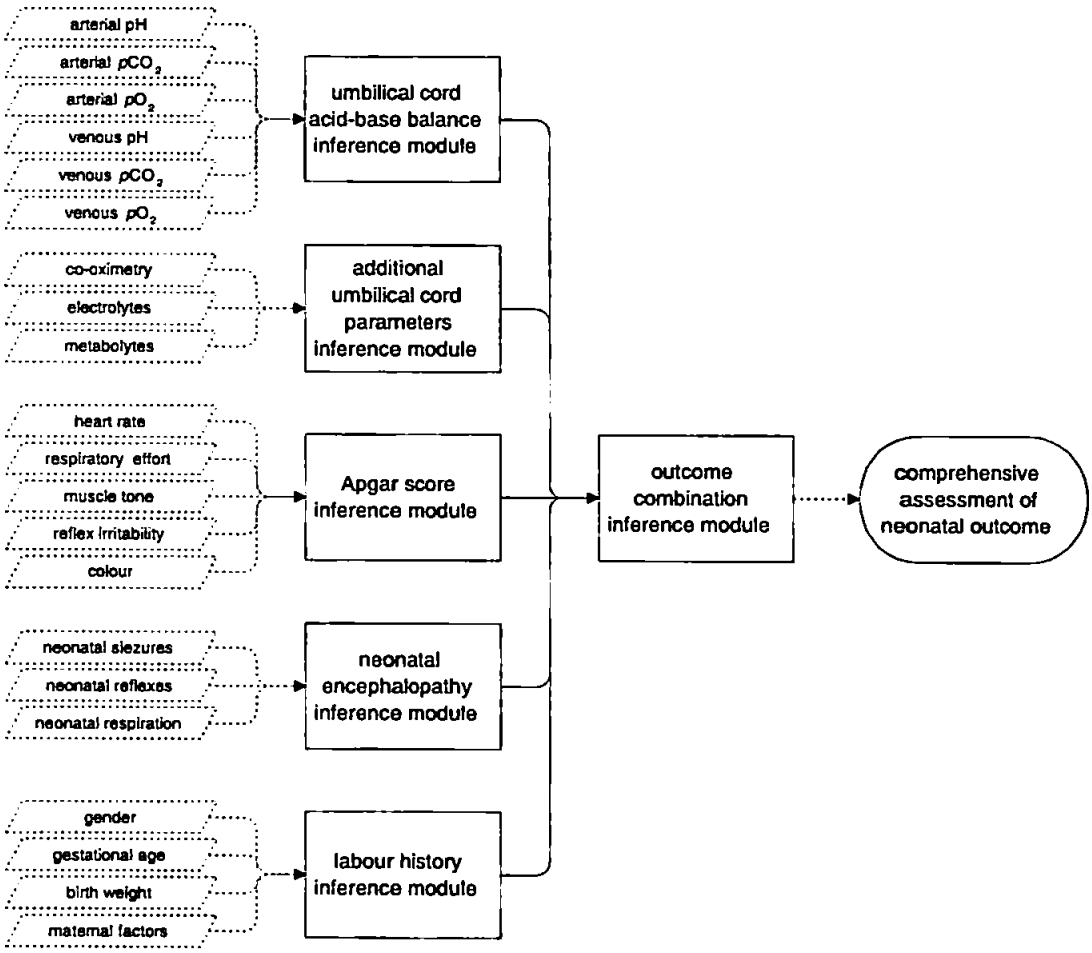


Figure 1.7: *a conceptual block diagram of an expert system for the comprehensive assessment of neonatal outcome*

their system was compared against *acidosis* defined as umbilical arterial pH < 7.20, which is generally accepted as unsatisfactory [6]. Thus, it might be necessary to enhance this work to utilise $Apgar^1$ and $Apgar^5$, and to test the output against more satisfactory endpoints.

The goal of this research was to develop a model of expertise to encapsulate the *objective* and *consistent* interpretation of immediate neonatal outcome, based on the measurement of umbilical cord acid-base status. This model would incorparate uncertainty handling techniques in order to represent and manipulate the uncertainty known to be present in the domain. Such a model would allow accurate comparisons to be made of obstetric care between hospitals and between treatments. As the entire project was large in scope, it was broken down into a number of smaller, achievable stages:

1. Development of a crisp expert system for umbilical acid-base analysis,
2. Characterisation of the uncertainty in umbilical acid-base parameters,
3. Development of an expert system for umbilical acid-base analysis incorporating uncertainty handling,
4. Development of a comprehensive model of uncertainty, characterising and evaluating the uncertainty in each of the input parameters, and
5. Development of a comprehensive expert system for the assessment of neonatal outcome incorporating uncertainty handling.

This thesis details the work done to achieve the first three stages above, in which, initially a crisp expert system, and then a fuzzy-logic based expert system to interpret umbilical cord acid-base results was developed. Chapter 2 introduces the basic concepts of expert systems and formal (crisp) logic, and then goes on to review a variety of methods of uncertainty handling, before concluding with a detailed introduction to fuzzy logic. This chapter is essentially technical background, and is only necesary if the reader is unfamiliar with these topics. Chapter 3 describes the development of the crisp expert system, including knowledge elicitation and spftware design. The first half of Chapter 4 describes the verification, validation and clinical assessment that took place as part of the evaluation of the crisp expert

system. The second half of this chapter (Section 4.6 to 4.9) illustrates a variety of uses of the acid-base data collected from the crisp expert system since its release, and is strictly optional. The development and optimisation of a preliminary fuzzy expert system to interpret *only* paired arterial–venous samples that had been previously validated by the crisp expert system is described in Chapter 5. An automatic tuning algorithm based on the simulated annealing technique was implemented to improve the fuzzy expert system after its performance was found to be poor in a validation study of fifty cases. The tuning algorithm and methodology employed are described in detail.

The post-optimisation performance of the fuzzy expert system was found to be very good on the tuning data, but was not re-validated on fresh data which would be neccessary to ensure that its improved performance was generalisable. Instead, the preliminary fuzzy model was extended to incorporate vessel identification rules and to enable it to deal with missing or invalid data. The development of this extended system, termed the *integrated fuzzy expert system*, is described in Chapter 6. The integrated fuzzy expert system was tuned with the same set of fifty cases, using the knowledge gained from the preliminary fuzzy system and fresh knowledge elicitation from clinical experts. The final part of Chapter 6 describes the rigourous validation study that took place after the development of the integrated fuzzy expert system was completed, in which novel cases were examined by clinical acid-base experts. This enabled the performance of the integrated system to be directly compared to experts on a previously unseen set of data.

Finally, the achievements and limitations of this research work, and a variety of possibilities for future work are reviewed and discussed in Chapter 7.

Chapter 2

Theory of Expert Systems

2.1 Rule Based Systems

2.1.1 Knowledge Representation

"An expert system is a computer program that represents and reasons with knowledge of some specialist subject with a view to solving problems or giving advice" [57]. Such a computer program simulates human reasoning over representations of human knowledge, using heuristic or approximate techniques. Expert systems manipulate symbolic representations of knowledge, rather than employing conventional algorithms.

Many real world problems are too complex to solve exhaustively and deterministically, for example whereas checking a proof of a theorem of propositional calculus is tractable, finding a proof for a theorem of propositional calculus is exponential in the number of variables. Such problems are often solved by human experts though, through the application of domain knowledge and heuristics (rules-of-thumb) to direct the search of problem space. As an example, the rules of chess which govern the meaning of pieces and their legal moves generates a total problem space (set of all possible games) that is far too large to exhaustively calculate the optimum move in each position. However, human experts can still always beat human novices at the game, and computers can still be programmed to play a very high level (expert)

game through the use of heuristics such as '*the queen is worth nine pawns*'.

Any knowledge representation must formally describe the syntax (the allowable forms) and semantics (the meaning) of the expressions used. The knowledge embodied in human experts can be thought of as comprising two distinct parts. Procedural knowledge describes the type of knowledge that allows us to know *how* to do things, whereas declarative knowledge refers to knowledge of *what* things are. For example, to follow a recipe to bake a cake from flour, eggs, butter and sugar it is necessary to have the procedural knowledge of what order to combine the relative quantities of ingredients, and the declarative knowledge of what each of the ingredients is (an egg is a small ovaloid object than comes from a hen, etc.). A number of methodologies have been developed to formalise the representation and manipulation of knowledge such as formal logic, production or rule-based systems, frames and associative networks [134].

2.1.2 Formal Logic

Classical formal logic was first developed by the ancient Greek philosophers, Plato and Aristotle, when they began to formalise the rules of thought involved in argument. Their goal was to abstract the processes of proving or disproving syllogisms, such as the famous:

> All men are mortal
>
> Socrates is a man
>
> Therefore, Socrates is mortal

In general, logic was predominately considered to be a branch of philosophy up to the nineteenth century when mathematicians began to develop the principles further. Mathematicians such as Boole, Frege and Russell took the principles of Aristotle and embodied them into a formal framework of symbolic manipulation to represent concepts such as truth, proof and inference.

The first concept is that all statements are either '*true*' or '*false*', represented by the symbols T and F, or sometimes 1 and 0. Lower case letters, traditionally $p, q, r, \ldots$, are used

to represent simple expressions or propositions like 'All men are mortal' or 'Socrates is a man'. Symbols are then introduced to represent the connectives of '*not*' (negation), '*and*' (conjunction), '*or*' (disjunction) and '*implies*' (implication), as shown in Table 2.1.

connective	*symbol*
not	$\neg$
and	$\wedge$
or	$\vee$
implies	$\Rightarrow$

Table 2.1: *logical connectives and their standard symbols*

The meaning of these connectives is most easily represented in *truth tables* such as Table 2.2.

p	q	$\neg p$	$p \wedge q$	$p \vee q$	$p \Rightarrow q$
T	T	F	T	T	T
T	F	F	F	T	F
F	T	T	F	T	T
F	F	T	F	F	T

Table 2.2: *truth table definitions for the primitives of classical logic*

From these, Aristotle's law of non-contradiction and law of the excluded middle can immediately be derived. The law of non-contradiction states that something cannot be both *true* and *false* at the same time. This is represented by the logic statement $(p \wedge \neg p) = F$. The law of excluded middle states that something must either be *true* or *false*, represented by $(p \vee \neg p) = T$. This is illustrated in truth table form in Table 2.3.

p	$\neg p$	$p \wedge \neg p$	$p \vee \neg p$
T	F	F	T
F	T	F	T

Table 2.3: *truth table of tautologies of non-contradiction and excluded middle*

This type of logic is called *propositional logic*, but the rules so far do not allow the proposition 'Socrates is mortal' to be derived from 'All men are mortal' and 'Socrates is a man'. To do this, the logic must be extended through the introduction of special operators called

quantifiers and special functions called *predicates*, to form a type of logic called *predicate logic*. In predicate logic simple variables $p, q, r, \ldots$, represent objects or things and predicate functions map their object arguments into *true* or *false* values. So for example, the predicate MAN(p) is representing whether the object p is a man or not. So MAN(Socrates) is *true* whereas MAN(Banana) is *false*.

The quantifiers are the universal $\forall$, meaning 'for all' and the existential, $\exists$, meaning 'there exists'. These allow statements such as $\exists p$: MAN(p) to represent the concept 'there exists an object such that the object is a man' (something is a man) or $\forall q$: MAN(q) to represent the concept 'for all objects the object is a man' (everything is a man). Using these constructs the original argument can be formalised as:

$$\forall p : \text{MAN}(p) \Rightarrow \text{MORTAL}(p) \tag{2.1}$$

$$\text{MAN(Socrates)} = T \tag{2.2}$$

$$\text{MORTAL(Socrates)} = T \tag{2.3}$$

2.1.3 Automatic Inference

The semantics of predicate logic have now been defined to place a meaning on each of the symbols, but it is still not yet possible to perform the reasoning to produce the conclusion (Equation 2.3) from the premises (Equations 2.1 and 2.2). To do this the syntax rules of how to produce statements from other statements in the formal language of predicate logic must be defined. Syntactic rules on how to transform the symbols in one statement into an alternative valid statement are known as *rules of inference*. A *proof* of a statement Φ in a formal system is a finite sequence of statements $\Phi_1, \Phi_2, \ldots \Phi_{n-1}, \Phi_n$ such that $\Phi_n = \Phi$ and each of which is either an *axiom* (a fact given as *true*) or follows from a lower numbered statement by a rule of inference.

There are two rules governing the introduction of qualifiers:

$$\text{P}(x) \rightarrow \forall x : \text{P}(x) \tag{2.4}$$

$$\text{P}(p) \rightarrow \exists x : \text{P}(x) \tag{2.5}$$

and two rules governing the elimination of qualifiers:

$$\forall x : \mathrm{P}(x), p \rightarrow \mathrm{P}(p) \tag{2.6}$$

$$\neg\exists x : \mathrm{P}(x), p \rightarrow \neg\mathrm{P}(p) \tag{2.7}$$

where P is a sentence of the system, p is a simple variable, x is a free variable, and the symbol $\rightarrow$ means '*derives*'. So Equation 2.6 for example allows 'MAN(Socrates) $\Rightarrow$ MORTAL(Socrates)' to be derived from '$\forall p$: MAN(p) $\Rightarrow$ MORTAL(p)' and the object 'Socrates'. A rule of inference known as *Modus Ponens* is needed to complete the derivation:

$$\Phi \Rightarrow \Psi, \Phi \rightarrow \Psi \tag{2.8}$$

which represents that if the implication $\Phi \Rightarrow \Psi$ is known and the fact Φ is known, then the fact Ψ can be derived. This finally allows the conclusion 'MORTAL(Socrates)' to be derived from the premises 'MAN(Socrates) $\Rightarrow$ MORTAL(Socrates)' and 'MAN(Socrates)'. Hence by following the formal rules of inference, allowing various syntactic substitution of symbols, the conclusion has been derived from the original premises. There are two other major rules of inference known as *Modus Tollens* and the *Resolution Principle*:

$$\Phi \Rightarrow \Psi, \neg\Psi \rightarrow \neg\Phi \tag{2.9}$$

$$\Phi \Rightarrow \Psi, \Phi \vee \Theta \rightarrow \Psi \vee \Theta \tag{2.10}$$

These can be illustrated with a simple example based on the implication 'if raining then cloudy'. With this implication and the fact 'it is raining', using *Modus Ponens* leads to the conclusion 'it is cloudy'. With the same implication, given the negation of the conclusion 'it is not cloudy', *Modus Tollens* works backwards to lead to the negation of the premise 'it is not raining'. The *Resolution Principle* gives 'it is sunny or cloudy' from 'it is raining or sunny' and 'if raining then cloudy'. This final example illustrates the point that although the reasoning may be formally logical and entirely consistent from the given facts and inference rules, the statement 'it is raining or sunny' does not actually reflect the real world.

2.1.4 Production Systems

Although the mathematical framework of predicate logic can be used to formalise the process of describing problems and performing deduction using the rules of inference, it is generally thought to be too complex and inflexible for real world problem solving. Human problem solving appears to be less formal and more adaptable than predicate logic, yet still manages to solve problems successfully. Consequently, the majority of expert systems also adopt a more informal approach, which although based on the principles of logic, attempts to match the more flexible human approach. The most popular approach is the *rule-based* or *production system* formalism. In a production system, knowledge is represented as a sequence of rules consisting of an *if* part and a *then* part, such as:

Rx	If	condition 1
		condition 2
		....
	Then	action 1
		action 2
		....

The general principle is that if all of the *conditions* or *premises* are satisfied then each of the *actions* or conclusions are performed. A working memory holds the currently known facts and an *inference engine* takes the known facts and generates new facts by determining which rules should be actioned and in which order. There are two alternative approaches to the control of rule-actioning termed *forward chaining* and *backward chaining*. With forward chaining the inference engine adopts the principle of *Modus Ponens* to infer new conclusions from the initial premises. Each new conclusion is added to the existing list of known facts and the generation of new facts continues until a goal state or terminal conclusion is reached. Conversely with backward chaining the inference engine starts with a hypothesis which is trying to be proven and works backwards using the principles of *Modus Ponens* and *Modus Tollens* to attempt to infer the initial premises.

In production system language *Modus Ponens* can be written as:

Fact	*A*	*true*	(it is raining)
Rule	If *A*	Then *B*	(if it is raining then it is cloudy)
Result	*B*	*true*	(it is cloudy)

and *Modus Tollens* as:

Fact	*B*	*false*	(it is not cloudy)
Rule	If *A*	Then *B*	(if it is raining then it is cloudy)
Result	*A*	*false*	(it is not raining)

Although *Modus Ponens* appears to be an obvious and almost trivial result, *Modus Tollens* appears to be harder to comprehend and less obvious, especially when written in the abstract form of a formal language. Forward chaining is more appropriate in situations where facts are few in quantity and the rule combinations will not produce too many potential alternatives, whereas backward chaining is more appropriate where facts are numerous and conclusions or hypotheses are available which are trying to be justified. In fact, most backward chaining algorithms employ a process of hypothesis testing. Given the rule 'If *A* Then *B*', the conclusion *B* is guessed (hypothesised) and the database is searched for fact *B*. If fact *A* is known, then the conclusion was valid.

The process of applying *Modus Ponens* and *Modus Tollens* is known as logical *deduction*. Logical *abduction* is the looser process (technically not valid) whereby given the rule 'If *A* Then *B*' and the fact *B* leads to the conclusion *A*. To see that this is not correct, consider a bird where the fact *Colour is Black* is known and the rule 'If *Bird is Crow* Then *Colour is Black*'. Clearly the conclusion *Bird is Crow* is premature as the bird could be another type that is black, such as blackbird! However logical abduction is often misused in practice, and indeed a form of abduction is used in backward chaining. A third form of logical reasoning is known as *induction*, which is the generalisation of large amounts of data to infer new rules. For example, 100 swans have been examined and each one was white, the rule 'If *bird is*

swan Then *colour is white*' can be induced. Clearly again, induction is not always logically valid, as the 101st swan may still be black!

As described, in general a rule may consist of any number of conditions and any number of conclusions. Most production systems allow the combination of multiple conditions via the connectives '*and*' and '*or*', corresponding to the $\wedge$ and $\vee$ of predicate logic, so that 'If *A and B* Then ... ' is only *true* if *both A* and *B* are *true*, whereas 'If *A or B* Then ... ' is *true* if *either A* or *B* (or both) are *true*. The operator '*not*' ($\neg$) to negate the sense of the condition is also allowed. A multiple conclusion indicates that each of the conclusion facts are added to the knowledge base in parallel if the conditions evaluate to *true*. Using these connectives complex rules may be constructed such as:

If ((*A and B*) *or* ((*not C*) *and D*) *or E*) Then *F and* (*not G*) *and H*

At any stage in the inference process it is possible that more than one rule may evaluate to *true* at any one time, in which case a decision as to which rule(s) to action, in which order, must be made. For example, in the following system:

Facts	*A*, *B*, *C*	
Rule1	If *A and B*	Then *D and not E*
Rule2	If *C or E*	Then *not D and F*
Rule3	If *A and D*	Then *G and H*

Both Rule1 and Rule2 can be actioned from the known facts, so that if Rule1 is actioned, either Rule2 or Rule3 could then be actioned (because *A*, *C*, and *D* would all be facts). However, if Rule2 is actioned before Rule1, Rule3 could not then also be actioned (as *D* is no longer a fact). There are a number of alternative methods of resolving such a situation, termed a *conflict-resolution strategy*, and the choice of strategy may cause a significant change in behaviour of the system. There are a number of possible strategies, for example:

- *Refractoriness*: A rule should not be allowed to fire more than once on the same data. This prevents a single rule firing multiple times.

- *Specificity*: The rule having the most number of conditions is the hardest to satisfy and should therefore be actioned first.

- *Rule ordering*: The rule at the top of the list has the highest priority and is actioned first.

2.1.5 Non-classical and Multi-Valued Logics

From the very earliest days of ancient Greek philosophy, there has been argument over the adequacy of the essential Aristotelian property that all statements must either be *true* or *false*. As discussed by Graham and Llewelyn-Jones [43], Aristotle himself maintained that future events are neither actually *true* or actually *false* but are potentially either and hence their truth value is indeterminate. The sophists school of philosophy was founded by Zeno a hundred years later. Zeno imagined a number of famous paradoxes to contradict the Aristotelian viewpoint [106]. For example, consider a pile of thousands of millions of grains of sand. Everyone would consider the pile to be a sand heap — in logic terms, the predicate SANDHEAP(Pile) = *true*. Similarly, everyone would agree that a single grain of sand does not constitute a heap, SANDHEAP(Grain) = *false*. But if grains of sand are removed one at a time, at what point does the pile change from a sand heap to NOT a sand heap?

In the 1930's a Polish mathematician, Lukasiewicz, introduced a three-valued logic with a new truth value, *indeterminate*, represented by the symbol *i*, ? or 1/2. Although the meaning of *not* is fairly easily defined in the new three-valued logic to be *not*(?) =?, the meaning of *and*, *or* and *implies* is more difficult to define. In fact there are many possible alternative definitions. Other systems of three-valued logic have also been proposed with different interpretations on the symbol ?. Bochvar interpreted ? as *meaningless*, Kleene as *undecided* and Heyting as *unknown* [46]. The truth tables for these logics are given in Table 2.4 [65].

Although each of these logics conform to the usual definitions for each operator when p and q are restricted to the classical values of T and F, they differ in their definition of the implication operator ($\Rightarrow$). In fact there are $(3^3)^3 = 19683$ possible combinations to choose

		Lukasiewicz			Bochvar			Kleene			Heyting		
p	q	$\wedge$	$\vee$	$\Rightarrow$	$\wedge$	$\vee$	$\Rightarrow$	$\wedge$	$\vee$	$\Rightarrow$	$\wedge$	$\vee$	$\Rightarrow$
F	F	F	F	T	F	F	T	F	F	T	F	F	T
F	?	F	?	T	?	?	?	F	?	T	F	?	T
F	T	F	T	T	F	T	T	F	T	T	F	T	T
?	F	F	?	?	?	?	?	F	?	?	F	?	F
?	?	?	?	T	?	?	?	?	?	?	?	?	T
?	T	?	T	T	?	?	?	?	T	T	?	T	T
T	F	F	T	F	F	T	F	F	T	F	F	T	F
T	?	?	T	?	?	?	?	?	T	?	?	T	?
T	T	T	T	T	T	T	T	T	T	T	T	T	T

Table 2.4: *conjunction, disjunction and implication operators of some of the more common three-valued logics*

from, each with a different interpretation and producing different results [43]. Also none of these three-valued logics satisfies the law of non-contradiction, $(p \wedge \neg p) = F$, or the law of excluded middle, $(p \vee \neg p) = T$, for all values. For example, in each case $(? \wedge \neg ?) = ?$ and $(? \vee \neg ?) = ?$. Lukasiewicz defined his truth table in terms of mathematical operations using 0 to represent *false*, 1 to represent *true* and $1/2$ to represent *indeterminate*, as shown in Table 2.5.

connective	operation
not	$\neg p = 1 - p$
and	$p \wedge q = \min(p, q)$
or	$p \vee q = \max(p, q)$
implies	$p \Rightarrow q = \min(1, 1 - p + q)$

Table 2.5: *connective definitions for Lukasiewicz logic*

Lukasiewicz then went on to generalise this further to include first n-valued logics with truth values labelled by rational numbers in the interval $[0, 1]$, obtained by evenly dividing the interval between 0 and 1 into n parts, and finally to infinite-valued logic where truth values can be any real number in the interval $[0, 1]$. Using the operators defined above the standard Lukasiewicz infinite-valued logic still reduces to classical logic at the extremes of 0 and 1. However there are an infinite variety of infinite-valued logics, each using a different form of implication operator [65].

2.2 Uncertainty in the Real World

2.2.1 Sources of Uncertainty

The real world is far from certain. In everyday experience it is rare to encounter facts that are absolutely *true* or *false* in the Aristotelian sense. In Section 2.1.5 the concept of several different types of non-certainty have been introduced, namely *indeterminate*, *undecided*, *unknown* and *meaningless* as well as the concept of degree of truth as represented by a truth value between 0 and 1. In general, uncertainty can be split into the two distinct notions of *vagueness*, representing concepts such as fuzziness, haziness, unclearness or indistinctiveness and *ambiguity*, representing concepts such as non-specificity, variety, generality or diversity.

In 1927 Heisenberg published his now famous *Uncertainty Principle* of quantum mechanics, which states that there is an ultimate limit in the accuracy to which an object's position and momentum (velocity multiplied by mass) may be known simultaneously. In mathematical terms:

$$\Delta x \Delta p_x \geq \frac{h}{4\pi} \tag{2.11}$$

where Δx is the uncertainty in the position, Δp_x is the uncertainty in momentum and h is Plank's constant ($\approx 6.626 \times 10^{-34}$Js) [34]. The principle states that the more exactly an object's position is measured, the less exactly the object's momentum can be known and *vice versa*.

Note that this startling result is not a reflection on the measuring device used, but is an incontrovertible fact of the quantum nature of the universe. It is not possible to 'beat' Heisenberg's Uncertainty Principle, whatever measuring device is used, and so there is always imprecision in the real world. Although this theoretical uncertainty is so small as to rarely affect real life measurements, it is nevertheless important as it establishes the fundamentally uncertain nature of the universe. In addition, the modern mathematics of Chaos Theory can be used to show that even minute changes in the underlying world can lead to dramatic alterations in

gross phenomena. The (often quoted) example is of how the beating of a butterfly's wings causes enough disturbance of the air around it to have sufficient 'knock-on' effect such that the earth's weather cannot be accurately predicted.

Human beings manage to reason and make decisions despite this uncertainty and it is important to mimic this imprecise reasoning in computer expert systems if they are to simulate human reasoning in a particular domain. The sources of this uncertainty are many, but in the context of expert systems the chief sources are [43]:

- unreliable sources of data and information
- abundance of irrelevant information
- imprecision of sensory apparatus
- faulty sensory equipment
- imprecision of natural language
- lack of understanding
- conflicting or complementary sources of facts
- hidden variables to produce apparent randomness

2.2.2 Precision, Accuracy and Reliability

In natural language the terms *precision*, *accuracy* and *reliability* are often used interchangeably, but technically they have quite distinct meanings. *Precision* refers to how closely or exactly a measurement is made. For example, stating a person's height to be 1829 mm is a more precise than stating they are 6 feet tall, even though the two statements are expressing the same height. Similarly, saying that they are 6 feet 0 inches tall is more precise than simply stating that they are 6 feet tall. Precision can be thought of as the number of decimal places to which a measurement is made.

Accuracy on the other hand refers to how well a measurement reflects the real world. For example, if a person is actually 185 cm (6 feet 1 inch) tall, then the statement 'the person is

about 6 feet tall' is more accurate than 'the person is 5 feet 3.5 inches tall' even though the latter is more precise. A measurement may be extremely precise, but completely inaccurate or on the other hand imprecise but accurate.

Finally, *reliability* refers to how stable a measurement is over time. For example, the measurement of a person's height is likely to be more reliable than their weight. Although both height and weight vary through each day, weight varies far more due to food intake, exercise, etc. On a longer time scale, as an adult, weight usually varies more through life than height. Unreliability may be inherent in the data itself (as in the weight and height example) or may be a reflection of the measurement device. For example, measuring a person's weight by the use of an electronic scale is likely to be more reliable than measuring weight by a beam balance, because the beam balance method will be affected by the exact positioning of the person and counter weights.

2.2.3 Uncertainty in Data

The examples above all describe uncertainty in data — that is, uncertainty in the facts that the (expert) system must deal with. This uncertainty may be any combination of imprecision, inaccuracy and unreliability. Data may be missing, for example sometimes it is only possible to take blood from a single vessel in the umbilical cord, rather than both the artery and vein, or the blood gas analyser may fail to produce a stable reading for a parameter due to air bubbles in the vicinity of the electrode. The blood gas analyser will have an inherent measurement precision and accuracy. Blood samples may be unreliable in that changes may occur as gaseous exchange with air takes place so that the sample degrades with time. As described earlier, there are considerable degrees of uncertainty in almost all the clinical data, other than the blood acid-base data, such as the vagueness in the subjective Apgar score or the considerable inaccuracy in the gestational age of the fetus (possibly reduced if confirmed by ultrasound examinations).

2.2.4 Uncertainty in Knowledge

In addition to the uncertainty present in the facts (or data base) of the expert system there may be uncertainty in the rules (or knowledge base) of the expert system. For example, uncertainty is present in the data if a person's height is known to be approximately 6 feet, whereas uncertainty is present in the rule if there is a statement such as 'if the person is tall then they usually have large feet'. Such rules with vagueness in the proposition often form a part of normal human reasoning, including expert knowledge. For example, in the case of umbilical acid-base analysis, most experts will give a rule like 'if all the values from the two samples are close then the samples probably came from the same vessel'.

Alternatively, the knowledge base may be incomplete, so that there are exceptions to rules or the rules do not cover the entire problem space. The rules may be contradictory, or at least various evidence may be simultaneously suggestive of different diagnoses. It might be difficult to formulate a set of rules that accurately represent what is going on in the mind of the expert. One set of input variables may lead to several possible conclusions.

2.2.5 Uncertainty Handling in Expert Systems

It is vital that an expert system that is ultimately to synthesise the performance of experts in a particular domain must address these issues of uncertainty management. An expert system should be 'aware' of the uncertainty in its conclusions to the same degree as the human experts are, and it should be able to represent this uncertainty to the user when presenting its conclusions. In general, a good mechanism for representing the uncertainty in an expert system should have the following properties [43]:

- consistent and natural semantics
- allow appropriate assumptions about independence
- easy and intelligible tracing of aggregation and propagation of uncertainty
- ability to explain the support for conclusions
- possess a conflict resolution strategy

There are a number of different techniques in existence for handling uncertainty, but the appropriate choice of method may well be context dependent. The focus of this work was to formulate a model for representing and handling the uncertainty present in the assessment of outcome of labour. It was originally envisaged that it might be necessary to formulate either a novel approach or a novel combination of existing approaches in order to achieve this successfully. Consequently, some of the existing methods of representing and manipulating uncertainty in expert systems will be reviewed.

2.3 Techniques for Representing Uncertainty

2.3.1 Probability Theory

Probability theory was first developed in the seventeenth century by Pascal and Fermat, who studied gambling games and games of chance. The work was principally extended by Bernoulli and de Moivre in the eighteenth century and later by Laplace, with contributions from many other eminent mathematicians including Gauss, Lagrange, Legendre and Poisson. In 1933 Kolmogorov first provided axioms to define probabilities in terms of sets of events, which underpinned the results already developed. A basic knowledge of set theory is assumed, summarised in Table 2.6.

Property	*Notation*	*Meaning*
Set	Ω	A collection of objects
Member	ω	An elementary object of Ω
Subset	A	A collection of objects contained within Ω
Complement	A^c	All objects in Ω that are not in A
Intersection	$A \cap B$	The collection of objects that are in both A and B
Union	$A \cup B$	The collection of objects either in A or B (or both)
Empty Set	$\emptyset$	A set containing no objects
Membership	$\omega \in \Omega$	ω is a member of Ω
Non-membership	$\omega \notin \Omega$	ω is not a member of Ω
Subsetship	$A \subseteq B$	A is a subset of B

Table 2.6: *notation for concepts of conventional set theory*

The set of all possible outcomes to an experiment is called the sample space and is denoted by Ω [45]. For example if a coin is tossed once there are two possible outcomes, heads or tails (denoted by H and T), so that $\Omega = \{H,T\}$, or if a six sided die is thrown once, then $\Omega = \{1,2,3,4,5,6\}$. Questions concerning the actual outcome of such an experiment can be formulated in terms of subsets of Ω. For example, the question 'is the result even' is equivalent to 'does the outcome lie in the subset $\{2,4,6\}$'. Thus, a set of questions that might be asked about the outcome of an experiment can be written as a set of subsets of the sample space Ω. Each such subset is referred to as an event.

So for example in the die throwing experiment, The following questions concerning the outcome are each given a representation with an event set:

is the outcome the number 1?	$A = \{1\}$
is the outcome an even number?	$A = \{2,4,6\}$
is the outcome even and 3 or less?	$A = \{2,4,6\} \cap \{1,2,3\}$
is the outcome not an even number?	$A = \{2,4,6\}^c$

The collection of subsets which is currently of interest is called the event space (denoted by $\Im$) and must have the following properties:

$$\emptyset \in \Im \tag{2.12}$$

$$\text{if } A_1, A_2, \ldots, A_n \in \Im \quad \text{then } \bigcup_{i=1}^{n} A_i \in \Im \tag{2.13}$$

$$\text{if } A \in \Im \quad \text{then } A^c \in \Im \tag{2.14}$$

In simple cases, such as the examples above, the event space $\Im$ is usually taken as the set of all possible subsets of Ω, called the power set of Ω and denoted by $\{0,1\}^{\Omega}$. For example in the coin tossing, $\Omega = \{H,T\}$ and $\Im = \{\emptyset, H, T, \Omega\}$.

This has defined the experiment and possible events that might be of interest, but has not yet defined a probability. To do this a probability measure **P** is defined to be a function taking

elements of the event space $\Im$ and producing a real number in the interval [0, 1], such that:

$$\mathbf{P}(\Omega) = 1 \tag{2.15}$$

$$\mathbf{P}\left(\bigcup_i^n A_i\right) = \sum_i^n \mathbf{P}(A_i) \quad \text{where } A_1, A_2, \ldots, A_n \text{ is a collection of disjoint members of } \Im, \text{ such that } A_i \cap A_j = \emptyset \text{ for all pairs } i, j \text{ satisfying } i \neq j \tag{2.16}$$

The triple $(\Omega, \Im, \mathbf{P})$ consisting of the sample space Ω, the event space $\Im$, and a probability measure $\mathbf{P}$ on $(\Omega, \Im)$ is called a *probability space*. For example in the coin tossing, $\Omega = \{H, T\}$ and $\Im = \{\emptyset, H, T, \Omega\}$ and a possible probability measure $\mathbf{P}$ is given by:

$$\mathbf{P}(\emptyset) = 0,\ \mathbf{P}(H) = p,\ \mathbf{P}(T) = 1 - p,\ \mathbf{P}(\Omega) = 1 \tag{2.17}$$

This axiomatic definition immediately leads to some theorems of classical probability. Firstly that the probability of any event occurring is one minus the probability of it not occurring:

$$\mathbf{P}(A) = 1 - \mathbf{P}(A^c) \tag{2.18}$$

and secondly that the probability of event A or event B occurring is equal to the probability of event A plus the probability of event B minus the probability of event A **and** event B occurring:

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) \tag{2.19}$$

Although this approach has defined the axioms of probability theory, it has not yet placed any meaning on the probability values in the probability space. There are in fact several philosophical views of the meaning of probabilities.

- The Frequency Interpretation. Probability is viewed as being meaningful only in cases where an experiment may be repeated. The probability of an event can then be defined as the proportion of number of times that the event occurs in repeated trials of the experiment. The assumption is then that the probability indicates the chance of a future repetition of the same experiment producing the required result. Proponents of this view refuse to allow probabilities to be assigned to unrepeatable experiments such as, for example, the probability of being run over and killed by a bus!

- The Objective Interpretation. Probability is thought of as an inherent physical characteristic, so that for example on the quantum mechanical level the probability distributions of particles are real. The dice or coin are made up of atoms and have a certain composition and symmetry such that it will fall on a certain face with an inherent probability. This approach assumes that probabilities exist independently from an observer and that these probabilities may be determined by experience. Objectors to this viewpoint have included Albert Einstein, as exemplified by his famous quote "God does not play dice with the universe".

- The Subjective Interpretation. Real gamblers do make probabilistic judgements as to whether a football team will win the FA Cup each year, and insurance companies as to whether a person may get run over by a bus. This approach views all probabilities as purely personal subjective beliefs that an event will occur. A probability is then simply the odds that a person would accept if asked to play an abstract game of gambling on the event.

Whatever the physical interpretation of probability, these axioms lay down a strict mathematical framework for manipulating and combining probabilities. Most advocates of probability theory accept that real humans (including experts) rarely use these rules rigorously when dealing with probabilities in decision making. Nevertheless these advocates do suggest that any computer system dealing with uncertainty ought to employ formal probabilities. Many opponents of probability, on the other hand, frequently make the mistake of thinking that probability must be based on either the frequency interpretation or the objective interpretation. Neither is true: deFinetti derives the axioms of probability theory from a subjective viewpoint [26] and he and others [87] have investigated second order probabilities (probabilities of probabilities) to allow for situations where a subjective view of a probability is changing with evidence or is not point valued.

2.3.2 Possibility Theory

An alternative to probability theory is *possibility* theory, first introduced by Zadeh in 1978 [143]. Zadeh stated that:

> "Intuitively, possibility relates to our perception of the degree of feasibility or ease of attainment whereas probability is associated with a degree of likelihood, belief, frequency or proportion"

In possibility theory, the conditions to define a probability measure (Equations 2.15 and 2.16) are modified so that a possibility measure, Π, is defined by [30]:

$$\Pi(\emptyset) = 0, \quad \Pi(\Omega) = 1 \tag{2.20}$$

$$\Pi\left(\bigcup_i^n A_i\right) = \max_i^n \Pi(A_i) \qquad \text{where } A_1, A_2, \ldots, A_n \text{ is any collection of subsets of } \Omega \tag{2.21}$$

Conceptually, events can have a high possibility and a low probability, but not *vice versa*. Mathematically, this is expressed by the relationship:

$$\mathbf{P}(A) \leq \Pi(A) \tag{2.22}$$

For example, consider the statement "Jon ate x eggs for breakfast" for different values of x. An illustration of the associated possibility distribution and probability distribution for this statement for various values of x is shown in Table 2.7. Whereas the probability distribution could have been obtained from a frequency perspective by observing the person at (say) 100 breakfasts and noting the number of eggs that was eaten at each, the possibility distribution could (theoretically) be obtained by feeding the subject with an increasing number of eggs each day until eating more became impossible.

	x	1	2	3	4	5	6	7	8	9	10
Possibility	$\Pi(x)$	1.0	1.0	1.0	1.0	0.8	0.6	0.4	0.2	0.1	0.0
Probability	$\mathbf{P}(x)$	0.1	0.8	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 2.7: *example showing the difference between* possibility *and* probability

2.3.3 Bayesian Reasoning

The mechanics of using probability theory in expert system decision making usually rests on an early theory by Bayes in 1763. If A and B are events ($A, B \in \Im$) and it is known that event B occurs, then the probability that A occurs may no longer be $\mathbf{P}(A)$. Clearly, A only occurs if and only if A **and** B occurs. Hence the conditional probability that A occurs, given that it is known that B occurs, is defined as:

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} \tag{2.23}$$

This definition immediately leads to *Bayes' Theorem* which states that:

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(B|A)\mathbf{P}(A)}{\mathbf{P}(B)} \tag{2.24}$$

This can be expressed as stating that the *posterior* probability of A, $\mathbf{P}(A|B)$, is given by the *prior* probability of A, $\mathbf{P}(A)$, multiplied by the ratio of the conditional probability of B occurring given A, $\mathbf{P}(B|A)$, to the overall probability of B, $\mathbf{P}(B)$. In order to apply this theorem usefully, the *Partition Theorem* is also needed. A partition of $(\Omega, \Im, \mathbf{P})$ is a collection of disjoint events B_i ($B_i \in \Im$ such that $B_i \cap B_j = \emptyset$ for all pairs i, j satisfying $i \neq j$) with union $\cup_i B_i = \Omega$. The partition theorem states that:

$$\mathbf{P}(A) = \sum_i \mathbf{P}(A|B_i)\mathbf{P}(B_i) \quad \forall A \in \Im \tag{2.25}$$

As an example of Bayesian reasoning consider an experiment in which four balls have been placed in a bag, either three green and one red or one green and three red. Initially there is no information as to which is the case, so it is assumed that the prior probability of either is 0.50.

Hypothesis A (3 Green and 1 Red): $\mathbf{P}(A) = 0.50,\ \mathbf{P}(G|A) = 0.75,\ \mathbf{P}(R|A) = 0.25$

Hypothesis B (1 Green and 3 Red): $\mathbf{P}(B) = 0.50,\ \mathbf{P}(G|B) = 0.25,\ \mathbf{P}(R|B) = 0.75$

So by the partition theorem:

$$\mathbf{P}(G) = 0.50 \times 0.75 + 0.50 \times 0.25 = 0.50$$

$$\mathbf{P}(R) = 0.50 \times 0.25 + 0.50 \times 0.75 = 0.50$$

Now assume that a green ball is drawn from the bag, labelled event E_1. By Bayes Theorem:

$$\mathbf{P}(A|E_1) = \mathbf{P}(A) \times \mathbf{P}(E_1|A)/\mathbf{P}(E_1) = 0.50 \times 0.75/0.50 = 0.75$$

$$\mathbf{P}(B|E_1) = \mathbf{P}(B) \times \mathbf{P}(E_1|B)/\mathbf{P}(E_1) = 0.50 \times 0.25/0.50 = 0.25$$

The probability (or belief) that hypothesis A was initially correct has now been increased due to the fact that a green ball was picked from the bag. In the light of this belief the overall chance of picking a green or red ball has also been modified so that the partition theorem now gives:

$$\mathbf{P}(G) = 0.75 \times 0.75 + 0.25 \times 0.25 = 0.625$$

$$\mathbf{P}(R) = 0.25 \times 0.75 + 0.75 \times 0.25 = 0.375$$

The ball is now replaced and the experiment repeated a second time. Another green ball is drawn from the bag and this is labelled event E_2. Applying Bayes Theorem a second time gives:

$$\mathbf{P}(A|E_2) = \mathbf{P}(A) \times \mathbf{P}(E_2|A)/\mathbf{P}(E_2) = 0.75 \times 0.75/0.625 = 0.90$$

$$\mathbf{P}(B|E_2) = \mathbf{P}(B) \times \mathbf{P}(E_2|B)/\mathbf{P}(E_2) = 0.25 \times 0.25/0.625 = 0.10$$

Each time that this was done the belief in hypothesis A would increase. However, this belief would never reach certainty (one), but would approach it asymptotically. A person or expert system would have to set an arbitrary limit of certainty above which it would be accepted that sufficient evidence had proven a hypothesis. Another problem with Bayes Theorem is that it can be difficult to calculate the conditional probabilities required, e.g. given a patient with chest pains, what is probability of a heart attack? In order to do this **P** (heart attack), **P** (chest pain) and **P** (heart attack AND chest pain) are all needed. Bayes theory gets difficult when considering a number of events (pieces of evidence) because all the joint probabilities are required. If joint independence is not assumed, a large number of probabilities is needed and the calculation becomes computationally expensive but if independence is assumed, the calculation is invalid if the assumption is incorrect.

2.3.4 Certainty Factors

To overcome these problems with Bayesian reasoning, the MYCIN expert system to identify bacterial infection [112] utilised a novel method of uncertainty reasoning termed *certainty factors*. Each hypothesis had an associated certainty factor ranging from -1 to +1, where -1 represented the belief that the hypothesis was wholly *false*, 0 represented *unknown* and +1 represented the belief that the hypothesis was wholly *true*. The certainty factor was itself computed as the difference of two measures, the current measure of belief (MB) and the current measure of disbelief (MD), so that:

$$CF(H|E) = MB(H|E) - MD(H|E) \tag{2.26}$$

for each hypothesis H given evidence E, where measures of belief and disbelief lay in the range $[0, 1]$.

The MYCIN system consisted of a number of rules, each of which added belief to a hypothesis in the light of evidence. Each rule had an inherent *tally factor* which represented the certainty of the inference itself. Rules could comprise multiple antecedents where the operations of conjunction (*and*) and disjunction (*or*) were implemented as maximum and minimum operators of multi-valued logic, respectively. The calculated belief in the rule was multiplied by the tally factor to give the overall truth value of the rule. When two rules added evidence E_1 and E_2 the combined belief was calculated by:

$$MB(H|E_1, E_2) = MB(H|E_1) + MB(H|E_2)[1 - MB(H|E_1)] \tag{2.27}$$

For example, consider a system with two rules:

IF engine IS NOT turning THEN battery flat (T 0.6)

IF horn IS NOT blowing THEN battery flat (T 0.9)

and assume that the state of knowledge was that (engine IS NOT turning = 0.7) and (horn IS NOT blowing = 0.3). After the first rule was applied $CF = 0.7 \times 0.6 = 0.42$, and then after the second rule was applied $CF = 0.42 + (0.3 \times 0.9) \times (1 - 0.42) = 0.58$. This result applies

irrespective of the order in which the rules were fired. However, this can *not* produce the same result if the two rules were combined to give:

IF engine IS NOT turning AND horn IS NOT blowing THEN battery flat (T ?)

In this case the truth of the antecedent would be $\min(0.7, 0.3) = 0.3$, and the truth of the consequence would be 0.3 times the tally factor. As the tally factor must lie in the range $[0, 1]$, this cannot give the same result as previously (0.58). Consequently, it was important how rules were formulated and it may be totally unobvious to the experts or knowledge engineer which form is correct. Another problem with certainty factors was that they have no formal basis, statistically or probabilistically, so that, for example, certainty factors for mutually exclusive hypotheses need not sum to unity. It has also been shown [2] that it is possible for certainty factors to radically depart from probability theory. For example, if some evidence E supports two hypotheses H_1 and H_2, it is possible that $CF(H_1|E) < CF(H_2|E)$ despite the fact that $\mathbf{P}(H_1|E) > \mathbf{P}(H_2|E)$ and $\mathbf{P}(H_1) > \mathbf{P}(H_2)$.

2.3.5 Dempster-Shafer Theory of Evidence

In the Dempster-Shafer theory of evidence [109] the event space is always defined to be the power set of the sample space. Whereas in probability theory the axioms force the probabilities assigned to the elementary members ω_i to sum to unity, the Dempster-Shafer theory loosens this restriction, allowing some of the probability effectively to be 'unassigned'. Instead, it is required that the probability assignments of all the possibilities sum to unity. For example, for a single coin toss, in probability theory, $\Omega = \{H, T\}, \Im = \{\emptyset, H, T, \Omega\}$ and the probability might be assigned $\mathbf{P}(\emptyset) = 0$, $\mathbf{P}(H) = 0.5$, $\mathbf{P}(T) = 0.5$, $\mathbf{P}(\Omega) = 1$, whereas in Dempster-Shafer, $\Omega = \{H, T\}, \Im = \{\emptyset, H, T, \Omega\}$ and the probability might be assigned $\mathbf{P}(\emptyset) = 0$, $\mathbf{P}(H) = 0.2$, $\mathbf{P}(T) = 0.2$, $\mathbf{P}(\Omega = \{H, T\}) = 0.6$.

In Dempster-Shafer theory the event $\{H, T\}$ indicates that the outcome was heads or tails, but it is not known which. Thus, the theory maintains a distinction between the event $\{H, T\}$

representing a head or a tail but it is not known which, and the events $\{H\} \cup \{T\}$ representing the known outcome of a head *or* the known outcome of a tail. The assignment of probability given to each subset is termed the *basic probability assignment* or *bpa* denoted by the function m, and must obey $m(\emptyset) = 0$ and $m(\Im) = 1$. The belief in and the plausibility of a hypothesis A, are given by:

$$Bel(A) = \sum m(B) \quad \text{where } B \subseteq A \tag{2.28}$$

$$Pls(A) = \sum m(B) \quad \text{where } B \cap A \neq \emptyset \tag{2.29}$$

or equivalently:

$$Pls(A) = 1 - Bel(A^c) \tag{2.30}$$

For example, if a patient has one and only one of hepatitis (H), cirrhosis (C) or pancreatic cancer (P), then:

$$\Omega = \{H, C, P\}$$

and

$$\Im = \{\{\emptyset\}, \{H\}, \{C\}, \{P\}, \{H,C\}, \{H,P\}, \{C,P\}, \{H,C,P\}\}$$

If each of these alternatives has a *bpa* as follows $\{0, 0.2, 0.2, 0.2, 0.1, 0.1, 0.1, 0.1\}$, then:

$$\begin{aligned} Bel(\{H\}) &= m(\{H\}) = 0.2 \\ Pls(\{H\}) &= 1 - Bel(\{H\}^c) \\ &= 1 - Bel(\{C,P\}) \\ &= 1 - \left[m(\{C\}) + m(\{P\}) + m(\{C,P\})\right] \\ &= 1 - (0.2 + 0.2 + 0.1) \\ &= 0.5 \end{aligned}$$

From these definitions, it is always the case that:

$$0 \leq Bel(A) \leq Pls(A) \leq 1 \tag{2.31}$$

Thus the interval given by $[Bel(A), Pls(A)]$ represents the interval between the amount of belief committed to a hypothesis and the amount of belief not committed to the negation of the hypothesis. This size of this interval ($[0.2, 0.5]$ in the example above), termed the belief interval, reflects the amount of uncertainty in the underlying beliefs. If, in the example above the *bpa* were assigned as $\{0,0,0,0,0,0,0,1\}$, then $Bel(H) = 0$ and $Pls(H) = 1$, and similarly for C and P, so that the belief interval for each would be $[0, 1]$. This would indicate that there was no evidence at all as to which one of the diseases the patient had.

The Dempster-Shafer theory then provides rules for the combination of *bpa*'s from the combination of two pieces of evidence, by multiplying through each power set. Unfortunately, the requirement to represent the power set of alternatives means that the combination rule is exponential. For instance, even in the small example above there are three diagnoses and therefore $2^3 = 8$ elements in the hypothesis power set. Excluding the empty set, whose *bpa* is always 0, there are $7 \times 7 = 49$ calculations of joint evidence required for each of the 7 hypotheses. Hence the complete calculation requires over $7 \times 49 = 343$ multiplications (there are more because there is also a normalisation factor involved)! Another problem is that there is no intuitive method of interpreting (or ranking) the belief intervals to decide which hypothesis is ultimately most favoured.

2.3.6 Fuzzy Sets

In 1937 the quantum philosopher Max Black had published a paper which defined the key concepts of a *vague* set [13]. *Fuzzy* sets were first defined in their present form by Lotfi Zadeh in 1965 [138] to provide a method for constructing numerical controllers for complex electronic equipment. Zadeh summarised his motivation in his *Principle of Incompatibility* [139]:

> "as the complexity of a system increases, our ability to make precise and yet significant statements about its behaviour diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics"

Zadeh extended the work to include the concept of a linguistic variable, which has values that are words or sentences in natural language, and the concept of fuzzy logic [140, 141, 142], which extended the Lukasiewicz multi-valued logic. More recently, Zadeh stated that fuzzy logic provides a systematic framework which 'makes it possible to deal with different types of uncertainty' [144]. Fuzzy theory has generated widespread interest and the theoretical work has since been extended to show that fuzzy theory contains both classical probability theory and the Dempster-Shafer belief and plausibility measures [65, 66]. Fuzzy theory is probably now the most popular form of including uncertainty handling within expert systems.

2.4 Theory of Fuzzy Sets and Fuzzy Systems

2.4.1 Fuzzy Sets

In conventional set theory, the set of positive integers that are less than 10 may be written as:

$$\Omega = \{1,2,3,4,5,6,7,8,9\}$$

Consider the subset of even integers, given by $A = \{2,4,6,8\}$. The complement of A, A^c is the set of all members of the universal set that are not in A, so that $A^c = \{1,3,5,7,9\}$. Each individual element is either a member of a set or is not a member of a set. This leads to results such as the union of a set and its complement is the universal set, $A \cup A^c = \Omega$, and the intersection of a set and its complement is the empty set, $A \cap A^c = \emptyset$. Thus classical set theory is conceptually and mathematically similar (isomorphic) to classical logic. Sets may be finite, such as the even integers below ten in the example above, or infinite, such as the set of all integers.

Consider, however, the set of tall people or the set of fast cars. Such concepts seem to have an intuitive meaning, but how can they be defined? Whereas it seems clear that a 6 feet 6 inches tall person would be a member of the set of tall people, what about a person who is 5 feet 10 inches, or 5 feet 8 inches? In classical set theory a boundary would be defined such

that people above the boundary height would be considered members of the set and those below the boundary would not be members. This can be thought of in terms of each person having a truth value of *true* or *false* as to their membership of the set. The term *crisp* has now been introduced to refer to classical non-fuzzy sets or non-fuzzy logic.

In the same way as classical logic is extended to multi-valued logic through intermediate truth values, so classical set theory can be extended to fuzzy set theory through the use of intermediate membership values. Now each person can be thought of as having a membership grade between 0 and 1 which indicates their degree of membership of the set of tall people (Figure 2.1). Thus a 6 feet 6 inches tall person may have a membership grade of 1 (they are certainly in the set of tall people), a 5 feet 0 inches tall person may have a membership grade of 0 (they are certainly not in the set of tall people) and a 5 feet 6 inches tall person may have a membership grade of 0.5 (they are considered half in the set and half not in the set of tall people).

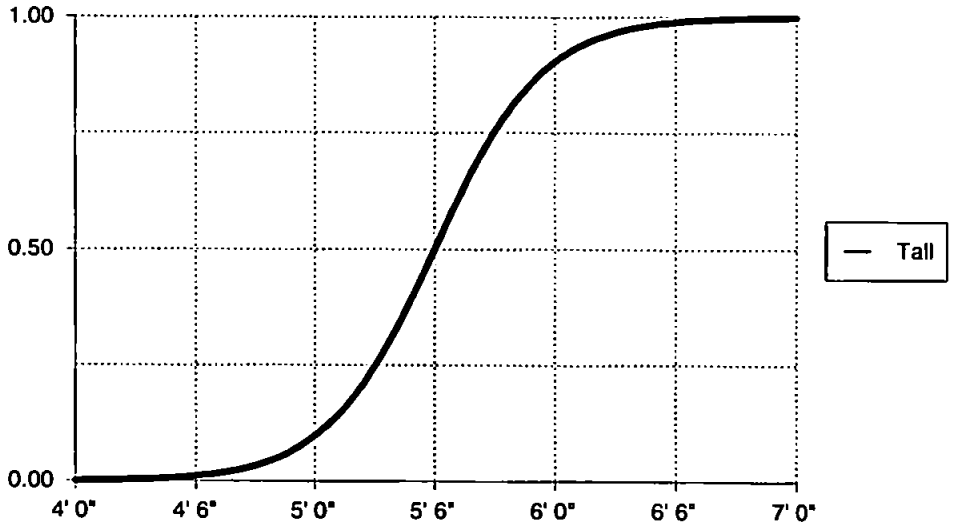


Figure 2.1: *a fuzzy set representation of tall people*

2.4.2 Membership Functions

The *universe of discourse* for a fuzzy set is defined to be the universal set of values over which a *base* variable ranges. In the example of the height of people, the base variable is height and the universe of discourse may be 3 feet 0 inches to 8 feet 0 inches. The membership

grade, μ, of a fuzzy set relates each value in the universe of discourse to its membership of that particular fuzzy set. A finite non-fuzzy set such as $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ may be written as $\Omega = \omega_1 + \omega_2 + \ldots + \omega_n$, where the + symbol represents the union operator rather than the arithmetic sum. In the case of finite fuzzy sets this can be written as $\Omega = \mu_1/\omega_1 + \mu_2/\omega_2 + \ldots + \mu_n/\omega_n$, where μ_i represents the membership grade of each element ω_i in the fuzzy set. This can equivalently be written as $\Omega = \sum_{i=1}^{n} \mu_i/\omega_i$. For example, as above the crisp set of positive integers less than ten is:

$$\Omega = \{1,2,3,4,5,6,7,8,9\} = 1+2+3+4+5+6+7+8+9 = \sum_{i=1}^{9} i \tag{2.32}$$

The fuzzy set of positive integers that are *near five* may be:

$$\Omega = 0/1 + 0/2 + 0.1/3 + 0.33/4 + 1/5 + 0.33/6 + 0.1/7 + 0/8 + 0/9 \tag{2.33}$$

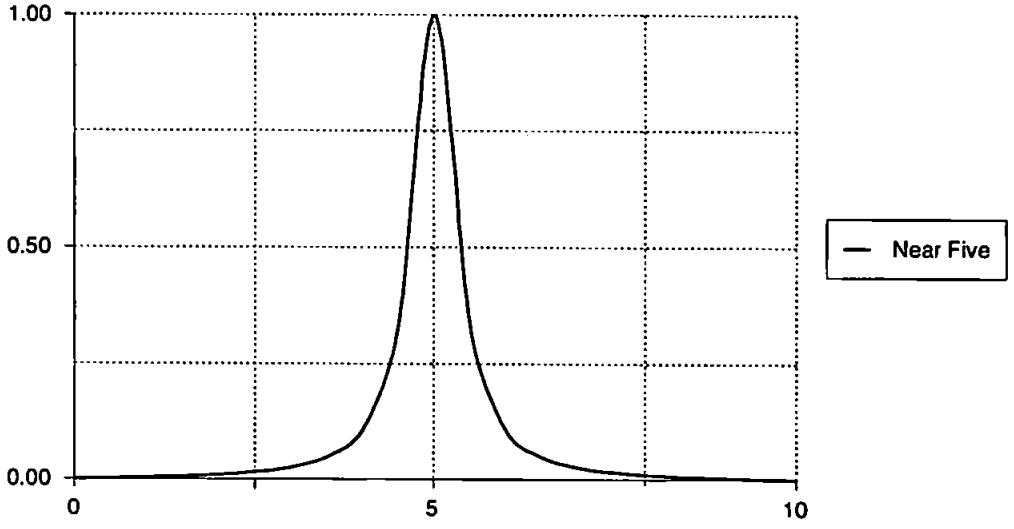


Figure 2.2: *a fuzzy representation of the number* near five

The membership grade may itself be a function, termed a *membership function*, of the base variable. For example, in the fuzzy set above it could have been defined as:

$$\Omega = \sum_{i=1}^{9} \mu(i)/i \qquad \text{where } \mu(x) = \frac{1}{1+2(x-5)^2} \tag{2.34}$$

For infinite sets, such as the set of real numbers, the fuzzy set may be written as $\Omega = \int_x \mu(x)/x$. In this case the resulting fuzzy set would be as shown in Figure 2.2. A fuzzy set

is termed *normal* if its membership grade reaches one in at least one point in its universe of discourse and is termed *regular* if it has a single peak in its membership function. Two fuzzy sets are considered equal if they share the same universe of discourse and $\mu_1(x) = \mu_2(x)\ \forall x$. In general the shape of the membership function can be any shape, but some of the more common shapes in practice are triangular, Gaussian, Sigmoid or 'S-shaped'. Each of these is demonstrated in Figures 2.3(a) – (d).

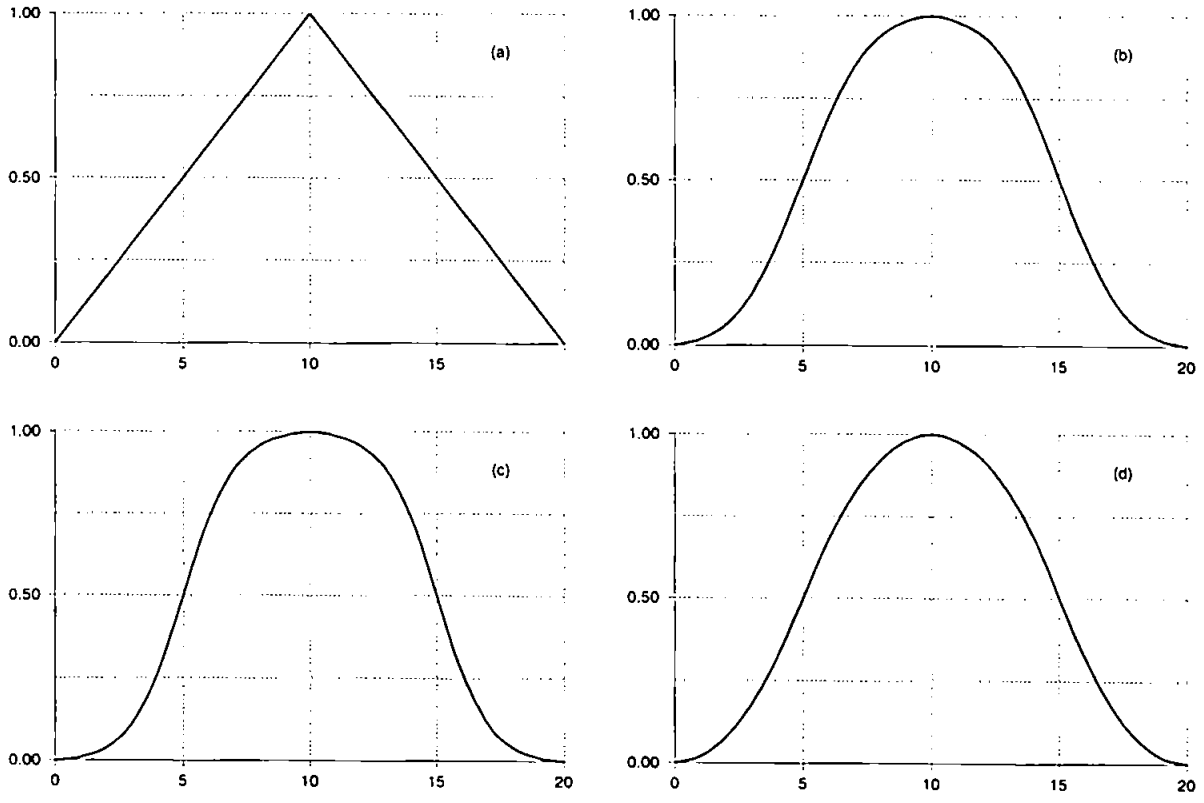


Figure 2.3: *membership functions of shape (a) triangular, (b) Gaussian, (c) Sigmoid and (d) S-shaped*

It is important to realise that although fuzzy membership grades may be given a probabilistic interpretation, it is by no means necessary or even desirable. For instance, if 100 people were asked 'would you describe the number four as *near five*?' and 33 out of the 100 said 'yes', then from the frequency interpretation it could be said that the probability of 'four' being '*near five*' was 0.33 and that therefore 'four' has a membership of 0.33 in the fuzzy set '*near five*'. Alternatively, it could be viewed that the number 'four' is 0.33 compatible with the concept of '*near five*'. In the height of a person, there is no probability involved as to what height they are, but their height can still be mapped onto the fuzzy sets 'tall',

'medium' or 'short', so somebody whose height is 5 feet 6 inches is 0.5 compatible with the description 'tall'. Another point is that membership grades may be subjective and/or context sensitive. For example, a person whose height is 5 feet 0 inches might generate a different set of membership grades to a person whose height was 6 feet 6 inches, and similarly the set of grades might differ if used to describe European males and Japanese females.

2.4.3 Linguistic Variables

A concept such as a person's height can be thought of as a linguistic variable [140, 141, 142] — that is, a variable that takes the value of linguistic terms such as 'tall' or 'short'. The set of allowable terms, each defined as a fuzzy set over the variable's universe of discourse, is called the *term set*. Thus, in the height example the linguistic variable is *height* with term set *short* + *medium* + *tall*. Operators, such as *and* and *or*, and hedges such as *very*, *fairly* or *somewhat*, can then be defined to modify these linguistic terms to produce other terms that map onto the equivalent natural language terms.

Such linguistic variables can then provide a basis for *approximate reasoning*, in which the mode of reasoning is no longer exact. This allows the creation of a framework for reasoning that may be more representative of real human reasoning than traditional two-valued or even multi-valued logic [140].

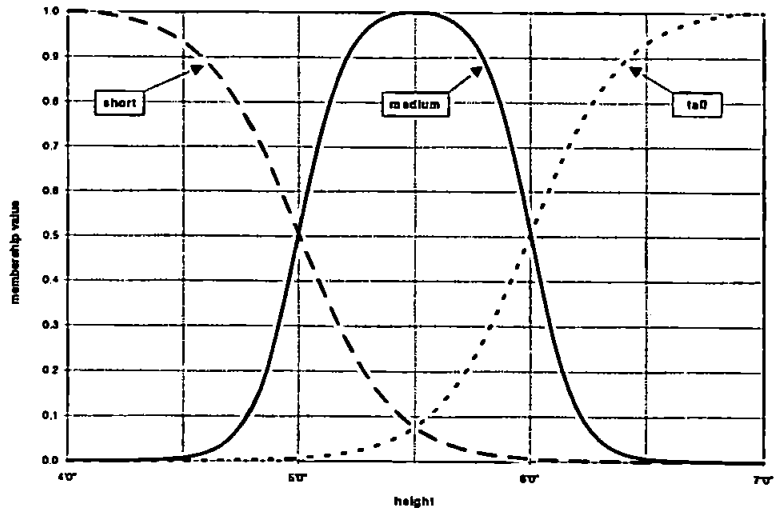


Figure 2.4: *terms in an example of a linguistic variable* height

2.4.4 Fuzzy Operators

Having defined the basic concept of a fuzzy set, it is possible to define a number of operations that can be performed on such sets. The main operators are the complement, the union and intersection. These are usually defined by the 'Zadeh' family as follows:

Complement	A^c	$\mu_A(x) = 1 - \mu_A(x)$	(2.35)
Intersection	$A \cap B$	$\mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)]$	(2.36)
Union	$A \cup B$	$\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)]$	(2.37)

Thus, the complement corresponds to the *not* of multi-valued logic, the union corresponds to *or* and the intersection corresponds to *and*. However, there are a number of other possible alternative definitions for each of these operators, for example the 'probabilistic' family:

Intersection	$A \cap B$	$\mu_{A \cap B}(x) = \mu_A(x) \cdot \mu_B(x)$	(2.38)
Union	$A \cup B$	$\mu_{A \cup B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x) \cdot \mu_B(x)$	(2.39)

The properties that each class of such operators should obey have been defined [65] and it can be shown that the standard *max* and *min* operators define the 'strongest' (most restrictive) union and the 'weakest' (least restrictive) intersection. With these operators it is possible to calculate the fuzzy sets corresponding to 'short *or* tall' or 'short *and* tall', for example as shown in Figures 2.5(a) and (b).

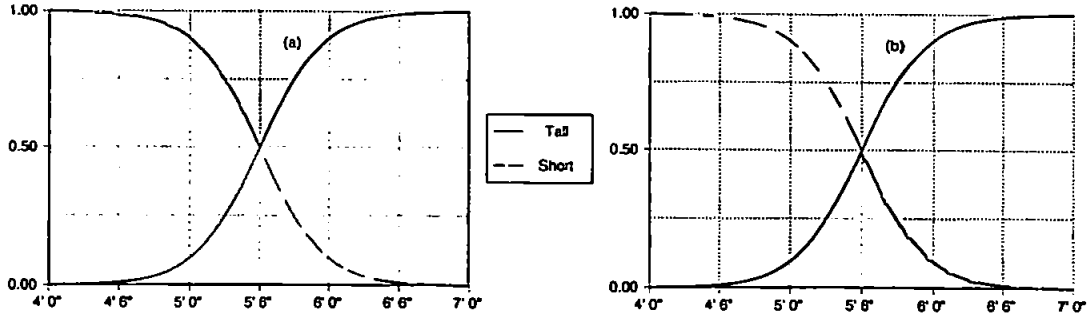


Figure 2.5: *fuzzy set (a) union 'short* or *tall', (b) intersection 'short* and *tall'*

2.4.5 Fuzzy Hedges

As stated above the complement, union and intersection operators define the connectives *not*, *or* and *and*. A number of other operators may also be defined to correspond to linguistic hedges such as the terms '*very*', '*fairly*', '*somewhat*' as used in everyday natural language. The main operators are the concentration, dilation, normalisation and intensification. These are defined as:

$$\text{Concentration} \qquad (\mu_A)(x) = \mu_A^2(x) \tag{2.40}$$

$$\text{Dilation} \qquad (\mu_A)(x) = \mu_A^{1/2}(x) \tag{2.41}$$

$$\text{Intensification} \qquad (\mu_A)(x) = \begin{cases} 2\mu_A^2(x) & \text{if } 0.0 \leq \mu_A(x) \leq 0.5 \\ 1 - 2(1-\mu_A(x))^2 & \text{if } 0.5 < \mu_A(x) \leq 0.5 \end{cases} \tag{2.42}$$

$$\text{Normalisation} \qquad (\mu_A)(x) = \frac{\mu_A(x)}{\max(\mu_A(x))} \tag{2.43}$$

Each of these operators has an equivalent meaning in natural language: concentration corresponds to *very*, dilation corresponds to *fairly* or *somewhat*, and intensification corresponds to *indeed*. Either the third or fourth power (A^3 or A^4) is taken as *extremely* depending on the particular application context. The effect of the main hedges are shown in Figures 2.6(a) and (b). Again, these definitions are not the only ones available. The appropriate choice of hedge definition will normally be context/application dependent, but the standard definitions above have been shown experimentally to correspond fairly well to actual usage [50, 92].

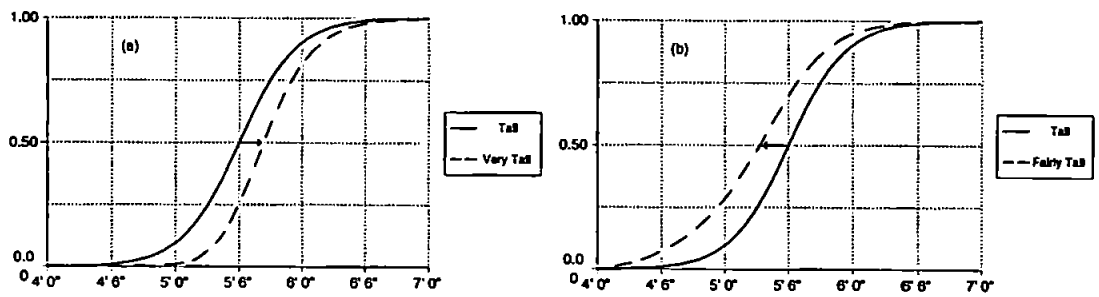


Figure 2.6: *effect of hedge operators (a)* very *and (b)* fairly

2.4.6 Linguistic Fuzzy Logic

In the full implementation of fuzzy logic as envisaged by Zadeh the truth of propositions are represented by a linguistic variable, *truth*. *truth* is a linguistic variable for which *true* and *false* are the primary terms in its term set, rather than a pair of values at the extremes of the unit interval [0, 1]. In multi-valued logic truth values are represented by a single real number in the interval [0, 1], where 0 represents *false*, 1 represents *true* and values between 0 and 1 represent partial truth, whereas in fuzzy logic *true* and *false* are represented by fuzzy subsets over the interval [0, 1] with arbitrary fuzzy subsets representing other intermediate truth values. Thus the linguistic fuzzy logic is actually a superset of multi-valued logic, just as multi-valued logic is a superset of classical two-valued logic. Connectives and hedges are allowed on the primary terms so that the full term set of the *truth* variable might be:

$$\begin{aligned} truth = & true + not\ true + very\ true + fairly\ true + \ldots \\ & + false + not\ false + very\ false + fairly\ false + \ldots \end{aligned} \tag{2.44}$$

The extension principle [140] states that if A is a fuzzy subset over the universe of discourse U, such that $A = \mu_1/\omega_1 + \mu_2/\omega_2 + \cdots + \mu_n/\omega_n$, and that f is a function mapping from U to V, then:

$$\begin{aligned} f(A) &= f(\mu_1/\omega_1 + \mu_2/\omega_2 + \cdots + \mu_n) \\ &= \mu_1/f(\omega_1) + \mu_2/f(\omega_2) + \cdots + \mu_n/f(\omega_n) \end{aligned} \tag{2.45}$$

This extension principle can be used to extend the meaning of *not*, *and*, *or* and *implies* to linguistic truth values. Specifically, if $\nu(A) = \mu_1/\nu_1 + \mu_2/\nu_2 + \cdots + \mu_n/\nu_n$, is a fuzzy subset in [0, 1] representing the linguistic truth value of the fuzzy proposition "x is A" (or simply A), then by applying the extension principle the linguistic truth value of *not A* is given by:

$$\nu(not\ A) = \mu_1/(1-\nu_1) + \mu_2/(1-\nu_2) + \cdots + \mu_n/(1-\nu_n) \tag{2.46}$$

In particular, if the linguistic truth value of A is *true*, that is $\nu(A) = true$, then the linguistic truth value of *false* may be defined as $false \equiv \nu(not\ A)$, so that:

$$false \equiv \nu(not\ A) = \mu_1/(1-\nu_1) + \mu_2/(1-\nu_2) + \cdots + \mu_n/(1-\nu_n) \tag{2.47}$$

Note that this result differs from the fuzzy complement operator, as defined above, so that the complement of A (A^c), which represents *not true*, is:

$$not\ true = A^c = (1-\mu_1)/\nu_1 + (1-\mu_2)/\nu_2 + \cdots + (1-\mu_n)/\nu_n \tag{2.48}$$

The difference between the two results can be seen more clearly in Figure 2.7:

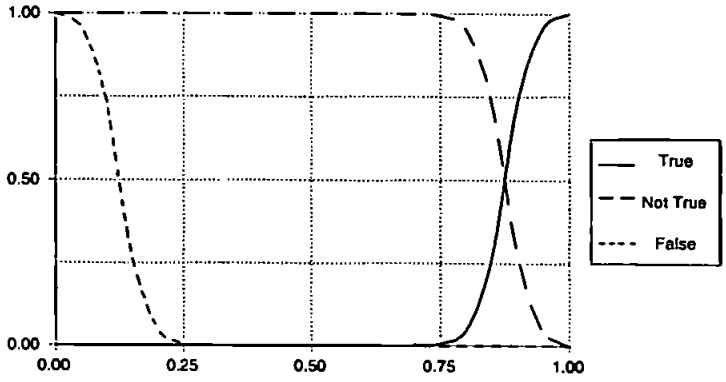


Figure 2.7: *fuzzy truth-values of* true, false *and* not true

Similarly, if A and B are two fuzzy propositions, with linguistic truth values $\nu(A) = \alpha_1/\nu_1 + \alpha_2/\nu_2 + \cdots + \alpha_n/\nu_n$ and $\nu(B) = \beta_1/\eta_1 + \beta_2/\eta_2 + \cdots + \beta_n/\eta_n$ respectively, then the extension principle can again be used to define meanings for the linguistic truth values of $\nu(A\ and\ B)$, $\nu(A\ or\ B)$ and $\nu(A\ implies\ B)$, so that:

$$A \wedge B = \nu(A\ and\ B) = \sum_{i,j} \min(\alpha_i, \beta_j)/\min(\nu_i, \eta_j) \tag{2.49}$$

$$A \vee B = \nu(A\ or\ B) = \sum_{i,j} \min(\alpha_i, \beta_j)/\max(\nu_i, \eta_j) \tag{2.50}$$

$$A \Rightarrow B = \nu(A\ implies\ B) = \sum_{i,j} \min(\alpha_i, \beta_j)/\max[(1-\nu_i), \min(\nu_i, \eta_j)] \tag{2.51}$$

Note again that there is a difference between the linguistic truth values $A \wedge B$ and $A \vee B$ and the fuzzy set operators of intersection $A \cap B$ and union $A \cup B$ as demonstrated in Figures 2.8 and 2.9. This distinction is only necessary when there is a need to represent qualified sentences such as " 'x is A' is τ", where τ is a linguistic truth value such as *true*, *not false*, *fairly true*, etc. For example, the implication operator above allows the representation of " '$A \Rightarrow B$' is *fairly true*" or " 'If x is A Then y is B' is *fairly true*". There is still controversy as to whether such truth qualification is necessary, and the theory has been criticised both philosophically and mathematically.

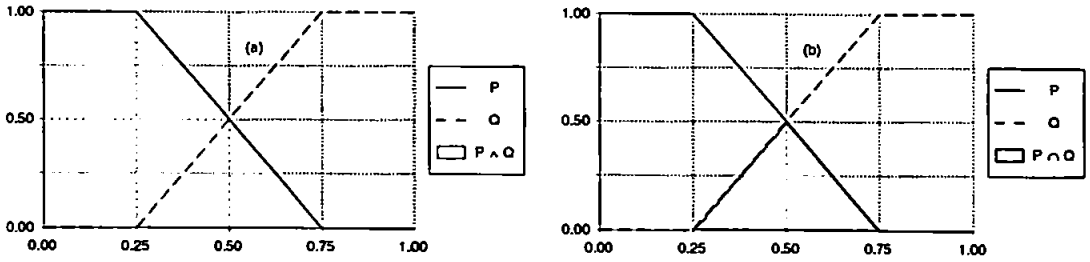


Figure 2.8: *illustration of the difference between fuzzy set operations and linguistic truth-value modification of two truth-values P and Q, for (a)* conjunction *and (b)* intersection

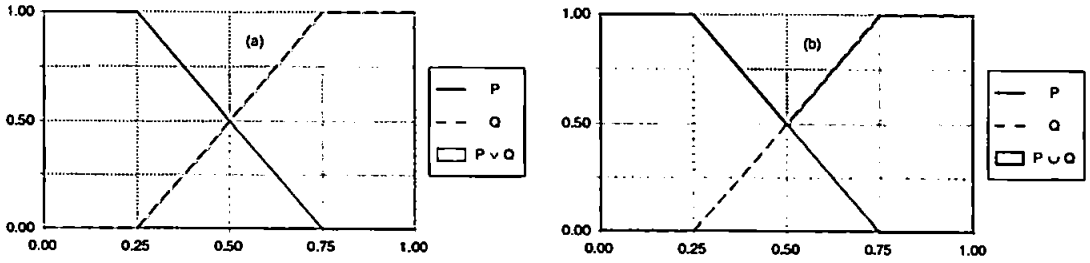


Figure 2.9: *illustration of the difference between fuzzy set operations and linguistic truth-value modification of two truth-values P and Q, for (a)* disjunction *and (b)* union

2.4.7 Generalised Modus Ponens

In his original paper, Zadeh proposed an alternative method of calculating the result of fuzzy implication without resorting to the truth-value domain. It has since been shown [122] that, given some basic assumptions for how the value of τ in 'A is τ' is calculated given an actual instance x of A, then the two methods are equivalent. Consequently, most current fuzzy expert systems do not implement linguistic truth values, but instead restrict propositions to the simpler form of 'If x is A Then y is B' where A and B are plain fuzzy sets.

With this restriction the fuzzy set operators reduce to the simpler form defined over two-dimensional Cartesian space:

$$A \wedge B = \sum_{i,j} \min(\alpha_i, \beta_j)/(\nu_i, \eta_j) \tag{2.52}$$

$$A \vee B = \sum_{i,j} \max(\alpha_i, \beta_j)/(\nu_i, \eta_j) \tag{2.53}$$

$$A \Rightarrow B = \sum_{i,j} \max[(1-\alpha_i), \min(\alpha_i, \beta_j)]/(\nu_i, \eta_j) \tag{2.54}$$

As an alternative approach, if u and v are elements of the universes of discourse U and V, a binary fuzzy relation R is defined as:

$$R = \sum_{i,j} \mu_R(u_i, v_j)/(u_i, v_j) \tag{2.55}$$

The composition (denoted $\circ$) of two binary relations R, from U to V, and S, from V to W, is then:

$$R \circ S = \sum_{i,j} \vee_k [\mu_R(u_i, v_k) \wedge \mu_S(v_k, w_j)]/(u_i, w_j) \tag{2.56}$$

If the fuzzy implication operator $(A \Rightarrow B)$ is represented as a binary fuzzy relation and the fuzzy proposition A^* is known, the *Generalised Modus Ponens* can now be written as:

$$A^*$$

$$A \Rightarrow B$$

$$B^* = A^* \circ (A \Rightarrow B) \tag{2.57}$$

to derive the fuzzy conclusion B^*, where Zadeh defined the implication operator as:

$$A \Rightarrow B = \sum_{i,j} [(1 - \alpha_i) \vee (\alpha_i \wedge \beta_j)]/(\nu_i, \eta_j) \tag{2.58}$$

2.4.8 Alternative Implication Operators

If it assumed that the membership grade at a particular time in A is a and the membership grade in B is b, then the notation can be simplified so that the result of implication for this pair is given in the same form as multi-valued logic as:

$$(1 - a) \vee (a \wedge b) \tag{2.59}$$

However, as for the union and intersection operators, there are many possible alternative fuzzy implication operators. The operator above is termed the Zadeh operator, but there are many other forms including the original Lukasiewicz implication, Bochvar, Heyting and so on. At least eleven different types of implication operator have been investigated, with no

conclusive answer as to which is best. As Zadeh himself pointed out [142], his implication operator does not obey the classical property of *Modus Ponens* that $(p, p \Rightarrow q) \rightarrow q$, so that in fuzzy form $(A, A \Rightarrow B) \rightarrow B^*$. So, even when the known fuzzy fact is identical to the antecedent, the implication leads to the conclusion of a fuzzy set that differs from the consequence set!

2.4.9 Special Fuzzy Sets

Fuzzy set definitions have been proposed [141] for the various different concepts of uncertainty *undefined*, *indeterminate* and *unknown*. The membership grades for these are 0.0, 0.5 and 1.0 (respectively) across the universe of discourse as shown in Figure 2.10. These sets, in which the membership grade is fixed, or level, across the universe of discourse are termed *level sets*.

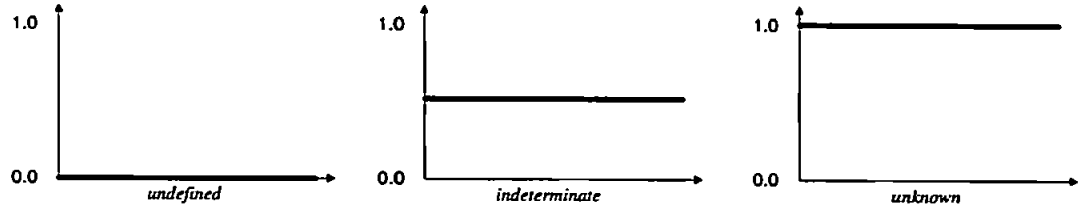


Figure 2.10: *fuzzy set representations of* undefined, indeterminate *and* unknown

A further special fuzzy set is the representation of a crisp value — a spike of (theoretically) zero width, with a membership of 1.0 at the crisp value within the universe of discourse, and zero membership elsewhere. This is termed a *fuzzy singleton*, and is illustrated in Figure 2.11.

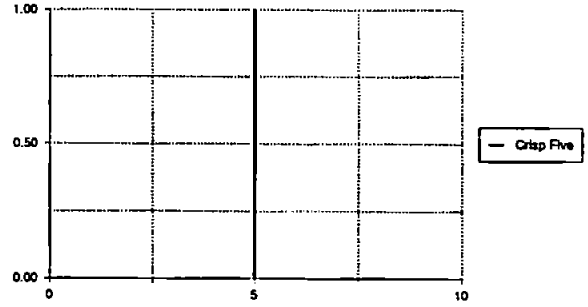


Figure 2.11: *a fuzzy singleton representation of the crisp value* five

2.4.10 Defuzzification

Once the fuzzy reasoning has been completed it is usually necessary to present the output of the reasoning in a human understandable form, through a process termed *defuzzification*. There are two principle classes of defuzzification, *arithmetic defuzzification* and *linguistic approximation*. In arithmetic defuzzification a mathematical method is used to extract the single value in the universe of discourse that 'best' (in some sense) represents the arbitrarily complex consequent fuzzy set (Figure 2.12). This approach is typically used in areas of control engineering where some crisp result must be obtained. In linguistic approximation the primary terms in the consequent variable's term set are compared against the actual output set in a variety of combinations until the 'best' representation is obtained in natural language. This approach is typically used in expert system advisory applications where human users view the output.

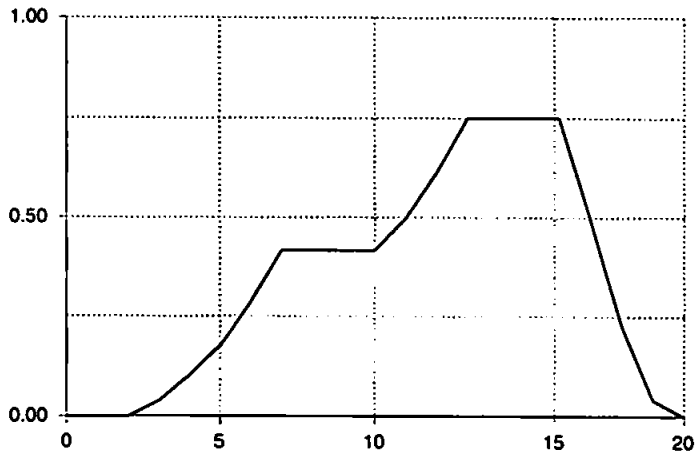


Figure 2.12: *a typical arbitrarily complex fuzzy output set*

Arithmetic Defuzzification

The two most popular methods of arithmetic defuzzification are the *centre-of-gravity* (*centroid*) algorithm and the *mean-of-maxima* algorithm. For the consequent set $A = \mu_1/\omega_1 + \mu_2/\omega_2 + \cdots + \mu_N/\omega_N$, the centre of gravity algorithm provides a single value by calculating

the imaginary balance point of the shape of the membership:

$$x_g = \frac{\sum_{i=1}^{N}(\mu_i \cdot \omega_i)}{\sum_{i=1}^{N}\mu_i} \tag{2.60}$$

The mean-of-maxima algorithm finds the point in the universe of discourse with maximum membership grade:

$$x_m = \max_i \mu_i \tag{2.61}$$

and calculates the mean of all the maxima if more than one maximum is found. An illustration of the output of these two methods is shown in Figure 2.13. Unfortunately, both these methods have problems. Firstly, they both obviously lose information by trying to represent a complex fuzzy shape as a single scalar number. The centre-of-gravity method is insensitive to the overall height of the fuzzy consequent set, and the mean-of-maxima is prone to discontinuities in output, as only a small change in shape (for instance if there are two similar sized peaks) can cause a sudden large change in output value.

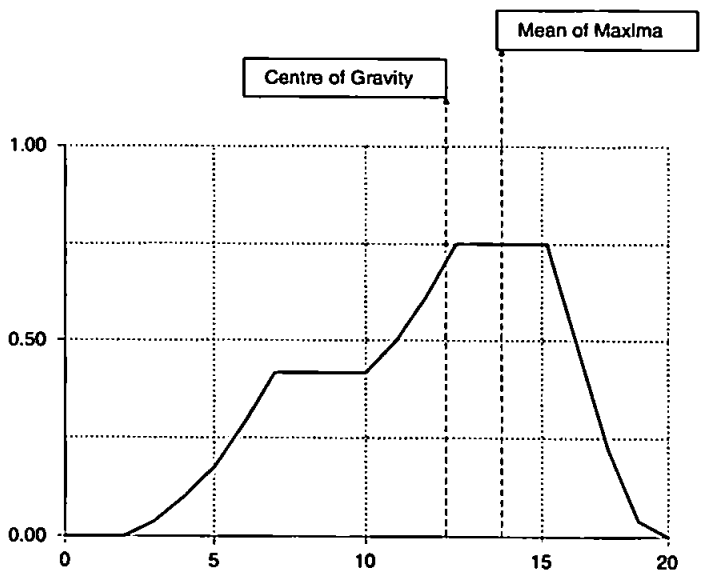


Figure 2.13: *an illustration of the difference between* centre-of-gravity *and* mean-of-maxima *defuzzification*

A number of alternative parameters can also be calculated to provide more information on the shape of the output as well as its location. A variety of such parameters are described, and illustrated through the example fuzzy output sets, A, B, C, D, E, and F, as shown in Figure 2.14.

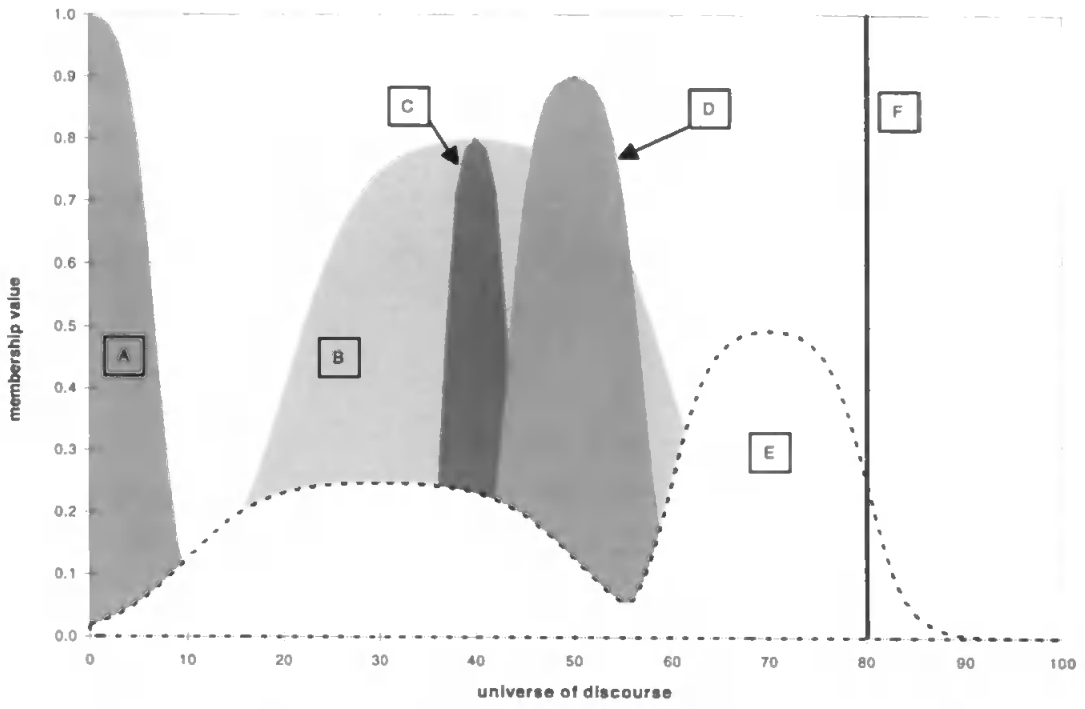


Figure 2.14: *illustration of a variety of fuzzy output sets, A, B, C, D, E, and F, with different defuzzification parameters*

Membership Grade at the Defuzzification Point

The membership grade of the output set at the centre-of-gravity, μ_g, or mean-of-maxima, μ_m, provides an indication of confidence in the result. In Figure 2.14 the output fuzzy sets D and E have the same centre-of-gravity ($x_g = 50$), but set D has a higher confidence in the result ($\mu_g = 0.90$) than set E ($\mu_g \approx 0.16$).

Maximum Membership Grade

The maximum membership grade attained by the consequence fuzzy set, μ_{max}, provides a direct measure of the maximum strength that the antecedents fired a rule. It especially useful for edge sets such as set A in Figure 2.14, as the centroid cannot be at the maximum point, so that $\mu_g < \mu_{max}$.

Normalised Area

The area of the output set normalised to its maximum value is given by:

$$area = \frac{\sum_{i=1}^{N} \mu_i}{N} \tag{2.62}$$

This gives a value of 1.0 for the *unknown* set ($\mu = 1$ across the universe), a value of 0.0 for the *undefined* set ($\mu = 0$ across the universe), and would give a minimal value (≈ 0) for a fuzzy singleton. In Figure 2.14 the output fuzzy sets B and C have the same centre-of-gravity ($x_g = 40$) and the same membership at this point ($\mu_g = 0.80$), but set B has a larger area and hence a larger uncertainty.

Fuzzy Entropy

Yet another measure, termed the *entropy* of a fuzzy set is defined by:

$$S = \frac{\sum_{i=1}^{N} \left(-\mu_i \log_2(\mu_i) - (1-\mu_i) \log_2(1-\mu_i) \right)}{N} \tag{2.63}$$

This is normalised to its maximum value to give a value between zero and one, and provides an indication of the lack of information contained in the output set in terms of the distance away from the extremes of $\mu = 0.0$ and $\mu = 1.0$. It therefore gives a value of 0.0 for the *unknown* and *undefined* sets, and gives a value of 1.0 for the *indeterminate* set ($\mu = 0.5$ across the universe). Similarly to the normalised area, it too gives a minimal value for fuzzy singletons.

Summary of Arithmetic Defuzzification

Table 2.8 summarises the various arithmetic defuzzification parameters defined above for sets A, B, C, D, E, and F, in Figure 2.14, and for the level sets *unknown*, *indeterminate* and *undefined*. It can be seen that for the fuzzy singleton F, which represents a crisp output from a fuzzy system, the values for μ_g and μ_{max} are both high (1.00) and the values for *area* and *entropy* are both low (0.00). The same tendencies can be seen for the example sets A, C, and D (only). There are a whole range of possibilities for other measures, such as the span of the

fuzzy set (percentage of the set that is non-zero), the percentage of total area within a certain distance of the centroid point, the number of peaks in the fuzzy set, and so on.

Set	x_g	μ_g	μmax	*area*	*entropy*
A	3	0.95	1.00	0.07	0.06
B	40	0.80	0.80	0.32	0.50
C	40	0.80	0.80	0.05	0.08
D	50	0.90	0.90	0.12	0.15
E	50	0.16	0.50	0.19	0.60
F (fuzzy singleton)	80	1.00	1.00	0.00	0.00
unknown (1.0/x)	50	1.00	1.00	1.00	0.00
indeterminate (0.5/x)	50	0.50	0.50	0.50	1.00
undefined (0.0/x)	50	0.00	0.00	0.00	0.00

Table 2.8: *summary of various arithmetic defuzzification parameters for sets A, B, C, D, E, and F, in Figure 2.14, with values obtained for level sets* unknown, indeterminate *and* undefined

Linguistic Approximation

In linguistic approximation a similarity measure is used to compute the distance between the actual output set and an arbitrary collection of primary terms, connectives and hedges. For example, a shower control variable with primitive terms such as *cold*, *medium* and *hot*, and allowable hedges of *fairly* and *very*, might produce a composite linguistic term such as *medium and fairly hot*. One such similarity metric is the Euclidean distance between fuzzy sets, given by:

$$\delta = \sum_i (\mu_i - \eta_i)^2 \tag{2.64}$$

where μ_i is the membership of the output set and η_i is the membership grade of the currently considered linguistic approximation — the minimum value of δ will determine the best match. Alternatively, the degree of overlap, γ, of two fuzzy sets, A and B, can be calculated by dividing the area of intersection by the area of the union of the sets:

$$\gamma = \frac{A \cap B}{A \cup B} \tag{2.65}$$

to give a value between zero for disparate sets and one for coincidental sets — the maximum value of γ will determine the best match.

A search is then begun in which the best match is sought whilst attempting to limit the complexity of the combination of terms, in order to produce comprehensible output. For example, although the linguistic combination *not extremely cold and fairly medium and medium and fairly hot* might produce a better match than *medium and fairly hot* for Figure 2.13, the latter term would be preferred due to its relative simplicity.

2.5 Summary of Uncertainty Handling Techniques

Certainty factors have waned in popularity after an initial interest generated by the success of MYCIN, probably as a result of the theoretical difficulties mentioned in Section 2.3.4. It is widely recognised that the Dempster-Shafer theory of evidence, although theoretically sound, is too cumbersome to implement for practical expert systems due to the problem of the vast number of calculations required for real problems. This leaves probability theory and fuzzy theory as the main contenders for practical uncertainty handling.

Debate on the pros and cons of these two techniques for uncertainty handling has often been emotive and personalised. Advocates of probability theory and conventional statistical descriptions have criticised fuzzy theory for being *un-mathematical* and essentially *unnecessary* [51, 70, 132]. Such advocates argue that probability theory can be utilised to handle any form of uncertainty in a rigorous mathematical framework, and reject the use of fuzzy theory as a sloppy convenience. Supporters of fuzzy logic argue that it encompasses probability theory, and allows for more natural representations of different forms of uncertainty [32, 64, 67].

It is this author's belief that there is merit in both sides of this argument. Often proponents of fuzzy logic seem unfamiliar with, or even unaware of, the *subjective interpretation* of probability theory, and make the mistake of assuming that probability theory requires the *frequency* or *objective interpretation*. Whilst this is *not* true, so that subjective probabilities in conjunction with second order probabilities can almost certainly handle all forms of uncertainty, it is almost certainly true, too, that rigorous implementation of probability theory

(possibly in the form of Bayesian reasoning) requires a significant effort on behalf of the expert system engineer. It is questionable as to whether the human expert reasons according to the strict mathematical rules of formal probability theory. Therefore, even if the correctness of fuzzy logic is questionable, the important point is that this might not matter. If the goal is to create an expert system that can simulate the human expert, then mathematical rigour may be deliberately relinquished in exchange for a comprehensible, attainable, working system.

Fuzzy theory does provide a convenient framework for representing uncertainty, both in data and knowledge, in a manner which can be appreciated by the non-mathematical domain expert. For this reason, fuzzy theory was chosen as the best overall technique for handling the uncertainty in this medical domain.

Chapter 3

The Formulation and Design of a Crisp Acid-Base Expert System

3.1 Introduction

Assessment of the acid-base status of umbilical cord blood has recently been recommended by the Royal College of Obstetricians and Gynaecologists [104]. Apart from the problem of fully understanding the fetal acid-base physiology, there are a number of difficulties with the procedure. The cord arteries are very small in comparison to the vein, which can lead to difficulties in obtaining an arterial sample of adequate volume. Due to the narrow diameter of the artery, it is also possible to stick the needle right through the arterial wall and accidentally sample the vein. Two samples, supposedly from each of the artery and vein, can thus actually be from the same vessel, which is usually the vein due to its much larger size. Once the samples are taken it is possible for the pO_2 and pCO_2 values to alter through exposure to air. Blood gas analysis machines require regular internal calibration and external quality control checks to ensure continuing accuracy and precision to the manufacturer's specifications, and failure to perform this routine maintenance can lead to erroneous results.

Umbilical cord acid-base analysis has provoked much debate. Many people have proposed a correlation between the acid-base status at delivery and other measures of neonatal condition,

including long-term neurological development [40, 42, 79], and others have disputed this [29, 105]. These differing opinions are possibly caused by the failure to recognise sampling or measurement errors, the failure to distinguish arterial samples from venous samples or to distinguish metabolic acidosis from respiratory acidosis [54, 48], and a lack of consensus as to what constitutes significant cut-off points for low pH or high BD_{ecf} [6, 133]. Unfortunately, few of these studies have addressed the problem of the quality of the basic data.

Selective umbilical cord blood sampling was initiated during a randomised control trial for ST waveform analysis in Plymouth in 1991 [130]. Careful retrospective analysis of the acid-base results highlighted a 25% failure rate to obtain arterial and venous paired samples with all parameters [129]. This sampling error rate is broadly in line with other studies in which the importance of paired samples was recognised, and this is despite the fact that the sampling took place within a research study which featured regular staff training sessions. The study also highlighted the facts that clinical staff were not very good at identifying measurement errors (for example asterisks alongside a parameter to indicate it was unreliable were frequently ignored), did not recognise the occurrence of two samples from the same vessel and were poor at interpreting the results.

Many medical expert systems have been developed and described in the last thirty years [112, 19], but the number reaching routine clinical use have been few [86, 4]. Some of the main reasons for the lack of clinical implementation include:

1. failing to address the need of the users,
2. appearing to remove the decision making process from the clinician, and
3. failing to address concerns about adequate validation.

A computer program for the diagnosis of complex acid-base disorders [108] and an expert system for the interpretation of blood gas analysis in the intensive care unit [145] have previously been described, but both these systems deal only with adult acid-base status. The fetus and neonate has substantially different blood characteristics from the adult, and therefore umbilical cord acid-base analysis requires specific knowledge of fetal physiology [123].

This chapter describes the formulation of the model of expertise required for the validation and accurate interpretation of the acid-base status of blood taken from the umbilical cord of the neonate immediately after delivery, and the design of an expert system to encapsulate this knowledge. The knowledge elicitation and development of the expert system rules are described in detail, and the stages of feasibility study, software design, and development of the user-interface are also described.

3.2 Expert System Design

3.2.1 Expert System Specification

In 1990 – 1991 a randomised control trial took place in Plymouth in which intrapartum monitoring was compared between patients monitored with cardiotocography (CTG) alone and patients monitored with CTG in conjunction with analysis of the ST segment of the electrocardiogram (ECG) waveform [130]. Umbilical cord blood acid-base assessment was introduced during the initial phase of this trial, and a database of approximately 2 100 acid-base results were collected from the 2 400 deliveries entered into the trial. The acid-base data were to be used to evaluate the difference in outcome between the two arms of the trial, but it was quickly realised that despite the strict trial protocols there were still numerous errors in the data. The careful retrospective review of all acid-base results that had to be carried out before the data could be used provided the original inspiration for the acid-base expert system.

The expert system has two main purposes;

- to validate the results, and
- to interpret the results.

One or two syringes of blood are sampled from the umbilical cord. The umbilical vein is large and almost invariably is filled with sufficient blood to enable a 2 ml blood sample to be

obtained. The arteries are much smaller and often only around 0.2 ml of blood is obtained. In around 2–5% of cases, the arterial vessels contain so little blood that no reasonable sample can be obtained. The samples are introduced by the operator into the blood gas analyser, and the results of each sample's analysis are passed to the expert system.

Each sample is initially validated separately, to ensure that its parameters are physiologically plausible within the within the limits of fetal acid-base. The expert system then identifies which blood sample originates from which vessel by examination of the input parameters — the operator does *not* indicate vessel origin to the system. If two samples are obtained, the expert system performs another level of validation to ensure that the results make physiological sense *when viewed as an arterial-venous pair.* Most blood gas analysis machines are intended for adult blood analysis, and therefore calculate the base deficit of the blood compartment (BD_{blood}), but it has been shown that the base deficit of the extra-cellular fluid compartment (BD_{ecf}) is more appropriate for the infant during the perinatal period [102]. The base deficit of the extra-cellular fluid is re-calculated therefore, if possible, from the input parameters according to Equation 1.13. Finally, an expert system interpretation is derived from the results.

3.2.2 Knowledge Elicitation

After a brief period of irregular umbilical acid-base assessment for the first 400 trial patients, a strict protocol for sampling and analysing the blood samples was introduced. The single umbilical vein runs through the middle of the umbilical cord and the two umbilical arteries spiral intertwined around the outside. The blood in the umbilical vein provides support for the two surrounding arteries, and the trial staff were trained to sample the umbilical artery first. Either or both arteries were sampled until a reasonable volume of blood was obtained, or more usually until there was no arterial blood remaining. The umbilical vein was then sampled which, in the exception of a very few cases, resulted in a full 2 ml syringe of blood. The syringes were marked and the samples were placed through the analyser in the same order; arterial sample first, followed by venous sample. Any deviation from this protocol

was clearly marked on the results printed by the analyser and entered onto the trial proforma. Thus the database indicated the vessel origin (artery or vein), or at least the intended or believed vessel origin, for each sample. Members of staff involved with the trial were trained to carry out this protocol, and had the training reinforced with weekly sessions throughout the trial period. The research staff who organised the trial were available at any time for guidance and assistance with any problems that were encountered.

The development of the expert system was an iterative process that took place in close collaboration with several clinicians experienced in the interpretation of perinatal acid-base data. The process was initiated during the retrospective review of the trial database, in which it was realised that, despite the clear and strict trial protocols, there remained numerous errors in the acid-base data. During the retrospective review the experts carefully examined each set of acid-base data in conjunction with knowledge of the events of the labour as captured for the trial. These events included the CTG recordings (with or without ST waveform analysis), any fetal scalp blood sample samples taken, analgesia (pain relief) and other treatments administered to the mother, and any other intrapartum or immediate postpartum events recorded in the mother's or infant's notes. This trial database was then used to formulate and verify the expert system rules through statistical analysis of the data in conjunction with the experts' knowledge of fetal physiology.

3.2.3 Validation Rules

The first task that the expert system undertakes is data validation. This can be divided into two main sections; identification of parameter errors and identification of the vessel origin of each sample. Two classes of parameter errors are detected; *analyser errors* and *physiological errors*. Analyser errors are those reported by the analyser at the time of sampling such as caused by the electrodes failing to reach a stable reading, indicated as double asterisks next to parameters. These are intended to represent unreliable readings, to be treated with extreme caution. However, it was noted that many of these results had been entered into the patient's notes without any regard for the asterisks, and occasionally the extremely high BD_{ecf} that

resulted from an invalid low pCO_2 was remarked on within the notes. The expert system disregards any asterisked parameter and treats it as an error.

Physiological errors are an additional class of error detected specifically by the expert system by examining whether the results are consistent with the range of possibilities for cord blood. There is a strong relationship between the pH and the pCO_2 of blood, represented by the Henderson-Hasselbalch Equation 1.7 and the Siggaard-Andersen Base Deficit Equation 1.13, and this fact was used to formulate two rules for validating the 'physiological plausibility' of results. A graph of pH against pCO_2 for both arterial and venous validated samples was plotted, as shown in Figure 3.1, and the 99.9% confidence of prediction intervals were calculated by regression analysis.

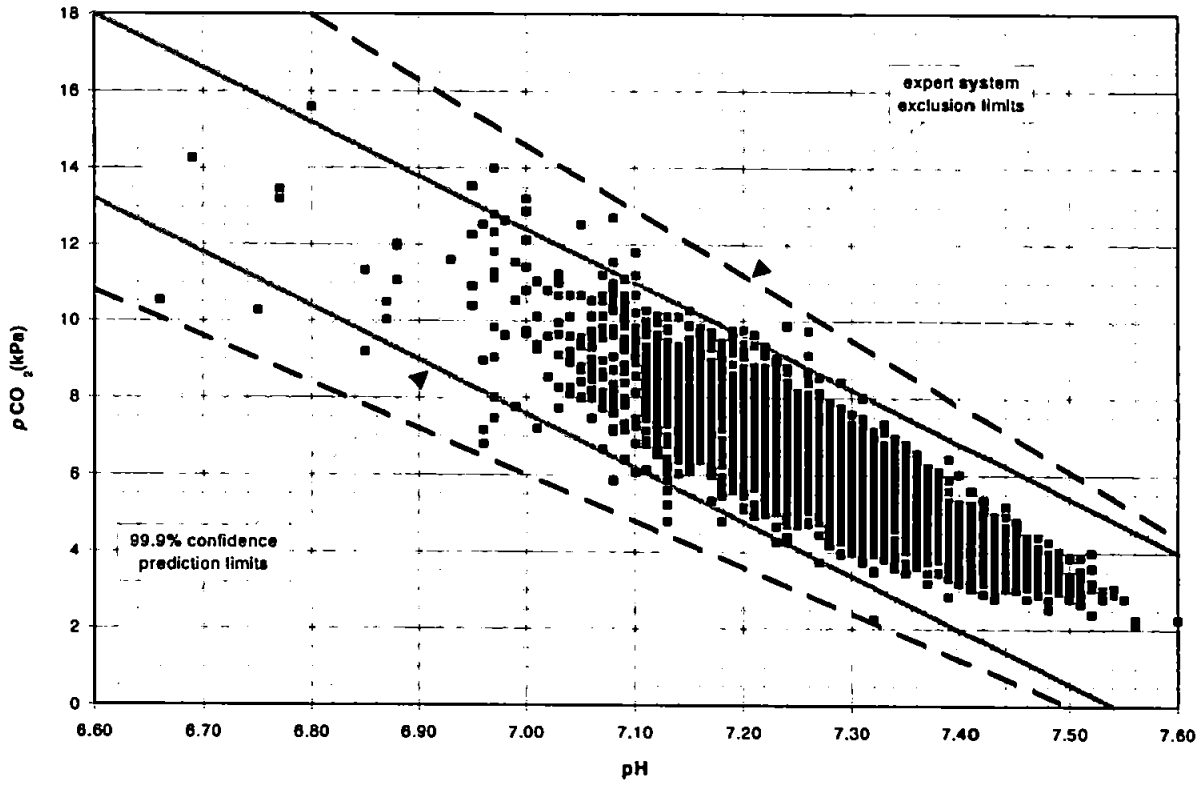


Figure 3.1: *scatter diagram of cord blood pH against pCO_2 with 99.9% prediction limits and expert system exclusion limits*

Analysis of the residuals showed that the variances of the data across the pCO_2 axis are not uniform for all pH, indicating that the linear regression prediction intervals are not strictly valid. Hence, the exclusion limits were constructed beyond these intervals, widening with the

increased variance in pCO_2 as pH decreases. Results that fall outside these exclusion limits — caused, for example, by the presence of non-blood fluid in the sample — are reported as errors. The rules are:

IF $pCO_2 < 90.0 - 12.0\,\text{pH}$

THEN mark the pCO_2 parameter as invalid

IF $pCO_2 > 133.6 - 17.0\,\text{pH}$

THEN mark the pCO_2 parameter as invalid

The next step was to develop rules for the assessment of vessel origin. After the introductory phase, acid-base results were taken from 1942 of 2013 deliveries (96.5%) during the study period, with staff forgetting or being unable to obtain samples in the remaining 71 (3.5%). A single blood sample only was obtained in 54 cases (2.7%) and in another 73 cases (3.6%) an unreliable pH (as indicated by the blood gas analyser) was obtained from one or both vessels. This left 1815 paired results with a reliable pH in each sample. Of these, the blood gas analyser identified 17 unreliable pCO_2 readings (0.9%), leaving 1798 pairs of results with apparently accurate pH and pCO_2 values. Similarly there were 64 unreliable pO_2 readings (3.5%), leaving 1751 pairs of results with apparently accurate pH and pO_2 values. The venous–arterial differences for pH and pO_2 and the arterial–venous difference for pCO_2 for these results were plotted as shown in Figures 3.2 to 3.4. Note that the differences are calculated in such a way that each should be positive physiologically. This is emphasised in the definitions in Table 3.1.

Difference	*Definition*
ΔpH	venous pH − arterial pH
ΔpCO_2	arterial pCO_2 − venous pCO_2
ΔpO_2	venous pO_2 − arterial pO_2
ΔBD_{ecf}	arterial BD_{ecf} − venous BD_{ecf}

Table 3.1: *the definition of Δ for acid-base parameters used throughout this thesis*

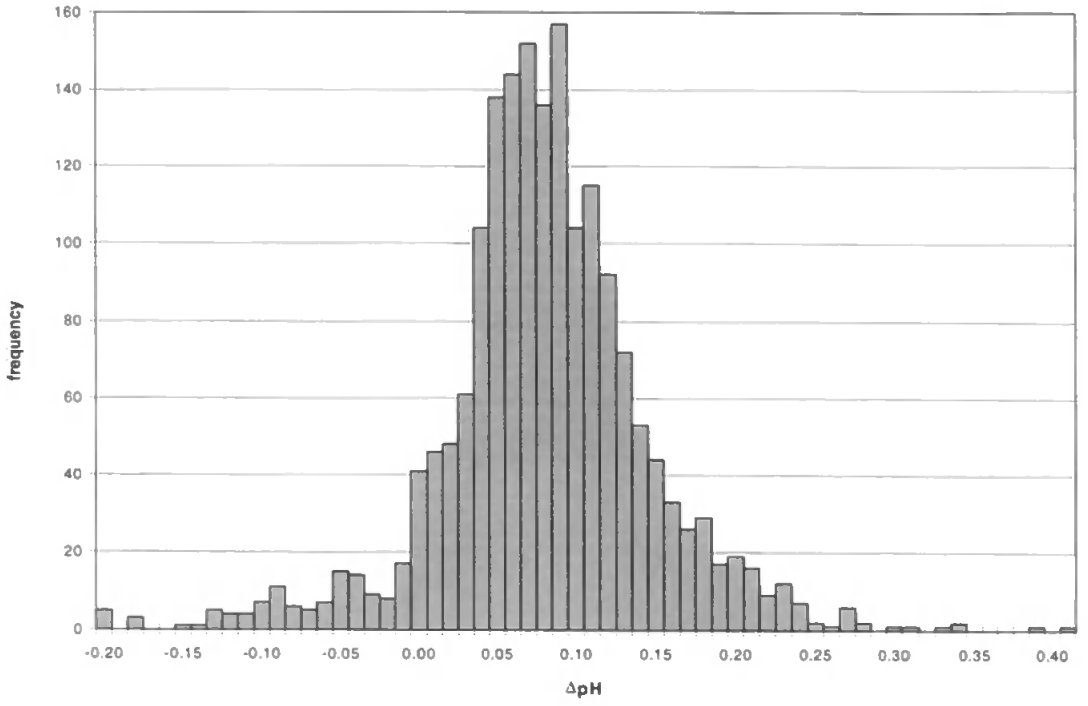


Figure 3.2: *distribution of ΔpH (venous − arterial) for $n = 1815$ cases with two pH's*

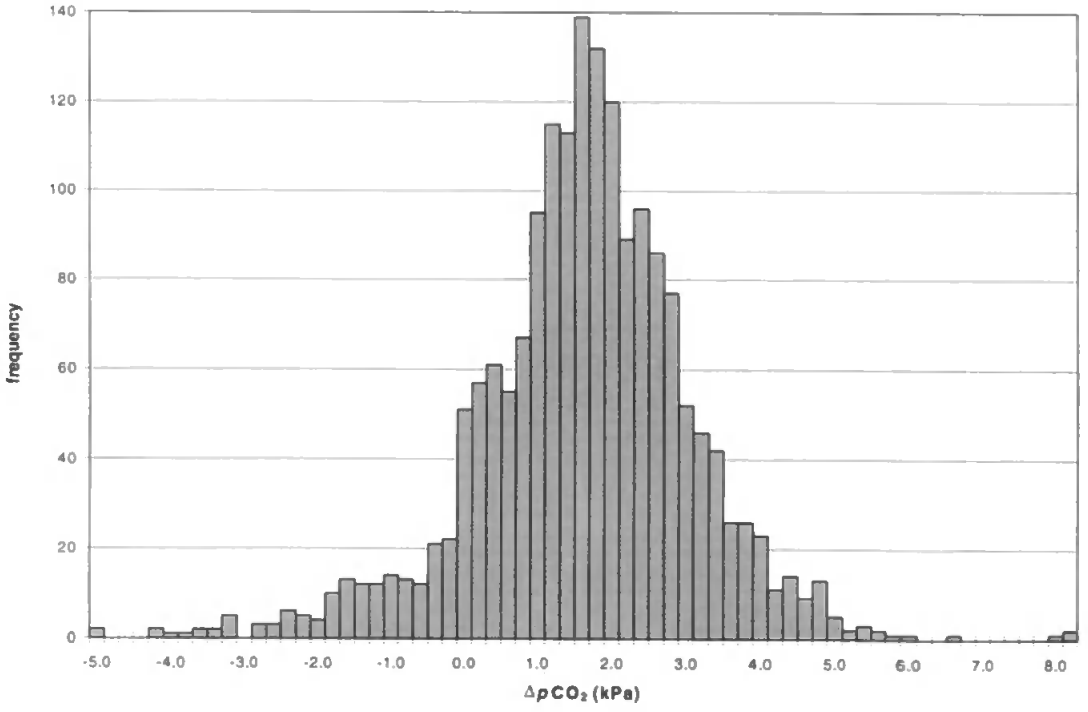


Figure 3.3: *distribution of ΔpCO_2 (arterial − venous) for $n = 1798$ cases with two pH's and two pCO_2's*

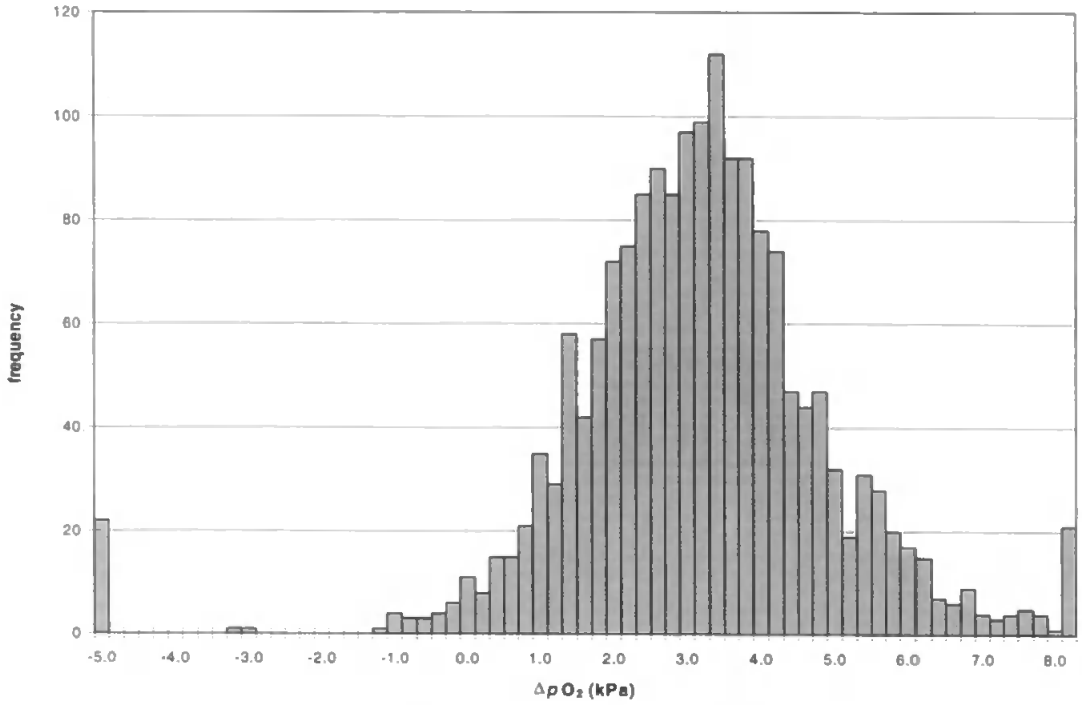


Figure 3.4: *distribution of* ΔpO_2 *(venous − arterial) for* $n = 1751$ *cases with two pH's and two* pO_2*'s*

There were a number of paired samples in which the differences were zero, very small, or opposite to the expected values (i.e. arterial pH/pO_2 higher than venous pH/pO_2, arterial pCO_2 lower than venous pCO_2). These results could not be explained physiologically and the most likely causes were thought to be:

1. inadvertent sampling from the same vessel twice (i.e. artery–artery or vein–vein instead of artery–vein), or
2. transposing the vessels either when taking the samples or on introduction into the analyser (i.e. vein–artery instead of artery–vein).

The former would have caused a *normal distribution* of differences centred at zero (with a width governed by the size of errors of the parameter) and the latter would have caused a mirror-image of the real arterial-venous distribution, reflected about zero. If, as seems probable, both occurred in some small proportion during the study, three distributions would result. An illustration of three possible such distributions is shown in Figure 3.5, consisting of a large positively skewed distribution, a normal distribution of relative proportion of 5%,

and a reflection about zero of the first distribution of relative proportion of 5%. These distributions summated, would form the distribution illustrated in Figure 3.6, which can be seen to approximate the actual pH distribution of Figure 3.2.

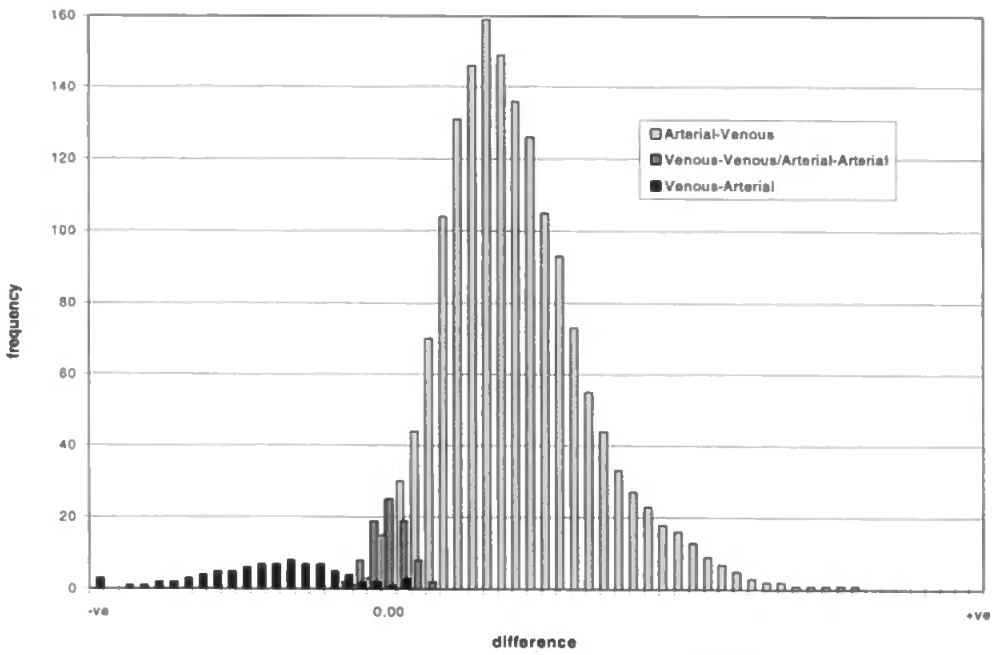


Figure 3.5: *hypothetical distributions of ΔpH (venous – arterial) for arterial – venous samples, mixed samples and swapped samples*

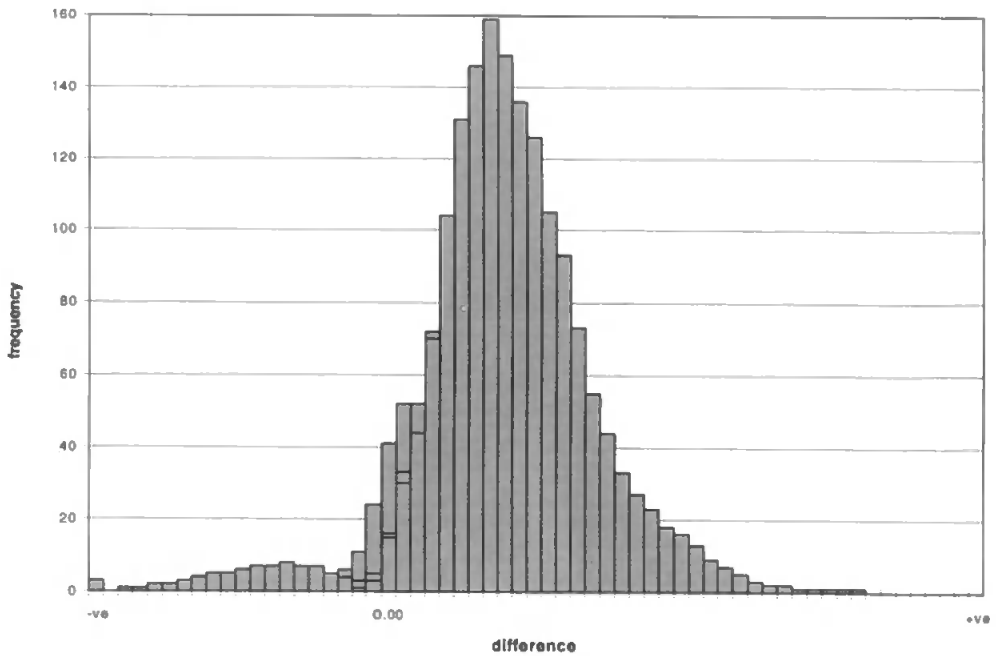


Figure 3.6: *hypothetical distributions of ΔpH combined — compare to Figure 3.2*

After lengthy discussion it was decided that ΔpH alone would be used to identify the vessel origin of the samples. As there is no physiological explanation for a negative venous–arterial ΔpH, the lowest pH is labelled as the arterial sample and the highest pH as the venous sample. The rule:

IF (venous − arterial pH) < 0.025
THEN mark both samples as 'same vessel: probably vein'

is then applied to mark the samples as originating from the same (or mixed) vessels if there is insufficient difference. This cut-off point was chosen for several reasons:

1. it is the point on Figure 3.2 where the distribution appears to deviate from the expected distribution,
2. it excluded 160 (8.8%) of the 1 815 paired results, which was deemed a reasonable proportion, and
3. it is difficult physiologically to justify differences smaller than this.

Paired results undergo a further, more sophisticated, stage of validation to ensure that they make 'physiological sense' when viewed as arterial and venous blood. Given that the pH of the artery must be lower than the vein, physiologically the pCO_2 **and** the BD_{ecf} of the artery should be higher; if this is found to not be the case, an error will be marked against the invalid results. The rules to exclude pCO_2 as invalid are:

IF (venous − arterial pH) < 0.06 AND (arterial − venous pCO_2) < 0.2 kPa)
THEN mark the arterial and venous pCO_2 parameters as invalid

IF (venous − arterial pH) ≥ 0.06 AND (arterial − venous pCO_2) < 0.5 kPa)
THEN mark the arterial and venous pCO_2 parameters as invalid

If the pH and pCO_2 values for a sample are accepted as valid, the base deficit of the extracellular fluid (BD_{ecf}) is calculated from Equation 1.13, and is then validated by the rule:

IF (arterial − venous BD_{ecf}) < -4 mmol.l^{-1}

THEN mark the venous BD_{ecf} parameter as invalid

This rule was designed to exclude cases in which an unusually low venous pCO_2 causes an unphysiologically high BD_{ecf}. An example from the database is shown in Table 3.2.

	pH	pCO_2 *(kPa)*	BD_{ecf} *(mmol.l^{-1})*
Arterial	7.20	8.3	3.1
Venous	7.27	2.9	15.0
Difference	0.07	5.4	-11.9

Table 3.2: *an example of a physiologically consistent pair of pH and pCO_2 results, which together represent an unphysiological pair of BD_{ecf} results*

3.2.4 Interpretation Rules

Once validated, the pH and BD_{ecf} of both vessels are examined to categorise the results into one of 54 interpretations, ranging from 'normal' to 'severe metabolic acidemia'. Results consistent with respiratory acidosis are distinguished from those indicating metabolic acidosis and the differences between the two vessels, if available, are used to further refine the diagnosis. An interpretation is performed on single samples as well as paired samples, although the information is very much more limited and the user is advised to retry with a paired sample.

Frequency distributions of the arterial and venous pH, pCO_2 and BD_{ecf} were plotted to establish the median values, lower 2.5^{th} centile and upper 97.5^{th} centile ranges of each; means and standard deviations cannot be used on the data as all the distributions are skewed and not *normal*. These are shown in Figures 3.7 to 3.12. The populations were checked against other published data to ensure that they were not specific to our data [36, 137, 121]. The results are summarised in Table 3.3.

Vessel	pH	pCO_2 (kPa)	BD_{ecf} (mmol.l^{-1})
Artery	7.26 (7.05 to 7.38)	7.3 (4.9 to 10.7)	2.4 (-2.5 to 9.7)
Vein	7.35 (7.17 to 7.48)	5.3 (3.5 to 7.9)	3.0 (-1.0 to 8.9)

Table 3.3: *the median and 2.5^{th} to 97.5^{th} centile range in parenthesis for umbilical cord arterial and venous pH, pCO_2 and BD_{ecf} ($n = 1448$)*

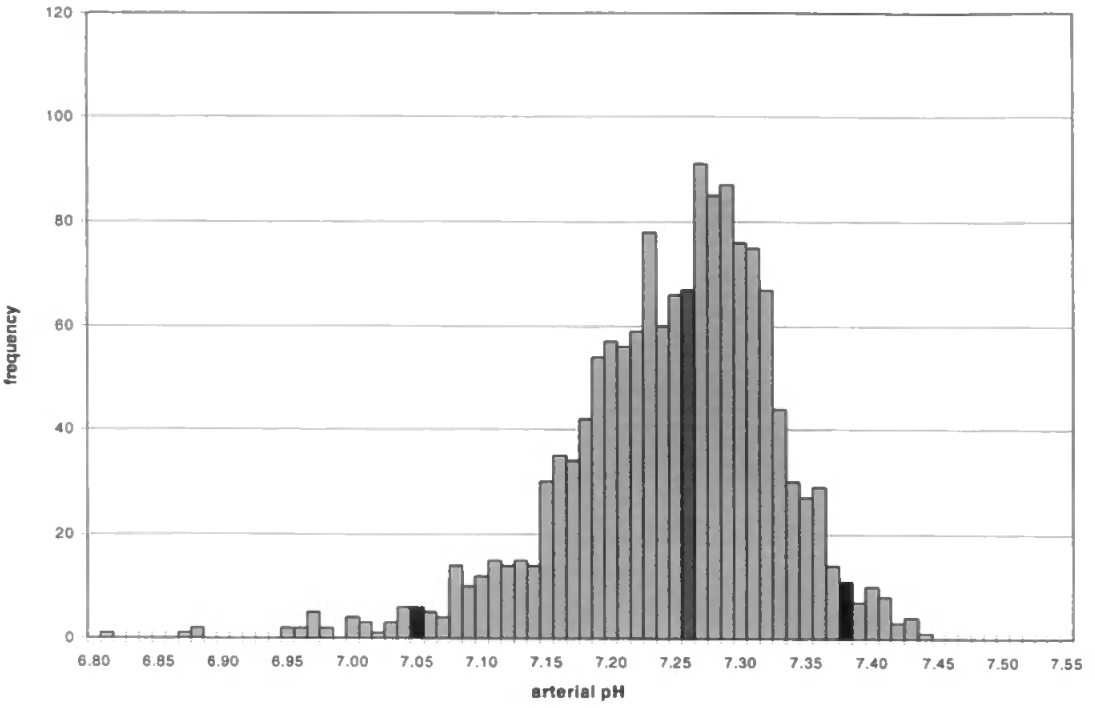


Figure 3.7: *distribution of arterial pH for n = 1448 validated results, showing median, 2.5th centile and 97.5th centile*

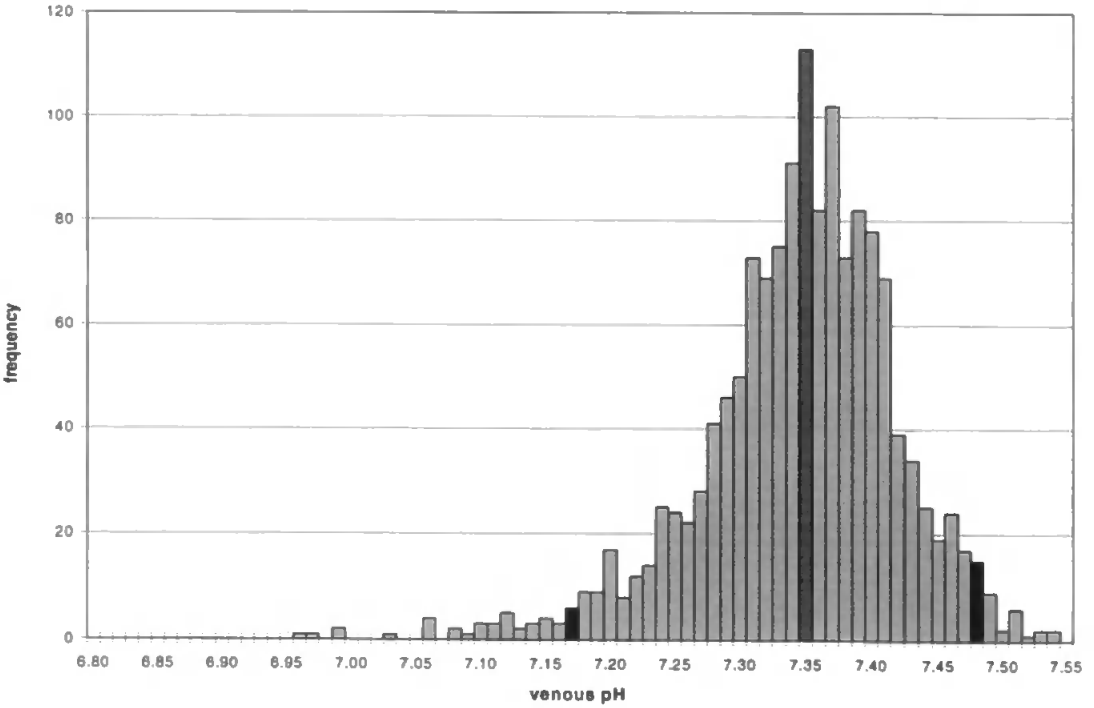


Figure 3.8: *distribution of venous pH for n = 1448 validated results, showing median, 2.5th centile and 97.5th centile*

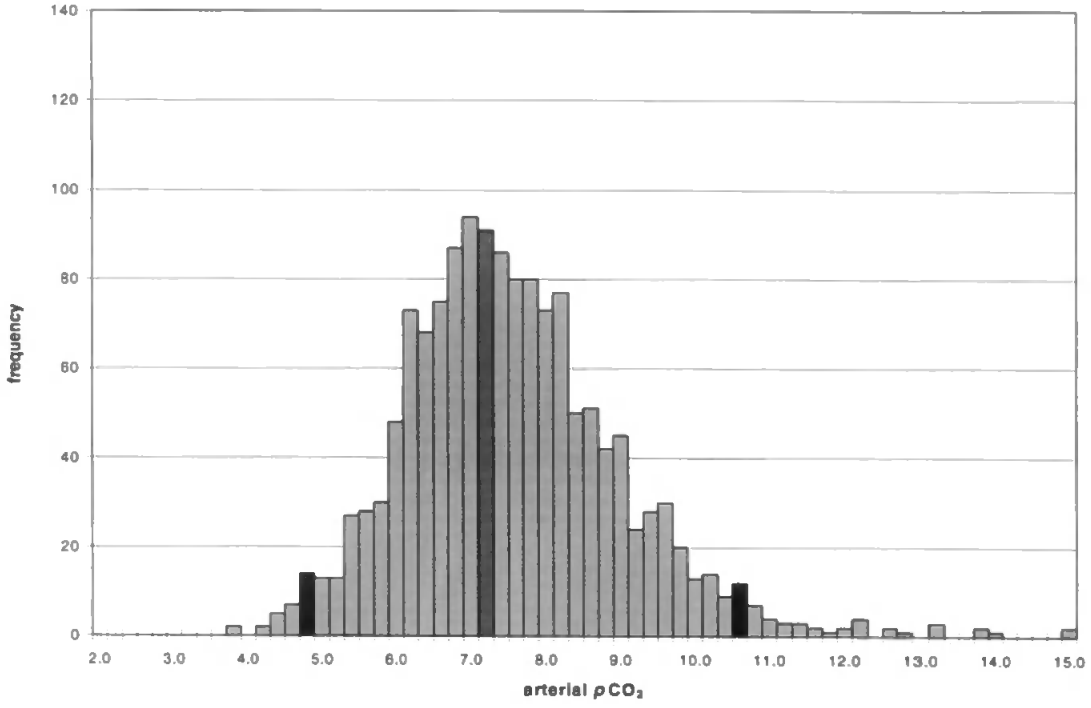


Figure 3.9: *distribution of arterial pCO_2 for n = 1448 validated results, showing median, 2.5th centile and 97.5th centile*

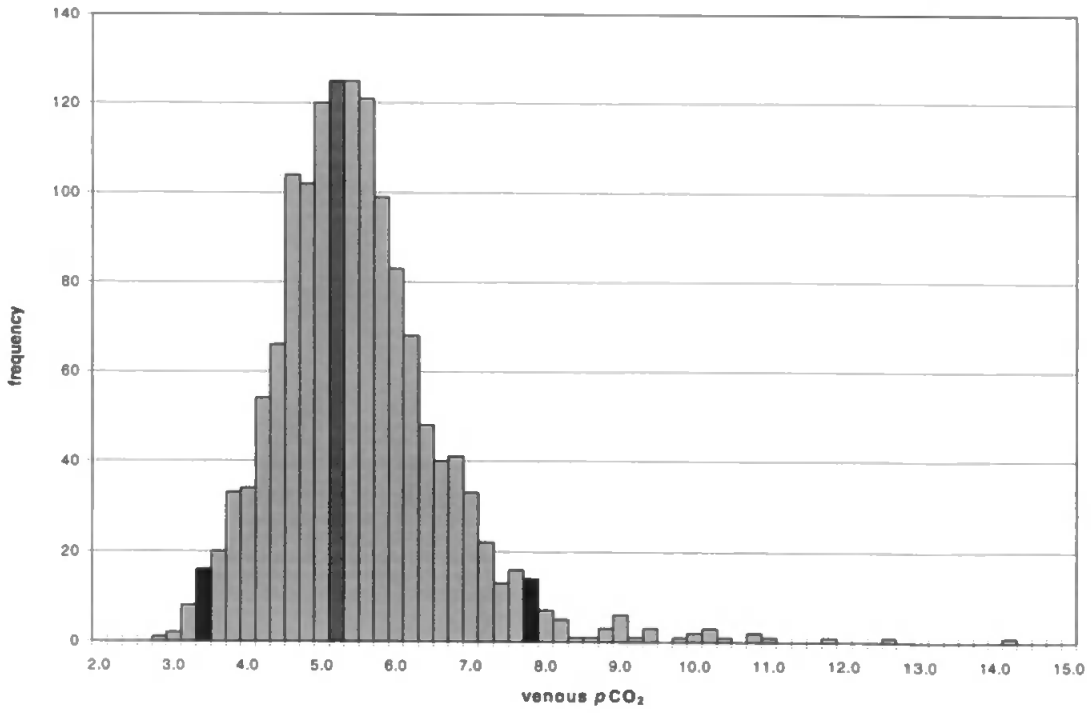


Figure 3.10: *distribution of venous pCO_2 for n = 1448 validated results, showing median, 2.5th centile and 97.5th centile*

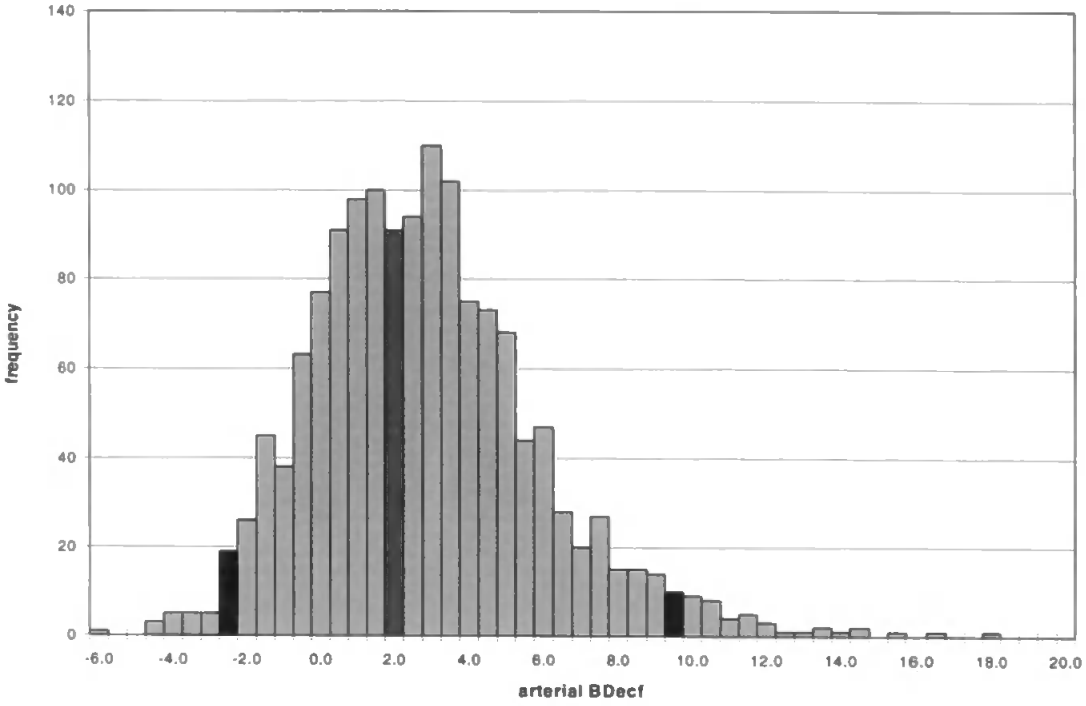


Figure 3.11: *distribution of arterial BD_{ecf} for n = 1448 validated results, showing median, 2.5th centile and 97.5th centile*

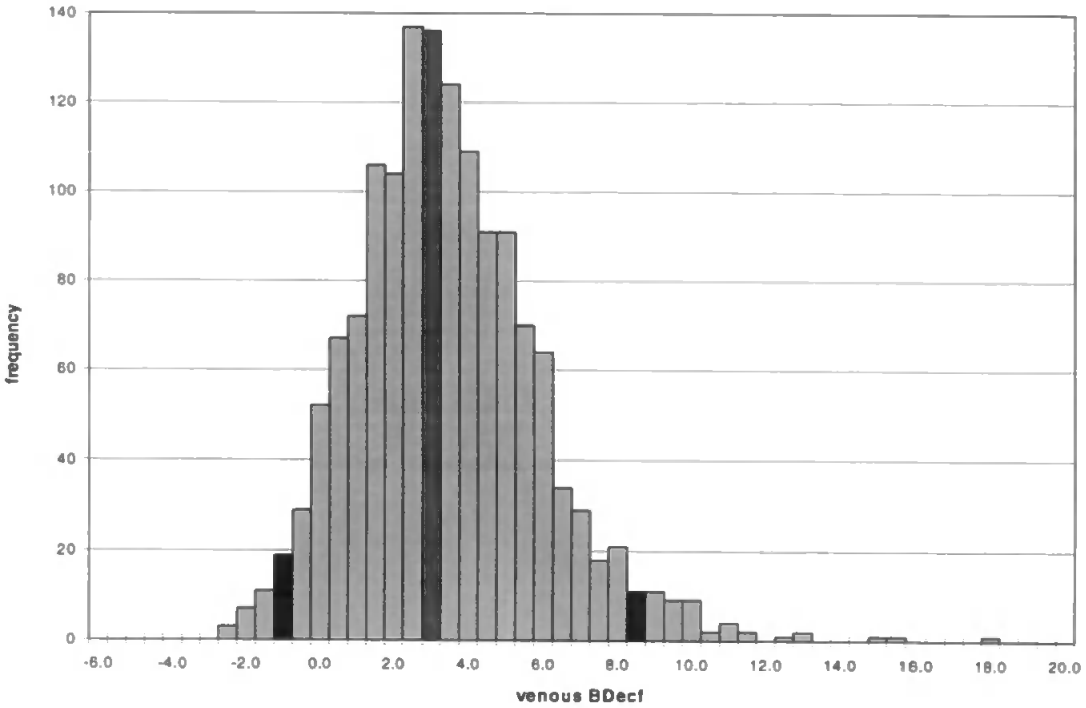


Figure 3.12: *distribution of venous BD_{ecf} for n = 1448 validated results, showing median, 2.5th centile and 97.5th centile*

The experts considered these statistics, other published statistics, published studies on neonatal outcome with acid-base data, published animal studies and their own knowledge of perinatal acid-base balance to formulate the rule cut-offs for the expert system interpretation. These cut-offs were chosen, using the experts' knowledge, with consideration of all four input parameters in conjunction. In particular the cut-offs were not designed to divide the four dimensional input space into statistically 'equal' regions, but rather to concentrate interpretation decision boundaries around the critical regions. The main cut-offs used are:

arterial pH:	7.05, 7.10 and 7.15
venous pH:	7.10, 7.15 and 7.25
arterial BD_{ecf}	8, 10 and 12 $mmol.l^{-1}$
venous BD_{ecf}	8 and 10 $mmol.l^{-1}$

A set of rules was generated by two of the clinicians after a knowledge elicitation session. The interpretation is basically a four-dimensional classification task of arterial pH (pH_A), arterial BD_{ecf} (BD_A), venous pH (pH_V) and venous BD_{ecf} (BD_V). Each rule was assigned an arbitrary node number, termed the *Expert System Number* (*ESN*), to identify the region — a full list of the final *ESN*'s can be found in Appendix D. It was felt that full coverage of the four-dimensional space was an essential attribute of the rules. Thus, a scheme was devised whereby the partitioning of the four-dimensions could be represented visually.

Initially a two-dimensional partition was fixed in pH_A and BD_A, and a two-dimensional partition graph of pH_V against BD_V was drawn by the experts to construct the classifications for this sub-space. As an example, the first partition was for $pH_A < 7.05$ *and* $BD_A \geq 12$ $mmol.l^{-1}$ with the partition graph of pH_V and BD_V shown in Figure 3.13. This process was then repeated by varying the BD_A partition first, for example $BD_A \geq 10$ *and* $BD_A < 12$ $mmol.l^{-1}$, and so on until the entire four-dimensional space was classified. This process effectively broke down the difficult problem of four-dimensional partitioning into multiple sets of two simultaneous two-dimensional partitioning.

The same process was carried out for each of the situations where one or more of the input parameters were missing (through errors). As any single vessel is considered to be a vein

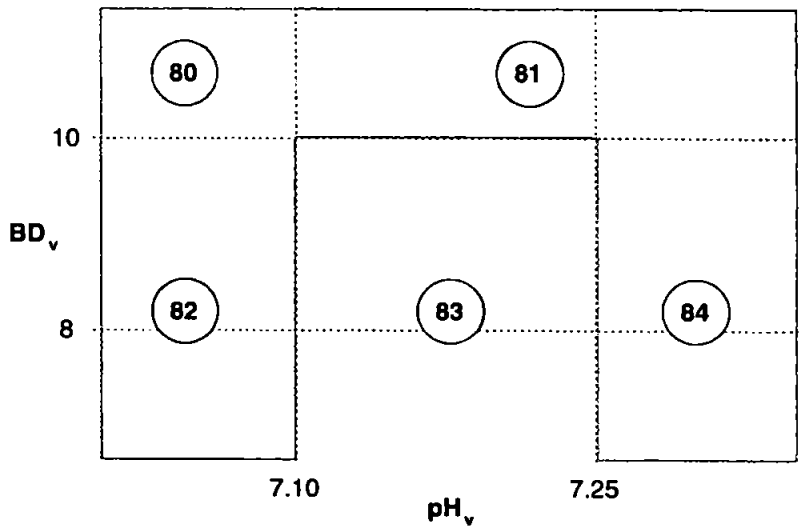


Figure 3.13: *partition graph of pH_V against BD_V for $pH_A < 7.05$ and $BD_A \geq 12$ $mmol.l^{-1}$, showing the* ESN *assigned to each area of input space*

— the vein is far larger than the arteries, and thus is far more likely to be the origin of the blood if only one sample is obtained — and the base deficit can only be obtained if the pH is obtained (as it is derived from the pH and pCO_2), the combinations of parameters available for interpretation are:

- pH_A, BD_A, pH_V, BD_V
- pH_A, BD_A, pH_V
- pH_A, pH_V, BD_V
- pH_A, pH_V
- pH_V, BD_V
- pH_V

The partitioning procedure was carried out for each of these cases separately, to generate a set of rules for the complete classification of all input parameters.

These rules were then encoded and applied to the database in a variety of ways. Firstly, the full results were passed to the expert system and the interpretations recorded. This produced a set of around 2 000 cases where the interpretation was 'normal' and around 100 cases where the interpretation was not normal ('non-normal'). For the non-normal cases, each input parameter was then marked as containing an error in all combinations itemised above

and these results were also passed to the expert system. This generated the interpretations of the expert system with successively less information and enabled the internal consistency of the rules to be checked. For example, Table 3.4 shows the categorisations and interpretations obtained from the expert system when a set of results are passed through the system with successively less information. It can be seen that the interpretation becomes vaguer, but in a consistent manner, as information is removed.

pH_A	BD_A	pH_V	BD_V	*ESN*	*ES Interpretation*
6.95	15.0	7.15	9.1	83	Acute-Moderate acidemia
6.95	15.0	7.15	*	61	Moderate acidemia, venous pH lowish, no venous BD_{ecf}
6.95	*	7.15	9.1	43	Questionable significance, arterial pH low, no arterial BD_{ecf}
6.95	*	7.15	*	31	Questionable significance, arterial pH low, no BD_{ecf}'s
-	-	7.15	9.1	25	Results OK, but arterial acidemia cannot be excluded
-	-	7.15	*	11	Questionable results, pH only, arterial acidemia cannot be excluded

Table 3.4: *an illustration of the interpretations obtained when results are passed through the expert system with successively less information: '*' indicates that the parameter contains an error; '-' that the parameter is not present*

A number of other techniques, such as generation of plausible random numbers, and gradual variation of one or both pH's or BD_{ecf}'s whilst the other parameters remained constant, were used to examine the behaviour of the expert system rules. The output generated was examined by the clinicians and the rules modified to eliminate inconsistencies and refine interpretations. This process continued iteratively until the rules were deemed acceptable. The complete rule set for the crisp expert system is shown in Appendix D.

3.3 Software Design

3.3.1 System Requirements and Specification

A feasibility study was initially carried out in the 'C' language on an IBM compatible computer running DOS, in which the communication protocols between the computer and blood gas analyser were investigated and developed. Once the practicality of an on-line connection was established, a high level specification was drawn up to detail the functioning of the program to the informal description of the clinicians. A modular description of the system is

shown in Figure 3.14. At this stage, the program development was moved to the *Microsoft Windows* environment, as this was seen as being more suited to the non-sequential aspects of the program identified during the feasibility study and more advanced as a user-interface. As the program was intended to be used by clinical staff, it was designed to be as straight forward and simple to use as possible. The content and appearance of each screen was developed in conjunction with the clinician involved with the project and each was prototyped and refined to the clinician's specifications.

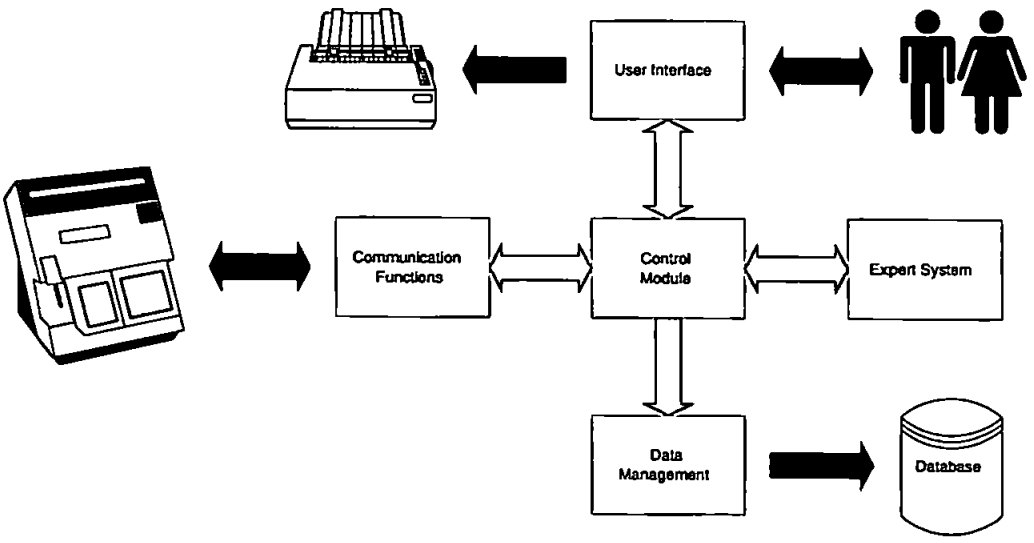


Figure 3.14: *conceptual block diagram of expert system*

3.3.2 Knowledge Representation

Internally, the dynamic knowledge representation of the expert system is conceptually organised as a set of frames, implemented as 'C' structures, with attributes such as pH, pCO_2, BD_{ecf}, and originating vessel (artery or vein) for each sample. The frames were:

- **patient** information
    1. patient (mother) surname
    2. patient (mother) first name
    3. patient (mother) hospital identification number
    4. infant number (for multiple births) — this is a Roman numeral 1...9 assigned by the attending clinicians to identify each infant in a multiple birth according to its actual delivery sequence, as each infant has neither a name or a hospital number until its birth is officially registered several hours (or more) after birth

- **date/time** information
    1. day
    2. month
    3. year
    4. hour
    5. minute
    6. second
- **sample** information
    1. sample number — a unique sample number issued by the blood gas analyser
    2. pH
    3. pCO_2
    4. pO_2
    5. HCO_3^-
    6. BD_{ecf}
    7. validation flags
    8. sample *date/time*
- **calibration** information
    1. old pH
    2. new pH
    3. old pCO_2
    4. new pCO_2
    5. old pO_2
    6. new pO_2
    7. barometric pressure
    8. calibration *date/time*
- **operator** information
    1. operator personal identification number (PIN)
    2. operator name

3.3.3 Inference Engine

To simplify the exchange of information with other modules in the system the expert system inference engine was tailor written, again in the 'C' language. The inference engine features a forward-chaining algorithm, which is suitable for the classification-type rules for both the validation and interpretation. The inference engine itself is relatively simple, occupying only 500 lines of code out of the total for the system of around 7 500 lines. The rules are encoded into 'C' internally to prevent unauthorised modification or reproduction.

3.3.4 Ancillary Functions

The user initially selects the type of sample to be placed through the blood gas analyser on the main menu. The program was primarily aimed at umbilical acid-base analysis, but in order to make the program more general, options were added to the menu system so that samples of other types could be entered. The main sample types identified were:

- paired cord samples
- single cord samples
- fetal blood samples
- neonatal blood samples
- adult blood samples
- quality control samples

Every sample is written to a database of 'raw samples' data when it is initially received, and to a specific database once identified. It was specified that logging of quality control samples and internal calibrations was essential, but that automatic detection of quality control results falling outside acceptable limits was beyond the scope of the project. The user simply inputs the quality control batch number and, when the results are displayed, selects **Passed** or **Failed** according to whether the results fall within the ranges for the batch. All calibrations are automatically detected and logged to database; the user is informed that a calibration is taking place and temporarily suspended from proceeding with data entry until the calibration sequence is completed. While the analyser is sampling or calibrating, a progress box is displayed indicating the amount of time expected for the sequence to complete and the amount of time elapsed in the form of a moving bar which grows from 0% to 100%. The actual time taken for each sequence is recorded and is used to form a rolling average of elapsed time, which becomes the expected time for the next sequence.

Having selected the type of sample, the user inputs the mother's surname, first name, hospital identification number, and possibly the infant number for multiple births, and then is

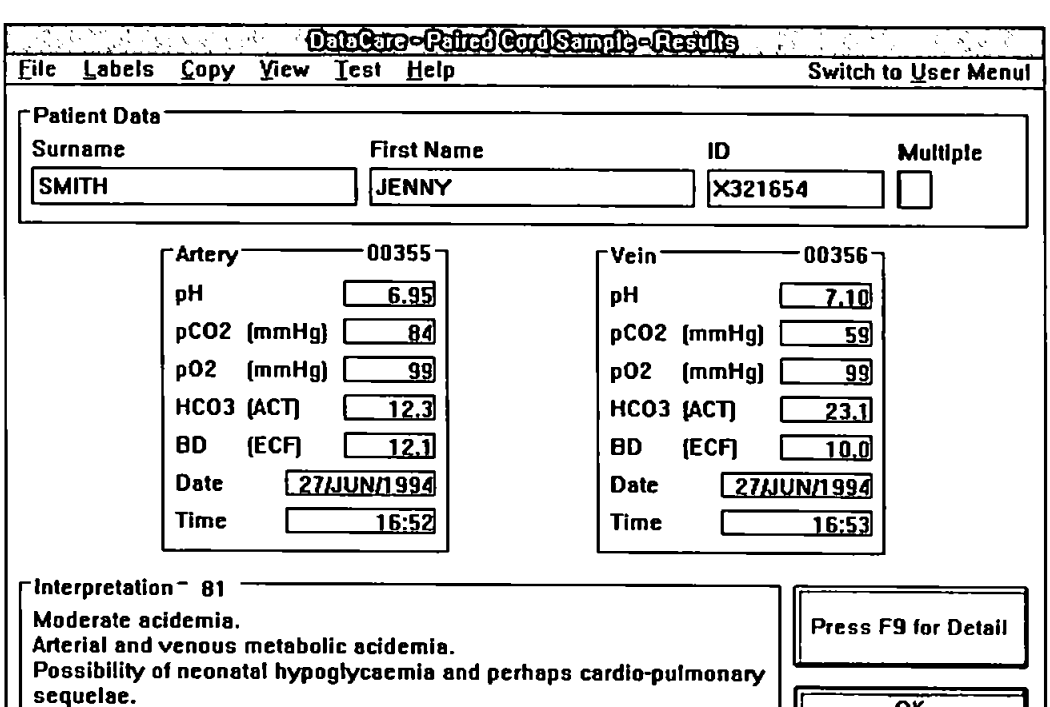


Figure 3.15: *an example screen showing paired cord results and expert system interpretation*

prompted to place the first sample into the analyser. The sample is measured by the analyser and the results transmitted to the computer. These are immediately passed to the expert system module for preliminary error checking. If an error is detected, the user is immediately presented with visual and optional audio feedback. The user is given the choice of *retrying* the sample if more blood is immediately available, *ignoring* the error if no more blood is available or *abandoning* the current measurement if more blood is to be obtained from the cord at a later stage. The user is prompted again if a second sample is required; once sampling is completed the results are again passed to the expert system for the second stage of error checking and interpretation.

The results and interpretation are presented on a screen which identifies which results came from which vessel and displays a brief form of the expert system interpretation (Figure 3.15). A more detailed explanation of the interpretation is available on request. The user now has a chance to correct any errors in the patient details and then can print the results to multiple sticky labels (Figure 3.16), specifically designed for the mother's and infant's medical notes.

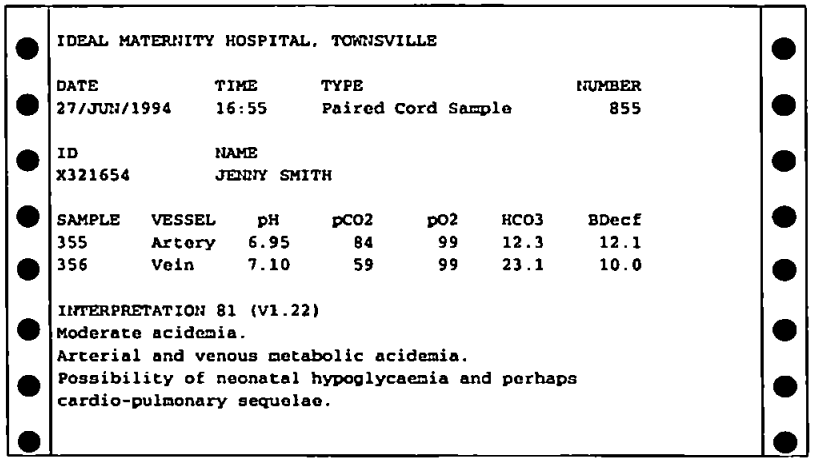


Figure 3.16: *an example label showing paired cord results*

An optional personal audit facility was also included in the design of the system. If this feature is selected, each user must enter a personal identification number (PIN) of up to four characters prior to being allowed to use the system. This allows the individual usage of the system to be tracked, and could for example be used to determine which members of staff are responsible for the most errors. PIN numbers could be assigned individually, or could be assigned to staff groups to determine their status.

3.4 Documentation

3.4.1 User Documentation

A complete online manual and help system was written, which contained an overview of the main functions of the system, detailed descriptions of each screen, descriptions of each 'Technical' command, the layout and specification of each of the databases, as well as overviews on how to use the mouse and keyboard for the novice user. The online help system features a context-sensitive user-interface, which will display the manual page appropriate to the current system screen, full bit-mapped graphics of screens and dialog boxes, and a hypertext link system to allow the user to navigate the manual. The manual was produced in printed form as well as in online form.

A set of flow diagrams detailing the required operator input were also produced to guide

the novice user through the main operations of the expert system. These were designed to be used as ready reference material, and are included in reduced format in Appendix C. An educational brochure was produced to explain the clinical background to umbilical cord acid-base analysis, the physiology on which it relies, the basis of how the information is interpreted and an explanation of the facilities provided by the expert system.

3.4.2 System Documentation

A comprehensive set of documentation was required for compliance to the quality control standards of the collaborative commercial partner. In addition to the conventional laboratory day book, a number of other documents were collated into a project file. The high level functional specification of the expert system that was produced during the initial feasibility study was updated to reflect the final software, as shown in Appendix A. A set of *operational sequence documents* (OSD) were compiled which describe the functioning of the software in terms that a non-specialist could understand, for example as a set of flow-charts or pseudo-code. In this case pseudo-code was chosen to represent the program logic, as the non-sequential aspect of a message driven operating system, such as *Microsoft Windows*, made the creation of meaningful flow-charts extremely difficult. The complete OSD for the released crisp expert system is shown in Appendix B. *Problem report* forms and *modification request* forms were designed, to be used to control changes or updates to the system.

It is generally accepted that 'C' is such a flexible language, especially in code layout, that it is liable to be abused to produce totally unreadable code. To avoid this, the code was commented extensively throughout and was laid out according to a strict convention for indentation. Test sheets to record the results of all test procedures carried out prior to any release of the software were created. Procedural documentation was written on how to compile the entire system, how to create the online and offline versions of the user manual, how to run the manual and automatic testing procedures and how to produce a master copy of the release software on floppy disk. Finally, complete hard-copy listings of all program code were produced.

3.5 Discussion

This chapter has described the formulation of a model of expertise for the validation and interpretation of umbilical cord blood acid-base data, and the encapsulation of this model in a crisp expert system. The expert system combines knowledge of the errors likely to occur in cord blood gas analysis, physiological knowledge of the reasonable results, and the knowledge of the data collected during a large randomised trial in Plymouth, to validate and then interpret results in a consistent and intelligent manner. The system also provides an automatic database of all results along with the calibration and quality control results of the machine. The stages of verification, validation and clinical assessment that took place before the commercial release of the system are described in the next chapter.

To date, far too little weight has been placed upon the importance of checking the quality of the data produced from umbilical cord acid-base analysis. In particular, the frequent occurrence of significant sampling errors in conjunction with machine errors leads to erroneous data, which are then interpreted (often poorly) without the full information of the pH's and base deficits of **both** vessels. The emphasis placed within the expert system on the data validation has been an important contribution. The blood gas analyser used in the original study had a number of design faults which contributed to the number of errors in the study; and these have been reduced by the introduction of newer technology machines. The problems of sampling blood from a wide variety of shapes and sizes of umbilical cords, the difficulty of handling samples once obtained and the relative delicacy of blood gas analyser electrodes ensure that errors are a feature of umbilical acid-base analysis, however experienced the clinical staff. The concept of using expert knowledge to define a set of 'physiologically plausible' conditions that samples must conform to is a novel aspect of this work within the field, which has now become accepted.

Several aspects of the expert system validation and interpretation rules warrant discussion. The scatter plot of pH against pCO_2 to establish the relationship between these parameters might be better represented either by plotting $[H^+]$ against pCO_2 or by plotting pH against $\log pCO_2$. Given a constant bicarbonate concentration and absence of the more complex

buffering systems, the relationship of $[H^+]$ to pCO_2 would be a straight line. In reality, a scatter would still be obtained, but on such a graph a contour of constant BD_{ecf} would be close to a straight line. Although this was attempted as it was thought likely to produce a *better* linear regression fit, in fact the resultant regression coefficient (R^2) was lower and the boundaries were less well defined. In fact there is a good argument for using $[H^+]$ directly in all clinical interpretation, as it is thought to be the hydrogen ions that are harmful to cells. However pH has been favoured historically in clinical use in most parts of the world, and its use has been retained within the expert system.

The hypothetical distributions of Figure 3.5 in Section 3.2.3 are in no way intended to be rigorous proof, or even statistical evidence, of the origin of the actual distributions seen in Figures 3.2 to 3.4. The distributions were constructed in such a way that the resultant summated distribution of Figure 3.6 was a close fit to the actual ΔpH distribution in Figure 3.2. However, they are an illustrative device derived from the known physiology of the acid-base status of arterial and venous umbilical blood and knowledge of the likely errors in sample acquisition.

The rule for the establishment of a probable single vessel when $\Delta pH < 0.025$ may falsely reject a small number of paired samples with a real physiological small difference, but this is relatively unimportant. In such a case the expert system will assume that the two samples are both from the vein and will warn that arterial acidemia cannot be excluded. As it assumes that the arterial results *will* be worse than the venous results, it will always *over-estimate* the significance of the results for the neonate. The opposite problem of failure to recognise two samples from the same vessel is more serious, as a real case of a large arterial-venous difference in which two venous samples were obtained rather than the intended arterial sample would result in a misdiagnosis of normality. The most notable such large difference from the trial database had a venous pH of 7.38 and venous BD_{ecf} of 4 $mmol.l^{-1}$, but with an arterial pH of 6.88 with an arterial BD_{ecf} of 16 $mmol.l^{-1}$— the infant unfortunately died. Had two venous samples been inadvertently obtained and been misinterpreted as a very close arterial-venous pair, then no relationship between the acid-base results and outcome would have been seen.

There is an argument for modification of the rules in Section 3.2.3 for the exclusion of parameters with negative ΔpCO_2 and negative ΔBD_{ecf} such that the venous parameters are always accepted and only arterial parameters are rejected. The justification is that the venous sample is usually large and often bubble free, whereas the arterial sample is often small and poor quality in comparison. Thus any error is far more likely to lie in the arterial sample. In such a situation the usually elevated arterial pCO_2 is likely to be artificially low due to leakage to the atmosphere. Whilst this would support the argument for retaining the venous pCO_2 (and hence BD_{ecf}) in the case of a small ΔpCO_2, the utility of having venous BD_{ecf} is limited *without* the arterial BD_{ecf}. A large negative ΔBD_{ecf}, as in the case of Table 3.2, must be the result either of a falsely low venous pCO_2 or a falsely high arterial pCO_2. As neither of these situations is compatible with the loss of arterial pCO_2 it was decided that loss of venous pCO_2 was the most likely cause and thus the venous BD_{ecf} was rejected.

The various levels of pH, pCO_2 and BD_{ecf} used in the expert system were derived from a combination of expert knowledge, statistical analysis and other published work. Although they were carefully considered and reviewed, they are by no means absolute or rigorous. More evidence from carefully managed clinical studies, in which umbilical acid-base results are collected and validated, is needed before more certain judgements of physiologically significant levels can be made.

The field of obstetrics in general, and acid-base interpretation in particular, is characterised by the large amount of uncertainty in the domain. The input data suffers from the inherent imprecision in the analyser measurement, together with the further imprecision added by poor handling of the samples. During the knowledge elicitation process it became clear that the acid-base knowledge was uncertain, even among respected experts, causing certain combinations of variables to be perplexing and difficult to interpret. Many authors have previously identified the need for uncertainty handling in medical expert systems [52, 53, 47], and the development of the crisp expert system presented here confirmed this requirement. Although this system functions at an acceptable level for clinical use, the lack of internal uncertainty handling in the knowledge and data has limited further improvement and future expansion of the system.

Chapter 4

Evaluation of the Crisp Expert System

4.1 Introduction

In this chapter the evaluation of the crisp expert system for umbilical acid-base assessment is presented. The requirements for the evaluation of medical expert systems in general are reviewed, followed by an examination of the requirements for this particular expert system. The stages of verification, validation and field testing that took place prior to the commercial release of the system are described in detail. The crisp expert system was officially released in July 1994. Since then, it has been installed at twenty two hospitals in the UK, and these are listed in Section 4.4.

A brief review is included of some of the possible sources for additional perinatal data that might be used in clinical assessment in the future. This is followed by three small studies of the clinical use of the expert system. Firstly, the acid-base results of several years from Plymouth are used to examine the frequency of errors, the classifications of the expert system and the reactions of the clinical users to the expert system. Secondly, various statistical analyses of the Plymouth acid-base data in relation to some other perinatal factors are presented. Lastly, the acid-base results in Plymouth and a nearby hospital in Exeter are compared. The evaluation processes are then summarised and some factors which have contributed to the successful implementation and user-acceptance of the system are discussed.

4.2 Evaluation of Medical Expert Systems

Many authors have used the terms *verification*, *validation*, *assessment* and *evaluation* in a differing and inconsistent manner in the literature [93, 94]. In this thesis the following terminology, designed specifically for the European AIM project [35], is adopted:

- *verification* is the process of ensuring that the expert system is functioning according to its specification,
- *validation* is the process of ensuring that the knowledge embedded within the expert system is an accurate representation of the domain[1],
- *assessment* is the process of determining the effect that the expert system has in the clinical setting — this can be further split into two further sub-tasks
    1. *human factors assessment* - determining whether the system is useful and usable to its clinical users, and
    2. *clinical assessment* - determining whether the system makes a measurable difference (improvement) to clinical care
- *evaluation* is a global term that refers to the collective processes of *verification*, *validation* and *assessment*.

Although most authors assert that thorough evaluation of medical expert systems is an essential pre-requisite to their routine use in the clinical situation, it is widely acknowledged that this is very difficult in practice [85, 135]. A formal clinical evaluation should either demonstrate that the new treatment, technique or technology improves patient care, or show that it maintains patient care whilst decreasing cost. The usual method of evaluating novel medical treatment, the double-blinded randomised control trial (RCT), in which neither the administering clinicians or treated patients are aware of which arm of the trial the patients are in, would be extremely difficult to implement for an expert system for a number of reasons:

[1] note that this use of the term *validation* is entirely separate to its use in Section 3.2.1, in which it refers to the process that the expert system carries out in performing checks on the input data — the different meanings should be clear by context

- the lack of external criteria against which to measure the expert system
- to 'blind' the clinician an independent third-party would have to interact with the expert system, thus adding an additional level of interpretation and indirection
- the effect of the expert system will depend on the initial skill levels of the clinicians involved in the trial
- the transfer of knowledge from the expert system to clinicians through interaction with the system over time may influence results

An alternative to the RCT is the less demanding 'test of no harm' or 'safety-test', in which the safety of the expert system is considered, and the establishment that the system cannot harm a patient is sufficient — full clinical assessment is not necessary from a safety point of view [59].

Traditionally expert systems have been characterised as:

- *decision making systems* — the expert system reaches decisions on patient care and presents the decisions as the correct patient management, for example an intelligent anaesthetic control system, or
- *decision support systems* — the expert systems reaches decisions on patient care and presents its recommendation to the human clinicians, who then reach their treatment decision based on the expert system's recommendations, their own judgements and other clinical factors.

The acid-base expert system presented here does not fall naturally into either of these categories. The expert system takes a set of data and performs validation and interpretation of the data, but does *not* offer (even a suggestion of) a decision for clinical action. Thus it effectively transforms the four-dimensional numerical input data into a single textual interpretation. This puts it into a category that is *less interventionist* than even a *decision support system.* It is an expert system by the definition at the beginning of Chapter 2, in that it represents and reasons with knowledge of a specialist subject (umbilical cord acid-base analysis)

with a view to solving the problem of validating and interpreting the raw data. However, as it neither makes nor suggests a decision, such an expert system might be termed an *interpretation support system*. Hence, this new category can be characterised as:

- *interpretation support systems* — the expert system performs an intelligent analysis of raw data, and presents processed data to the clinician in a form which is more natural, but does not recommend any specific clinical action.

Many other (non expert system) technologies have been introduced into clinical use in the last 30 – 40 years, usually without the stringent evaluation requirements that have been advocated for expert systems. Many of these technologies are microprocessor based, for example the CTG (Section 1.1.3), yet had little or no formal evaluation before their introduction. Although it might be argued that they have suffered as a result, their introduction has been implemented and is often widespread. As many of these technologies are very specialised and also involve significant amounts of data-preprocessing, the differences between these systems and an *interpretation support* expert system are small and blurred. Thus, it is argued, the evaluation requirements for clinical assessment of each of the expert system types is more accurately represented by Table 4.1.

Expert System Category	*RCT*	*Safety Test*
decision making	required	required
decision support	desirable	required
interpretation support	optional	required

Table 4.1: *expert system types and evaluation requirements for clinical assessment*

The final aspect of evaluation that must be addressed is the medico-legal consideration. It is still not clear whether an expert system will be viewed as a 'product' or a 'service' by the courts, if it is subject to litigation [17]. If it is viewed as a product, then it would be subject to product liability laws, which are particularly strict in the USA. However, if the expert system is considered to be a service, then it must reach the standard expected of an 'informed and sensible body of opinion' [135].

Umbilical cord blood acid-base analysis is a classic example of a domain where *no* gold-standard exists. As stated in the medical introduction, it is currently difficult to establish the degree of brain damage, even in extreme circumstances, and totally impossible (with current technology) to identify absolutely whether any diagnosed damage actually occurred *during* labour. Consequently, it is only possible to validate the performance of the expert system against the opinions of respected clinical experts.

4.3 Evaluation for Commercial Release

The industrial partner collaborating in this project was a British company certified to conform to the requirements of the BS5750 quality assurance standard. In essence, this standard simply states that within an organisation all procedures should be specified and that adherence to the specifications should be provable. In practice, this implies that each procedure should have a specification document which identifies what tasks should be carried out and the documentation that should be produced as a result. As the collaborative partner was a blood gas analyser manufacturer whose products feature complex electronics and embedded software destined for critical clinical use, the procedures for software testing were already established. These procedures concentrated on ensuring that the software was clinically safe. In addition, their BS5750 requirements implied that any third-party software developer had to comply with these existent testing procedures.

Thus, a number of specific tasks were carried out in compliance with the company's BS5750 requirements, and to allow the release of the expert system. These tasks were to:

1. ensure that the system was safe,
2. ensure that the interpretations agreed with respected experts, and
3. demonstrate the potential for economic benefit.

The tasks carried out in the evaluation of the crisp expert system will each be described. Table 4.2 shows how each of these tasks relates to the evaluation terminology presented at the beginning of this section.

Task	*Verification*	*Validation*	*Assessment*
Subsystem Validation	•		
Face Validation	•	•	
Hazard Analysis	•		
Sensitivity Analysis	•	•	
Economic Assessment			•
Field Tests in Plymouth		•	•
Field Tests in Exeter		•	•

Table 4.2: *how each task relates to components of evaluation*

4.3.1 Subsystem Validation

Subsystem validation was carried out to ensure that the software development cycle complied with BS5750 quality standards. This involved extensive 'destruction testing' of the software in which, as far as possible, every aspect of the software was tested. Specifically, each line of code was examined to ensure that its behaviour was well determined. The code was structured such that the use of `goto` was eliminated and the use of `break` to prematurely exit from control loops was avoided. The code was compiled under the highest level of warning messages, and was refined until it produced *no* warnings. In addition, each time any function (including library functions) was called, all the possible return values for the function were anticipated and the calling code was amended to take appropriate action in each case.

The principal is that each subsystem (function) should not be able to exhibit any behaviour other than anticipated. Any non-anticipated behaviour is catered for through the use of a software exception routine, such that a message is displayed to the user screen with a description of the exception condition, and an instruction for the user to call the technical support department. If this occurs the expert system is immediately halted. Conceptually, it is better for the system to halt with an exception condition, rather than to continue to run in a state that could result in an ill-founded expert system interpretation. The few minor problems that this process highlighted were corrected.

4.3.2 Face Validation

During face validation project team members, potential expert system users, and people knowledgeable about the application domain, subjectively compare the performance of the expert system against human expert performance [93]. Face validation was partially integrated into the development phase, by the processes of rule development described in Section 3.2.2. Once the rules had been established, the complete rule set was given to a number of other experienced clinicians. Each clinician was asked to highlight any interpretation rules that they would disagree with. Additionally, all 'non-normal' results that occurred during the initial field trials in Plymouth and Exeter (see below) were regularly reviewed by the resident experts.

The result of this face validation was the minor modification of one rule, with it being split into two sub-rules. At the end of this process, no cases of non-trivial disagreement between clinicians and expert system had been discovered. This was then taken to be sufficient for an adequate demonstration of the legal criterion of reaching the standard expected of an 'informed and sensible body of opinion' described above.

4.3.3 Hazard Analysis

The process of *hazard analysis* was prescribed as part of the BS5750 requirements of the industrial partner. In contrast to the 'white-box' approach of subsystem validation, in which the code itself is examined to anticipate failures, in *hazard analysis* a 'black-box' approach is used, in which the behaviour of the system is observed in response to all conceivable external events. Each potential hazard is identified and documented, and the appropriate behaviour of the system is specified. The hazard is then instigated or simulated (as far as possible) and the actual behaviour of the system is recorded. The specific behaviour of the system is not particularly important — the specified behaviour to an unlikely hazard could be an immediate system crash with unintelligible error message — although commercial considerations usually impose the requirement of graceful and predictable behaviour to **all** hazards.

As an example, consider the expert system function to allow a download of all data to floppy disk. Table 4.3 shows the hazards that were identified for this procedure, and the anticipated program behaviour in each case.

Action	*Behaviour*
select database download to A:	databases are copied to floppy disk
select database download to C:	databases are copied to hard disk
user selects 'Cancel' button	no databases are copied
floppy disk is write-protected	warn user and prompt for new disk
floppy disk is unformatted	warn user and prompt for new disk
floppy disk is nearly full	prompt for new disk when disk is full
floppy disk contains previous download	warn user and abandon download immediately
floppy disk removed during download	prompt user to re-enter disk in drive

Table 4.3: *an example of hazard analysis for the process of database download onto floppy disk*

Such manual hazard analysis tested all user-selectable functions and commands, but did not address either the screen sequences which result from normal user input or the functioning of the expert system module itself. Both of these aspects were considered too time consuming and too complex to test manually, due to the large number of combinations of each. Therefore, a suite of automatic test procedures was created to attempt to ensure the correct functioning of these software components. As an additional advantage this test suite was designed to be utilised both in the development of the system, and when the software had to be updated, for example, to communicate with a new blood gas analyser developed by the industrial partner.

The automatic test program has the ability to simulate communications functions and user inputs. The communication functions of the blood gas analyser were encapsulated in a simulation program, which could be connected via a loop-back to the test machine and run in one instance of the test program. A second program was created to run the expert system and to simulate user input, which was run through a second instance of the test program. A set of bespoke databases were created to drive simultaneously both test programs with a set of key strokes to simulate, a set of sample results to transmit and a set of target output states of the system. The target output states could comprise not only the system screen that should

be on view, but also the expected expert system interpretation. A set of target databases that should result from a complete run of the test suite was also created.

The complete test suite attempted to verify every 'path' through the system, from the initial 'Main Menu' to the final results screen, and back to the 'Main Menu'. As the control structure allowed looped paths, theoretically there are an infinite number of paths through the system. For example, if a bad sample is detected, the operator is prompted to select *Retry*, *Ignore* or *Abandon*. If *Retry* is selected, the user returns to the same sample screen, and inputs another sample — which could be bad — etc. Rather than following these potentially infinite loops, the test program tried each looping branch once, and then twice *only*, and each optional looping branch zero times, once, and twice *only*. After all these tests had been completed, and satisfactorily passed, the system was deemed to be functionally correct according to specification.

4.3.4 Sensitivity Analysis

O'Keefe [93] defines sensitivity analysis as "systematically changing expert system input variable values and parameters over some range of interest and observing the effect upon system performance". The test suite described above was utilised to perform a comprehensive sensitivity analysis on the expert system categorisations.

The same principal as described above was adopted, with one test program simulating a blood gas analyser by reading pre-defined results from a database and transmitting them as if they were actual sample results, while a second instance of the test program simulated a user running the expert system and verifying that the obtained results matched those expected.

As there were too many possibilities of input data to test exhaustively, the partition graphs described in Section 3.2.2 were again utilised to select data samples that lay on the corners, borders and middle of each partition. Figure 4.1 shows the same partition graph with the data points that lie on corners, borders and middles highlighted. Appropriate values of the pH and pCO_2 for each vessel that would produce closest possible values of pH and BD_{ecf} to the required point were calculated independently, and entered into a database. As can

be seen from Figure 4.1, there are 26 sample points for this partition, which corresponds to $pH_A < 7.05$ *and* $BD_A \geq 12$ mmol.l^{-1}. Hence, with a pH_A value on the border (7.04) and far away (6.80), and a BD_A value on the border (12 mmol.l^{-1}) and far away (15 mmol.l^{-1}), there are 104 ($26 \times 2 \times 2$) sample points.

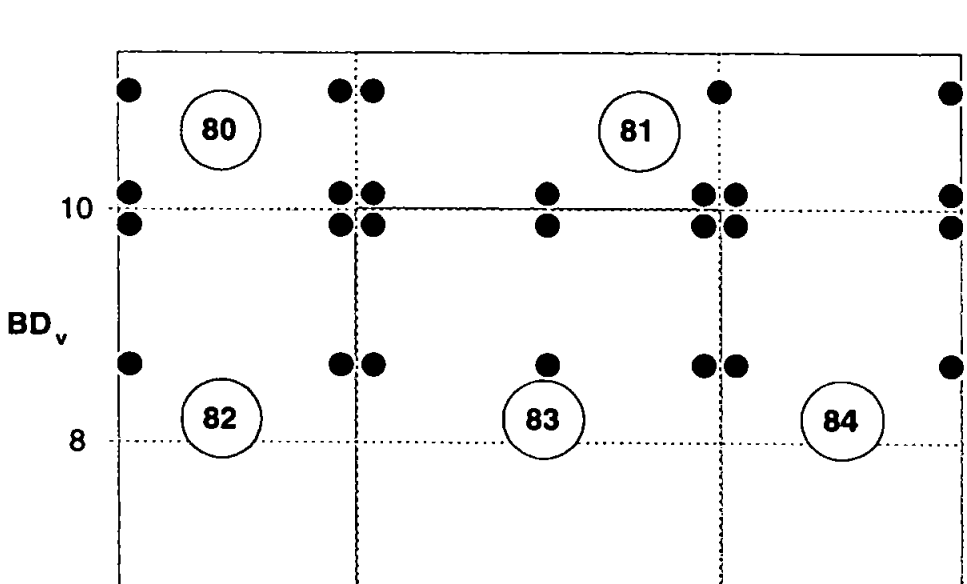


Figure 4.1: *partition graph showing data points chosen for simulated sampling in sensitivity analysis of expert system categorisation, with* ESN *for each region (see Appendix D for full list of categories with* ESN*)*

Altogether almost 1 000 samples were created to systematically test the expert system categorisation across the entire ranges of each parameter. In each case the expected ES category was forecast, and the test program verified that the specified result was obtained. The process did highlight a small number of cases (six) in which the interaction of validation and categorisation rules produced a different output to that expected. These cases were closely examined and it was judged that the actual output was more 'reasonable' than the anticipated output. Hence the anticipated output was adjusted and the test continued. These tests are now run at each software modification/update to re-verify and re-validate the functioning of the expert system.

4.3.5 Economic Assessment

If a full randomised control trial of an expert system is not possible, or not required, then it has not necessarily been shown to be of any benefit. Given that the expert system can be shown to be safe ('do no harm'), an *economic* assessment of the benefits of the expert system may be enough to justify its use [85].

Umbilical cord acid-base (UAB) assessment has the potential to be of large economic benefit. To address the economic factors, it is necessary to discuss the problems of litigation in more detail. As stated in Section 1.4.3, it is estimated that only around 10–20% of neurological handicap can be attributed to intrapartum events — either intrapartum asphyxia or traumatic damage, caused for example, by instrumental (forceps) delivery.

In the current litigious climate, the parents of a brain-damaged infant are often encouraged to sue, and the defendant is likely to be the obstetric clinician. For the plaintiff to be successful, the prosecution must prove three things:

1. the presence of brain damage,
2. causality, and
3. negligence.

Although this would apparently favour the defendant, it is not uncommon for the judge (or jury) to find in favour of the unfortunately brain-damaged child simply because there were some problems with the birth and *the clinician has no adequate defence* — despite the normal legal requirement to prove *guilt*. As it is relatively straight forward to find problems with almost *any* birth, a clinician without defence is always susceptible.

Normal UAB results will greatly increase the credibility of a defendants case, and are often sufficient in themselves to defend a case successfully [120]. In Plymouth, several cases have been deflected from litigation (the case dropped by the prosecution before going to court) on the basis of *normal acid-base results alone*. Each lost case currently costs an estimated £2 million in settlement and legal fees (split fairly evenly). In the UK, this money comes out of the general health-care budget rather than any specific legal fund.

Table 4.4 shows a breakdown of deliveries according to whether the infant has cerebral palsy (CP), and the possible influence that UAB results might have on any litigation. It may be assumed that cases in the left-hand column, in which the infant does have CP as a result of intrapartum events, are made no more costly to settle through the presence of adverse UAB results. Even if more cases are settled as a consequence (which is doubtful) this would most probably be offset by the reduction in legal fees due to early settlements. For cases in the middle column, in which the infant has CP as a result of other non-intrapartum events, the vast majority (around 98% or greater) will have normal UAB results, and hence will stand a much larger chance of being successfully defended. The 2% which may have abnormal UAB results, even though intrapartum events did *not* cause the CP, might be thought to make defence harder. In fact, these would probably make the situation little worse than if no UAB results are present, and are hugely outnumbered by those with normal UAB results.

	CP **200 / 100 000**		no CP 99 800 / 100 000
	intrapartum: 20% 40 / 100 000	other: 80% 160 / 100 000	
normal results	trauma: 10%? 4 / 100 000 saved/settled?	98% **157 / 100 000 saved**	98% 97 800 / 100 000 no litigation
abnormal results	asphyxia: 90%? 36 / 100 000 settled?	2% 3 / 100 000 false litigation?	2% 2 000 / 100 000 no litigation

Table 4.4: *the likely effect that umbilical acid-base assessment would have on litigation for cerebral palsy (CP)*

The right-hand column represents the cases that clinicians often worry about. There is no cerebral palsy, but abnormal UAB results were found. Although this group is large (2 000 per 100 000), in fact they do not cause a problem because litigation will not be undertaken (and certainly would not be successful) due to the failure to satisfy criterion 1 above through the absence of CP.

Given a conservative CP rate of around 2 per thousand, there will be around ten cases each year in a hospital such as Plymouth, with 5 000 deliveries per year. As eight of these will

probably have an antenatal or postnatal cause, and $\approx 157/160$ of these will have normal UAB results, they are all likely to be defended. UAB analysis costs roughly £20 000 per year in equipment and maintenance costs. Therefore, it is only necessary to save one additional CP litigation case every 100 years (£2 000 000 / £20 000) out of the 800 potential cases to justify the cost. This of course is only possible if the UAB results were obtained and were reliable. As the expert system costs a mere £200 per year at most (£2 500 spread over at least $12\frac{1}{2}$ years), the expert system must only improve reliability rates of UAB analysis by 1% (£200 / £20 000) to ensure that its additional cost over 'manual' UAB analysis is justified! Although the reliability rates have not been formally investigated in isolation, an improvement of 4% reliability has been shown in Plymouth in the year since the introduction of the expert system (see Section 4.6 below). Note that the costs incurred in the research development have not been included in this cost-benefit analysis. Thus, the analysis demonstrates the potential benefit to the obstetric unit that purchases the expert system, but does not demonstrate the financial benefit of the project as a whole to the University in developing the system.

It should be remembered that the reduction of unwarranted litigation is desirable not only to save money, but also to avoid the increase in 'defensive' medical practice which is ultimately harmful to the patient: the obstetric clinician should not be held responsible for events that were outside their control. As a final comment, there is a good argument for considering 'no-fault' compensation for any individual unfortunate enough to have cerebral palsy, but this is (and should be) distinct from the argument over current trends in litigation.

4.3.6 Field Tests in Plymouth

Derriford Hospital (originally Freedom Fields Hospital), Plymouth is a District General Hospital with almost 5 000 deliveries per year. Cord blood gas analysis had been performed routinely on every delivery since March 1992. The expert system was introduced in July 1993 and caused no additional clinical work. The only minor change in working practice required was for the nursing auxiliaries, who normally perform the sampling, to note the mother's name and hospital identification number before analysis.

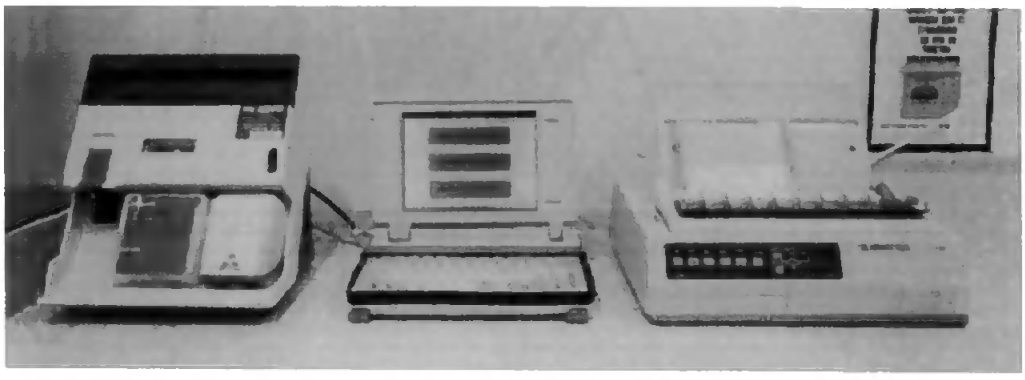


Figure 4.2: *the crisp expert system implementation at Plymouth*

The system was designed to be simple to use, and the users' quick acceptance of the system endorsed this. The multiple sticky labels for patient notes eliminated the time consuming and error prone process of transcription, and this helped significantly with user acceptability. During the trial, deliveries were carried out on two floors, with a blood gas analyser on each. The expert system was linked to only one of the analysers as the maternity department was soon to be moved to a new site at Derriford Hospital; the system proved popular enough that very quickly *all* cord samples were passed through the expert system on the instigation of the users. Minor changes and enhancements were made to the program in response to users' requests, problems or queries.

A specially designed table was constructed to secure the system with a lockable metal cradle for the PC and space for the blood gas analyser and printer. The whole table was fixed to the floor to prevent removal. It was found that after less than two weeks in its clinical location an attempt had been made to alter the system, albeit in a relatively harmless way. The Microsoft Windows *clock* and *calculator* applications had been launched and were left running in the background. Although this was a minor incident, the fact that it occurred so soon after installation highlighted that it would be dangerous to leave *any* potentially harmful 'back doors'. The operating system software was configured to prevent unauthorised users from running any programs other than the expert system or making any changes to the operating environment. A facility is available to download the databases to floppy disk for permanent archiving of results. This requires a software password and the key to unlock the bracket covering the floppy disk drive.

4.3.7 Field Tests in Exeter

The Heavitree Maternity Unit, Exeter, a District General Hospital with around 3 000 deliveries per year, 45 miles from Plymouth, was chosen as an external test site to complete the implementation and clinical testing. They wished to introduce cord blood sampling, but had previously tried a blood gas analyser without expert system, which they had found unsatisfactory. Three midwives with responsibilities for the organisation and management of the labour ward visited Plymouth to view a demonstration of the expert system in clinical use. They discussed the usability and usefulness of the system with the clinical staff at Plymouth and received tutorial and training sessions from the research staff.

The expert system was introduced towards the end of September 1993. In contrast to Plymouth, the midwives (as opposed to the nursing auxiliaries) were responsible for the sampling and analysis of the cord blood. Despite the potential pitfalls of combining the introduction of cord blood acid-base analysis with the introduction of an unfamiliar computerised expert system, the system proved popular with the staff and gained a high user acceptance. By the end of October 1993, just over one month after the system's introduction, a very high percentage (86.3%) of all deliveries were being put through the system.

4.4 Clinical Installations

The crisp acid-base expert system has been commercially licenced to *Chiron Diagnostics, Ltd* (formerly *Ciba-Corning Diagnostics, Ltd*) and sold as **Expert DataCare**. For a commercial product, possibly the best indication of its success is the number of sales. To date **Expert DataCare** has been placed at the following twenty two hospitals in the United Kingdom:

- Derriford Hospital, Plymouth, Devon
- Royal Devon and Exeter Hospital, Exeter, Devon
- Kings College Hospital, London

- St Marys Hospital, London
- Arrowe Park Hospital, The Wirral, Merseyside
- Liverpool Womens Hospital, Liverpool, Merseyside
- South Cleveland Hospital, Middlesborough, Cleveland
- Alexandra Hospital, Reddich, Worcestershire
- St Hellier Hospital, Carshalton, Surrey
- Nevill Hall Hospital, Abergeveny, Gwent
- Royal Oldham Hospital, Oldham, Greater Manchester
- Queen Elizabeth The Queen Mother Hospital, Margate, Kent
- St Johns Hospital, Chelmsford, Essex
- Kings Mill Hospital, Sutton-in-Ashfield, Nottinghamshire
- Birmingham Womens Hospital, Birmingham
- Wycombe General Hospital, High Wycombe, Buckinghamshire
- Hillingdon Hospital, London
- Mount Vernon & Watford Hospital, Watford, Hertfordshire
- Royal Cornwall Hospital, Truro, Cornwall
- Bristol Maternity Hospital, Bristol
- Dr Grays Hospital, Elgin, Morayshire
- Forth Park Hospital, Kirkcaldy, Fife

4.5 Review of Available Databases

4.5.1 The Importance of Additional Data

The crisp expert system for the analysis of umbilical acid-base assessment was implemented in Plymouth in July 1993, as described above. In the period to the end of December 1995,

over 11 000 real patient samples had been collected, of which over 10 000 consist of a pair of samples. This *Master Cords* database would form an essential role in the expansion of the crisp and fuzzy acid-base expert system to a comprehensive assessment of neonatal outcome expert system. However, in order to expand the system further to incorporate the other outcome factors such as Apgars, Neonatal Encephalopathy and background labour information, it would be necessary to obtain suitable databases of these factors.

Unfortunately, a comprehensive maternal and neonatal computerised database is not currently implemented in Plymouth. The only comprehensive collection of patient information is to be found in each patient's hand-written paper notes. Consequently, it is a non-trivial task to compile the data necessary for incorporating these additional factors. Fortunately, there are currently two disparate separately compiled databases which can be copied and linked to the umbilical acid-base database by means of the maternal details. These are the *Birth Register* database and the *Paediatric* database and each will be described in more detail. In addition, there is a large database of cord blood gas results with other neonatal, maternal and background information in existence at the Kingston Hospital in Ontario, Canada. As this *Kingston* database also has some long-term follow-up information available, it is hoped that it could be utilised in the external validation of the eventual neonatal outcome expert system.

4.5.2 Master Cord Blood Acid-Base Database

The *Master Cords* database is automatically generated by the crisp expert system. As described earlier, the *Paired Cord Samples* and *Single Cord Samples* are originally written to two separate databases. These databases are downloaded from the portable PC on labour ward to floppy disk roughly once per month, where they are transferred to a central file server. These are then combined and appended to the master database, and two extra fields are added. The first field, *TYPE*, is added automatically to indicate the origin of the record (P = Paired, S = Single) and the second field, *GOOD*, is added manually to label whether the record corresponds to a real patient (as opposed to a test sample, for example). The record layout is shown in Table 4.5.

Field Name	*Type*	*Description*
TYPE	Text	Sample Type
N	Number (Long)	Record Number
DATE	Date/Time	Record Date
TIME	Date/Time	Record Time
OPERATOR	Text	Operator Identifier
ID	Text	Maternal Hospital ID
SURNAME	Text	Maternal Surname
FIRSTNAME	Text	Maternal First Name
I	Text	Multiple Infant Number
NA	Number (Long)	Arterial Sample Number
A	Text	Arterial Error Flags
PHA	Number (Double)	Arterial pH
PCO2A	Number (Double)	Arterial pCO_2
PO2A	Number (Double)	Arterial pO_2
HCO3A	Number (Double)	Arterial HCO_3^-
BDA	Number (Double)	Arterial Base Deficit
DA	Date/Time	Arterial Sample Date
TA	Date/Time	Arterial Sample Time
NV	Number (Long)	Venous Sample Number
V	Text	Venous Error Flags
PHV	Number (Double)	Venous pH
PCO2V	Number (Double)	Venous pCO_2
PO2V	Number (Double)	Venous pO_2
HCO3V	Number (Double)	Venous HCO_3^-
BDV	Number (Double)	Venous Base Deficit
DV	Date/Time	Venous Sample Date
TV	Date/Time	Venous Sample Time
ESN	Number (Integer)	Expert System Category
ESI	Text	Expert System Interpretation
ESV	Text	Expert System Version
GOOD	Text	Real Patient Sample Flag

Table 4.5: *record layout of Master Cord Blood Acid-Base database*

The table consists of the record number with date and time of record, followed by the mother's details with maternal hospital identification number and the infant number for multiple births, followed by the arterial sample results (blank for single samples), followed by the venous sample results and finally with the expert system categorisation and interpretation.

4.5.3 Plymouth Birth Register Database

The Plymouth *Birth Register* database is created within the maternity department from manual records entered into the delivery book. The delivery book consists of one record per labour (twins are both entered under the same record) and is written immediately after delivery by the labour ward staff, mainly as a legal document of the birth. These entries are then periodically re-entered onto a computer database by senior labour ward staff to enable overall delivery statistics for the hospital to be compiled. At this stage it is re-keyed by maternal hospital identification number and date of delivery. Each infant in a multiple birth is given a unique record, distinguished by the infant number in a similar fashion to entries in the *Master Cords* database. The record layout is shown in Table 4.6.

The record holds details of maternal factors such as age, parity and medical problems and some background details of the labour such as mode of analgesia, induction and delivery mode and some basic neonatal information of gestational age, gender, birth weight and Apgar scores. A small number of labours have the length of the first and second stage (as estimated by the delivery midwife) entered. Unfortunately, the quality of data entry in the delivery book is fairly poor so that often the record of the data is imprecise or even incorrect. Consequently, there is sizeable additional uncertainty present, as well as the inherent uncertainty in data such as gestational age.

4.5.4 Plymouth Paediatric Database

Once the infant has been born and labour completed, the mother and child are transferred from the obstetric ward to the paediatric ward. At this stage, in Plymouth, each infant is

Field Name	*Type*	*Description*
YEAR	Number (Integer)	Year
N	Number (Long)	Record Number
ID	Text	Maternal Hospital ID
BOOKING	Text	Booking Code
CONS	Text	Consultant Code
GPTRANS	Text	GP Transfer Flag
AGE	Number (Integer)	Maternal Age
GRAVIDA	Number (Integer)	Maternal Parity
MEDPROBLEM	Number (Integer)	Maternal Medical Problems
INDUCTION	Number (Integer)	Labour Induced Flag
DELDATE	Date/Time	Delivery Date
I	Text	Multiple Infant Number
DELIVERY	Text	Mode of Delivery
PROBLEMS	Number (Integer)	Postpartum Problems
GA	Number (Integer)	General Anaesthetic Flag
EPIDURALCS	Number (Integer)	Epidural Caesarean Flag
EPIDURALLA	Number (Integer)	Epidural Labour Flag
PERINEUM	Number (Integer)	State of Perineum
BLOODLOSS	Number (Integer)	Amount of Blood Loss
GESTATION	Number (Integer)	Estimated Gestational Age
SEX	Text	Infant Gender
WEIGHT	Number (Integer)	Birth Weight
WARD	Number (Integer)	Special Ward Transfer Flag
A1	Text	1 Minute Apgar Score
A5	Text	5 Minute Apgar Score
SB	Number (Integer)	Still Birth Flag
MECONIUM	Number (Integer)	Meconium Presence Flag
STAGE1	Date/Time	Length of First Stage
STAGE2	Date/Time	Length of Second Stage

Table 4.6: *record layout of Plymouth Birth Register database*

entered onto a *Paediatric* database, to record details of its birth and progress through to discharge from hospital. Some of this information duplicates the information on the *Birth Register* database and some is additional. Even the duplicate information may be useful as this data is entered by dedicated data-entry clerks and is multiply checked to ensure its accuracy. Hence the combination of data may reduce the overall uncertainty in parameters. The record layout is shown in Table 4.7.

Field Name	*Type*	*Description*
MO_PATID	Text	Maternal Hospital ID
POSTCODE	Text	Maternal Post Code
INF_PATID	Text	Infant Hospital ID
D_OF_ADM	Date/Time	Date of Admission
DELIVERY	Text	Mode of Delivery
I	Text	Multiple Infant Number
B_WT	Number (Double)	Birth Weight
GEST_N	Text	Gestational Age
BTH_OFC	Number (Double)	Crown Circumference
LENGTH	Number (Double)	Infant Length
APGAR_1	Text	1 Minute Apgar Score
APGAR_2	Text	5 Minute Apgar Score
MAT_AGE	Number (Integer)	Maternal Age
PARITY	Text	Maternal Parity
GRAVID_Y	Text	Maternal Gravida
DR_AT_DEL	Text	Doctor Present at Delivery Flag
DR_AT_RES	Text	Doctor Present at Resuscitation Flag
NICU_ADM	Text	Admission to NICU Flag
HIE23_YN	Text	HIE Grade 2/3 Diagnosed Flag
CONV_YN	Text	Neonatal Convulsions Flag

Table 4.7: *record layout of Plymouth Paediatric database*

The database also includes important flags indicating the diagnosis of grade 2 and 3 hypoxic ischaemic encephalopathy, and the presence of neonatal convulsions. Possibly the most important aspect of this database is the link from maternal hospital ID to the infant hospital ID, which may allow the follow-up of infants through childhood by use of the regional child development database. This has yet to be fully investigated, but may be a valuable method of eventual validation of the expert system.

4.5.5 Kingston Blood Gas and Outcome Database

A collaborative link has been instigated with the Kingston Hospital in Ontario, Canada through Professor James Low. Professor Low has maintained an interest in umbilical cord acid-base assessment for many years and is one of the leading world authorities on umbilical acid-base interpretation. His research group in Canada has compiled a large database of over 20 000 results, which is possibly the largest such database currently in the world. Recently, these results have been passed through the crisp expert system for validation and interpretation.

Unfortunately, it is not known at present exactly what the database consists, but it certainly comprises other maternal, labour and neonatal data in addition to umbilical cord acid-base results for a number of cases. As the database has been collated over many years, there is long term follow up data available on some of the most severely asphyxiated infants. Consequently, it is hoped that it could be utilised in the future development and external validation of a comprehensive neonatal outcome expert system.

4.6 Usage of the Expert System in Plymouth

4.6.1 Methods

The introduction of an expert system has allowed, for the first time, the systematic identification of the various errors present in umbilical cord acid-base results, and has provided an accurate distribution of their frequencies. The availability of computerised records for each delivery from the *Birth Register* database has also allowed the overall usage of the system to be determined. Some of the issues of user acceptability and user attitudes have been examined through the use of the expert system individual audit facility.

4.6.2 Results

At the time of writing, up to date information on the number of deliveries in Plymouth was available up to the end of 1995. In the period between 5th July 1993 and 31st December 1995 there were 11 746 babies delivered, including multiple births and excluding stillbirths. In the same period there were 11 492 patient samples placed through the expert system. Of these, 271 were subsequent repeats due to errors in the original samples, resulting in 11 221 distinct samples from actual patients (95.5% of all deliveries). In 120 cases these samples were originally placed through the expert system as two or more *Single* or *Paired samples*, which had to be manually combined into an arterial-venous pair for an infant. As the manual combination process takes place after the data has been downloaded from the system, these paired results do not get an expert system interpretation. One *Paired* sample produced no results at all, as a result of analyser problems, leaving 11 100 classifiable samples. The validation of these samples is represented in Figure 4.3. A terminology for the categorisation of umbilical cord acid-base errors is proposed in Table 4.8.

From Figure 4.3 it can be seen that 9 109 samples were categorised as *Full Paired* samples from the 11 746 deliveries. Thus 77.5% of all deliveries resulted in a validated arterial and venous paired cord sample. The full paired results receive the most comprehensive interpretation, ranging from category 80 (severe metabolic acidemia in both vessels) to 120 (all results normal), broadly ordered by severity in terms of likelihood of the infant having suffered asphyxial damage during labour. Categories 80 to 84 indicate that an arterial metabolic acidemia was present, of decreasing significance; 90 to 94 indicate that an arterial acidemia was present, but non-metabolic; and categories 100 to 113 indicate various states of mixed arterial and venous non-metabolic acidemia. The full breakdown of actual expert system categorisation is shown in Table 4.9. It can be seen that only 36 infants (0.4% of full paired, 0.3% of all deliveries) were classified in the most serious category 80 — severe metabolic acidemia in both vessels.

It had been thought that staff attitudes towards being monitored and audited would prevent the personal auditing feature of the expert system from being used, but when staff opinions

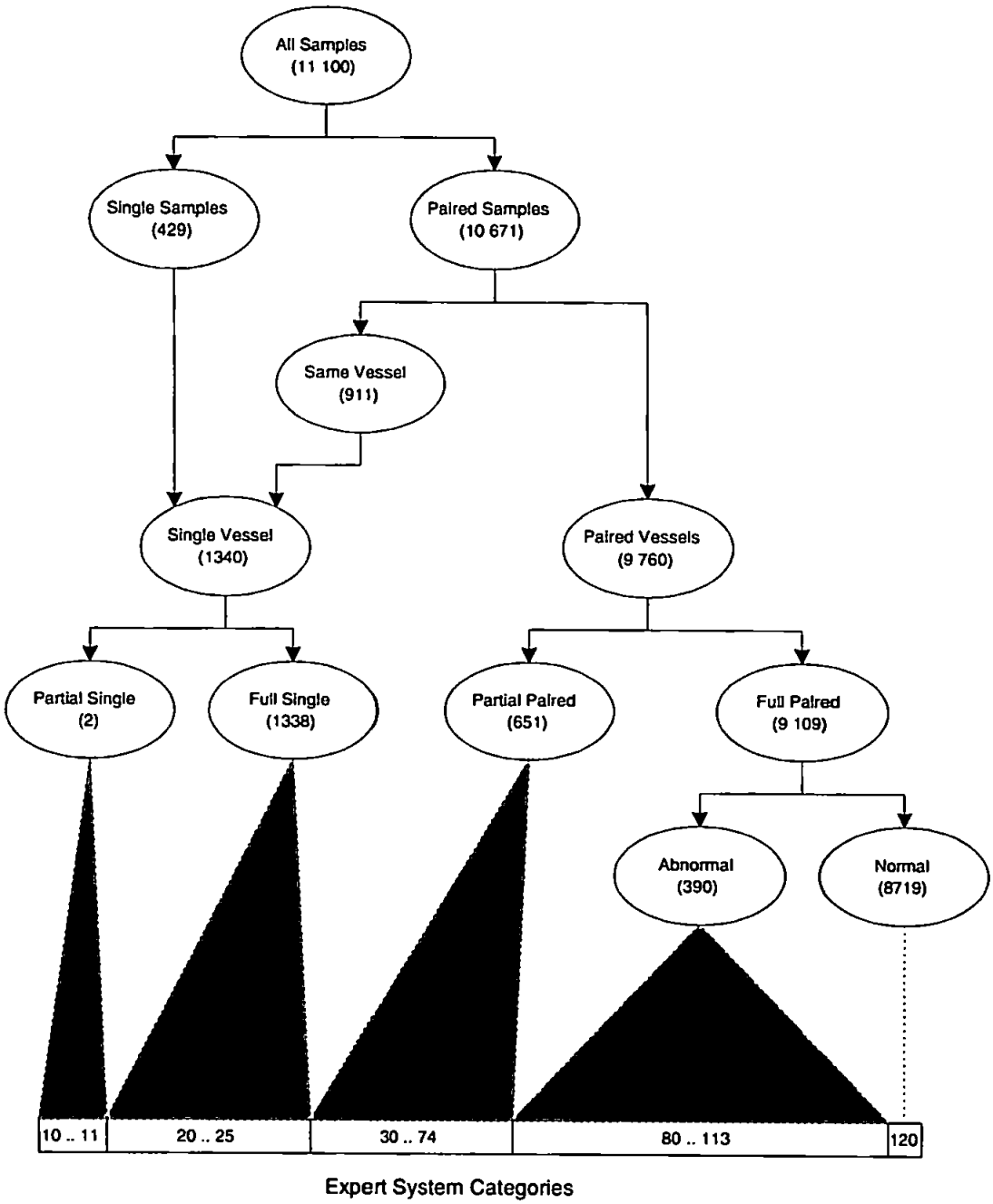


Figure 4.3: *a representation of the validation and categorisation process of the expert system, figures in brackets show the numbers in each group, and the mapping of the groups onto the expert system categories is shown at the bottom. (See Table 4.8 for a guide to the terminology used, and Appendix D for a full list of expert system categories with corresponding numeric* ESN*)*

Terminology	*Meaning*
Single Sample	A case for which only one sample of blood was taken, or more precisely only one sample was ever analysed
Paired Sample	A case for which two samples of blood have been taken (presumably intended to be arterial and venous blood, but not necessarily so)
Same Vessel	A **Paired Sample** in which very similar results from the two samples indicate that they are in fact from the same vessel, a mixed sample or extremely unlikely to have been from different vessels
Single Vessel	A **Single** or **Paired Sample** which is taken to be a single blood vessel, usually presumed to be the vein
Paired Vessel	A **Paired Sample** that appears to be both arterial and venous blood
Partial Single	A **Single Vessel** in which pCO_2 or BD_{ecf} errors are present
Full Single	A **Single Vessel** with fully validated pH, pCO_2 and BD_{ecf}
Partial Paired	A **Paired Vessel** in which one or more pCO_2 or BD_{ecf} errors are present
Full Paired	A **Paired Vessel** with fully validated pH, pCO_2 or BD_{ecf} for both arterial and venous samples
Abnormal	A **Full Paired** result in which one or more of the parameters indicates some abnormality
Normal	A **Full Paired** result in which all parameters indicate a normal (physiological) condition

Table 4.8: *a proposed terminology for the categorisation of umbilical cord acid-base analysis*

were gathered, it was found that the auxiliary staff were keen to receive individual performance feedback. Consequently the audit facility of the expert system was instituted from 5th October 1994, and anonymous league tables of performance were posted at roughly six-monthly intervals. In the year prior to auditing, from 5th October 1993 to 4th October 1994 (inclusive), there were 3 633 full paired samples collected from 4 774 deliveries (76.1%); in the year after auditing, from 5th October 1994 to 4th October 1995 (inclusive), there were 3 698 full paired samples collected from 4 619 deliveries (80.1%). This improvement is statistically highly significant (Yates corrected $\chi^2_1 = 21.27 : p \ll 0.001$) and is extremely encouraging, but it could just be an effect of the staff responsible for sampling simply improving over time. Nevertheless, the error rate has dropped steadily throughout the implementation and it is hoped that this progress will continue.

Category	*Class*	*Description*	*Number*
11	Partial single	normal pH	2
20	Full single	severe metabolic acidemia	7
21	Full single	moderate acidemia	12
22	Full single	non-significant acidemia	16
23	Full single	questionable significance	11
24	Full single	questionable significance	48
25	Full single	single normal result, arterial acidemia not excluded	1244
30	Partial paired	significant acidemia, but no base deficits	4
31	Partial paired	questionable significance	5
32	Partial paired	acute acidemia with large arterial-venous difference	1
33	Partial paired	questionable significance	2
34	Partial paired	results normal, but no base deficits	510
53	Partial paired	results normal, but no arterial base deficit	6
64	Partial paired	non-significant acidemia, but no venous base deficit	1
74	Partial paired	results normal, but no venous base deficit	122
80	Full paired	severe arterial and venous metabolic acidemia	36
81	Full paired	moderate arterial and venous acidemia	22
83	Full paired	acute moderate acidemia	6
84	Full paired	acute moderate acidemia with large arterial-venous difference	10
90	Full paired	non-acute moderate acidemia	14
91	Full paired	moderate acidemia	2
92	Full paired	significant acidemia	18
93	Full paired	significant acidemia, but non-metabolic	21
94	Full paired	non-significant acidemia	63
100	Full paired	moderate non-acute arterial and venous acidemia	2
101	Full paired	significant arterial acidemia	16
102	Full paired	questionable significance	29
104	Full paired	acute moderate acidemia, with normal venous results	8
105	Full paired	questionable significance	43
110	Full paired	significant acidemia, both vessels with venous metabolic	1
111	Full paired	significant acidemia, with venous metabolic	7
112	Full paired	non-acute mixed acidemia	40
113	Full paired	questionable significance	52
120	Full paired	all results normal	8719

Table 4.9: *the expert system categorisation of results*

4.7 Relationships Between Plymouth Acid-Base Data and Outcome Parameters

4.7.1 Methods

The Plymouth *Master Cords* database and *Birth Register* database collected between 3rd July 1993 and 31st December 1994 were examined to establish inter-relationships between various acid-base parameters and other perinatal parameters. In this period there were 6 928 deliveries for which 6 207 (89.6%) paired samples were collected. Of these samples, 4 994 (72.1% of deliveries, 80.5% of pairs) were classified as full paired samples by the expert system *and* had perinatal data available. The acid-base parameters considered were:

- Arterial pH (pH_A)
- Arterial Base Deficit (BD_A)
- Venous pH (pH_V)
- Venous Base Deficit (BD_V)
- Expert System Category Number (ESN)

The perinatal data considered were:

- Apgar score at 1 minute ($Apgar^1$) and at 5 minutes ($Apgar^5$)
- Gestational Age
- Birth Weight
- Length of Labour (data available in 1 102 cases only)
- Delivery Mode
- Maternal Problems
- Maternal Parity
- Gender

Of the acid-base parameters pH_A, BD_A, pH_V and BD_V are interval data, and ESN is categorical data. Of the perinatal parameters the Apgar scores are ordinal data; gestational age, birth weight and length of labour are ratio data; delivery mode, maternal problems, gender and parity are categorical data. Although the Apgar score is supposed to comprise five factors each scored 0 – 2, totalled to give an overall score of 0 – 10 (see Section 1.4.2), in practice it is a subjective impression of the infant. Very few scores of ten are ever assigned, especially by doctors, and clinically scores of nine or ten are considered interchangeable. Hence, for this study, the scores were combined to a single group labelled 9.

The various associations were examined by a number of parametric and non-parametric tests. There are two factors which may prevent parametric tests from being strictly valid: the parent populations are known to be skewed and non-normal, and the variance of parameters is thought to vary across their ranges. In general parametric tests have been used whenever the large numbers have indicated that the non-normality can be tolerated by the central limit theorem. Due to the slight uncertainty in the validity of assumptions and because of the large number of tests carried out, a relatively cautious value of $\alpha = 0.001$ was chosen to indicate significance. For each case, the test chosen, the null and alternative hypotheses, and the resultant test statistic and its significance are stated.

4.7.2 Results

pH_A, BD_A, pH_V and BD_V Grouped by Apgar^1 and Apgar^5

Analysis of variance (ANOVA) was used to examine the mean of each parameter grouped by Apgar scores. If μ_i represents the mean of the parameter for an Apgar score of i, then the null hypothesis, H_0, in each case was that $\mu_0 = \mu_1 = \mu_2 = \cdots = \mu_9$; the alternative hypothesis, H_1, was that $\mu_i \neq \mu_j$ for some groups i and j. The numbers of cases in each group are shown in Table 4.10 — unfortunately, due to the infrequency of low Apgar^5 scores, even with almost 5 000 results, the Apgar^5 of 0 group has no cases and the Apgar^5 of 1, 2 and 3 groups have only one case each. The results of the ANOVA are shown in Table 4.11, and are shown graphically in Figures 4.4 to 4.7 (Pages 134 and 135), with the mean and its 95% confidence

interval for each group. The assumption of homogeneity of variance across the groups may well be violated and, as some of the groups are small in number and the numbers in each group are very different, this could have serious consequences on the validity of the test [49]. Hence the process was repeated using Kruskal-Wallis ANOVA by ranks to test the location of the medians of each group — this also has the advantage of not assuming normality of the parent population. Although the full results will not be quoted, the Kruskal-Wallis test confirmed significance in all cases, albeit at slightly lower p-values.

	0	1	2	3	4	5	6	7	8	9
Apgar^1	2	5	10	35	37	96	130	227	524	3928
Apgar^5	0	1	1	1	4	13	19	29	95	4831

Table 4.10: *number of cases in each Apgar group ($N = 4994$)*

	Apgar^1		Apgar^5	
Parameter	$F_{(9,4984)}$	p	$F_{(5,4985)}$	p
pH_A	78.3	10^{-136}	35.3	10^{-35}
BD_A	32.3	10^{-55}	19.0	10^{-18}
pH_V	68.5	10^{-119}	40.8	10^{-41}
BD_V	26.4	10^{-44}	19.4	10^{-19}

Table 4.11: *results of ANOVA to test equal means for pH_A, BD_A, pH_V and BD_V by Apgar groups*

The single case with Apgar^5 of 2 shown in Figures 4.6 and 4.7 was of interest, as the low Apgar^5 appeared inconsistent with the normal pH's and Base Deficits. When the case was examined it was found that the Apgar^1 score was 9. Thus, the infant had normal acid-base status and normal Apgar score at 1 minute, but its condition worsened significantly to reach Apgar^5 of 2. The long term outcome of this infant is not currently known, but clearly the normal acid-base and Apgar^1 would seem to eliminate intrapartum asphyxia as the cause of the very low Apgar^5.

pH_A, BD_A, pH_V and BD_V Ordered by Apgar^1 and Apgar^5

The ANOVA (and Kruskal-Wallis) tests presented above simply test for differences in location of the groups. As the Apgar scores are ordered, the Jonckheere test for ordered alterna-

tives was used to test the null hypothesis $H_0 : \theta_0 = \theta_1 = \theta_2 = \cdots = \theta_9$, where θ_i is the median of the pH group with Apgar score i, against the alternative hypothesis $H_1 : \theta_0 \leq \theta_1 \leq \theta_2 \leq \cdots \leq \theta_9$, with at least one of the differences a strict inequality ($<$). The direction of the test was reversed for base deficit. This test was chosen as it is a non-parametric test requiring only that each sample is from the same population. The results in Table 4.12 show the J^* statistic which is approximately normally distributed with mean zero and standard deviation one. It can be seen that all parameters allow for the rejection of H_0 at $\alpha = 0.001$.

Parameter	*Ordered By*	J^*	p
pH_A	Apgar^1	18.8	10^{-79}
BD_A	Apgar^1	12.7	10^{-37}
pH_V	Apgar^1	16.1	10^{-58}
BD_V	Apgar^1	10.1	10^{-24}
pH_A	Apgar^5	7.2	10^{-13}
BD_A	Apgar^5	5.5	10^{-8}
pH_V	Apgar^5	8.1	10^{-16}
BD_V	Apgar^5	4.8	10^{-6}

Table 4.12: *results of the Jonckheere test for ordered alternatives to test ordered medians for pH_A, BD_A, pH_V and BD_V by Apgar groups*

ESN by Apgar^1 and Apgar^5

The relationship between the expert system categorisation (*ESN*) and Apgar scores was examined by the *chi-square* test. The null hypothesis, H_0, is that the row categories are independent of the events represented by the columns, i.e. that expert system categorisation is independent of Apgar score, and the alternative hypothesis, H_1, is that the events are related. Due to the very small numbers in low Apgar groups, especially for Apgar^5, Apgar scores of 0 to 6 were grouped together. The observed and expected frequencies are shown in Table 4.13 and 4.14, with a resultant chi-square statistic of $\chi^2 = 231.1$, $df = 3$, $p \approx 10^{-50}$ and $\chi^2 = 148.6$, $df = 3$, $p \approx 10^{-32}$ respectively.

	Apgar[1] score				
ESN	≤ 6	7	8	9	Total
$\neq 120$	53	20	29	69	171
	10.8	*7.8*	*17.9*	*134.5*	
$= 120$	262	207	495	3859	4823
	304.2	*219.2*	*506.1*	*3793.5*	
Total	315	227	524	3928	4994

Table 4.13: *observed and expected frequencies of non-normal* (ESN $\neq 120$) *and normal* (ESN $= 120$) *expert system categories by Apgar*[1]*:* $\chi^2_3 = 231.1$

	Apgar[5] score				
ESN	≤ 6	7	8	9	Total
$\neq 120$	12	6	13	140	171
	1.3	*1.0*	*3.3*	*165.4*	
$= 120$	27	23	82	4691	4823
	37.7	*28.0*	*91.7*	*4665.6*	
Total	39	29	95	4831	4994

Table 4.14: *observed and expected frequencies of non-normal* (ESN $\neq 120$) *and normal* (ESN $= 120$) *expert system categories by Apgar*[5]*:* $\chi^2_3 = 148.6$

Correlations with Apgar Scores

The last question to be addressed by examination of Apgar scores was whether the pH scale or the $[H^+]$ units scale was more closely associated with low Apgar scores. This caused a problem: as the Apgar is ordinal data, technically it should be correlated by using a rank order method such as Spearman's rank order correlation or Kendall's τ coefficient. However a rank order method will be insensitive to the conversion from pH units to $[H^+]$ as the data order is unaffected by the transform. Indeed the question is meaningless while the Apgar score is ordinal as the intervals between each Apgar score could be simply adjusted to give the same 'fit' with a parameter expressed in each of the scales. To overcome this, the question was reformulated as: 'if the Apgar score was *assumed* to be an interval scale, would it correlate better to a parameter in which hydrogen ions are expressed on the pH scale or the $[H^+]$ units scale?'

The Pearson product-moment correlation coefficient was calculated for each of pH_A, $[H^+]_A$, pH_V and $[H^+]_V$, to test the null hypothesis that the variables are linearly independent against

the alternative hypothesis that there is linear dependency, and the results are shown in Table 4.15. It was found that conversion from pH units to concentration of H^+ ions ($[H^+]$) increased the correlation coefficients found in all cases. These results are consistent with the underlying physiology whereby it is thought that it is the concentration of hydrogen ions that causes damage. It indicates that choice of the logarithmic pH scale rather than the natural $[H^+]$ scale, driven by aesthetic preference of the clinical users, may be inappropriate.

	Apgar^1		Apgar^5	
Parameter	*R*	*p*	*R*	*p*
pH_A	0.330	10^{-128}	0.181	10^{-38}
$[H^+]_A$	-0.367	10^{-159}	-0.248	10^{-71}
pH_V	0.309	10^{-111}	0.197	10^{-45}
$[H^+]_V$	-0.346	10^{-141}	-0.270	10^{-84}

Table 4.15: *Pearson product-moment correlation coefficient for pH_A, BD_A, pH_V and BD_V with Apgar scores*

In addition, the correlation coefficient was also calculated for all simple multiples of $[H^+]_A$, BD_A, $[H^+]_V$ and BD_V (15 combinations). Overall, although the correlation coefficients are all low, Apgar^1 was found to correlate best inversely to $[H^+]_A \cdot [H^+]_V$, whereas Apgar^5 was found to correlate best inversely to $[H^+]_A \cdot BD_A \cdot [H^+]_V \cdot BD_V$ as shown in Table 4.16. This again is consistent with the physiology whereby just the level of hydrogen ions in both vessels (indicative of *any* respiratory through to metabolic acidosis) are associated with immediate problems reflected in one minute Apgar score, whereas the hydrogen ions in conjunction with base deficit in both vessels (indicative of a *metabolic* acidosis) are associated with longer duration problems reflected in the five minute Apgar score. Note that the low value of R^2 shown in Table 4.16 indicates that the contribution of acid-base results to Apgar scores is low — consistent with the fact that many other factors also affect the Apgar score.

	Parameter	*R*	R^2	*p*
Apgar^1	$[H^+]_A \cdot [H^+]_V$	-0.380	0.144	10^{-171}
Apgar^5	$[H^+]_A \cdot BD_A \cdot [H^+]_V \cdot BD_V$	-0.416	0.173	10^{-208}

Table 4.16: *Pearson product-moment correlation coefficient for the best multiplicative combination of $[H^+]_A$, BD_A, $[H^+]_V$ and BD_V with Apgar scores*

Correlation of pH_A and BD_A with *Gestational Age*, *Birth Weight* and *Length of Labour*

The Pearson product-moment correlation coefficient was calculated for pH_A and BD_A with each of *Gestational Age*, *Birth Weight* and *Length of Labour*. This tested the null hypothesis that the variables are linearly independent against the alternative hypothesis that there is linear dependency. The results are shown in Table 4.17, and graphically in Figures 4.8 to 4.13 on Pages 136 to 138. The null hypothesis was not rejected for *Birth Weight*, and, although it could be rejected for *Gestational Age*, the correlation coefficient is so small that the effect is negligible. Linear correlation was found for *Length of Labour*, with the pH_A decreasing slightly and BD_A increasing slightly for longer labours.

	Gestational Age		*Birth Weight*		*Length of Labour*	
Parameter	*R*	*p*	*R*	*p*	*R*	*p*
pH_A	-0.076	10^{-8}	0.005	0.718	-0.178	10^{-7}
BD_A	0.060	10^{-5}	-0.038	0.008	0.178	10^{-7}

Table 4.17: *Pearson product-moment correlation coefficient for pH_A and BD_A with* Gestational Age, Birth Weight *and* Length of Labour

pH_A and BD_A Grouped by *Delivery Mode*, *Maternal Problems* and *Maternal Parity*

Analysis of variance (ANOVA) was used to examine the mean of pH_A and BD_A grouped by *Delivery Mode*, *Maternal Problems* and *Maternal Parity*. In each case, the null hypothesis tested was that there was no difference in mean for each group against the alternative hypothesis that at least one group had different mean. A description of the different groups and the number of cases found in each group is shown in Table 4.18. The results of the ANOVA are shown in Table 4.19 and are shown graphically in Figures 4.14 to 4.19 (Pages 139 to 141), with the mean and its 95% confidence interval for each group. As with the Apgar groups these were checked with Kruskal-Wallis because of the very unequal numbers in the groups and the possibility of unequal variance. All findings were confirmed except in the case of pH_A by *Maternal Problems*: for this group the Kruskal-Wallis statistic was 14.1 with $df = 5$, $p = 0.015$.

Delivery Mode		
Description	*Abbreviation*	*n*
Spontaneous Vertex Delivery (Normal)	SVD	3860
Vaginal Breech Presentation	Breech	34
Neville-Barnes Forceps Delivery	NBFD	224
Wrigleys Forceps Delivery	Wrigleys	36
Keenans Rotational Forceps Delivery	KRFD	35
Ventouse Vacuum Delivery	Ventouse	137
Emergency Caesarian Section	Emergency CS	381
Emergency Caesarian Section (Breech)	Emergency CS Breech	27
Elective Caesarian Section	Elective CS	212
Elective Caesarian Section (Breech)	Elective CS Breech	48
Maternal Problems		
Description	*Abbreviation*	*n*
None	None	4747
Placental Abruption	Abruption	23
Pre-eclampsia	Eclampsia	9
Diabetic	Diabetic	22
Previous Caesarian Section	Previous CS	175
Placenta Praevia	Praevia	18
Maternal Parity		
Description	*Abbreviation*	*n*
Primiparous (First Delivery)	Primip	2256
Multiparous (Second+ Delivery)	Multip	2738

Table 4.18: *description and number of cases in each group*

	Delivery Mode		*Maternal Problems*		*Maternal Parity*	
Parameter	$F_{(9,4984)}$	p	$F_{(5,4988)}$	p	$F_{(1,4992)}$	p
pH_A	7.7	10^{-11}	8.9	10^{-8}	234.3	10^{-51}
BD_A	11.4	10^{-17}	10.6	10^{-10}	153.5	10^{-34}

Table 4.19: *results of ANOVA to test equal means for pH_A and BD_A by* Delivery Mode, Maternal Problems *and* Maternal Parity

pH_A, BD_A, pH_V, BD_V and *ESN* Grouped by Gender

Finally an examination of the acid-base data by the gender of the infant was carried out. Gender was only recorded for 4 968 of the 4 994 deliveries, of which 2 554 were male and 2 414 were female. For pH_A, BD_A, pH_V, and BD_V, Student's t-test was used to examine the null hypothesis that there was no difference in the means, against the alternative hypothesis that there was a difference. The results are shown in Table 4.20 and are shown graphically in Figures 4.20 and 4.21 (Page 142), with the means and 95% confidence intervals. Chi-squared was used to examine the frequencies of normal and non-normal expert system categorisation by *Gender*, and these results are shown in Table 4.21, with a Yates corrected $\chi^2_1 = 4.48$, $p = 0.034$. Although only the pH_A and pH_V parameters showed a significant difference at the $\alpha = 0.001$ level, all tests showed a difference at a less restrictive $\alpha = 0.05$ level. As the two groups are large and roughly equal in size, a less restrictive significance level may be appropriate.

Parameter	t	p
pH_A	4.52	10^{-5}
BD_A	1.99	0.047
pH_V	3.65	0.00026
BD_V	2.71	0.0067

Table 4.20: *results of Student's t-test for difference in means of pH_A, BD_A, pH_V and BD_V by* Gender

	Gender		
ESN	*Male*	*Female*	Total
$\neq 120$	102	69	171
	87.9	*83.1*	
$= 120$	2452	2345	4797
	2466.1	*2330.9*	
Total	2554	2414	4968

Table 4.21: *observed and expected frequencies of non-normal* (ESN $\neq 120$) *and normal* (ESN $= 120$) *expert system categories by* Gender: $\chi^2_1 = 4.48$

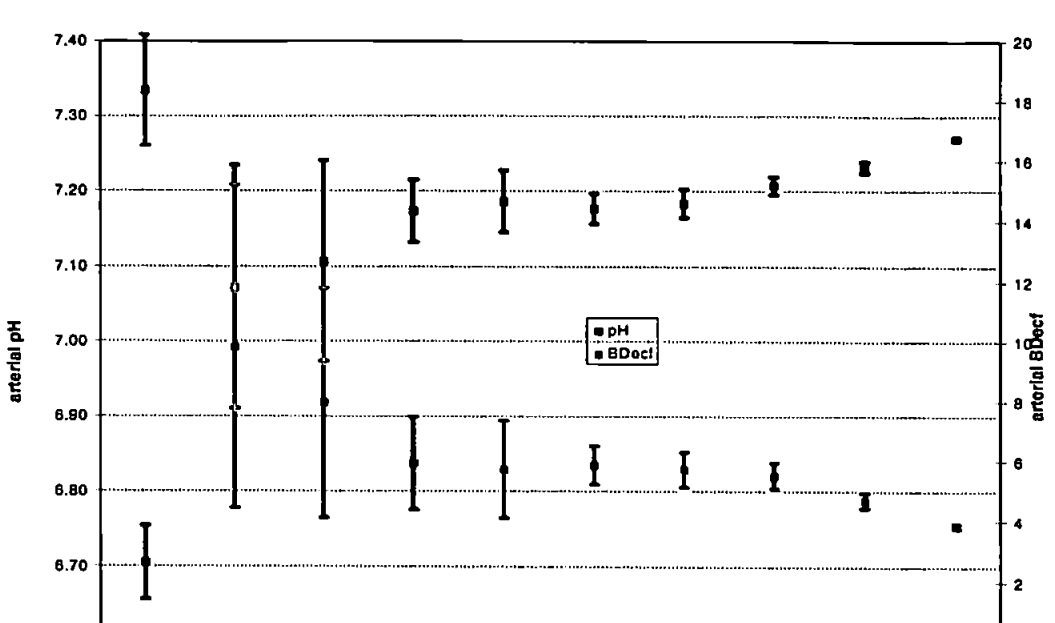


Figure 4.4: *mean and 95% confidence intervals of arterial pH and BD_{ecf} grouped by Apgar[1] (see Table 4.10 for numbers in each group)*

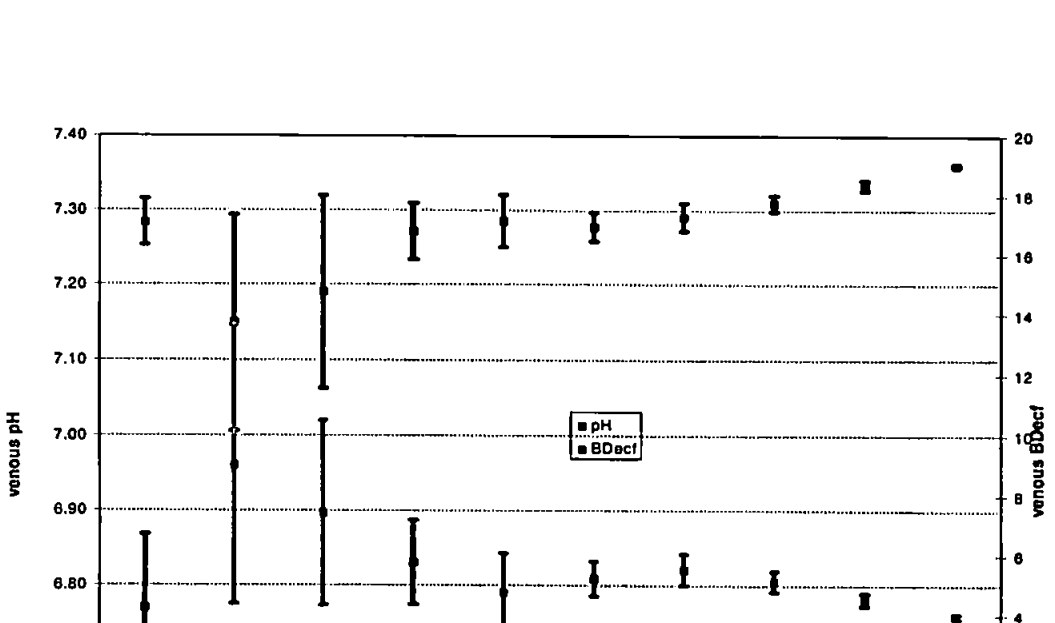


Figure 4.5: *mean and 95% confidence intervals of venous pH and BD_{ecf} grouped by Apgar[1] (see Table 4.10 for numbers in each group)*

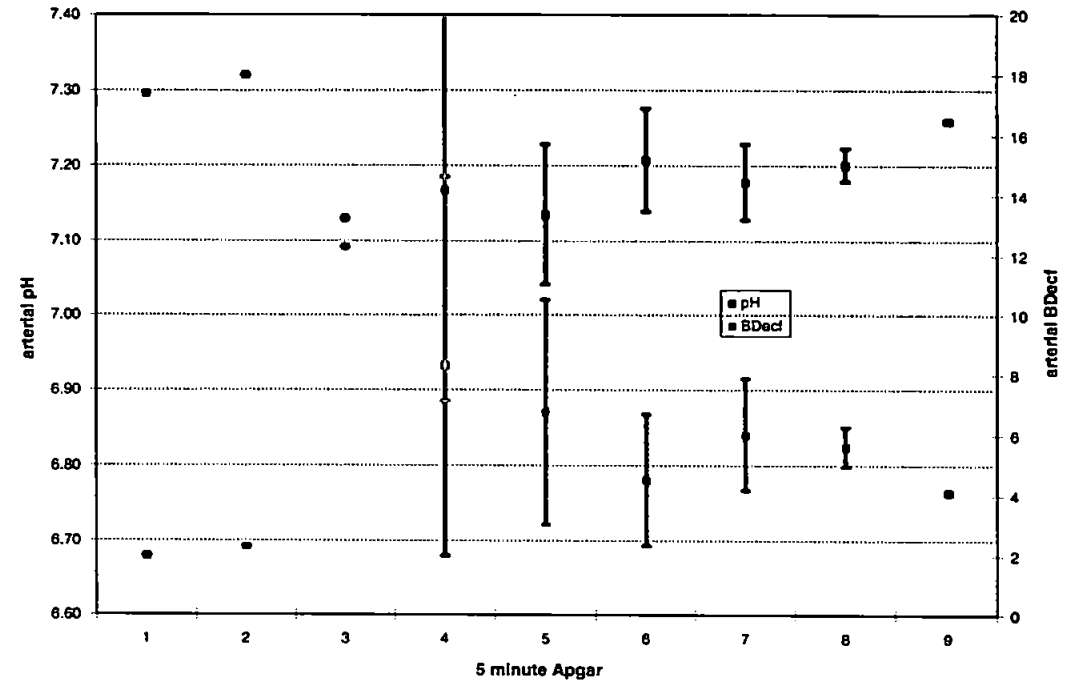


Figure 4.6: *mean and 95% confidence intervals of arterial pH and BD_{ecf} grouped by Apgar[5] (see Table 4.10 for numbers in each group)*

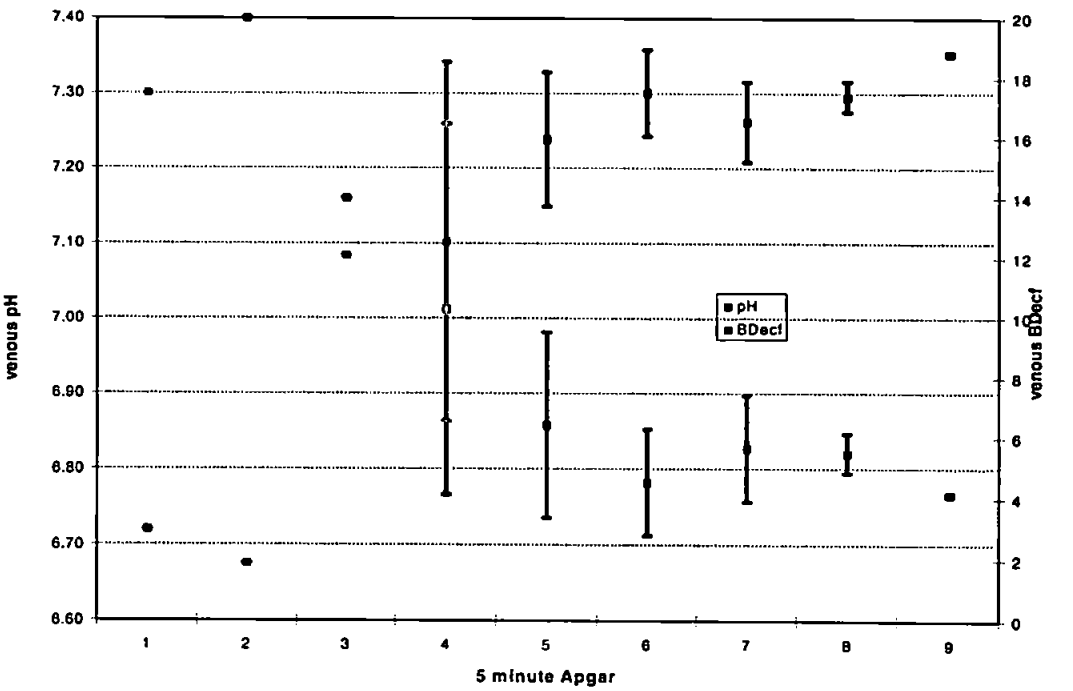


Figure 4.7: *mean and 95% confidence intervals of venous pH and BD_{ecf} grouped by Apgar[5] (see Table 4.10 for numbers in each group)*

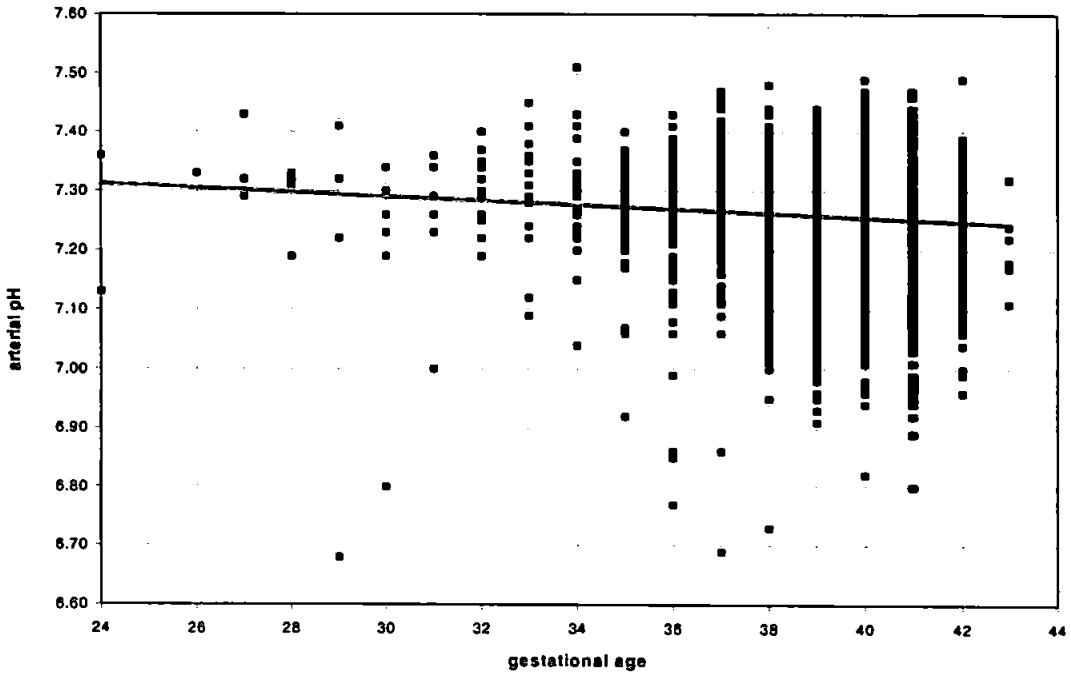


Figure 4.8: *scatter plot of arterial pH against* Gestational Age *with regression line*

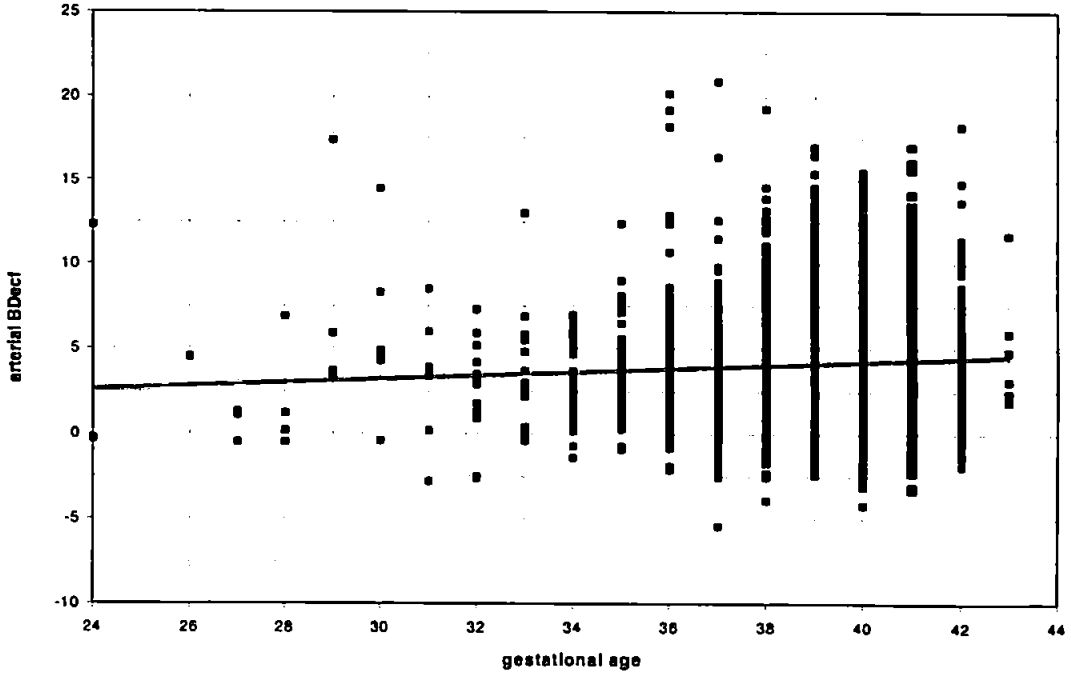


Figure 4.9: *scatter plot of arterial* BD_{ecf} *against* Gestational Age *with regression line*

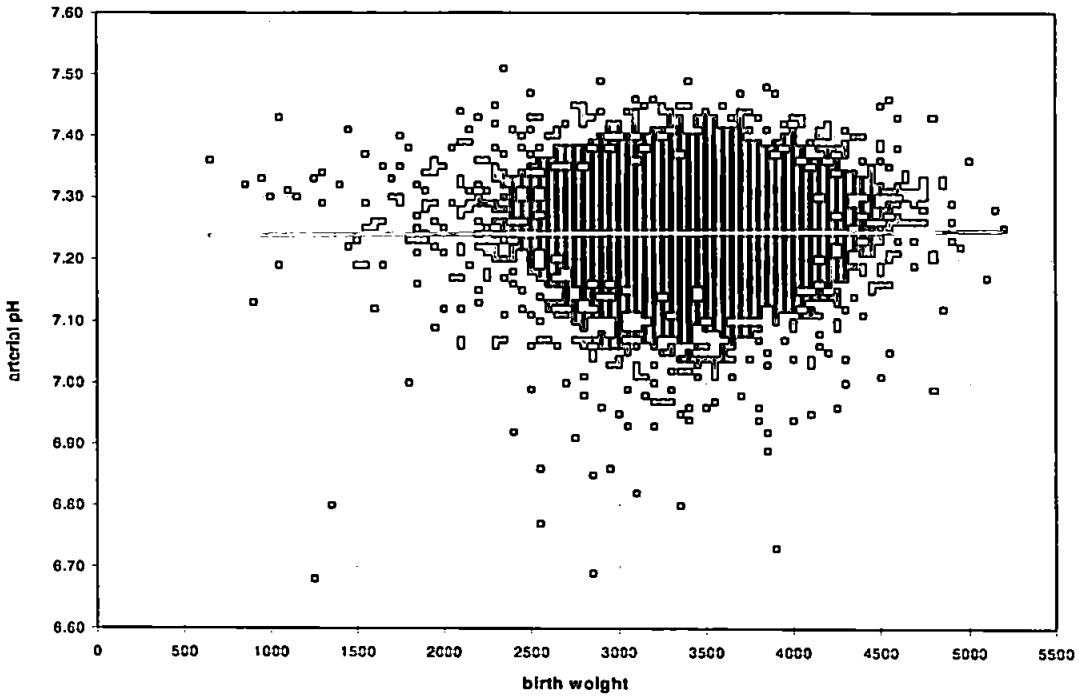


Figure 4.10: *scatter plot of arterial pH against* Birth Weight *with regression line*

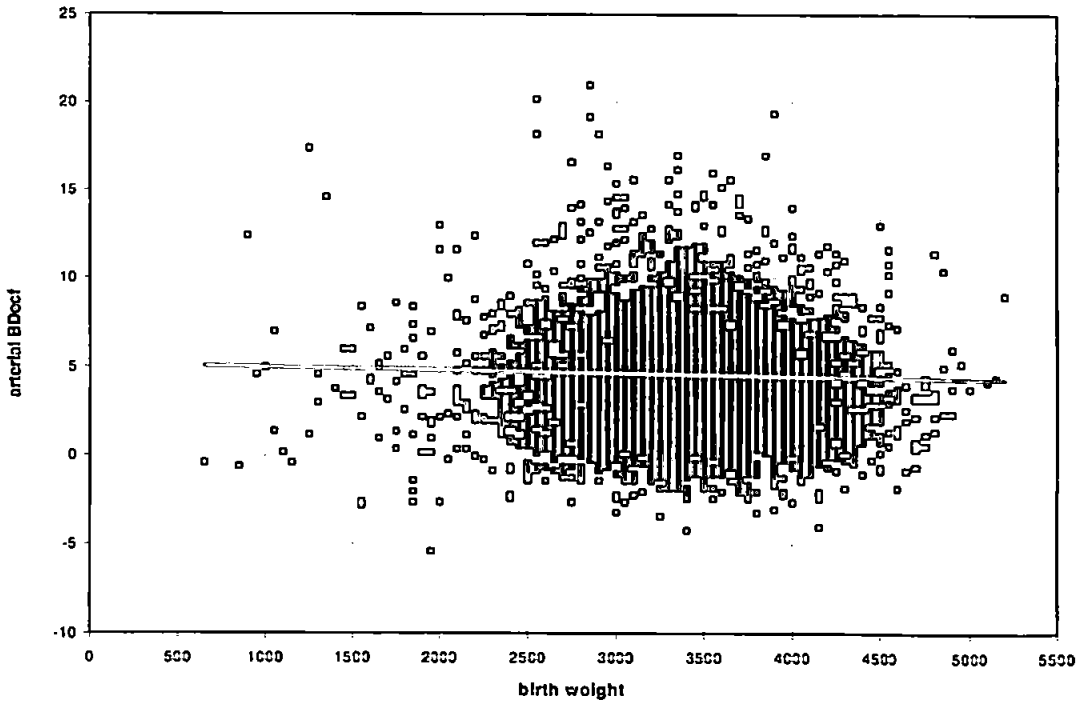


Figure 4.11: *scatter plot of arterial* BD_{ecf} *against* Birth Weight *with regression line*

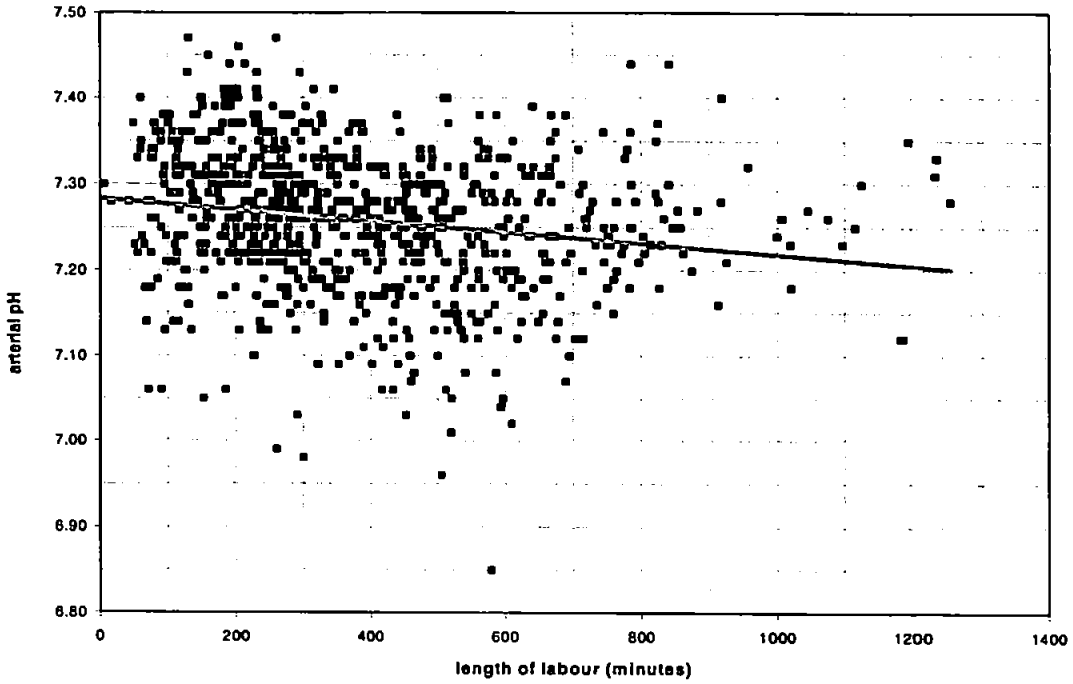


Figure 4.12: *scatter plot of arterial pH against* Length of Labour *with regression line*

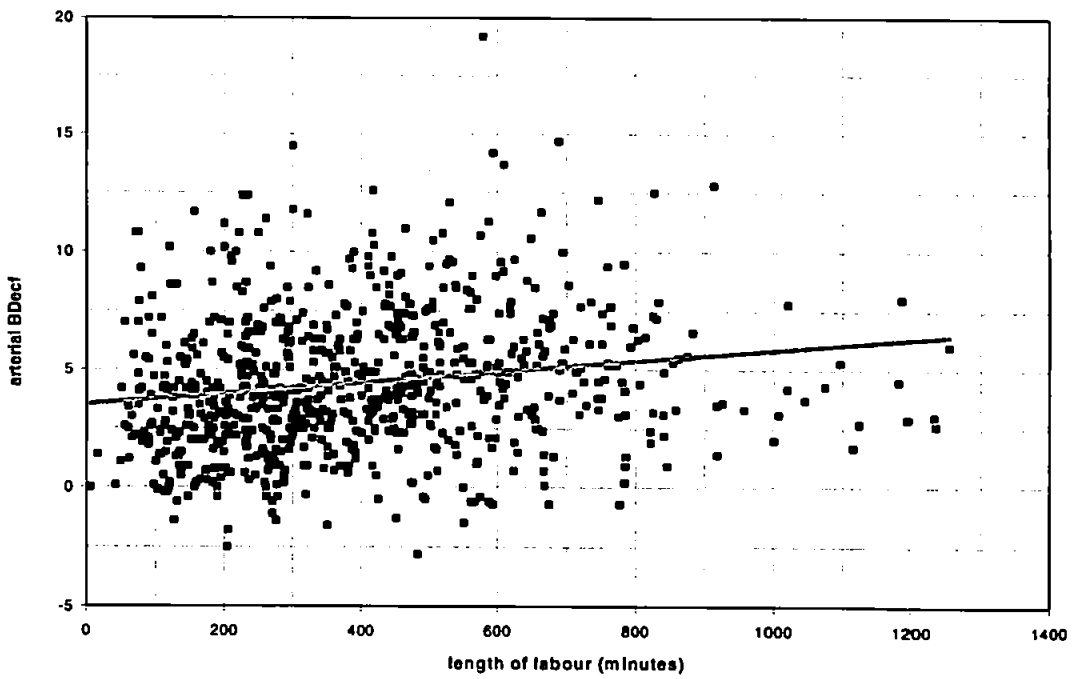


Figure 4.13: *scatter plot of arterial* BD_{ecf} *against* Length of Labour *with regression line*

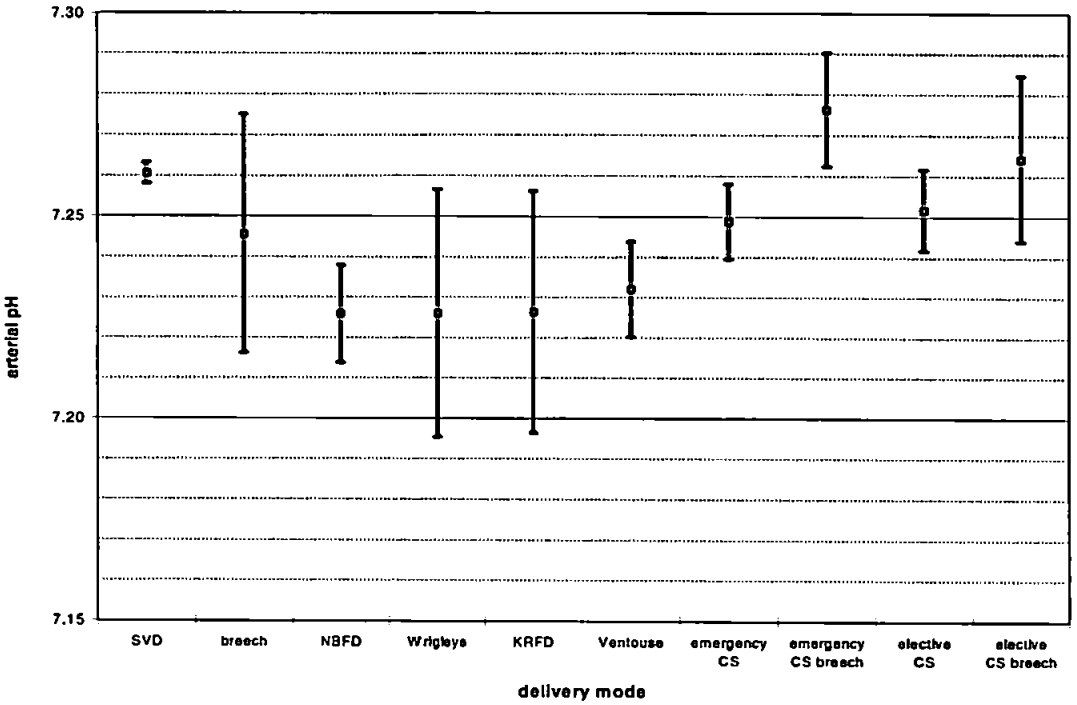


Figure 4.14: *mean and 95% confidence intervals of arterial pH grouped by* Delivery Mode *(see Table 4.18 for numbers in each group)*

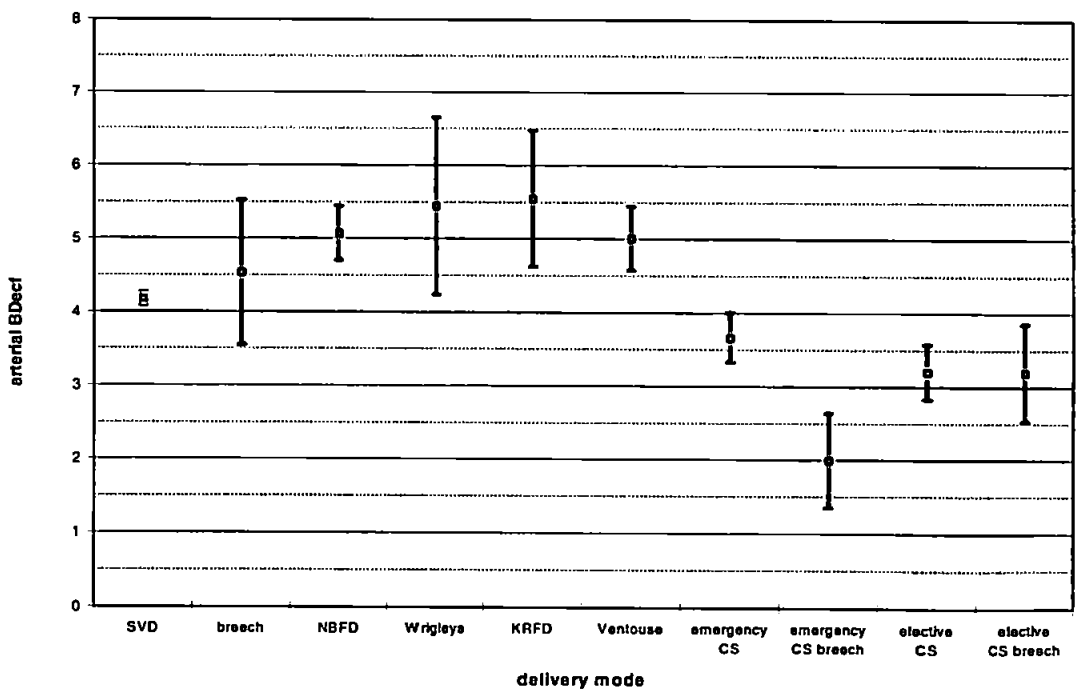


Figure 4.15: *mean and 95% confidence intervals of arterial* BD_{ecf} *grouped by* Delivery Mode *(see Table 4.18 for numbers in each group)*

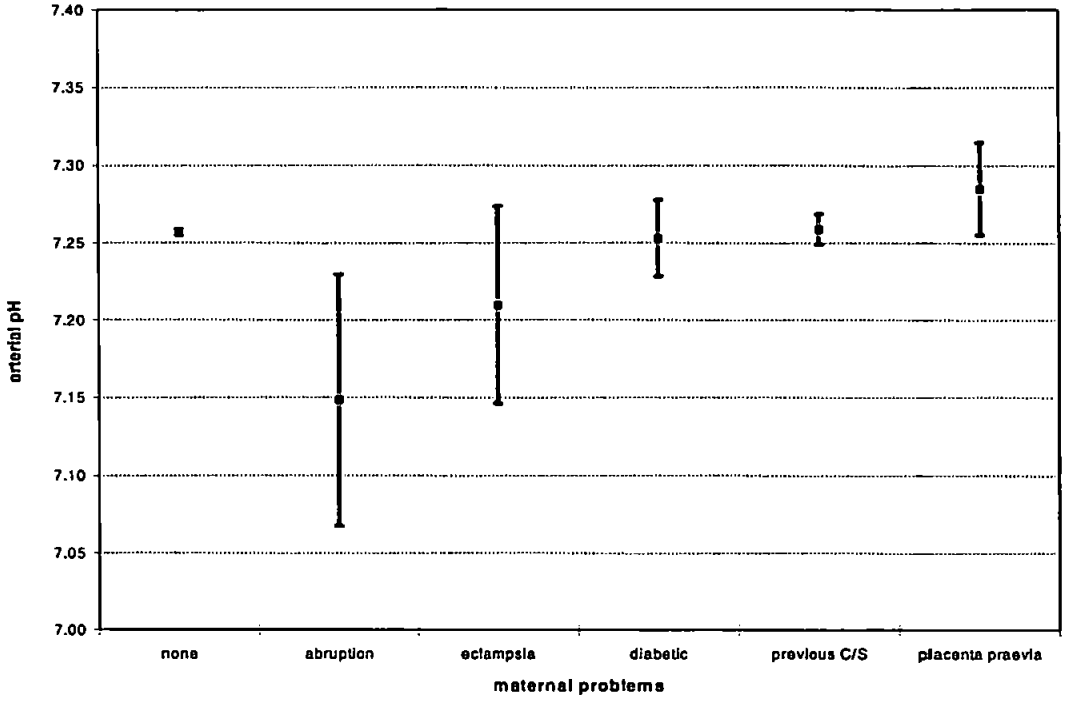


Figure 4.16: *mean and 95% confidence intervals of arterial pH grouped by* Maternal Problems *(see Table 4.18 for numbers in each group)*

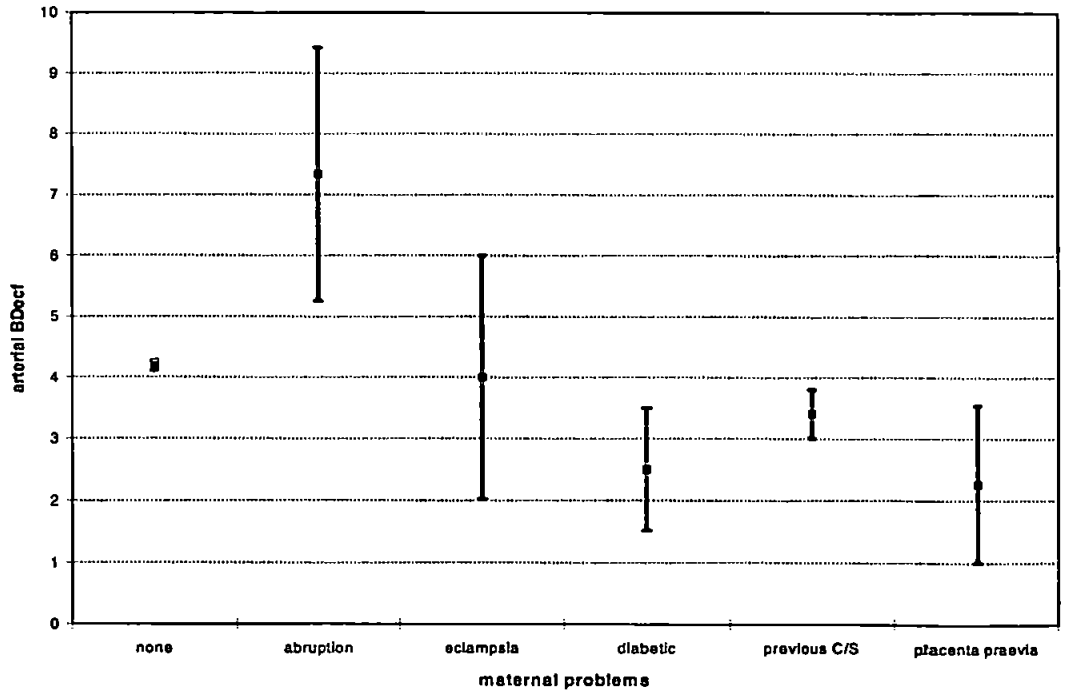


Figure 4.17: *mean and 95% confidence intervals of arterial* BD_{ecf} *grouped by* Maternal Problems *(see Table 4.18 for numbers in each group)*

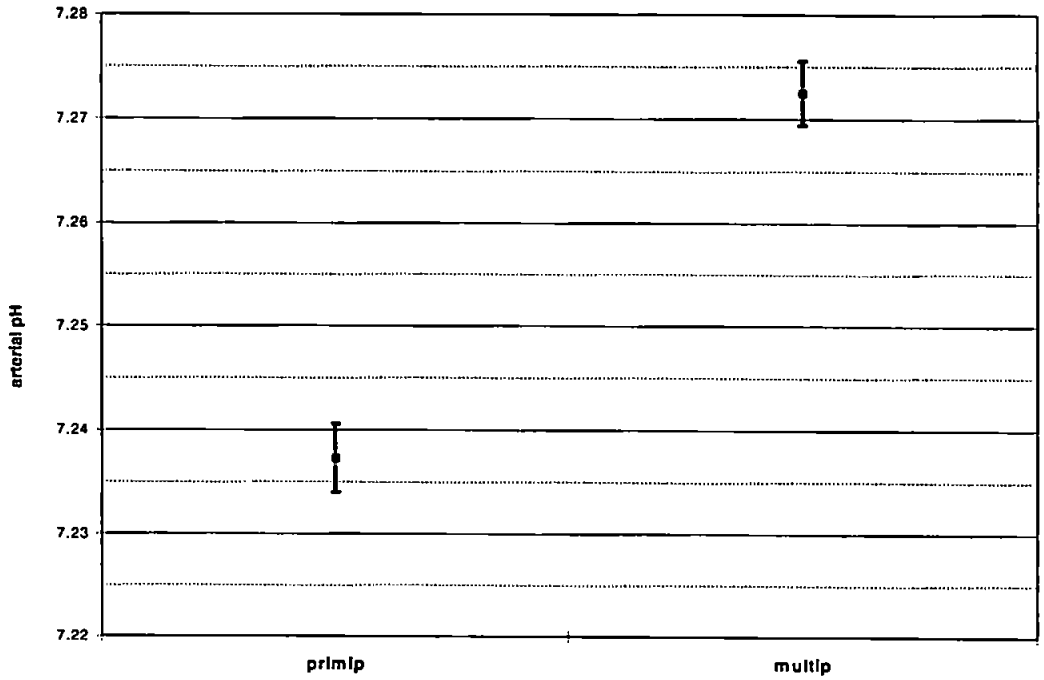


Figure 4.18: *mean and 95% confidence intervals of arterial pH grouped by* Parity *(see Table 4.18 for numbers in each group)*

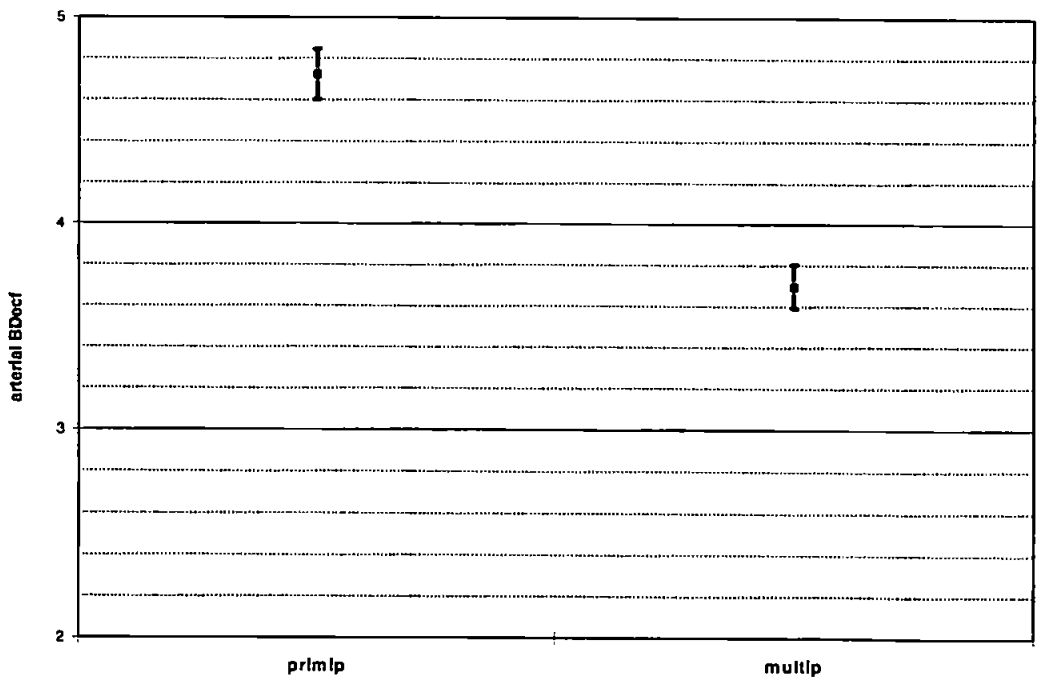


Figure 4.19: *mean and 95% confidence intervals of arterial* BD_{ecf} *grouped by* Parity *(see Table 4.18 for numbers in each group)*

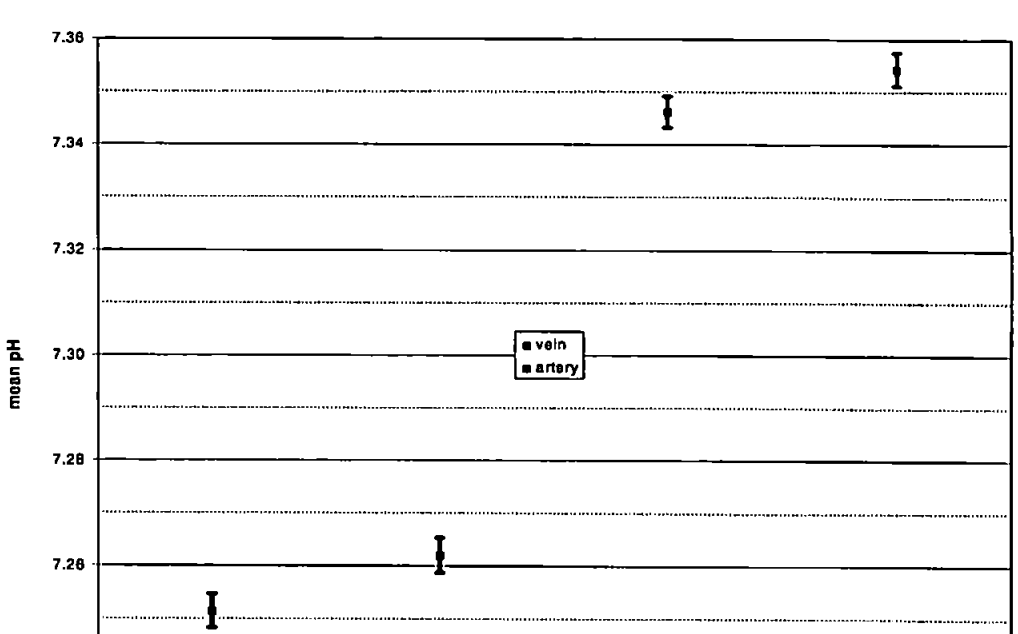


Figure 4.20: *mean and 95% confidence intervals of arterial and venous pH grouped by* Gender *(male $n = 2554$, female $n = 2414$)*

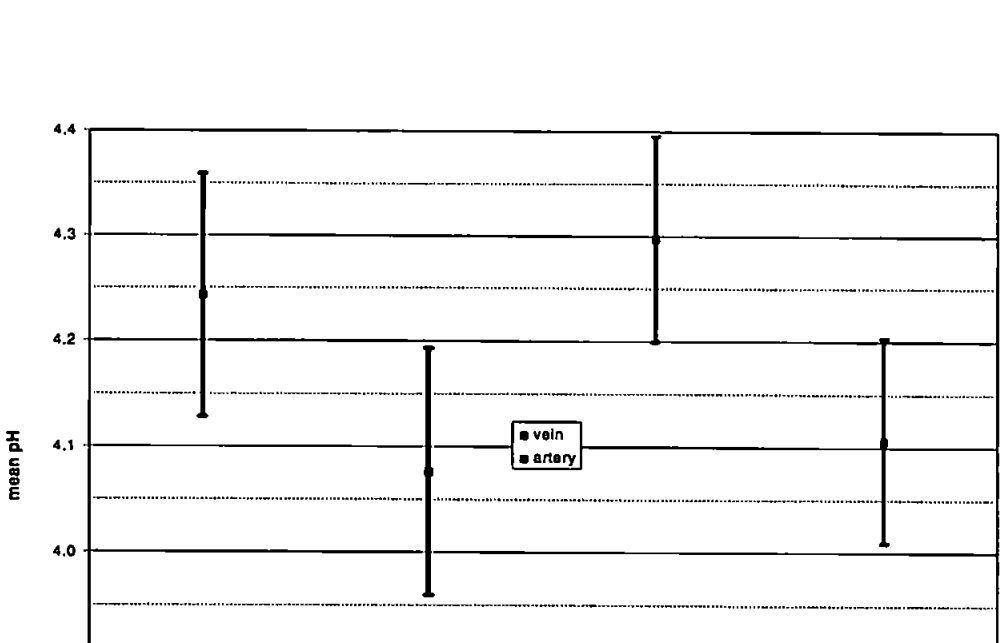


Figure 4.21: *mean and 95% confidence intervals of arterial and venous BD_{ecf} grouped by* Gender *(male $n = 2554$, female $n = 2414$)*

4.8 A Comparison of Plymouth and Exeter

4.8.1 Methods

Routine cord blood sampling for every delivery began in Exeter on the 1st October 1993, after a short period of familiarisation with the system. Data collection took place in the middle of July 1994, so data for the nine complete months from 1st Oct 1993 to 30th June 1994 were taken from both centres for comparison. Median and centile ranges were used to describe the populations as all have markedly skewed distributions. Comparisons of location were made using Student's t-test, as the high numbers in both groups ensured that the test was reliable, and the proportions of expert system categorisation were tested with χ^2, all at 5% significance level.

Every blood gas analyser will have its own particular calibration, which will depend on internal calibration parameters and individual electrodes. Consequently each machine will have minor performance differences. The quality control results from both centres were examined to standardise the pH and pCO_2 results so that the pH and BD_{ecf} results could be properly compared, without the minor machine differences influencing the statistics. Three sets of quality control material with preset levels of pH and pCO_2 parameters were measured regularly to ensure the correct functioning of the analysers, as part of the routine clinical maintenance of the machines. Comparison of these quality control results at each level allowed for compensation of differences in machine performance (cross-calibration). Regression analysis of the monthly means for each parameter at each level showed that there was no overall trend in machine calibration.

Examination of mean differences by Student's t-test showed that there was no difference in mean pH at each of the three levels, but that there was a significant mean difference in pCO_2 readings. Regression analysis on the means of each of the three levels at the two sites showed that the Exeter pCO_2 results could be corrected by:

$$pCO_2(\text{corrected}) = 1.07 pCO_2(\text{measured}) - 0.81 \tag{4.1}$$

which represents a correction of between approximately 0.0 and 0.5 kPa over the range of pCO_2 ($R^2 \approx 1.0$). Once this was applied, the mean differences in pCO_2 quality control results at each of the three levels were eliminated. This correction was then applied to all pCO_2 readings from Exeter and the results used to re-calculate the BD_{ecf} parameters, *before* further statistical description and comparison presented below.

4.8.2 Results

In Plymouth, 3 318 samples were taken from 3 544 deliveries (93.6%). Of these samples, 95.7% were intended to be from both artery and vein, and 4.3% were single samples. Of the intended artery and vein samples, the expert system classified 84.5% (2 684) as validated arterial-venous pairs. In Exeter, 1 848 samples were taken from 2 116 deliveries (87.3%). Of these samples, 88.7% were intended to be from both artery and vein, and 11.3% were single samples. Of the intended artery and vein samples, the expert system classified 74.5% (1 222) as validated arterial-venous pairs. Thus only 75.7% of deliveries in Plymouth and only 57.8% in Exeter resulted in validated paired samples. The median and 2.5th to 97.5th centile range for each parameter are given in Table 4.22.

		Plymouth ($n = 2684$)		
Vessel	pH	pCO_2 (kPa)	pO_2 (kPa)	BD_{ecf} (mmol.l^{-1})
Artery	7.27 (7.06 to 7.40)	6.7 (4.5 to 9.7)	2.1 (0.8 to 3.7)	3.6 (-0.9 to 11.0)
Vein	7.36 (7.16 to 7.49)	5.1 (3.3 to 7.6)	3.6 (1.9 to 5.3)	3.6 (-0.5 to 9.4)
		Exeter ($n = 1222$)		
Vessel	pH	pCO_2 (kPa)	pO_2 (kPa)	BD_{ecf} (mmol.l^{-1})
Artery	7.23 (7.03 to 7.36)	7.1 (4.6 to 10.3)	2.4 (0.9 to 4.3)	4.7 (-0.4 to 12.2)
Vein	7.34 (7.15 to 7.46)	5.3 (3.4 to 7.9)	4.1 (2.1 to 6.1)	4.1 (-0.6 to 10.3)

Table 4.22: *median and 2.5th to 97.5th centile range for acid-base parameters at Plymouth and Exeter*

Examination of the distributions of pH in the artery (Figure 4.22) and vein (Figure 4.23) showed a significant shift to lower pH's in both vessels in Exeter compared to Plymouth (pH_A: $t = 13.9$, $p \approx 10^{-42}$; pH_V: $t = 6.6$, $p \approx 10^{-11}$). Similarly, the distributions of BD_{ecf} in the artery (Figure 4.24) and vein (Figure 4.25) both showed a significant shift to higher

BD_{ecf}'s in Exeter compared to Plymouth (BD_A: $t = -10.8$, $p \approx 10^{-26}$; BD_V: $t = -3.6$, $p \approx 0.0003$).

As the differences between vessels are clinically significant [102], these too were examined by Student's t-test. The mean pH difference of 0.11 at Exeter was found to be significantly greater than that of 0.09 at Plymouth ($t = -11.7$, $p \approx 10^{-30}$), and the mean BD_{ecf} difference of 0.6 mmol.l^{-1} at Exeter was also found to be significantly greater than that of 0.0 mmol.l^{-1} at Plymouth ($t = -12.0$, $p \approx 10^{-31}$).

The expert system categorisations were examined and the results are shown in Table 4.23. As can be seen, the observed frequencies of the 'worst' *ESN* groups (80's and 90's) are higher at Exeter than at Plymouth, although the results did not reach significance ($\chi^2_4 = 6.43$, $df = 4$, $p = 0.169$). When cast in a 2×2 contingency table of *ESN* groups 80's and 90's (arterial pH < 7.05) compared to the rest, the 41 (3.4%) cases at Exeter were significantly more than the 57 (2.1%) cases at Plymouth (Yates corrected $\chi^2_1 = 4.71, p = 0.030$). Although the same trend was apparent when grouped by *ESN* 80's ($pH_A < 7.05$ and $BD_A \geq 12$ mmol.l^{-1}) compared to the rest, the difference (17 cases at Exeter and 22 cases at Plymouth) did not reach statistical significance (Yates corrected $\chi^2_1 = 2.23, p = 0.136$).

	ESN					
Centre	80's	90's	100's	110's	120	Total
Plymouth	22	35	18	12	2597	2684
	26.8	*40.5*	*19.2*	*10.3*	*2587.1*	
Exeter	17	24	10	3	1168	1222
	12.2	*18.5*	*8.8*	*4.7*	*1177.9*	
Total	39	59	28	15	3765	3906

Table 4.23: *observed and expected frequencies of expert system categories at Plymouth and Exeter:* $\chi^2_4 = 6.43$

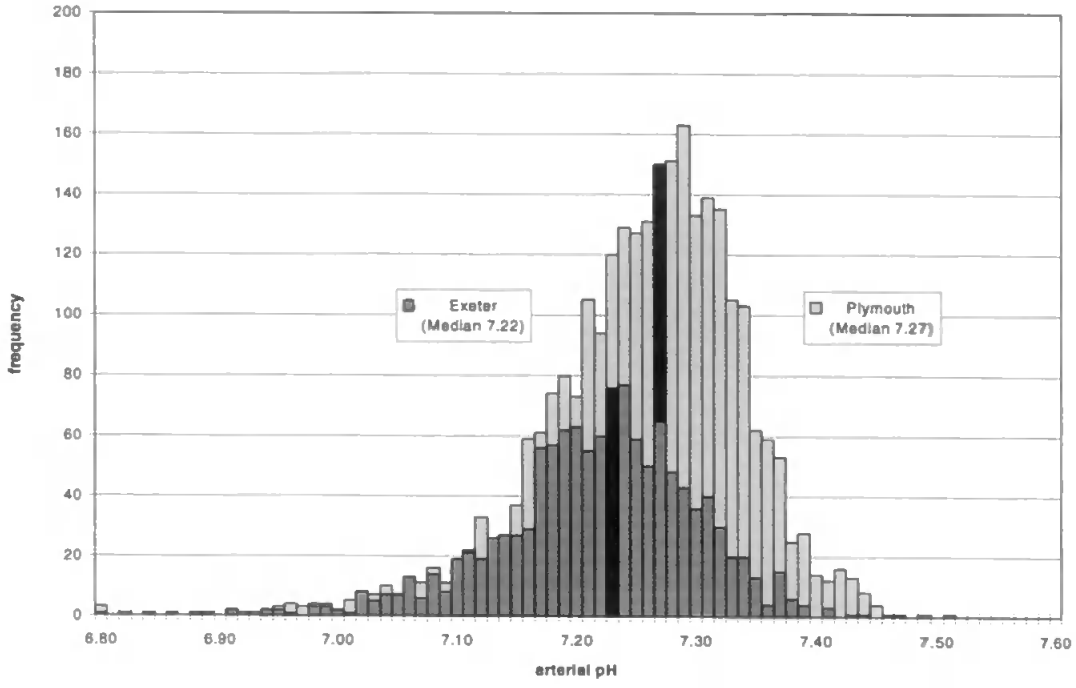


Figure 4.22: *distributions of arterial pH at Plymouth and Exeter*

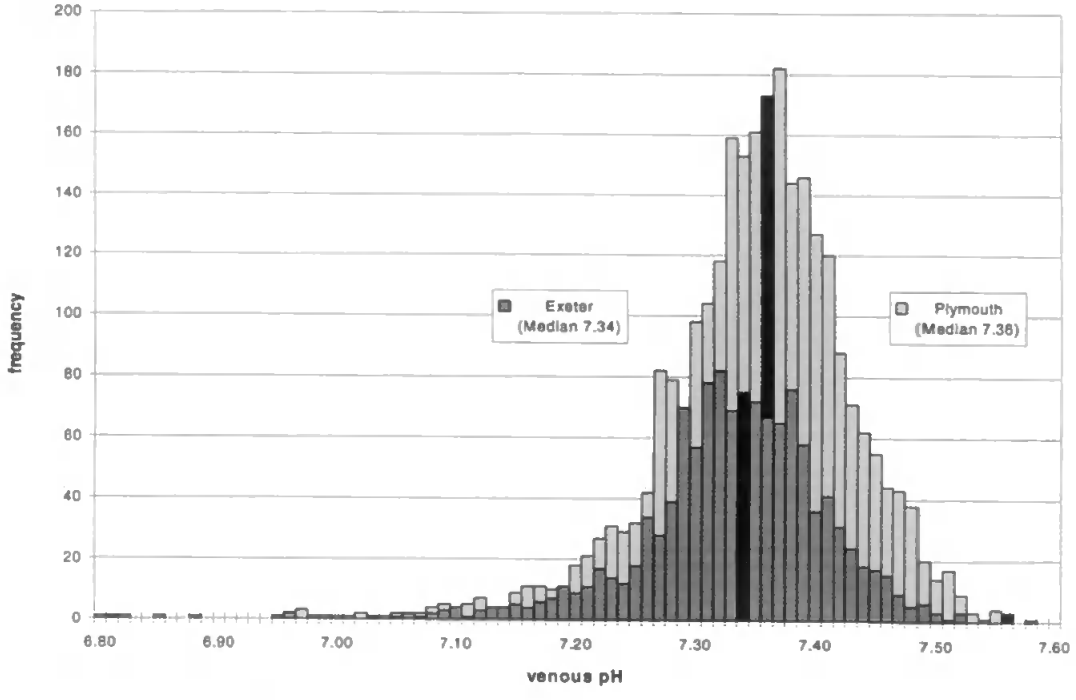


Figure 4.23: *distributions of venous pH at Plymouth and Exeter*

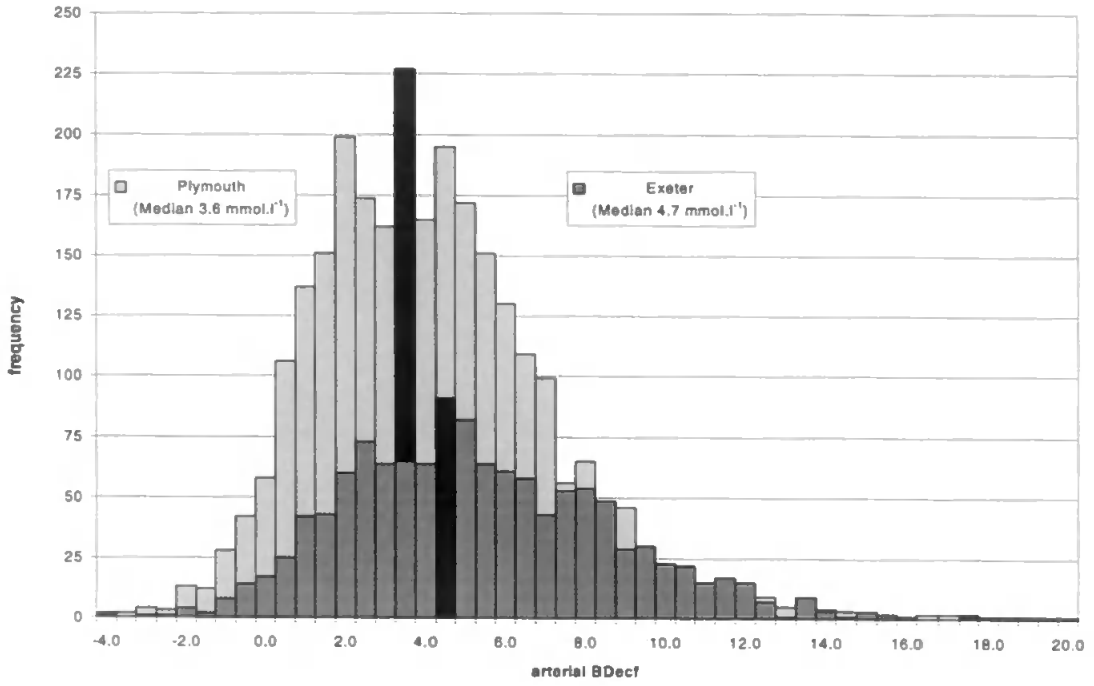


Figure 4.24: *distributions of arterial BD_{ecf} at Plymouth and Exeter*

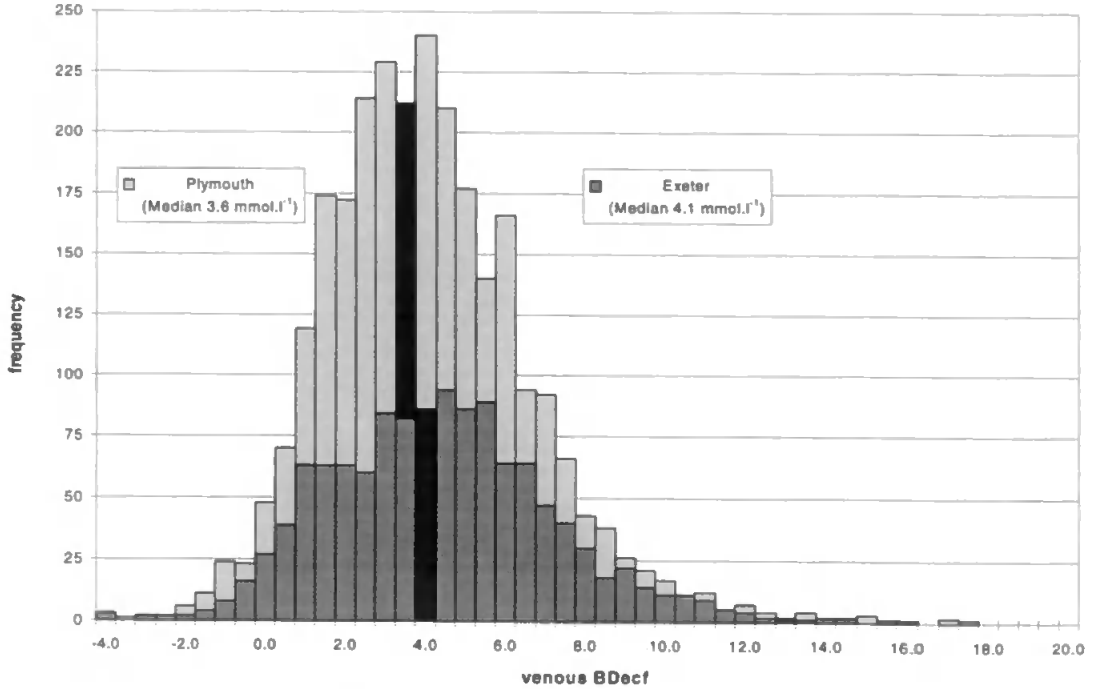


Figure 4.25: *distributions of venous BD_{ecf} at Plymouth and Exeter*

4.9 Umbilical Cord Acid-Base Charts

Siggaard-Andersen published an acid-base chart for adult and infant blood, in which the pH is plotted against the logarithm of pCO_2 such that lines of constant BD_{ecf} would appear as straight lines [115]. The chart also indicated ranges for various types of abnormal acid-base status, including *acute* and *chronic* forms of base deficit, in which the term *chronic* was defined to be over a period of 6 – 12 hours. Note that *adult* is used in this context to mean *non-fetal* or not in the immediate neonatal period.

The normal range for adult arterial acid-base was centred around a pH of 7.40, pCO_2 of 5.3 kPa (40 mmHg) and a BD_{ecf} of 0 $mmol.l^{-1}$. This is clearly not appropriate for the fetus at the end of labour, where the normal range for arterial pH is around 7.27 or below and the normal range for arterial pCO_2 is around 6.7 kPa (50 mmHg) or above. This corresponds to a BD_{ecf} of around 3.6 $mmol.l^{-1}$. Venous values for umbilical cord blood are nearer to the adult values, but still clearly different. In addition, there is probably no fetal equivalent of the *chronic base deficit* defined for adults in which the acids are not the result of anaerobic metabolism. These differences are illustrated in Figures 4.26 and 4.27, in which the values of the approximately 10000 arterial and venous results from paired vessels collected in Plymouth are superimposed onto the standard Siggaard-Andersen Acid-Base Chart.

It can be seen from these charts that there are no results in the positive base excess portion, and that the vast majority of results are outside the *normal range* defined for adults. Most lie in the range between *acute hypercapnia* (short term accumulation of pCO_2) equivalent to an acute respiratory acidosis and *acute base deficit* (short term accumulation of non-carbonic acid) equivalent to an acute metabolic acidosis. Again the term *acute* is defined for adults, and may not correspond to the clinical usage of the term in the perinatal period. Additionally, ten arterial points and four venous points lie off the top left-hand side of the charts. It is clear from these charts that the adult term *normal* is inappropriate for umbilical acid-base assessment: this reinforces the point that specific knowledge of fetal physiology is required for accurate interpretation of umbilical acid-base.

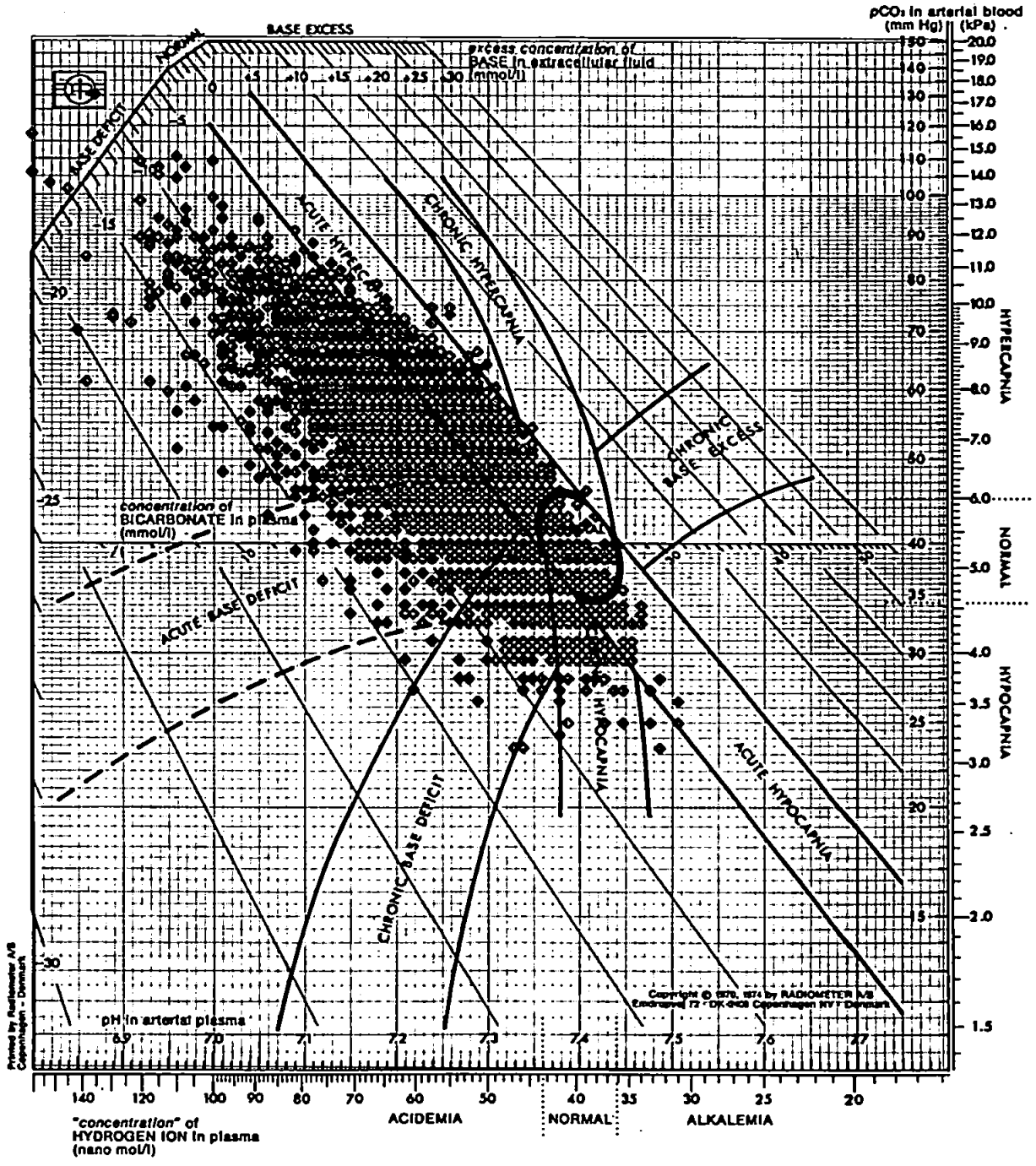


Figure 4.26: *umbilical arterial acid-base results superimposed onto the* Siggaard-Andersen Acid-Base Chart *for adults*

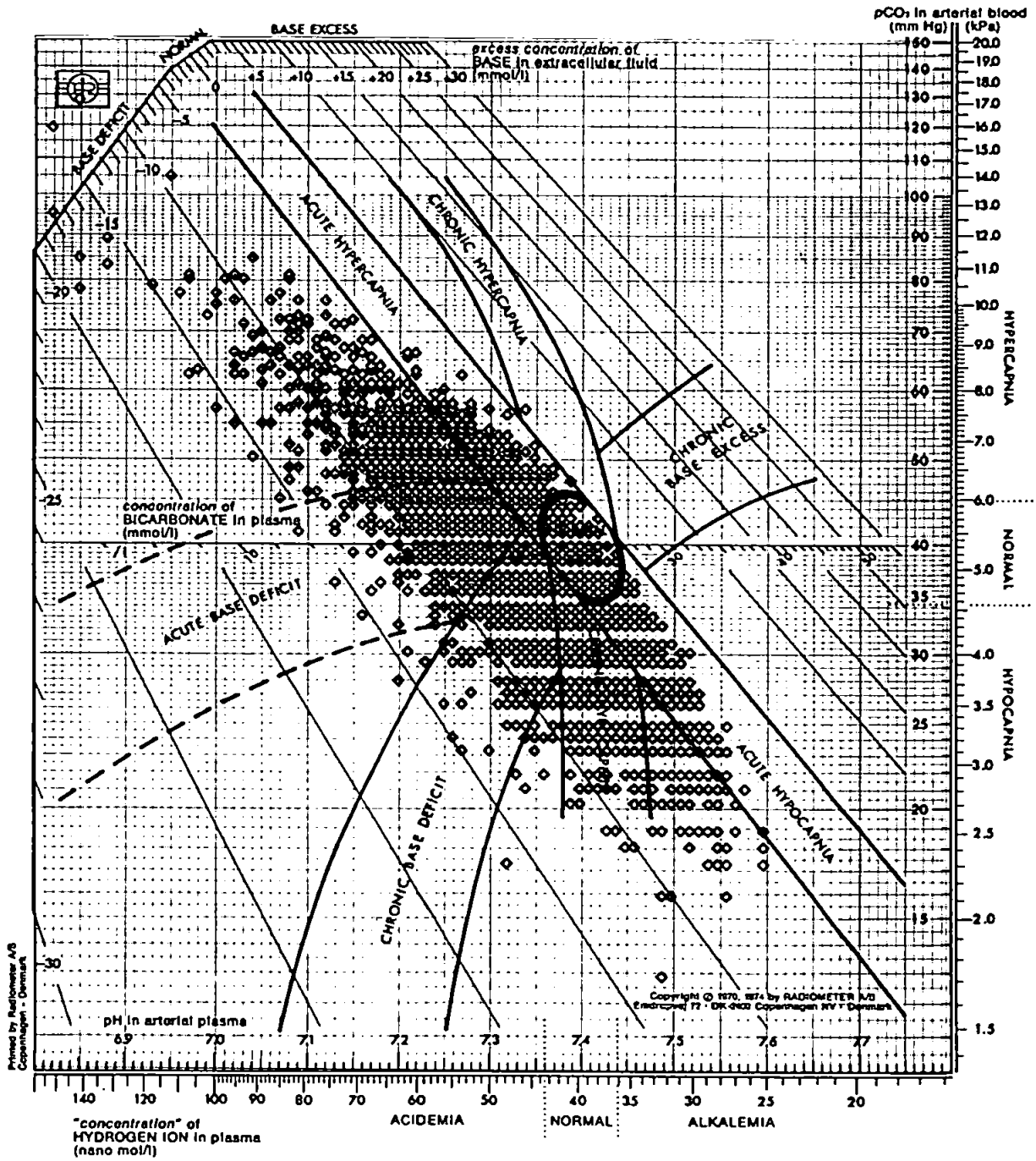


Figure 4.27: *umbilical venous acid-base results superimposed onto the* Siggaard-Andersen Acid-Base Chart *for adults*

4.10 Discussion

This chapter has detailed the various aspects of the evaluation of the crisp expert system. The evaluation process can be divided into two main sections:

1. evaluation required for commercial release, and
2. subsequent clinical assessment.

Although the evaluation process has been presented separately from the design and development process described in the previous chapter, it is important to emphasise that the two processes took place concurrently to a large extent.

Early on in the design phase, it was decided that the system would produce an interpretation of data *only*. This, it is believed, contributed to the relatively rapid transfer of the system from the research environment to commercial release. This was beneficial for two reasons:

1. it reduced the requirement for clinical assessment of the system, as discussed in Section 4.2, to little more than proof of clinical safety, and
2. it avoided the presentation of a threat to the clinicians, that a decision support system might.

Although clinicians continue to debate the necessity and utility of umbilical acid-base assessment, there has been wide recognition that stringent data validation is valuable if the procedure is to be undertaken. As data validation is viewed as a tedious task, the expert system is seen as a benefit. Questions have been raised on details of the interpretations offered, but as they are not presented as decisions, the clinicians feel happy to place their own interpretation on the data, if they disagree.

The system has been in live clinical use at Plymouth, Exeter and a number of other hospitals within the UK for several years. The system has been well accepted by the staff at Plymouth and Exeter. Indeed, the staff at Exeter found that the expert system considerably eased the introduction of routine cord blood gas analysis into a unit which previously had not performed

any, by providing an easily accessible output (a natural language interpretation in clinical terminology) whereas otherwise they would receive a set of numerical data. The extensive software testing and validation process, in conjunction with the overall structured project methodology, has greatly contributed to the software reliability and user-acceptance.

The system is now (July 1997) installed at twenty two hospitals around the UK, and as yet there have been no cases of software problems or maintenance required. In early 1994 the expert system was updated to connect to a new model blood gas analyser — 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 validation process. In terms of project time, at least as much time was occupied by the validation process as for the rest of the project (from specification through knowledge elicitation to coding) combined. Although the BS5750 requirements of the industrial partner initially seemed onerous and overly restrictive, in fact the effects of quality assurance on the project have been beneficial.

Most of the relationships among acid-base parameters that were found in Section 4.7 were broadly as expected physiologically, and are probably interrelated themselves. For example, increased gestational age implies increased birth weight implies increased length of labour; subsequent deliveries implies increased birth weight implies increased length of labour; male gender implies increased birth weight implies increased length of labour; and so on. Most of the relationships are complex and warrant possible future investigation with more sophisticated statistical methods or other methods such as neural networks. As most of the associations and correlations are weak (although statistically highly significant) it is dubious whether they have real clinical significance and thus need not be taken into account in the interpretation of umbilical acid-base status. As an aside, Figures 4.4 to 4.7 seem to support the notion that clinicians cannot accurately assign middle Apgar scores, as Apgars 3 to 6 showed no mean difference for *any* parameter. It is an interesting discovery that the male infant may have a significantly higher chance of having non-normal umbilical acid-base at delivery; this is possibly associated with the increased fetal heart rate in females, which would cause increased blood flow through the placenta with higher gaseous exchange.

There were significant differences in the population statistics between Plymouth and Exeter found in Section 4.8. This is in the context of two sets of identically validated results, collected on cross-calibrated machines. These results do not seem to be due to selective sampling and may be caused by a difference in practice in the two centres. Neither centre has sufficient other data to examine whether these acid-base differences are reflected in other outcome measures, such as long-term neurological development. Indeed, it would also be necessary to have far more cases in each centre before such a question could be adequately addressed. There was no obvious difference in possibly the two most important factors reflecting management of labour; caesarean section or epidural rates. The BD_{ecf} difference could be a result of longer second stages in Exeter as large arterial–venous BD_{ecf} differences are commonly the result of reduced umbilical blood flow in the second stage [102]. It is believed that the second stage has become shorter in Plymouth as a result of routine cord sampling, which commenced in March 1992. Unfortunately, data on duration of second stage was not available to examine this. The fact that the system has shown a significant difference between the two centres demonstrates its potential as an audit tool. It provides an exciting opportunity to explore the impact of change in clinical practice, both over time and between centres.

It is important to stress that these clinical studies have *not* fully assessed the expert system in the accepted sense. They simply demonstrate the potential of the system in terms of clinical comparisons of umbilical acid-base data. In order to fully assess the impact of the expert system, it is necessary to define how the information that the expert system provides is to be used, and then to design a clinical trial to measure the effect of this information. There are, in fact, four distinct main uses of the information:

1. individual feedback — the clinician(s) responsible for delivery use the acid-base information to assess their own management of a particular case retrospectively

2. group audit statistics — population data are used to compare changes over time or between centres as illustrated in the comparison of Plymouth and Exeter

3. neonatal guidance — paediatricians responsible for care of an infant in poor condi-

tion use the umbilical acid-base data as 'baseline' information to assess changes in condition or to guide feeding and resuscitation regimes

4. medicolegal protection — umbilical acid-base information is used to protect obstetricians from undeserved and unwarranted litigation (see Section 4.3.5)

Each of these uses implies different assessment criteria:

1. individual feedback — interviews with obstetric clinicians, performance data
2. group audit statistics — long term outcome variables, other perinatal data, large multi-centre trials
3. neonatal guidance — interviews with paediatric clinicians, neonatal data
4. medicolegal protection — records of legal cases or deflections

In conclusion, the crisp expert system was successfully evaluated to the requirements for commercial release. Full clinical assessment of the expert system, as with any other medical expert system, is a *much* harder problem, which would require the availability of appropriate data collected within a long term multi-centre randomised control trial. There is currently probably not sufficient political impetus or the necessary financial commitment for such a study to take place in this country.

Chapter 5

The Design and Optimisation of a Preliminary Fuzzy Expert System

5.1 Introduction

A number of problems were identified in the implementation of conventional crisp rules used in the initial system. The interpretation section of the crisp expert system utilised a number of rules of a form similar to:

> IF arterial pH < 7.05 AND arterial $BD_{ecf} \geq 12$ mmol.l^{-1}
>
> THEN severe arterial metabolic acidemia

Such rules feature sharp boundary cut-offs which are not representative of real decision making processes and do not employ any form of uncertainty representation in the conclusion to imply a less than certain diagnosis. An illustration of the problems caused by the rule boundaries encoded into the crisp expert system is shown in Table 5.1.

Arterial			*Venous*				
pH	pCO_2	BD_{ecf}	pH	pCO_2	BD_{ecf}	*ESN*	*Interpretation*
7.04	8.4	12.0	7.09	8.1	10.0	80	Severe metabolic acidemia
7.05	8.5	11.3	7.10	8.2	9.4	120	All results normal

Table 5.1: *illustration of the different crisp expert system interpretations obtained from two samples with minimal parameter differences*

It can be seen that a minimal change in the pH and pCO_2 parameters of each vessel can result in a dramatic change in the expert system categorisation of the results. It was felt that a fuzzy logic based expert system would offer more realistic and acceptable interpretation. In a fuzzy system, a rule such as above may be replaced by:

> IF arterial pH is low AND arterial BD_{ecf} is high
>
> THEN arterial acidemia is metabolic

The use of fuzzy logic allows for more gradual changes between categories and allows for a representation of certainty in the rule consequence through the ability to fire rules with varying strength dependent on the antecedents. Additionally, fuzzy logic can allow the results to be presented to clinicians in a more natural form. A preliminary investigation was performed to convert the crisp expert system directly into a fuzzy expert system (FES). The purposes of this study were:

- to determine the feasibility of converting the existing crisp rules into a set of fuzzy rules, without the necessity of additional expert knowledge elicitation sessions, and
- to investigate whether the fuzzy system would offer any improvement in performance over the crisp system in its interpretation of results.

Fuzzy logic and fuzzy set theory provide a good framework for managing uncertainty and imprecision in medicine [3, 24, 47] and have been successfully applied to a number of areas [5, 24, 128]. However, although the theoretical properties of fuzzy systems have been extensively investigated, the implementation of a fuzzy expert system in practice involves a great deal of pragmatic choices. This includes considerations for the type of inference methodology, rule set and fuzzy operators to determine an appropriate fuzzy model of the expertise for a particular application. Unfortunately, there is currently no theoretical method for determining the best model for a given domain, which is important in safety critical areas such as medicine. In this chapter, the principles of fuzzy modelling are discussed and a technique for the optimisation of fuzzy models by simulated annealing is described. The application of the technique to the tuning of a preliminary fuzzy expert system for umbilical acid-base analysis is then presented.

5.2 Principles of Fuzzy Modelling

The successful development of a fuzzy model for a particular application domain is a complex multi-step process, in which the designer is faced with a large number of alternative implementation strategies. The principal alternatives are in the selection of:

- inference methodology
- linguistic variables and fuzzy terms
- rule set
- fuzzy operators
- membership functions

The overall effect of a fuzzy system, viewed in terms of the resultant set of vectors representing the fuzzy output variables obtained from a set of vectors representing input variables, is governed by a combination of all these design choices. Even the simplest modification to the fuzzy system, such as the use of linear rather than non-linear membership functions, may alter the input-output mapping of the fuzzy model. In a real application there is the additional process of defuzzification to map the output fuzzy sets to the real world. Unfortunately, there is currently very little theoretical guidance as to which of the design choices are 'correct' for a particular domain.

5.2.1 Inference Methodology

A fuzzy inference method is a process by which a possibly imprecise conclusion is reached through the application of a set of fuzzy production rules to a set of input parameters. In forward-chaining crisp expert systems this is achieved by repeated application of the *Modus Ponens* rule of classical logic to the input propositions to generate a set of possible conclusions. In fuzzy systems the inference process is more problematic. Zadeh proposed the Compositional Rule of Inference (CRI) [139] as a method for modelling a single fuzzy implication 'IF X is A THEN Y is B', in which A and B are fuzzy sets, by considering implication

as the composition of a fuzzy relation R, representing $A \Rightarrow B$, with A. Although the CRI defines a method for performing a single fuzzy inference, it leaves the problem for combination of inference from a set of fuzzy rules.

Given a set of N fuzzy rules of the form 'IF X is A_i THEN Y is B_i', $i = 1 \ldots N$, there are two main approaches for performing combined inference on an observation A' [124]. Firstly, the CRI may be applied to each rule $i = 1 \ldots N$ to produce conclusion sets B'_i which are then combined to the single conclusion B' through the use of a combination operator; or secondly, by combining the relations R_i into a single relation R with a combination operator and then performing a single CRI of R with A' to form B'. In general, these two approaches do not give the same result. The result depends on the choice of composition, combination and implication operators. Once either of these two methods has been used, a defuzzification algorithm is usually applied to the fuzzy consequence set B' to produce a single value, b', in the universe of discourse of B.

The defuzzification process can be avoided through the use of fuzzy rules of the form 'IF X is A_i THEN Y is k_i', where k_i are constants, or more generally a linear combination of input variables x_i [119]. In this approach, the crisp input a' or fuzzy input set A' is resolved to a single membership value μ_{A_i} to produce the matching strength of the antecedent α_i. This is then combined with the output constant or linear combination and summed to give the crisp result $b' = \sum \alpha_i k_i$.

Even when the precise inference methodology has been chosen, there are further options available to refine the inference process. For example, although in most fuzzy expert systems every rule $i = 1 \ldots N$ is calculated and applied in parallel, rule precedence could be used to somehow select or prefer a subset of rules out of the total rule set.

5.2.2 Linguistic Variables and Fuzzy Terms

In a practical fuzzy expert system the linguistic variables to be used must be decided. A conventional knowledge elicitation process may reveal the natural linguistic variables used by the expert in a particular domain, or the input parameters may be determined by other factors

such as the data available from a certain machine. Similarly the linguistic output variables may be immediately apparent from the nature of the problem, but more often will need to be determined empirically based on the experts' knowledge and expectations. Additional internal linguistic variables may be introduced by the knowledge engineer in order to simplify the structure of rules or refine the inference processing.

The number of fuzzy terms in each of the linguistic variables will similarly be determined. It has been found that an odd number of fuzzy terms is usually preferable in fuzzy control type applications, but this heuristic is not necessarily true in other domains, such as medical expert systems. Finally, the universe of discourse of each of the linguistic variables will be established. This will usually be a trivial process for the input variables, once the number and form of the fuzzy terms in each variables' term-set has been settled, but will often be arbitrary for the output variables.

5.2.3 Rule Set

Once the knowledge elicitation process has determined the linguistic variables and their component fuzzy terms, a set of fuzzy rules must be constructed. This, again, is a non-trivial process, which may be made easier through the availability of a working set of rules elicited from experts within the domain. Often the rules must be constructed by the knowledge engineer, or generated automatically by one of the techniques of fuzzy rule induction. Whichever methods of generating the rules are used, the overall structure and the exact number, form and content of each must be determined. In general, two rules of the form 'IF X_1 is A_1 THEN Y is B, and 'IF X_2 is A_2 THEN Y is B', will give different results to the combined rule 'IF X_1 is A_1 OR X_2 is A_2 THEN Y is B' [124].

It is generally accepted that the use of linguistic hedges such as *very* or *slightly* should be avoided in the initial rule set, unless absolutely necessary, but in the later stages of tuning the fuzzy model such hedges may be introduced to refine the effects of specific rules. Hedges are frequently implemented as *power hedges*, where the membership function is raised to a certain power to form the hedged term; *very* is usually taken as the square of the mem-

bership function (μ^2), and *slightly* is taken as the square-root ($\mu^{1/2}$). Hedges may also be implemented as *shift hedges*, in which the entire membership function is shifted to the left or right by a (usually) constant offset. However, a wide variety of other general purpose hedge functions have been proposed to correspond to these and other hedges, and functions specific to a particular domain may also be more appropriate.

5.2.4 Fuzzy Operators

Many authors have examined the theoretical properties of the principal *complementation*, *union* and *intersection* operators on fuzzy sets. Although the principal axioms that these operators must obey have been established [65], there are still many functions that do satisfy these axioms. Indeed, several families of operator functions have been proposed that take an infinitely tunable parameter to alter the strength of the function whilst obeying the theoretical axioms, such as the Yager class [136] or Dubois & Prade [30] class of unions and intersections.

Once the choice of *complementation*, *union* and *intersection* operators has been made, there are a number of ways to choose a theoretically sound implication operator. In classical propositional logic the implication operator $(a \Rightarrow b)$ can be shown to be equivalent to $(\neg a \vee b)$, $\neg(a \wedge \neg b)$ and $\big(\neg a \vee (a \wedge b)\big)$. Dubois & Prade [31] have shown how various fuzzy implication operators can be derived in a consistent manner from the choice of the other fuzzy operators using these equivalences and others. Bandler & Kohout [12] and Whalen & Schott [131] have also investigated the theoretical and practical properties of some of the different implication operators and have concluded that different implication operators suit different application areas. Mamdani [82] adopted a slightly less rigorous approach to the problem by using the min operator for fuzzy implication. Although theoretical considerations prevent $\min(a,b)$ from being a valid implication operator, the performance of the Mamdani model has made it a popular choice for fuzzy controllers.

5.2.5 Membership Functions

The final part of constructing a fuzzy model is the precise location and shape of the membership functions for each of the fuzzy terms used. Assuming that the number of fuzzy terms in each linguistic variable has already been decided, it is simply a matter of deciding where in the universe of discourse each crossover point should be, and what degree of overlap should be allowed for each set. Various shapes of membership function have been proposed and used in practice, with the most popular choices being *triangular*, *sigmoid* or *gaussian*, but again many other choices are also possible. Triangular functions are often used in control applications, but human experts sometimes find them difficult to relate to (assuming they have been convinced of the fuzzy approach). In general, a membership function may have almost any shape, not necessarily modelled by any simple function, but such arbitrary defined functions are rarely found in practical applications.

In most applications the membership functions are always defined to be normal ($\mu_{\max} = 1$), or at least if sigmoid membership functions are used, then asymptotically normal. Indeed, the assumption of normal term-sets is widely used in both theoretical and practical work on fuzzy sets. In general, however, there is no absolute requirement for normality in fuzzy sets, even when they appear in the term-sets of linguistic variables used in a fuzzy expert system. In the model used in this chapter, the requirement for normality is relaxed and the height of fuzzy sets is an important adjustable parameter.

Membership functions for practical fuzzy systems are often determined through the knowledge elicitation process with human experts, or are determined empirically by altering the shape and location until the fuzzy system performs to an adequate level, appropriate to each application. There is no theoretical reason why membership functions should not vary between individual experts, vary by context, or vary over time as a result of experience.

5.3 N-Dimensional Optimisation

As can be seen from the above discussion, the formulation of a fuzzy model appropriate for a particular domain involves a large number of choices, many of which have no theoretical basis. In general, then, the formulation of an optimal fuzzy model in terms of its performance at a given task in the particular domain is a problem of N-dimensional non-linear optimisation, in which N is very large even for the most trivial of fuzzy systems.

A number of approaches have been proposed for the development, tuning and optimisation of fuzzy models. Many of these have been based on the integration of neural networks with fuzzy logic [76, 72], or hybrid neuro-fuzzy clustering methods such as F-ARTMAP [18] or ANFIS [58]. More general algorithmic approaches have included the Σ-PAFIO algorithm for model optimisation [126, 96], a framework for synthesising fuzzy rules [103] and *Genetic Algorithm* based optimisation methods [95]. The well known *Simulated Annealing* algorithm [61] has been applied to the tuning of fuzzy neurons [68], the combinatorial optimisation of fuzzy partitioning [11] and for multi-variable optimisation of a fuzzy relational model-based controller [116]. In this chapter, *Simulated Annealing* with an adaptation for continuous minimisation by the Simplex method [98] is applied in a novel manner to tuning of the fuzzy model. This general purpose algorithm can be used for both combinatorial and continuous parameter optimisation, and thus is suitable for both structure identification and parameter estimation of the model.

5.4 Development of an Optimisation Technique

5.4.1 Simulated Annealing

The *simulated annealing* algorithm is a general purpose algorithm for performing approximate optimisation in large dimensional problems [61]. It is generally useful for combinatorial optimisation problems, and/or problems where derivatives of the cost function being optimised are not available. The algorithm takes its name from the fact that it simulates

the physical process of annealing, in which atoms in a liquid freeze to form a solid in the minimal energy state by cooling at sufficiently slow rate that thermal equilibrium is reached at each temperature. Thermal equilibrium is characterised by the *Boltzmann distribution* in which the probability of being in a state i with energy E_i at temperature T is given by:

$$\mathbf{P}(i) \propto \exp\left(\frac{-E_i}{kT}\right) \tag{5.1}$$

where k is a physical constant known as the *Boltzmann constant* [83]. The simulated annealing algorithm views the cost function, $f(x)$, being minimised as equivalent to the energy state of a physical system (k is taken as 1), and the process of reaching equilibrium is equivalent to repeatedly accepting or rejecting changes in energy state from i to j according to the probability:

$$\mathbf{P}(\text{accept change to state } j \text{ from } i) = \begin{cases} \exp\left(\frac{f(i)-f(j)}{T}\right) & \text{if } f(j) > f(i) \\ 1 & \text{if } f(j) \leq f(i) \end{cases} \tag{5.2}$$

The algorithm depends on two control parameters; the temperature, T, and the number of state changes, L, considered at each temperature. An initial state is generated and then L state changes are considered according to the criterion in equation 5.2. The temperature T is then reduced according to a *cooling schedule* and the process is repeated until an appropriate stopping criterion is reached. Typical stopping criterion include; when the cost function is minimised to a certain value, when the changes in energy being considered are sufficiently small or when the temperature is sufficiently low. Theoretical convergence properties and overall efficiency of the algorithm depend on the cooling schedule chosen — it has been shown that the algorithm is guaranteed to optimise $f(x)$ in infinite time [1]. In practice appropriate values of T, L and the cooling schedule can be determined experimentally to return near-optimal solutions.

The Discrete Simulated Annealing Algorithm (Algorithm A)

Let $\mathbf{x}$ be a vector of m discrete variables, $\mathbf{x} = (x_1, x_2, \ldots, x_m)$, where each $x_i \in X_i$, the set of all possible values of the i^{th} discrete variable. Let $f(\mathbf{x})$ be the cost function being optimised (minimised).

A.1) Let $\mathbf{x}_o$ be an arbitrary initial (either random or specified) starting point, let T be the initial temperature, and let L_d be the initial number of repetitions;

A.2) Repeat L_d times;

(a) Generate a trial point $\mathbf{x}^*$ from the set of all possible solutions;

(b) If $f(\mathbf{x}^*) \leq f(\mathbf{x}_o)$, set $\mathbf{x}_o = \mathbf{x}^*$;

(c) If $f(\mathbf{x}^*) > f(\mathbf{x}_o)$, then if $\exp\left(\frac{f(\mathbf{x}_o)-f(\mathbf{x}^*)}{T}\right) > \text{random}[0,1)$, set $\mathbf{x}_o = \mathbf{x}^*$;

A.3) If *stopping criterion*, stop;

A.4) Decrease temperature T and adjust repetitions L_d according to annealing schedule, goto step 2.

5.4.2 Continuous Minimisation by the Simplex Method of Simulated Annealing

The simulated annealing algorithm implements a local 'hill-climbing' / 'valley-descent' algorithm in which downhill changes are always accepted and uphill changes are accepted initially, but with decreasing probability as the algorithm progresses. The technique has been applied most often to discrete combinatorial optimisation problems, but it has also been applied to problems with continuous variables. With continuous variables it is necessary to minimise a function $f(\mathbf{y})$, where $\mathbf{y}$ is an n-dimensional vector in $\Re^n$ with components $(y_1, y_2, \ldots, y_n)$. The change in state is taken as a move in hyperspace to $\mathbf{y} + \Delta\mathbf{y}$, so that the change in energy is $f(\mathbf{y}+\Delta\mathbf{y}) - f(\mathbf{y})$. A number of methods by which to choose the direction and length of the step $\Delta\mathbf{y}$ have been proposed [127, 16, 25], including a modification of the downhill simplex method of Nelder and Mead [88].

A simplex is a geometrical figure of $n+1$ vertices in n-dimensional space, for example a triangle in 2-dimensional space or a tetrahedron in 3-dimensional space. The simplex is initialised with a set of $n+1$ points, and the function is evaluated at each of these points. The simplex then makes one of the following steps:

1. a reflection of the vertex with the highest function evaluation through the opposite face to a point which conserves the 'volume' of the simplex,

2. a reflection of the highest vertex through the opposite face to a point which increases the 'volume' of the simplex,

3. a contraction of the highest vertex towards the opposite face of the simplex, or

4. a contraction along all faces towards the lowest vertex of the simplex.

The simplex algorithm specifies the series of these steps to take, in such a way that the simplex converges to a (local) minimum. In the simulated annealing adaptation of the simplex method [98] a positive, logarithmically distributed random variable proportional to the temperature T is added to the function evaluation at each vertex of the simplex when determining the highest and lowest points so that:

$$g(\mathbf{y}) = f(\mathbf{y}) - T \ln\left(\text{random}(0,1)\right) \tag{5.3}$$

and a similar value is subtracted from the function evaluation of each new point that is tried as a replacement point. Thus a downhill step is always accepted, and an uphill step is accepted with probability $\propto \exp(-\textit{uphill step-size}/T)$. In the limit $T \to 0$, this algorithm reduces to the downhill simplex and converges to a local minimum. As with any form of the simulated annealing algorithm the choice of initial temperature, step length (simplex size) and cooling schedule will determine whether the global minimum will be found successfully, and these parameters will usually be found by trial and error.

A problem occurs with the simplex algorithm when $\mathbf{y}$ is bounded, especially when the global function minimum is located on or very near to the boundary (as is often the case in continuous variables used in fuzzy models). The standard procedure when not using the simplex adaptation [16] is to simply reject any step that generates $\mathbf{y}$ outside its bounds. However, if the simplex adaptation is used and vertices that lie outside the bounds are rejected, then a global minimum that lies on a boundary will never be enclosed by the simplex and thus will not be successfully located. To overcome this problem, if a $\mathbf{y}$ is generated with at least one

y_i outside a bound y_i' a penalty term can be added, so that:

$$g(\mathbf{y}) = g(\mathbf{y}') + \gamma\left(\sum_{i=1}^{n}(y_i - y_i')^2\right) \tag{5.4}$$

for each y_i outside a bound y_i', where $g(\mathbf{y}')$ is the function defined by equation 5.3 evaluated at the boundary.

The Continuous Simulated Annealing Simplex Algorithm (Algorithm B)

Let $\mathbf{y}$ be a vector of n continuous variables, $\mathbf{y} = (\, y_1, y_2, \ldots, y_n \,)$, where each $y_i \in \Re$ with bounds $[y_i', y_i'']$. Let $f(\mathbf{y})$ be the cost function being optimised (minimised), let $\mathbf{y}_1$, $\mathbf{y}_2, \ldots, \mathbf{y}_n, \mathbf{y}_{n+1}$ be $n+1$ vectors representing a simplex in the n-dimensional space. Let $F_j = f(\mathbf{y}_j)$, $G_j = f(\mathbf{y}_j) - T \ln(\text{random}(0,1))$ (F_j increased by a thermal fluctuation factor), and $H_j = f(\mathbf{y}_j) + T \ln(\text{random}(0,1))$ (F_j decreased by a thermal fluctuation factor). If any $\mathbf{y}$ is generated with y_i outside bound y_i' or y_i'', use Equation 5.4 to evaluate the function.

B.1) Let $\mathbf{y}_o$ be an arbitrary initial (either random or specified) starting point, let T be the initial temperature, and let L_c be the initial number of repetitions;

B.2) Generate the initial simplex: let $\mathbf{y}_i = \mathbf{y}_o + \lambda \mathbf{e}_i$ for $i = 1 \ldots n$, where the $\mathbf{e}_i$'s are n unit vectors, and λ is a constant corresponding to the characteristic length of the problem, and let $\mathbf{y}_{n+1} = \mathbf{y}_o$;

B.3) Evaluate the $n+1$ cost functions, $F_j = f(\mathbf{y}_j)$ for $j = 1 \ldots n+1$;

B.4) Repeat L_c times;

(a) Sort simplex points in order from best to worse (with a thermal fluctuation), such that $G_1 \leq G_2 \leq \cdots \leq G_n \leq G_{n+1}$;

(b) Calculate the centroid of n best points in $n+1$-dimensional space,

$$\mathbf{c} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{y}_j;$$

(c) Reflect the worst point through the 'facing' centroid point, $\mathbf{y}_r = \mathbf{c} + (\mathbf{c} - \mathbf{y}_{n+1})$;

(d) If $H_r < F_1$: expand the reflected new best point, $\mathbf{y}_e = \mathbf{c} + 2(\mathbf{c} - \mathbf{y}_{n+1})$, then if $H_e < H_r$, set $\mathbf{y}_{n+1} = \mathbf{y}_e$, else set $\mathbf{y}_{n+1} = \mathbf{y}_r$;

(e) If $F_1 \leq H_r \leq F_n$: set $\mathbf{y}_{n+1} = \mathbf{y}_r$;

(f) If $F_n < H_r < F_{n+1}$: contract the reflected simplex, $\mathbf{y}_c = \mathbf{c} + \frac{1}{2}(\mathbf{y}_r - \mathbf{c})$;

(g) If $H_r \geq F_{n+1}$: contract the original simplex, $\mathbf{y}_c = \mathbf{c} + \frac{1}{2}(\mathbf{y}_{n+1} - \mathbf{c})$;

(h) If (and only if) a contraction was attempted: If $H_c < \min(H_r, F_{n+1})$, then the contraction succeeded, set $\mathbf{y}_{n+1} = \mathbf{y}_c$, else contract all points of the simplex towards the best point, set $\mathbf{y}_j = \mathbf{y}_1 + \frac{1}{2}(\mathbf{y}_j - \mathbf{y}_1)$, for $j = 2 \ldots n+1$;

B.5) If *stopping criterion*, stop;

B.6) Decrease temperature T and adjust repetitions L_c according to annealing schedule, goto step 4.

5.4.3 Simulated Annealing Optimisation of a Fuzzy Model

A fuzzy model may be viewed as a N-dimensional vector characterising a fuzzy expert system, with a set of m discrete variables and n continuous variables. Thus:

$$f(\mathbf{x}, \mathbf{y}) = f(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n) \tag{5.5}$$

where each x_i is a discrete variable and each y_j is a continuous variable, as above. Discrete variables represent structural choices such as which rules from the total set of possible fuzzy rules to use, or which of the set of operator families to implement; continuous variables represent parameters such as modal values or spreads of membership values, or the tunable parameter, w, in the Yager class of operators [136].

If a cost function can be found to evaluate the performance of a fuzzy model with each set of x_i and y_j then the simulated annealing (SA) algorithm can be applied to optimise the model. If no continuous variables are used, then the discrete SA algorithm (Section 5.4.1) may be used alone, and if no discrete variables are used, then the continuous SA algorithm (Section 5.4.2) may be used alone. However, if both discrete and continuous parameters are

to be optimised simultaneously, then a slightly more sophisticated SA must be used. In this case a decision must be made to either generate a new state through a change in the discrete variables, or via a simplex move in the continuous variables. This decision can be made simply as a random function of m and n (the number of discrete and continuous variables, respectively) or through a more sophisticated criteria dependent on $\Delta f(\Delta \mathbf{x}, \mathbf{y})$ and $\Delta f(\mathbf{x}, \Delta \mathbf{y})$; that is the relative changes in the cost function that result from changes in only the discrete variables or only the continuous variables.

Simulated Annealing Fuzzy Tuning Algorithm (Algorithm C)

Let $\mathbf{x}$ be a vector of m discrete variables, $\mathbf{x} = (x_1, x_2, \ldots, x_m)$, and let $\mathbf{y}$ be a vector of n continuous variables, $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. Let $f(\mathbf{x}, \mathbf{y})$ represent the cost function being minimised, obtained by running the fuzzy model parameterised by $\mathbf{x}$ and $\mathbf{y}$ on K sets of input data and evaluating the K sets of output data against required output (for example, through the use of an agreement measure as described in Section 5.4.4).

C.1) Collect K samples of training data with inputs and desired output;

C.2) Select the m discrete variables, $x_1, \ldots, x_m$, used to construct the fuzzy model;

C.3) Select the n continuous parameters, $y_1, \ldots, y_n$, used to adjust the fuzzy model;

C.4) Initialise the fuzzy model: let $\mathbf{x}_o$ be the initial starting configuration of discrete variables, let $\mathbf{y}_o$ be the initial starting configuration of continuous parameters, let T be the initial annealing temperature, let L_d be the initial number of discrete repetitions, and let L_c be the initial number of continuous (simplex) repetitions;

C.5) Initialise the simplex for continuous optimisation as in step 2 to 3 of Algorithm B;

C.6) Decide whether to make discrete or continuous adjustment to the fuzzy model;

C.7) If discrete adjustment, perform L_d discrete annealing repetitions to tune $\mathbf{x}$ in $f(\mathbf{x}, \mathbf{y})$, as in step 2 of Algorithm A;

C.8) If continuous adjustment, perform L_c simplex annealing repetitions to tune $\mathbf{y}$ in $f(\mathbf{x}, \mathbf{y})$, as in step 4 of Algorithm B;

C.9) If *stopping criterion*, stop;

C.10) Decrease temperature T and adjust repetitions L_d and L_c according to annealing schedule, goto step 6.

The selection of *stopping criterion* is a difficult problem. If the optimal value of the cost function, f_{opt}, is known *a priori*, and given an infinite amount of time such that the cooling schedule can be sufficiently slow, then the algorithm can be terminated when $f - f_{opt} < \varepsilon$, an arbitrarily low tolerance. In practice f_{opt} is usually unknown and a finite cooling schedule must obviously be used. The choice of reasonable cooling schedule and appropriate stopping criterion is the hardest part of successfully implementing the tuning algorithm, and the solution will be problem dependent. Theoretical issues have been investigated [1] and practical strategies have been developed [1, 25, 98]. For fuzzy model tuning, a relatively simple strategy may be implemented, such as:

- reduce T to ωT, where $0 < \omega < 1$, terminate when $T \leq T_0$; or
- reduce T to ωT, terminate when $\Delta f < \varepsilon$, i.e. when the improvement obtained in f in an iteration is sufficiently small.

5.4.4 Obtaining a Cost Function

In general *any* cost function can be used: it does not need to be smooth or even continuous, but it must be bounded [25]. In practice a measure of mean square error between actual and some desired target output would be appropriate to tuning a fuzzy model. Usually, the output of a fuzzy system will be arbitrarily scaled, making an absolute comparison between the fuzzy system output and target output difficult. In such a situation, a statistic which ignores the absolute values, but takes into account the ordering of results, may be appropriate. Spearman rank order correlation can be used to determine the degree of association between two sets of rank-ordered data. It is defined as:

$$\rho_s = \frac{\sum_{i=1}^{K}(r_i - \bar{r})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^{K}(r_i - \bar{r})^2 \sum_{i=1}^{K}(s_i - \bar{s})^2}} \tag{5.6}$$

where r_i and s_i are the two sets of rankings being compared, and K is the number of ranks [113]. If there are no ties in the rankings, then Equation 5.6 can be simplified to give:

$$\rho_s = 1 - \frac{6\sum_{i=1}^{K}(r_i - s_i)^2}{K(K^2-1)} \tag{5.7}$$

The ρ_s statistic will be $+1$ if the two rankings are identical (complete agreement) and -1 if the rankings are opposite (complete disagreement). The significance of the statistic can easily be obtained by:

$$z = \rho_s\sqrt{K-1} \qquad \text{or} \qquad t = \rho_s\sqrt{\frac{K-2}{1-\rho_s{}^2}}, \text{ with } df = K-2$$

if K is larger than about 20. If r_i are the desired rankings that have been determined for a particular set of input parameters to a fuzzy expert system and s_i are the actual rankings obtained from the fuzzy system, then $f = 1 - \rho_s$ is a suitable cost function to be minimised by SA. Note that this is effectively the same as minimising the mean square error between the desired rankings and the obtained rankings.

In practice a target set of rankings may be obtained in a number of ways. In many cases an objective *correct* ordering of a set of results may be available, but for situations where such a verified ordering is not available, a target set may be obtained by seeking the opinions of as many human domain experts as is feasible. If k different expert opinions are obtained, then the inter-expert agreement can be calculated from their rankings to ensure that there is an acceptable body of opinion. The pairwise average rankings may thus be calculated to be used as the target set for SA of the fuzzy expert system, or the individual rankings may be kept separate to allow the fuzzy system to be trained to certain experts. Note that if the average correlation between all pairs is denoted ave(ρ_s), then ave(ρ_s) can only range from $-1/(k-1)$ to $+1$, as it is not possible for k experts to all totally disagree with each other [113]. In this case the significance of ave(ρ_s) must be tested by:

$$\chi^2 = (K-1)\left((k-1)\,\text{ave}(\rho_s) + 1\right), \text{ with } df = K-1.$$

5.4.5 Aside: Rank Order Statistics in MATLAB

Rank order statistics, such as the Spearman rank order correlation coefficient described above, require the the data to be ranked in order of magnitude of the original observation. For example, the observation with the lowest value gets a rank of 1, the second lowest gets a rank of 2, etc. If two or more observations are tied (have the same magnitude), then they are assigned the *average* of the ranks that would have been assigned had the observations differed slightly. Thus, for example, if two observations are both second lowest, they each are assigned a rank of 2.5, instead of receiving ranks 2 and 3. As rank order correlations were to be used extensively in this project, a method for calculating ranks was desired. It was found that the *statistics* toolbox in MATLAB did not include a built-in function for calculating and assigning such ranks. Note that this non-parametric statistical use of *rank* is entirely distinct from the *rank* of a matrix (the order of the largest non-zero determinant that can be formed from the elements of the given matrix [118]), for which MATLAB does have a built-in function.

Initially a `sort` function followed by a `for` loop to search for tied values was utilised to implement a ranking function, but it was found that this was extremely slow for large data sets, such as the 9 000 cases shown in Table 5.11, (Section 5.7). Consequently, a MATLAB function utilising only primitive array functions was created, which performed the ranking in a fraction of the time of the 'brute-force' algorithm. This function, listed in Appendix J with an illustration of its use, was submitted to the official *MathWorks* MATLAB world-wide-web site (`http://www.mathworks.com`). This function was subsequently released within the statistics section of user-supplied m-files under the name `ranks.m`, currently from page (`http://www.mathworks.com/statv4.html`).

5.5 Formulation of the Initial Fuzzy Model

The crisp expert system is a forward-chaining classification system, which (after the validation phase) is based on four main input variables; arterial pH (pH_A), arterial base deficit

(BD_A), venous pH (pH_V), and venous base deficit (BD_V). It was decided to restrict the initial fuzzy expert system only to the interpretation of true paired samples (samples verified as being an arterial and venous pair with validated pH_A, BD_A, pH_V and BD_V parameters) as these rules represented a self-contained subset of the crisp system. There were 21 such crisp rules operating on the four input parameters which needed conversion into equivalent fuzzy rules. The initial choices were as follows.

5.5.1 Inference Methodology

As this restricted application followed the conventional fuzzy control application in having crisp input variables and no internal fuzzy chaining, the Mamdani model of inference was used. All fuzzy rules were of the form 'IF X_1 is A_1 THEN Y_1 is B_1', where A and B are fuzzy sets. The min operator was used for implication throughout. Although the production of complex fuzzy outputs with shape information that could be used to infer a confidence in the output was one of the original reasons for implementing a FES, it was necessary to obtain crisp outputs for the purposes of evaluation of the fuzzy model. As an arbitrary choice, centre-of-gravity (centroid) defuzzification was used to produce crisp values on an arbitrary scale of 0 to 100 for each fuzzy output variable.

5.5.2 Linguistic Variables and Fuzzy Terms

Each of the four input parameters was assigned a linguistic variable and examination of the crisp rules showed that each could naturally be divided into three fuzzy terms, corresponding to meanings of *low*, *medium* and *high*. Two output fuzzy variables were used, severity of acidemia (*acidemia*) and duration of acidemia (*duration*). From the crisp rules it was determined that the acidemia variable had five terms in its term-set: *severe*, *moderate*, *significant*, *mild* (non-significant) and *none*; and that the duration variable had three terms: *chronic*, *intermediate* and *acute*.

5.5.3 Rule Set

The rules for the crisp expert system were obtained as a result of knowledge elicitation sessions with several leading clinicians skilled in umbilical cord blood acid-base analysis, and had been carefully refined to form a complete and consistent set of classifiers. As the crisp system had reached clinical implementation, the rule set was taken to be acceptable and was therefore recoded directly into fuzzy equivalents: the rules are detailed in Appendix E. The *fairly* (or *somewhat*) hedge was taken as the square-root operator and the *very* hedge was taken as the square operator.

5.5.4 Fuzzy Operators

The probabilistic family of operators was chosen, in which conjunction is defined as $(a*b)$, disjunction as $(a+b-a*b)$ and negation as $1-a$. The reason for this was that all the fuzzy rules feature conjunction of the four linguistic variables, like:

> IF pH_A is *low* AND BD_A is *high* AND pH_V is *low* AND BD_V is *high*
> THEN *acidemia* is *severe*

If the min operator is used for conjunction, the overall truth value of such a rule will obviously be determined solely by its lowest membership. For example, if the rule above is presented with the values shown in Table 5.2, the min operator would produce the same overall truth for both cases, despite the fact that case B is evidently (to any clinician) worse than case A, as the arterial pH is lower and both the base deficits are higher. The probabilistic operator mimics the reasoning of the clinician in considering all the parameters, so that the overall truth for B is higher.

Case	pH_A	μ_1	BD_A	μ_2	pH_V	μ_3	BD_V	μ_4	$\min(\mu_1,\mu_2,\mu_3,\mu_4)$	$(\mu_1*\mu_2*\mu_3*\mu_4)$
A	7.05	0.5	11.5	0.9	7.10	0.5	9.5	0.9	0.50	0.20
B	6.95	1.0	15.0	1.0	7.10	0.5	11.0	1.0	0.50	0.50

Table 5.2: *an example to justify the use of the probabilistic conjunction operator* $(a*b)$, *rather than the minimum conjunction operator* $\min(a,b)$

5.5.5 Membership Functions

If the membership of any term plateaus at 1.0, then the same effect as shown in Table 5.2 can be demonstrated. The simplest way to avoid this problem is to have *sigmoid* memberships that approach 1.0 asymptotically. All fuzzy sets were modelled with *sigmoid* membership functions — left-edge sets were modelled by a decreasing *sigmoid*:

$$\mu(x) = \frac{1}{1 + e^{(x-m)/\alpha\sigma}} \tag{5.8}$$

right-edge sets were modelled by an increasing *sigmoid*:

$$\mu(x) = \frac{1}{1 + e^{(m-x)/\alpha\sigma}} \tag{5.9}$$

and middle sets were modelled by a combination of two *sigmoids*:

$$\mu(x) = \frac{1}{\left(1 + e^{(x-m-\sigma/2)/\alpha\sigma}\right)\left(1 + e^{(m-\sigma/2-x)/\alpha\sigma}\right)} \tag{5.10}$$

where m is the location of the $\mu = 0.5$ value of the *sigmoid* for left and right sets, or the location of the centre value for middle sets; σ is the width parameter of the *sigmoid*, corresponding to the width at $\mu = 0.5$ for middle sets; and α is a constant of ≈ 0.1 governing the slope of the *sigmoids* such that the maximum value of a middle set is near to unity.

These four fuzzy input variables had the position and width of their terms determined by the cut-offs encoded into the crisp rules. Thus, for example, arterial pH fuzzy variable had its transition from *low* to *medium* at 7.05 as this value was used throughout the crisp rules. The term-sets for each fuzzy input variable were generated from Equations 5.8, 5.9, and 5.10 using the universe of discourse and parameters shown in Table 5.3, and are shown in Fig. 5.1 and Fig. 5.2. The output variables were also modelled with *sigmoid* membership functions, with the base variable and cross over of each term determined arbitrarily on a universe of discourse of 0...100. The term-sets for the output variables are shown in Fig. 5.3.

Variable	*Universe*	m_{low}	σ_{low}	m_{mid}	σ_{mid}	m_{high}	σ_{high}
pH_A	6.60...7.60	7.05	0.15	7.10	0.10	7.15	0.15
BD_A	0...20	8	6	10	4	12	6
pH_V	6.60...7.60	7.10	0.15	7.15	0.10	7.20	0.15
BD_V	0...20	6	6	8	4	10	6

Table 5.3: *parameters used to generate initial* sigmoid *input membership functions*

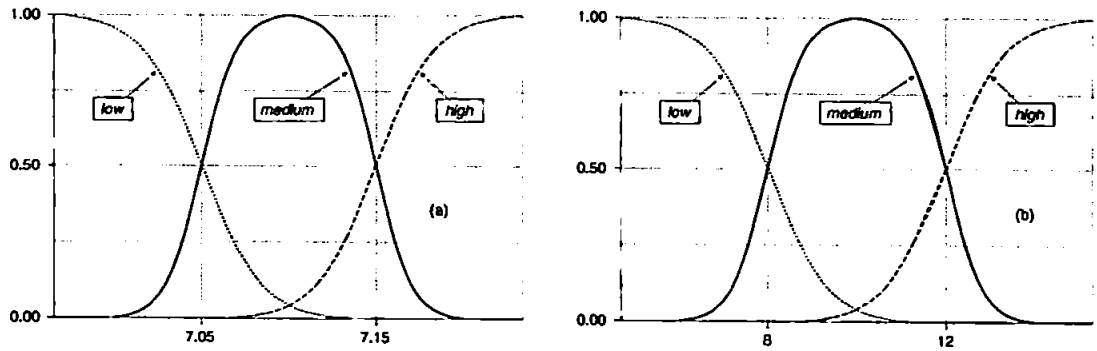


Figure 5.1: *term-sets of (a) the arterial pH (pH_A) and (b) the arterial BD_{ecf} (BD_A) fuzzy input variables*

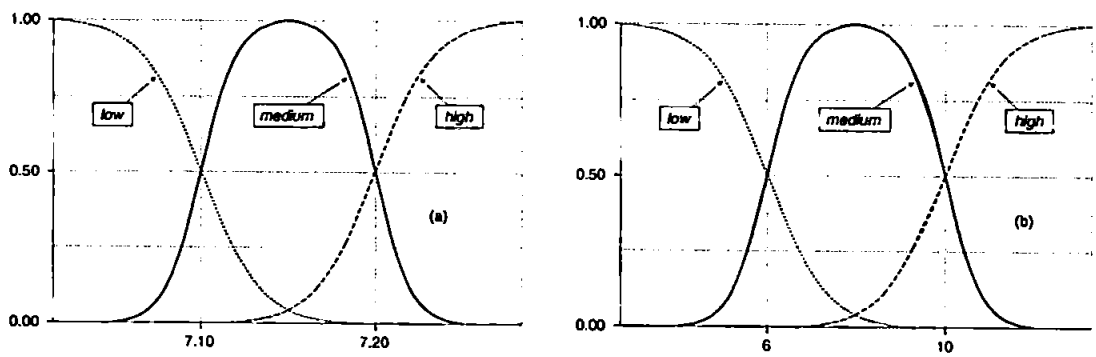


Figure 5.2: *term-sets of (a) the venous pH (pH_V) and (b) the venous BD_{ecf} (BD_V) fuzzy input variables*

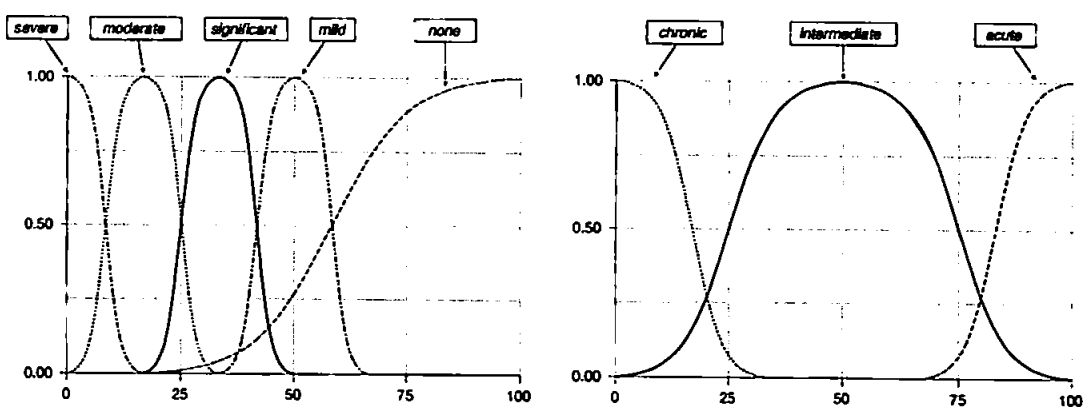


Figure 5.3: *term-sets of (a) the* acidemia *and (b) the* duration *fuzzy output variables*

5.6 Comparison of the Fuzzy System to the Crisp Expert System and Practising Clinicians

5.6.1 Determination of the Evaluation Function

The two crisp outputs of the FES were combined into a single number to represent that duration of acidemia is of secondary importance to severity of acidemia, as given by:

$$condition = acidemia + \frac{duration}{10} \tag{5.11}$$

The fuzzy expert system re-analysed the 10000 *true paired* samples from the database collected at Plymouth and the output was compared to the crisp system. The crisp classification had produced a category in the range 80 to 120. Although these category numbers were originally chosen arbitrarily, they were designed to correspond to an ordered scale, in which 80 was the worst outcome (severe metabolic acidemia in both vessels) and 120 was the best outcome (normal range in both vessels). It was immediately apparent that the classification produced by the fuzzy expert system differed from that of the crisp expert system, in that the ordering of results differed greatly.

An evaluation function utilising the Spearman rank order correlation coefficient described earlier was used to determine which order of results was the most appropriate by comparing the crisp and fuzzy expert systems to practising clinicians. As it was clearly impractical to get the clinicians to order all 10000 results, or even the approximately 400 abnormal results, an extract of 50 results from the 400 abnormals was chosen. The set of 50 results was selected to cover a wide range of categories by randomly selecting a number from each of the crisp categories. The clinicians were asked to rank these 50 results in order from 'worst' to 'best' in terms of likelihood of the infant having suffered asphyxial damage. Two clinicians involved in the creation of the rules and four clinicians considered experienced in the interpretation of cord acid-base results were recruited to take part in the comparison study. They consisted of one Professor of Physiology, one Consultant, one Senior Registrar, two Clinical Research Fellows and a Senior Midwife.

The difficulty of this task can be illustrated by the example in Table 5.4. In simple terms, as the pH in both vessels decreases the infant's condition worsens, and as the BD_{ecf} in both vessels increases the infant's condition worsens. However, the four dimensional data must be considered in parallel, so that in Table 5.4 result *A* might be considered worse as both the base deficits are higher than *B*, or *B* might be considered worse as the arterial pH is lower than *A*.

Results	pH_A	BD_V	pH_V	BD_V
A	6.94	11.6	6.97	11.8
B	6.87	8.5	7.11	9.4

Table 5.4: *an example of the difficulty of ordering acid-base results*

Additionally, the relationship of each of the expert system's ordering to another form of immediate outcome assessment was investigated by comparing the results ordered first by Apgar score at 5 minutes and then by Apgar score at 1 minute.

5.6.2 Performance of the Initial Fuzzy Expert System

The results of the agreement with clinicians of the initial fuzzy expert system ($fuzzy^0$) are shown in Table 5.5. It can be seen that the average inter-clinician correlation, ave(ρ_s), is very high (0.91), indicating that the clinicians agreed with each other very well on the order of results. The crisp expert system correlated reasonably well with the clinicians (0.80), but the performance of the fuzzy expert system was significantly worse.

Agreement	*Correlation*	*Significance*
clinicians $\Leftrightarrow$ clinicians	≈ 0.91	$(p \ll 0.001)$
crisp system $\Leftrightarrow$ clinicians	≈ 0.80	$(p \ll 0.001)$
$fuzzy^0$ system $\Leftrightarrow$ clinicians	≈ 0.67	$(p \ll 0.001)$

Table 5.5: *initial results of agreement with clinicians*

The agreement with outcome is shown in Table 5.6. Given that other factors affect the Apgar score, the precise level of clinicians' agreement with outcome is not particularly important, but the fact that there was significant correlation beyond chance indicates that the clinicians'

ordering did reflect actual clinical outcome. The important point is that the crisp expert system performed with a level close to the clinicians, but again the fuzzy system performed significantly worse.

Agreement	*Correlation*	*Significance*
clinicians $\Leftrightarrow$ outcome	≈ 0.47	$(p \ll 0.001)$
crisp system $\Leftrightarrow$ outcome	≈ 0.39	$(p \ll 0.001)$
fuzzy^0 system $\Leftrightarrow$ outcome	≈ 0.17	$(p \approx 0.001)$

Table 5.6: *initial results of agreement with outcome (Apgar^5, Apgar^1)*

5.7 Tuning the Fuzzy Expert System

5.7.1 Tuning Parameters

As the fuzzy expert system had performed badly in comparison with the clinicians, and worse than the crisp system, some of the fundamental assumptions that were made in the original formulation of the fuzzy model were reconsidered. The usual method of tuning the fuzzy model would be to adjust the shape, location and width of the membership function for each fuzzy term. Although arbitrary changes to individual membership functions are appropriate in fuzzy control type applications, the changes made to membership functions in this application were restricted, in order to keep the model interpretable from a clinical knowledge level. The following changes to the fuzzy model were investigated through application of the simulated annealing tuning algorithm to the initial (*fuzzy*^{0}) fuzzy expert system (changes marked with a • bullet character in this section were discrete variable changes, those marked with a $*$ bullet character were continuous variable changes):

- the standard 'Zadeh' family of operators was used with conjunction as $\min(a,b)$, disjunction as $\max(a,b)$ and negation as $1-a$, rather than the probabilistic family,
- linear membership functions with the same location and crossover as the sigmoid membership functions described in Equations 5.8 to 5.10, and

- the hedges *very* and *fairly* were introduced into the rules.

The reasons for the poor fuzzy performance were investigated by examination of which input variables were most highly correlated with the clinicians and each expert system. It was found that both the clinicians and the crisp expert system correlated most highly with pH_A, pH_V. In contrast, the fuzzy expert system correlated most highly with BD_A, BD_V. Thus it appeared that the fuzzy system weighted the base deficit parameters more than pH in comparison to the clinicians and crisp expert system. It was decided, therefore, to attempt to improve the fuzzy expert system's performance by adjusting the relative weighting of the base deficit parameters in comparison to the pH parameters. A number of ways to do this were investigated:

- one new rule of the form 'IF pH_A is *low* AND pH_V is *low* THEN *acidemia* is *severe*' was added, as all the other rules of the fuzzy expert system feature a combination of pH_A, pH_V, BD_A and BD_V, and so depend on both parameters

* the membership functions of each of the four input variables were shifted

* the height of pH and BD_{ecf} membership functions were decreased so that the fuzzy sets were not normalised, thus altering the influence of pH and base deficit in each rule (Fig. 5.4)

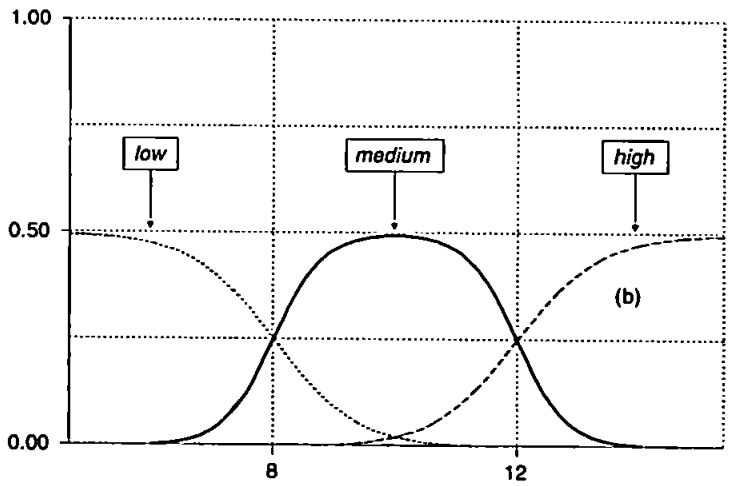


Figure 5.4: *fuzzy base deficit variable with reduced membership heights*

5.7.2 Application of the Tuning Algorithm

The effects of these changes were investigated by applying the simulated annealing algorithm detailed in Section 5.4.3. The $K = 50$ test cases were used as the target set for the FES, and the minimised evaluation function was $f = 1 - \rho_s$, where ρ_s is the Spearman rank correlation of the output of the FES against the clinicians' rankings by Equation 5.7 in Section 5.4.4. Thus, for this application:

$$\begin{aligned}
\mathbf{x} &= (\textit{operator family}, \textit{membership function shape}, \ldots \\
&\quad \textit{rule set}, \text{ very } \textit{hedging}, \text{ fairly } \textit{hedging}) \\
\mathbf{y} &= (\textit{pH}_A \textit{ offset}, \textit{BD}_A \textit{ offset}, \textit{pH}_V \textit{ offset}, \textit{BD}_V \textit{ offset}, \ldots \\
&\quad \textit{pH}_A \textit{ height}, \textit{BD}_A \textit{ height}, \textit{pH}_V \textit{ height}, \textit{BD}_V \textit{ height})
\end{aligned}$$

The variables were encoded as follows:

$$\begin{aligned}
\textit{operator family}: \quad & x_1 = \{\, 0 \equiv \text{probabilistic}, 1 \equiv \text{min-max} \,\} \\
\textit{mf shape}: \quad & x_2 = \{\, 0 \equiv \text{sigmoid}, 1 \equiv \text{linear} \,\} \\
\textit{rule set}: \quad & x_3 = \{\, 0 \equiv \text{original}, 1 \equiv \text{extra rule included} \,\} \\
\text{very } \textit{hedging}: \quad & x_4 = \{\, 0 \equiv \text{no } \textit{very} \text{ hedges}, \ldots \\
& \qquad \textstyle\sum_{i=1}^{21} 2^i \equiv \textit{very} \text{ hedge for acidemia in Rule } i \,\} \\
\text{fairly } \textit{hedging}: \quad & x_5 = \{\, 0 \equiv \text{no } \textit{fairly} \text{ hedges}, \ldots \\
& \qquad \textstyle\sum_{i=1}^{21} 2^i \equiv \textit{fairly} \text{ hedge for acidemia in Rule } i \,\} \\
\textit{pH}_A \textit{ offset}: \quad & y_1 = [\, 6.50, 7.50 \,] \\
\textit{BD}_A \textit{ offset}: \quad & y_2 = [\, 0, 20 \,] \\
\textit{pH}_V \textit{ offset}: \quad & y_3 = [\, 6.50, 7.50 \,] \\
\textit{BD}_V \textit{ offset}: \quad & y_4 = [\, 0, 20 \,] \\
\textit{pH}_A \textit{ height}: \quad & y_5 = [\, 0, 1 \,] \\
\textit{BD}_A \textit{ height}: \quad & y_6 = [\, 0, 1 \,] \\
\textit{pH}_V \textit{ height}: \quad & y_7 = [\, 0, 1 \,] \\
\textit{BD}_V \textit{ height}: \quad & y_8 = [\, 0, 1 \,]
\end{aligned}$$

The fuzzy model was then tuned automatically with Algorithm C by evaluating the cost function as follows:

1. Generate fuzzy model input term memberships as shown in Table 5.7;
   (a) If $x_2 = 0$ then use *sigmoid* membership functions, else use linear membership functions;
   (b) Reduce height of three pH_A terms by y_5, reduce height of three BD_A terms by y_6, reduce height of three pH_V terms by y_7, reduce height of three BD_V terms by y_8;
2. Run the fuzzy model on K input data;
   (a) If $x_3 = 0$ then use original rule set, else use rules set with additional rule;
   (b) If $x_4 > 0$ then insert a *very* hedge operator into specified rules;
   (c) If $x_5 > 0$ then insert a *fairly* hedge operator into specified rules;
   (d) If $x_1 = 0$ then use probabilistic operators, else use min-max operators;
   (e) Defuzzify output variables and calculate single crisp output by Equation 5.11;
3. Rank K outputs from $1 \ldots K$ in order from worst to best;
4. Calculate ρ_s for each clinician by Equation 5.7, and then calculate the average fuzzy expert system agreement $\bar{\rho}_s$;
5. Cost function $= 1 - \bar{\rho}_s$;

Variable	m_{low}	σ_{low}	m_{med}	σ_{med}	m_{high}	σ_{high}
pH_A	y_1	0.15	$y_1 + 0.05$	0.10	$y_1 + 0.10$	0.15
BD_A	$y_2 - 4$	6	$y_2 - 2$	4	y_2	6
pH_V	y_3	0.15	$y_3 + 0.05$	0.10	$y_3 + 0.10$	0.15
BD_V	$y_4 - 4$	6	$y_4 - 4$	4	y_4	6

Table 5.7: *parameters used to generate tunable* sigmoid *membership functions for input variables*

An approximation to the optimal cost function f_{opt} was obtained by calculating the sum of the ranks of the six clinicians for each of the K cases, and then re-ranking these (row-wise)

sums. As an example, if six clinicians each ranked three cases A, B, and C as shown in Table 5.8, the optimal ranking would be $A = 1$, $B = 3$, and $C = 2$, as shown in the right-hand column. This *ranking of the summed rankings* can be thought of as the best *consensus* view, and gave a value of $\rho_s(opt) \approx 0.97$, and therefore $f_{opt} = 1 - \rho_s(opt) \approx 0.03$. However, it is not necessarily the case that *any* configuration of the fuzzy model could attain f_{opt}, so in practice the average inter-clinician agreement, $\text{ave}(\rho_s) = 0.91$, was taken as $\rho_s(min)$, so that $f < 1 - \rho_s(min) \approx 0.09$ was taken as acceptable performance.

	Clinician							
Case	1	2	3	4	5	6	Σ_1^6	r_{opt}
A	1	1	2	2	1	2	9	1
B	2	3	1	3	2	3	14	3
C	3	2	3	1	3	1	13	2

Table 5.8: *an example of six separate rankings, with the row-wise sum and* optimal *ranking*

The initial temperature, number of repeats and λ parameter (step 2, Algorithm B) were found by trial and error. The tunable parameters were initialised in a random configuration, and the annealing algorithm was run using a cooling schedule whereby the temperature was reduced by a factor of $\omega = 0.95$ until it reached $\frac{1}{100}^{\text{th}}$ its initial value (90 iterations). The number of repeats was typically around 300 per iteration, so that a total of *approx* 27 000 annealings were performed. Each evaluation of the cost function by running the fuzzy model took ~ 1 second, and therefore each annealing trial took ≈ 8 hours. At the completion of each run, the final configuration of the fuzzy model and its performance in terms of the lowest cost function was stored. Multiple runs were undertaken from different random starting points, in case the annealing had located a local minima of the cost function.

5.7.3 Performance of the Tuned Fuzzy Expert System

The biggest effect on the performance of the FES was found to be:

1. the sub-normalisation of the base deficit term-sets: the sub-normalisation was carried out on each base deficit variable independently, affecting all the terms equally. Very

slight differences in the performance could be obtained through having the height of the two variables slightly unequal, but overall the best effect was achieved by reducing the height of both variables' term-sets to 0.4.

2. the introduction of the extra rule specified above: there are a very large number of other rule changes that could have been considered, such as creating new rules and/or deleting the existing rules, but this was not attempted.

The choice of probabilistic operators was found to have a small beneficial effect over the max / min operators, as was the use of *sigmoid* membership functions over *triangular* functions. The introduction of hedges was found to have a negligible effect. The tuned FES thus has membership functions identical in shape and location to those shown in Fig. 5.1 and 5.2, but with the maximum height of all the base deficit terms reduced to 0.4. No alteration of the output variable membership functions was attempted. After introduction of the single additional rule and reduction of the base deficit membership heights to 0.4, the final results shown as $fuzzy^1$ in Table 5.9 and 5.10 were obtained. The clinicians', crisp expert system's, and $fuzzy^0$ expert system's results are re-presented for ease of comparison.

Agreement	*Correlation*	*Significance*
clinicians $\Leftrightarrow$ clinicians	≈ 0.91	$(p \ll 0.001)$
crisp system $\Leftrightarrow$ clinicians	≈ 0.80	$(p \ll 0.001)$
$fuzzy^0$ system $\Leftrightarrow$ clinicians	≈ 0.67	$(p \ll 0.001)$
$fuzzy^1$ system $\Leftrightarrow$ clinicians	≈ 0.93	$(p \ll 0.001)$

Table 5.9: *results of agreement with clinicians after refinement of fuzzy expert system*

Agreement	*Correlation*	*Significance*
clinicians $\Leftrightarrow$ outcome	≈ 0.47	$(p \ll 0.001)$
crisp system $\Leftrightarrow$ outcome	≈ 0.39	$(p \ll 0.001)$
$fuzzy^0$ system $\Leftrightarrow$ outcome	≈ 0.17	$(p \approx 0.001)$
$fuzzy^1$ system $\Leftrightarrow$ outcome	≈ 0.51	$(p \ll 0.001)$

Table 5.10: *results of agreement with outcome ($Apgar^5$, $Apgar^1$) after refinement of fuzzy expert system*

It can be seen that the tuned fuzzy expert system ($fuzzy^1$) achieved an excellent agreement with the clinicians and matched the clinicians in its agreement with outcome. Its modified

performance was better than the crisp expert system and effectively indistinguishable from the clinicians. A graph of the six clinicians rankings plotted against the *fuzzy*^{1} rankings is shown in Fig. 5.5. If perfect agreement had been found, a straight line through the origin would have been obtained. For example, for the case that the *fuzzy*^{1} system ranked the worst (ranking 1), all the clinicians also ranked the worst, and all the points are superimposed. On the other hand, for the case that the *fuzzy*^{1} system ranked 25th worst, the clinicians ranked as 14th, 16th, 17th, 18th, 19th, and 20th.

The performance of the final (tuned) fuzzy expert system was then validated by calculating the agreement of the crisp expert system, the original fuzzy system (*fuzzy*^{0}) and the tuned fuzzy system (*fuzzy*^{1}) against the entire database of cases with *true paired* samples where the Apgar scores had been recorded ($n = 9003$). The results in Table 5.11 showed that the *fuzzy*^{1} indeed performed the best of the three systems, albeit with fairly low overall correlation. The agreement was also calculated against the Apgar scores for all cases with abnormal acid-base status ($n = 383$, defined by the crisp cut-offs $pH_A < 7.05$, $BD_A \geq 12$ mmol.l^{-1}, $pH_V < 7.10$, and $BD_V \geq 10$ mmol.l^{-1}) and the results in Table 5.12 were obtained. It can be seen that the *fuzzy*^{1} system again achieved the highest correlation of the three systems.

Agreement	Correlation	Significance
crisp system $\Leftrightarrow$ outcome	≈ 0.21	$(p < 0.001)$
fuzzy^0 system $\Leftrightarrow$ outcome	≈ 0.14	$(p \approx 0.001)$
***fuzzy*^{1} system $\Leftrightarrow$ outcome**	≈ 0.24	$(p < 0.001)$

Table 5.11: *validation of crisp, initial and tuned fuzzy expert systems with outcome (Apgar^5, Apgar^1) for all cases* ($n = 9003$)

Agreement	Correlation	Significance
crisp system $\Leftrightarrow$ outcome	≈ 0.44	$(p < 0.001)$
fuzzy^0 system $\Leftrightarrow$ outcome	≈ 0.26	$(p \approx 0.001)$
***fuzzy*^{1} system $\Leftrightarrow$ outcome**	≈ 0.52	$(p < 0.001)$

Table 5.12: *validation of crisp, initial and tuned fuzzy expert systems with outcome (Apgar^5, Apgar^1) for all abnormal cases* ($n = 383$)

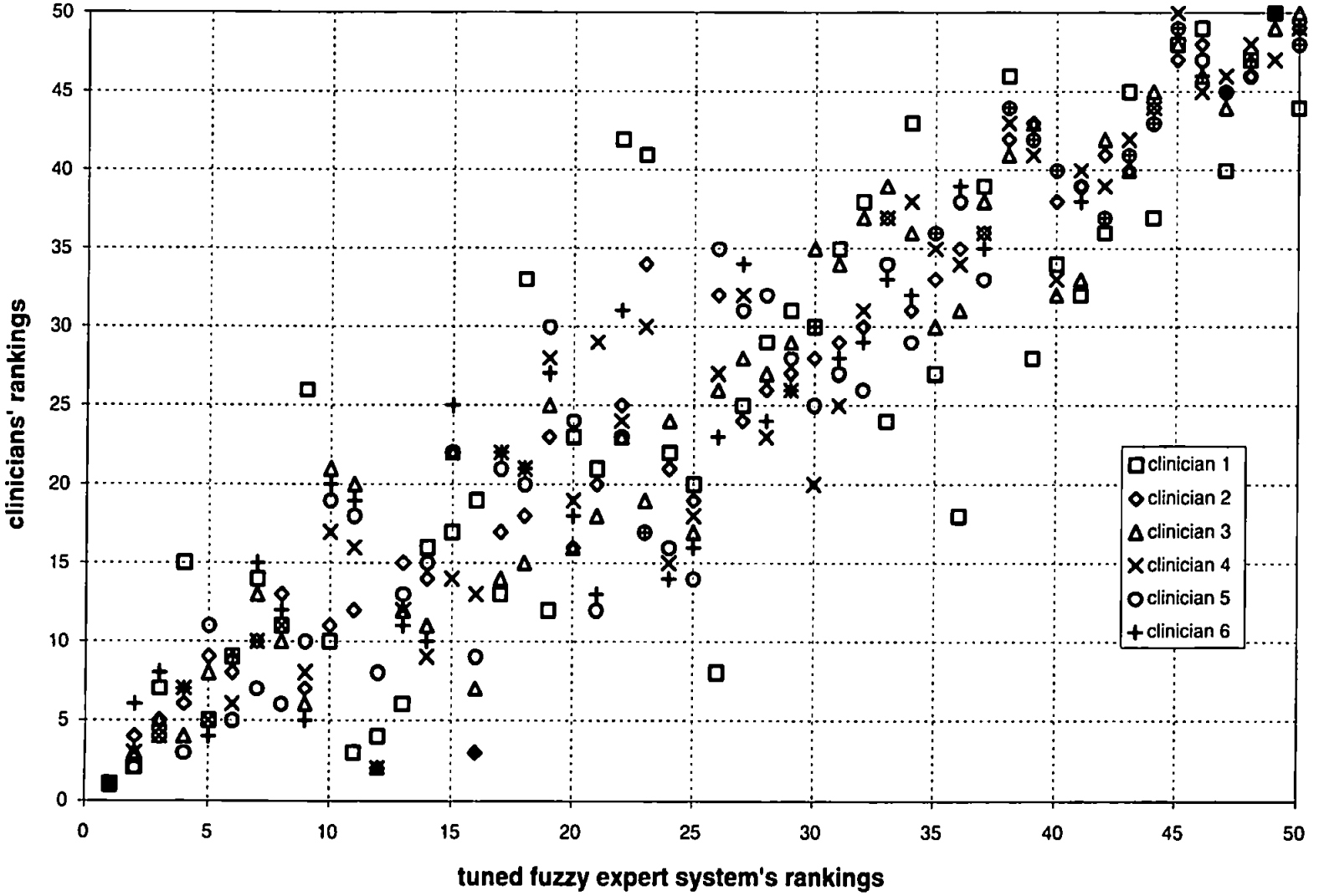


Figure 5.5: *graph of six clinicians rankings against the optimised fuzzy expert system*

5.7.4 Summary of Tuning Procedure

Although the theoretical basis of the simulated annealing algorithm has been presented at the start of this Chapter in order to separate the technical and implementation details, the actual development of the algorithm took place after the development of the initial *fuzzy*^{0} fuzzy expert system, and after the 'validation' data had been obtained. Specifically, the actual order of events was as follows.

1. The initial fuzzy expert system (*fuzzy*^{0}) was developed from the rules and cut-offs encapsulated in the crisp expert system.
2. The set of fifty validation cases was selected, and the six clinicians were recruited to rank the cases.
3. Examination of the validation data demonstrated that the initial FES (*fuzzy*^{0}) performed significantly worse than the clinicians, and worse than the crisp expert system.
4. The validation data, the crisp expert system output and the *fuzzy*^{0} output were examined to attempt to identify the reasons for the poor performance of *fuzzy*^{0}. This demonstrated that the *fuzzy*^{0} rankings differed in a non-simplistic manner.
5. Examination of the input data showed that the *fuzzy*^{0} system correlated most highly with BD_{ecf} whereas the crisp system correlated most highly with pH.
6. Manual methods of increasing the BD_{ecf} weighting were examined, but it was quickly realised that there were far too many parameters to systematically modify manually.
7. The automatic simulated annealing tuning method was developed to tune the system using the already obtained 'validation' data set as the target output.
8. Multiple simulated annealing runs were undertaken, using a variety of control parameters, to tune the system from a large number of random starting configurations.
9. The simulated annealing algorithm identified the configuration with one additional rule and height of the BD_{ecf} membership functions reduced to 0.4.
10. This 'tuned' system (*fuzzy*^{1}) was then fixed as the best preliminary FES, and was validated by comparison with the full data set of $n = 9003$ Apgar scores.

5.8 Discussion

The initial performance of the fuzzy expert system was poor — worse than the original crisp system — not only when compared to clinicians in the test that was designed, but also when compared to the most popular form of current overall neonatal outcome assessment (the Apgar score). The reason for this poor performance is not clear. In the crisp system the rule set is non-overlapping and exclusive; each set of input parameters will fire a single rule. It appeared that the multiple interaction of the fuzzy rules (especially in the intermediate cases) caused its performance to degrade, due to weighting the base deficit parameters too highly (see Section 5.7). Ultimately, the reason for the poor performance could only be investigated through lengthy discussions with the clinicians on a case by case basis to identify situations where the crisp system out-performed the fuzzy system, and then by examination of the internal processing to establish 'incorrect' behaviour. Unfortunately, lack of availability of the clinicians prevented such an investigation.

An encouraging aspect of this study was that the clinicians did form a good body of agreement between themselves — not always the case in studies involving *expert* clinicians. The definition of *expert* is rarely addressed in expert systems research, and is even rarer in medicine. Often it is assumed that the more senior the clinician, the more expert they invariably become. However, it is usually the more junior ranks that are performing decision making tasks on a day to day basis. In this study the term *clinician* has been preferred to the term *expert*, although many of those involved may deserve the latter tag. Part of the difficulty with this field of research is that there is currently no *gold-standard* against which to compare umbilical acid-base analysis. Not all infants that are born with abnormally low pH and high base deficit will be permanently damaged, and other forms of immediate neonatal assessment do not distinguish between damage that has occurred before labour to that which has occurred during labour. The ultimate gold-standard may be the long term neurological development of the neonate, but this is obviously confounded by many environmental and educational effects, which have swamped the initial effects of labour on the brain well before the infant's IQ can be accurately measured.

The performance of the optimised fuzzy expert system was excellent — effectively indistinguishable from the clinicians — and the best of the expert systems in the final validation test against Apgar scores. Obviously the FES has been tailored to perform well on the training set of 50 cases, and would have to be retested against clinician's opinion on novel cases to ensure that the results can be generalised. However, the advantage of the approach described over a neural network approach is that the final fuzzy system can be analysed and discussed with the (expert) clinicians to identify whether its rule set is acceptable on the meta-level. The two changes made to produce the final fuzzy model can both be justified from clinical knowledge, and the resultant fuzzy model is attractive through its ability to vary the relative importance of the pH and base deficit parameters simply through alteration of the height of the base deficit term sets.

The use of sub-normal membership functions to adjust the relative weighting of fuzzy terms is a novel and promising general purpose technique. The decision to change the height of membership functions was made in an attempt to rectify the apparent over emphasis that the fuzzy expert system placed on the base deficit parameters. In the vast majority of applications and in most theoretical work on fuzzy sets, it is assumed that membership functions should be normalised. Whilst it is true that most, if not all, clinicians would label a base deficit of, for example, 25 as *high* with a truth of 1.0, it is questionable whether the truth that the statement conveys is *as true* as the statement that a pH of 6.60 is *low* — especially when the statements are considered in the context of possible damage having occurred to a neonate. In other words the context in which a particular linguistic term is used and its relationship to other linguistic variables in the same context can alter its maximum attainable truth value.

If one accepts the central tenet of fuzzy logic — that truth can have different linguistic values, with shape and form to represent different human concepts of truth — then there should be no problem with accepting that certain statements will never be *as true* as others within a specified domain. The use of sub-normal fuzzy sets in a linguistic variable allows the truth of statements concerning that variable to be weighted within a domain in a natural and consistent manner. Any internal change to a fuzzy model that alters the input-output mapping can be said to change the meaning of the rules embodied in the model. There is

almost certainly an alternative formulation of the fuzzy model that would achieve as good a performance as the system presented here, but without needing to use sub-normal sets. Such a model may feature a set of rules in which the pH parameters and the BD_{ecf} parameters are separated out, with rules concerning base deficits having an overall weighting factor, but this representation may not be as appealing to the clinician.

The simulated annealing algorithm is a simple, general purpose technique that can optimise a fuzzy model used in a fuzzy expert system. Other optimisation techniques could have been used, and would have been equally valid. Simulated annealing is not necessarily the quickest or most efficient form of global optimisation in terms of function evaluations, but it is easy to apply to both discrete and continuous variables. If pragmatic considerations of speed (in terms of coding effort) and ease of getting a fuzzy system working to a high performance level outweigh the theoretical drawbacks of the simulated annealing approach, then the method is extremely useful.

Chapter 6

An Integrated Fuzzy Expert System

6.1 Introduction

In the previous chapter, the development and optimisation of a preliminary fuzzy expert system was described. Although this preliminary fuzzy expert system was tuned to achieve a high level of agreement with clinicians and to associate with Apgar scores, it was characterised by a number of restrictions:

- *crisp input variables* — the variables to the fuzzy expert system were represented as fuzzy singletons, equivalent to the crisp value obtained from the blood gas analyser;
- *adapted crisp expert system rules* — the rule set was adapted directly from the crisp rule set, rather than from a rule set specifically designed for a fuzzy expert system;
- *restricted rule set* — the rule set was restricted to the 21 crisp rules that dealt with the interpretation of **full paired** results only, i.e. samples which the crisp expert system had previously validated as comprising an error-free pH and BD_{ecf} from both artery and vein; and
- *crisp output variables* — the two output variables *acidemia* and *duration* elicited from the crisp rule set, although represented internally as fuzzy sets, had been centre-of-gravity defuzzified and combined to give a single crisp output.

In this chapter, the design and development of a more advanced fuzzy expert system, referred to as the *integrated fuzzy expert system*, is described. The integrated fuzzy expert system was enhanced from the previous system in several ways:

- *fuzzy input variables* — the input variables have a width to explicitly represent that this input value is an estimate of the 'true' parameter value;
- *fuzzy vessel identification* — the simplistic crisp rule for determination of vessel identification (see Section 3.2.3, Page 82) was refined through the introduction of a fuzzy rule set for a decision based on the three primitive input parameters (pH, pCO_2 and pO_2) from both samples;
- *new fuzzy rules* — a fresh knowledge elicitation was undertaken with the same experts used for the crisp expert system rules to create a set of fuzzy rules specifically designed for the fuzzy expert system;
- *interpretation of all results* — unknown values were represented such that results with invalid or missing parameters in all combinations (see Section 3.2.4, Page 88) could be processed by the fuzzy expert system;
- *fuzzy output variables* — three fuzzy output variables (*acidemia*, *component*, and *duration*) were utilised in rule consequences, with the availability of graphical output of the consequence fuzzy sets, in addition to several alternative numerical representations of uncertainty in the interpretation; and
- *linguistic approximation* — linguistic approximation of the fuzzy output variables was also introduced to allow textual output from the fuzzy expert system.

The integrated fuzzy expert system was developed with knowledge gained from the preliminary fuzzy expert system (Chapter 5), and its performance was evaluated and refined with the validation data already obtained. When the development of the integrated system was complete, a comprehensive validation process was undertaken to re-evaluate the numeric and linguistic outputs of both the vessel identification and interpretation aspects of the system.

The inter-expert agreements and the comparative performance of the expert system obtained in the validation study are presented and discussed.

6.2 Experimental Determination of Umbilical Cord Acid-Base Errors

6.2.1 Sources of Uncertainty in Umbilical Cord Acid-Base Parameters

The first enhancement was to incorporate fuzziness into the input variables of the preliminary fuzzy expert system, to represent the fundamental uncertainty present in the acid-base parameters. To do this, it was necessary first to characterise and determine the uncertainty inherent in each of the measured parameters. Three main sources of uncertainty were identified:

- Machine Error
- Sampling Error
- Time Delays

Experiments were designed to quantify the contribution of each source of error to the typical overall uncertainty in each parameter. The factors that comprise each source of uncertainty and the experiments designed to quantify them are described in detail below.

6.2.2 Machine Error

Any measurement device has inherent uncertainty in its results, generally referred to as *measurement error*. Without going into details, a blood gas analysis machine determines the pH, pCO_2 and pO_2 by measuring changes in the electrical potential across a membrane when fluid is introduced into a chemical electrode [115]. The performance of these electrodes is governed by the quality of the membrane, the temperature of the fluid, the stability of the

reference electrode, etc. The machine performs a regular internal calibration, roughly once per hour, with fluid of known composition which has calibration gas bubbled through it. This is known a *one-point calibration* or simply a *calibration*. Roughly once per day, a second form of calibration is performed in which two samples of fluid are measured and the difference between them governs the slope of the calibration curve. This is known as a *two-point calibration* or simply a *slope*.

In addition, roughly once per day, maintenance personnel ensure the accuracy of the machine by measuring three types of quality control fluid which has been factory preset to certain levels for each parameter. If the quality control results fall outside these limits maintenance procedures are carried out, such as the renewal of old electrodes, removal of protein deposits that build up on the electrode membranes, etc. The manufacturers of the blood gas analysis machine quote the nominal laboratory measurement errors, but in practice these can be enlarged through poor maintenance or quality control. Quality control of blood gas machines is a complex field in its own right, but in clinical practice the errors can be measured and allowed for.

These *machine errors* can be determined by repeatedly measuring a stable quality control material and by pairwise measurement of blood samples. Experiments were carried out to repeatedly measure samples of an artificial blood substitute quality control material that is pre-calibrated and known to stay stable for up to four hours in the syringe. Experiments were also carried out in which large samples of blood were taken from the umbilical vein and introduced pairwise into the analyser.

6.2.3 Sampling Error

In actual clinical practice a number of errors can be introduced through the sampling technique. Blood is sampled by inserting a small needle into the desired vessel and drawing the blood into a syringe. The umbilical arteries are small vessels, which are often difficult to sample and sometimes can be empty of blood. When sampling the artery, it is especially easy to stick the needle right through the vessel so that venous blood, mixed arterial-venous

blood, or other fluid is inadvertently drawn into the syringe. Often air is drawn into the syringe along with the blood, which will cause errors due to gaseous exchange with the air. An as yet unresolved issue is the homogeneity of the blood within the vessels. There is a natural gradient in blood parameters across the placenta and across the infant, but there may be a gradient along the length of the vessel. Additionally, there may be a difference between the blood in the two umbilical arteries.

This class of *sampling error* is especially difficult to quantify as it depends on the quality of cord segment that is obtained by the clinical staff and on the ability of the person responsible for the sampling. Any experiment to quantify it is subject to the *Hawthorn Effect* whereby the performance of the staff is improved purely through the process of being measured. Experiments were performed by a trained clinician and an inexperienced engineer to attempt to simulate a difference in clinical performance. A number of experiments were carried out in which:

- blood was taken from sections of umbilical cord that had been isolated through the use of additional clamps,
- blood was taken from both umbilical arteries,
- blood samples of various size were taken, and
- samples had varying quantities of air deliberately introduced.

6.2.4 Time Delays

The third major class of uncertainty is caused by various time delays that are encountered in the procedure in actual clinical practice. This can be subdivided into three broad categories. Firstly, there is the precise moment at which the umbilical cord is clamped after delivery. If an infant has made breathing efforts and/or has begun to cry before the cord is clamped, then the parameters may be affected by the onset of neonatal respiration. This effect was measured through a series of experiments in which a catheter (tube) was inserted into an umbilical vessel and blood was repeatedly drawn from the vessel as the infant began respiration. The

cord was then clamped after a period of several minutes and a normal pair of cord samples was then taken.

Blood may be left in the umbilical cord for several minutes until a clinician is available to perform the sampling. Blood does not clot whilst the cord is left intact and so it may be left for a significant time (more than an hour) before the sampling takes place. Although the cord is sealed from air during this time, it is possible that slow gaseous exchange takes place through the internal vessel walls or that the cells within the cord continue with residual metabolism. The cord may be placed on ice to prevent the effect of residual cell metabolism. These effects were quantified by a series of experiments in which one sample was performed and then cord was left at room temperature and on ice for a variable period before a repeat sample was performed.

Once the blood has been sampled, it may be left in the syringes for a variable amount of time before measurement. Although the blood would normally coagulate in this time, an anti-coagulation agent, heparin, is usually added to the syringe to prevent this. Sometimes the blood gas machine will perform an internal calibration, which may take several minutes, after the samples have been taken but before the have been introduced into the machine. Occasionally an emergency might occur so that the clinician must attend to more urgent matters before the measurement is performed. The parameters may then change by gaseous exchange with air in the syringe or via the syringe walls. This was quantified by sampling blood, performing a measurement immediately and then repeating the measurements at intervals.

6.2.5 Summary of Uncertainty Quantification

It was found that the machine errors for each parameter agreed closely with the manufacturer's specification. This was probably a reflection of the fact that the machine in Plymouth received regular high quality maintenance from the medical physics department. However it was found that the sampling errors were significantly larger than the machine error, indicating that in actual clinical practice the machine imprecision is almost negligible. It was found

that for each parameter the arterial error was significantly higher than the venous error. This was probably due to the difference in sample volume and the difficulty of arterial sampling. The umbilical vein is large and easily accessible so that a sample volume of over 2ml of blood containing little air was usually obtained. In comparison, the arteries are small and difficult to sample, so that the combined total arterial sample volume was usually between 0.2ml and 0.5ml and often contained a significant volume of air.

Time delay errors were found to be small if time delays were reasonable. Specifically, it was found that the cord may be left at room temperature for up to one hour or on ice for up to four hours without significant change. Similarly, a syringe with good volume ($\approx$ 2ml) of blood and no air may be left at room temperature for over 30 minutes or on ice for over an hour without significant change. Change due to delayed clamping of the cord was again found to be small (in comparison to sampling errors) if the cord was clamped within two minutes. In Plymouth, protocols have been instigated in clinical practice to ensure that these time delays are not exceeded.

The results from each source of uncertainty were combined and the *maximum likely* uncertainty found in each parameter (where *maximum likely* was defined as two standard deviations from the mean) is summarised in Table 6.1. These uncertainties can be used to generate the fuzzy sets for each input parameter from the crisp result obtained from the blood gas machine. Initially, each parameter can be fuzzified to a fuzzy number with width governed by the larger uncertainty of the arterial parameter. Once the samples have been identified as to their vessel of origin, the venous parameters can have their uncertainty reduced.

	δpH	δpCO_2	δpO_2
Arterial	0.025	0.61 kPa	0.31 kPa
Venous	0.010	0.24 kPa	0.15 kPa

Table 6.1: *the combined* maximum likely *uncertainty (two standard deviations from the mean) in each umbilical cord acid-base parameter*

6.2.6 Derived Uncertainty in Base Deficit

Having determined the likely errors in each of the primary input parameters, it is necessary to calculate the likely error in the BD_{ecf}, by the combination the errors in the pH and pCO_2 from which is derived. Taylor's theorem in two-dimensions states that:

$$f(x+h,y+k) = f(x,y) + h \cdot \frac{\partial f(x,y)}{\partial x} + k \cdot \frac{\partial f(x,y)}{\partial y} + \ldots \tag{6.1}$$

Combination and simplification of the Siggaard-Andersen base deficit equation (1.13 and the Henderson-Hasselbalch equation (1.7) gives:

$$BD_{ecf}(pH, pCO_2) \approx 131.80 - 14.83\,pH - 0.21\,pCO_2\,10^{(pH-6.10)} \tag{6.2}$$

From this:

$$\frac{\partial BD_{ecf}(pH, pCO_2)}{\partial pH} \approx -14.83 - 0.48\,pCO_2\,10^{(pH-6.10)} \tag{6.3}$$

and:

$$\frac{\partial BD_{ecf}(pH, pCO_2)}{\partial pCO_2} \approx -0.21 \cdot 10^{(pH-6.10)} \tag{6.4}$$

Combining Equations 6.3 and 6.4 by Equation 6.1, and changing sign as it is only the magnitude which is required, gives:

$$\delta BD_{ecf} \approx \left(14.83 + 0.48\,pCO_2\,10^{(pH-6.10)}\right)\delta pH + \left(0.21 \cdot 10^{(pH-6.10)}\right)\delta pCO_2 \tag{6.5}$$

This gives the uncertainty in the BD_{ecf}, δBD_{ecf}, at a given pH and pCO_2, when the values of δpH and δpCO_2 are known. Note that this is strictly valid only with independent variables, in which case Equation 6.5 should be used with the errors expressed as variances rather than standard deviations. However, ignoring the known dependence of pH and pCO_2 and using the errors expressed as standard deviations from Table 6.1 both cause the δBD_{ecf} calculated by Equation 6.5 to be an *over-estimate* of the true error. As an over-estimate was prefered to an under-estimate, these two problems will be ignored.

Application of Equation 6.5 to some example samples, using the empirically determined values of δpH and δpCO_2 from Table 6.1, are shown in Table 6.2.

		pH	pCO_2	BD_{ecf}	δpH	δpCO_2	δBD_{ecf}
Arterial	A_1	7.25	8.0	0.5	0.025	0.61	3.3
	A_2	7.25	4.0	12.4	0.025	0.61	2.6
	A_3	7.00	16.0	1.3	0.025	0.61	2.8
	A_4	7.00	9.6	12.0	0.025	0.61	2.2
Venous	V_1	7.25	8.0	0.5	0.010	0.24	1.5
	V_2	7.25	4.0	12.4	0.010	0.24	1.2
	V_3	7.00	16.0	1.3	0.010	0.24	1.2
	V_4	7.00	9.6	12.0	0.010	0.24	1.0

Table 6.2: *examples of typical arterial and venous samples with empirical uncertainties in pH and pCO_2, and the associated calculated uncertainty in BD_{ecf}*

6.3 Design of the Integrated Fuzzy Model

As the previously described preliminary fuzzy expert system had achieved a high level of performance (after tuning) compared to the clinicians, most of the fundamental fuzzy model assumptions adopted for the preliminary system were retained for the integrated fuzzy expert system. The changes introduced into the fuzzy model specification were as follows.

6.3.1 Inference Methodology

The Mamdani model of inference was retained as previously, with the min operator used for implication. Two sets of fuzzy rules were employed; the *vessel identification* rules and the *interpretation* rules. Initially the lowest pH is labelled as the artery, or if the pH's are the same, the highest pCO_2 is labelled as the artery, or if the pH's and pCO_2's are the same, the lowest pO_2 is labelled as the artery. If two samples are present, these parameters are then passed through the vessel identification rules to determine whether they represent an arterial-venous pair. Once vessel identification has been carried out, the samples (possibly containing *unknown* values) are passed through the interpretation rules. The processes involved are detailed in Sections 6.5 and 6.6.

Centre-of-gravity (centroid) defuzzification was performed on the fuzzy output variables, and three other techniques were employed to provide an indication of uncertainty in the centroids (see Section 2.4.10):

1. membership grade at the centroid, μ_g,
2. normalised area of the output set, A,
3. *entropy* of the output set, S.

Linguistic approximation was also performed on all the fuzzy outputs, and this is detailed in Section 6.4.

The integrated system required the representation of missing values, in cases where parameters were missing, only one sample was available, or the vessel identification rules established that samples were from the same or mixed vessels. The representation of *missing* was implemented by setting the parameter value to one of the three level sets: *unknown*, *indeterminate*, or *undefined*. In the Mamdani style of fuzzy inference the output sets are initialised with *undefined*, and so the system starts in a state where μ_g, A and S are all zero. Each rule that then fires *adds* information to the output sets, to produce a result that varies from the ideal singleton output, to the extreme of *unknown* across the universe. It was found that only by using the *unknown* level set to represent *missing* was the information and uncertainty in the output, calculated by the summation of the area and entropy of the output sets, increased for cases with missing values as required. The methodology used is detailed in Section 6.6.

6.3.2 Linguistic Variables and Fuzzy Terms

The vessel identification rules operate on the differences between two samples, ΔpH, ΔpCO_2 and ΔpO_2. In the conventional fuzzy approach, the natural scheme would have been to create three linguistic variables, one for each of the Δ differences, and create a set of terms for each one. The term-sets for the variables might then include terms such as *zero*, *low*, *mid*, and *high*, which were designed to correspond to appropriate values independently for each parameter. Rules would have been constructed of the form 'IF ΔpH IS *low* THEN *origin* IS *same*'. This approach would have involved the creation of three linguistic variables, each with three or four terms in the term-set, and the possible implementation of fuzzy arithmetic

to calculate the Δ's. To avoid this, a set of fuzzy rules were developed to operate directly on comparisons of the input parameters, pH, pCO_2 and pO_2, for the two samples, where each had been fuzzified to possess a width commensurate to its uncertainty. Thus no additional terms were required for the pCO_2 and pO_2 parameters used in vessel identification.

The output of the vessel identification was a single linguistic variable, *origin*. The *origin* fuzzy output consisted of three terms: *same*, *mixed* and *different*, to respectively indicate two samples from the same vessel (near identical results), one or more samples from mixed vessels (vessel differences inconsistent or contradictory), and two samples from an arterial-venous pair (vessel differences as expected physiologically).

It had previously be noted informally that the preliminary fuzzy expert system, and indeed the original crisp expert system, made no provision for abnormally high (alkalotic) results. Thus it was decided to introduce an additional term, *normal*, to each of the pH variables, and to shift the *high* term to the right to correspond to an abnormally high pH. A new consequence set, *alkalotic*, was also added to the right hand side of the *acidemia* variable term-set, and the term previously labelled as *none* was re-labelled as *normal*.

After consultation with the experts, it was decided to re-order the terms *significant* and *moderate* for the *acidemia* linguistic variable, in comparison to the preliminary system. It was felt that the term *significant* carried a more ominous connotation than the term *moderate* when applied to the degree of acidemia, and should therefore appear further to the left in the acidemia term-sets. Thus in the integrated system the terms are: *severe*, *significant*, *moderate*, *mild*, *normal*, and *alkalotic*, in order from the highest degree of acidemia (most acidotic) to the least.

It was noted that a major omission from the preliminary fuzzy expert system was the lack of representation of metabolic component of the acidemia. As the differentiation of *metabolic* acidemia from *respiratory* acidemia, with their different clinical implications, was a central theme to this work, this was rectified by the inclusion of a third linguistic variable — the *component* of acidemia. This *component* variable had three terms in its term-set: *metabolic*, *mixed* and *respiratory*.

6.3.3 Rule Set

As stated in the introduction to this chapter a new set of fuzzy rules was developed, for both the vessel identification and the interpretation capabilities. Fresh knowledge elicitation sessions were undertaken with the same experts that had developed the crisp rules. The formulation of each of the rule sets is described in more detail in Sections 6.5 and 6.6.

6.3.4 Fuzzy Operators

The probabilistic operator family was retained as previously, although the standard 'Zadeh' operators (min and max) were implemented as an alternative. In elicitation sessions with experts, the fuzzy output sets produced with the probabilistic operators were favoured as they avoided the 'plateau' produced with the 'Zadeh' operators. It was found that the probabilistic family generated smoother transition surfaces for the vessel identification rules and produced higher performance for the interpretation rules.

6.3.5 Membership Functions

Fuzzy sets were modelled with *sigmoid* membership functions, modified slightly from the preliminary system — left-edge sets were modelled by a decreasing *sigmoid*:

$$\mu(x) = \frac{1.01}{1 + e^{5(x-c)/w}} \tag{6.6}$$

right-edge sets were modelled by an increasing *sigmoid*:

$$\mu(x) = \frac{1.01}{1 + e^{5(c-x)/w}} \tag{6.7}$$

and middle sets were modelled by a combination of two *sigmoids*:

$$\mu(x) = \frac{1.01^2}{\left(1 + e^{15(x-c)/w-5}\right)\left(1 + e^{15(c-x)/w-5}\right)} \tag{6.8}$$

where c is the centre point of the fuzzy set (0.5 for left and right-hand sets, and 1.0 for middle sets); w is the width of the fuzzy set; and $\mu(x) = \min(\mu(x), 1.0)$. Note that the numerator was

slightly above 1.0 to ensure that all sets were normal ($\mu_{max} = 1.0$). These functions were designed to create left hand sets which transfer from α to $1-\alpha$, right hand sets which transfer from $1-\alpha$ to α, and middle sets which transfer from α to 1 and back to α over the range $x - c/2$ to $x + c/2$, where $\alpha = 1/(1+e^{2.5}) \approx 0.076$. This is illustrated in Figure 6.1, where left, middle and right sets have all been generated with $c = 50$ and $w = 40$ on a universe of discourse of $(0, 100)$.

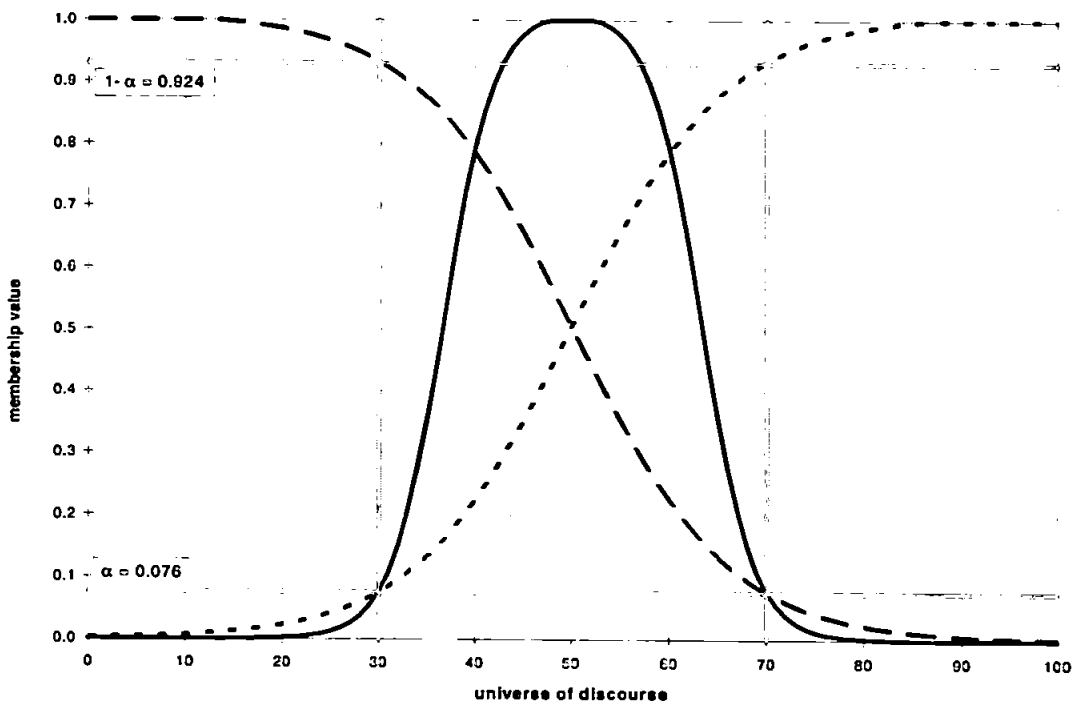


Figure 6.1: *illustration of* left, right *and* mid sigmoid *fuzzy membership functions*

Note that the use of the width parameter, w, is somewhat different from the σ used in the previous chapter. Previously the width had corresponded to the width of the *sigmoid* at height 0.5, but this was changed so that the width corresponded to a large majority of the area of the *sigmoid* being included within this width. This was so that it could be more directly related to the maximum likely uncertainties quantified in Section 6.2. The universe of discourse of each parameter is given in Table 6.3, and the parameters used to generate the membership functions of each term in each linguistic variable are given in Table 6.4. The resultant term sets are shown in Figures 6.2 to 6.9.

Variable	*Universe of Discourse*
pH_A	6.60...7.60
BD_A	0...20
pH_V	6.60...7.60
BD_V	0...20
origin	0...100
acidemia	0...100
component	0...100
duration	0...100

Table 6.3: *universe of discourse for fuzzy input and output variables*

Inputs	*Fuzzy Term*					
	low	*mid*	*normal*	*high*		
	left	middle	middle	right		
pH_A	7.05(0.30)	7.10(0.30)	7.30(0.60)	7.45(0.15)		
	low	*mid*	*high*			
	left	middle	right			
BD_A	10(10)	11(6)	12(4)			
	low	*mid*	*normal*	*high*		
	left	middle	middle	right		
pH_V	7.10(0.30)	7.15(0.30)	7.35(0.60)	7.50(0.15)		
	low	*mid*	*high*			
	left	middle	right			
BD_V	8(10)	9(6)	10(4)			
Outputs						
	same	*mixed*	*different*			
	left	middle	right			
origin	40(40)	50(60)	60(40)			
	severe	*significant*	*moderate*	*mild*	*normal*	*alkalotic*
	left	middle	middle	middle	right	right
acidemia	10(10)	15(20)	30(20)	50(40)	60(40)	90(10)
	metabolic	*mixed*	*respiratory*			
	left	middle	right			
component	25(30)	50(80)	75(30)			
	chronic	*intermediate*	*acute*			
	left	middle	right			
duration	25(30)	50(80)	75(30)			

Table 6.4: *centre (and width) parameters used to generate* sigmoid *membership functions for fuzzy input and output variables*

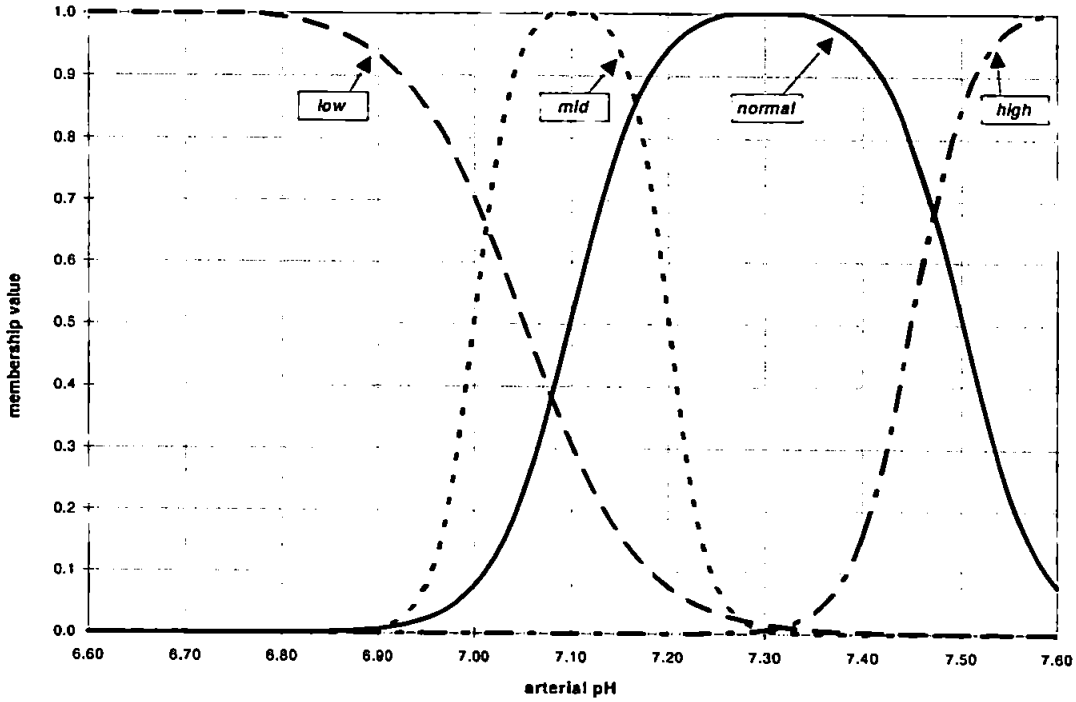


Figure 6.2: *term sets of* arterial pH *fuzzy input variable*

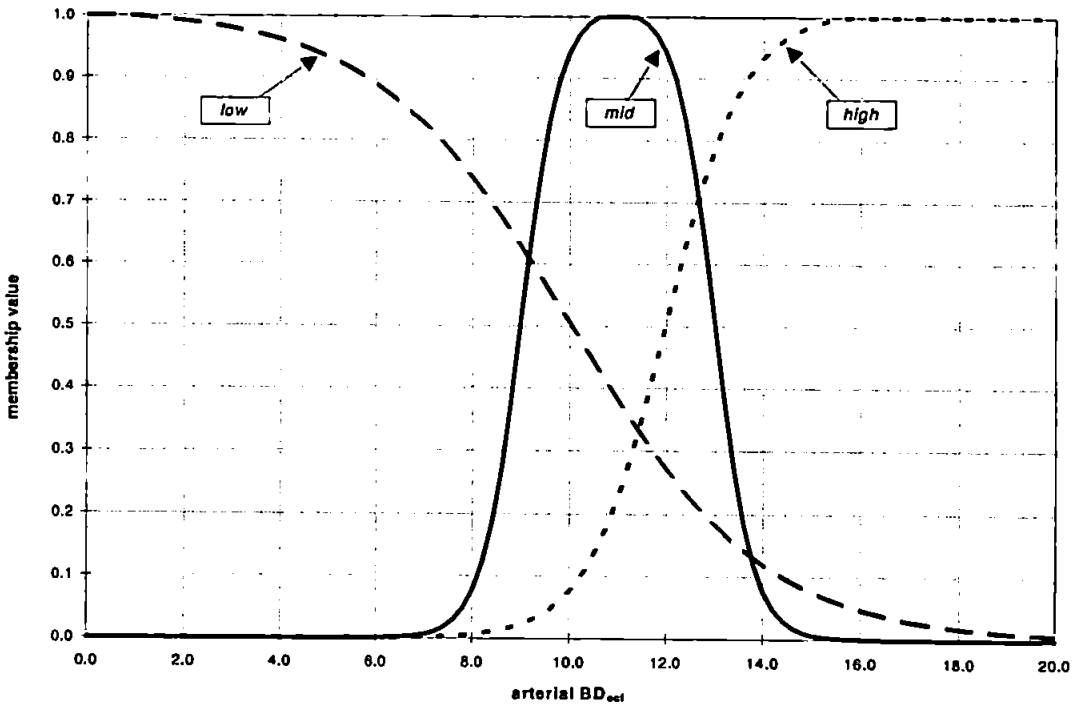


Figure 6.3: *term sets of* arterial BD_{ecf} *fuzzy input variable*

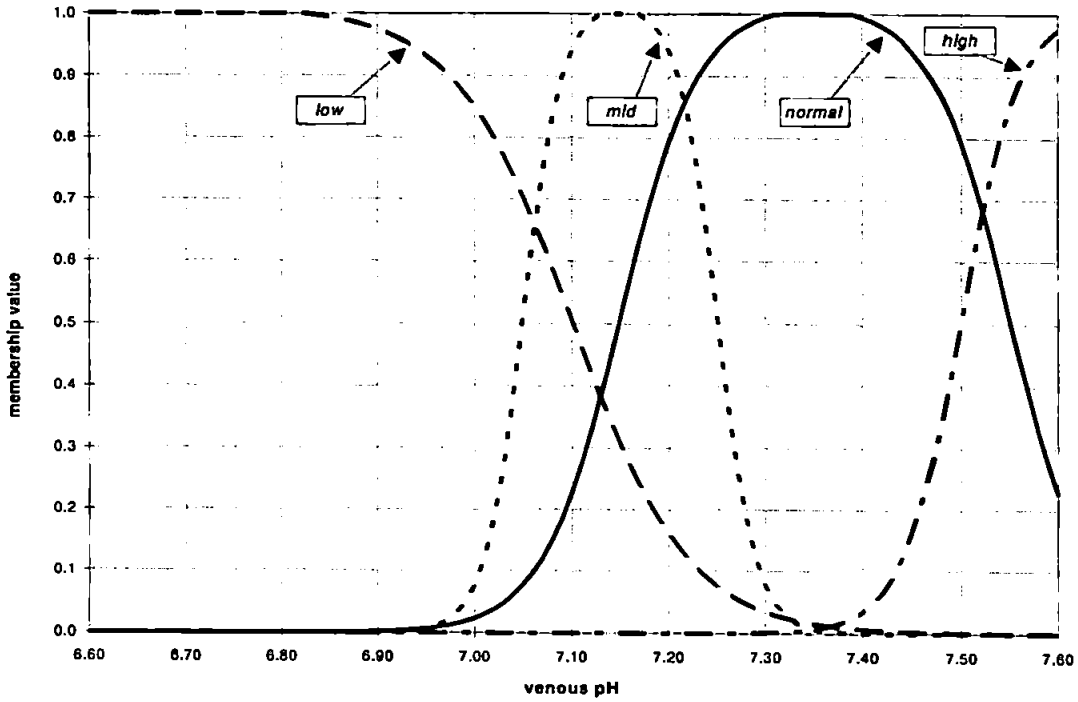


Figure 6.4: *term sets of* venous pH *fuzzy input variable*

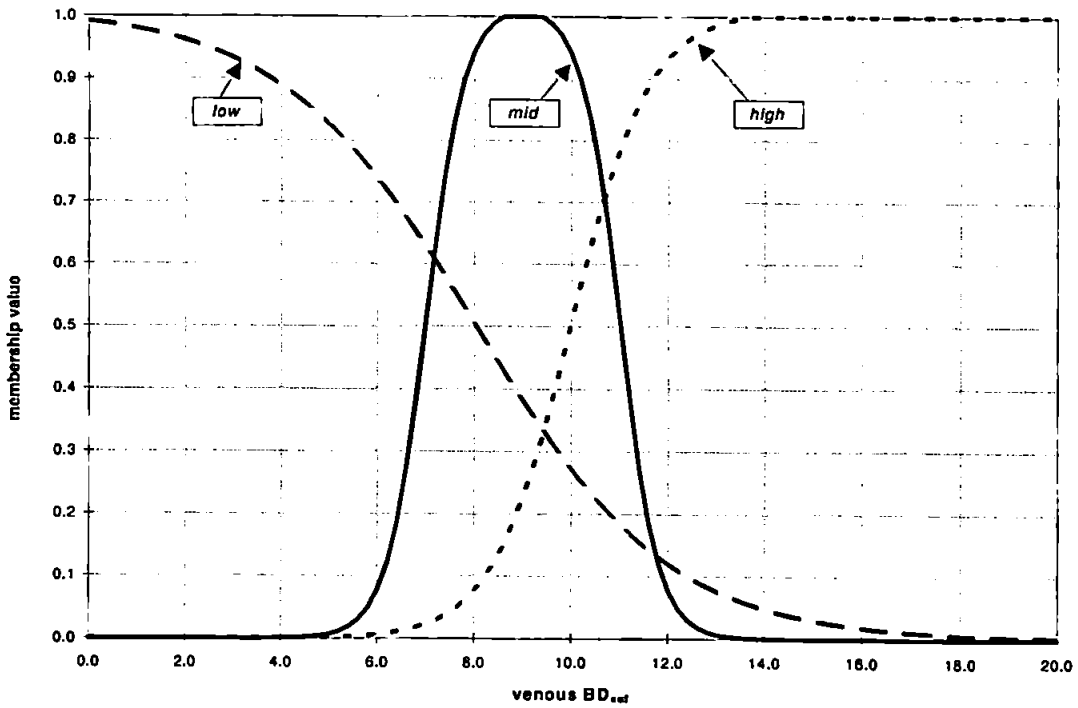


Figure 6.5: *term sets of* venous BD_{ecf} *fuzzy input variable*

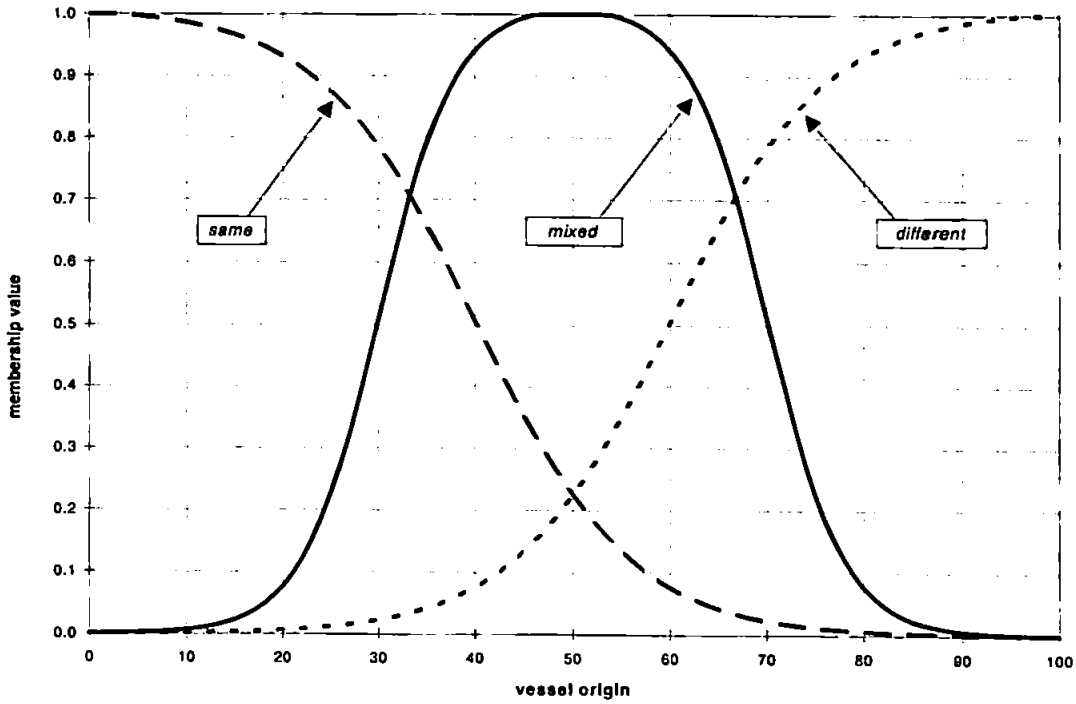


Figure 6.6: *term sets of* origin *fuzzy output variable*

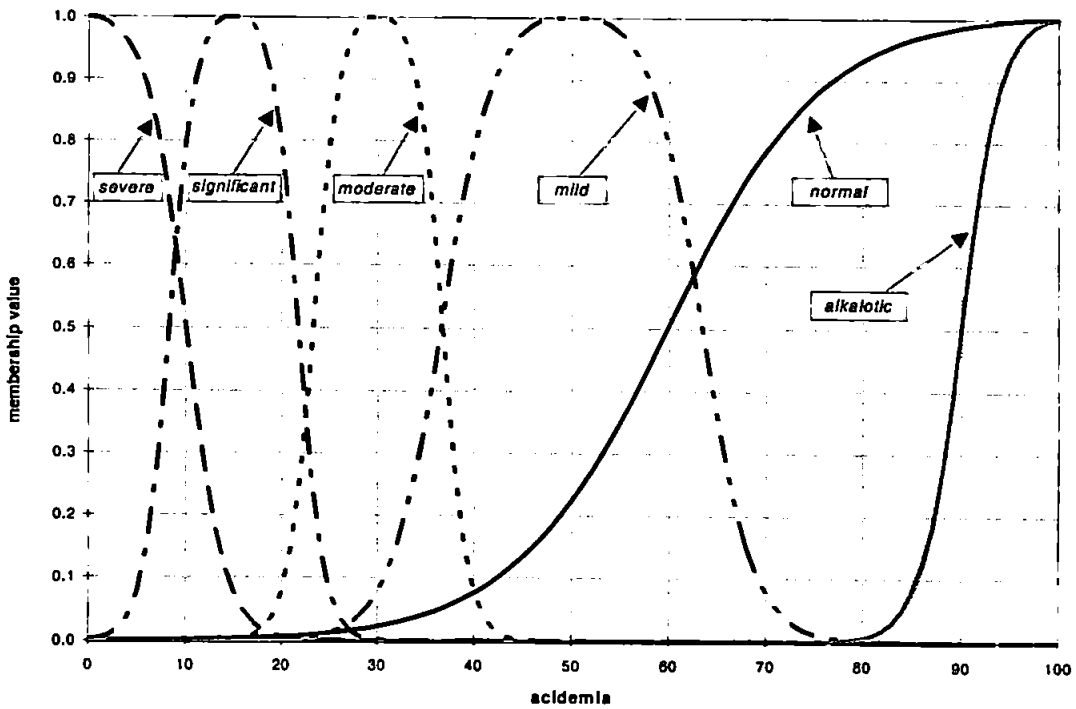


Figure 6.7: *term sets of* acidemia *fuzzy output variable*

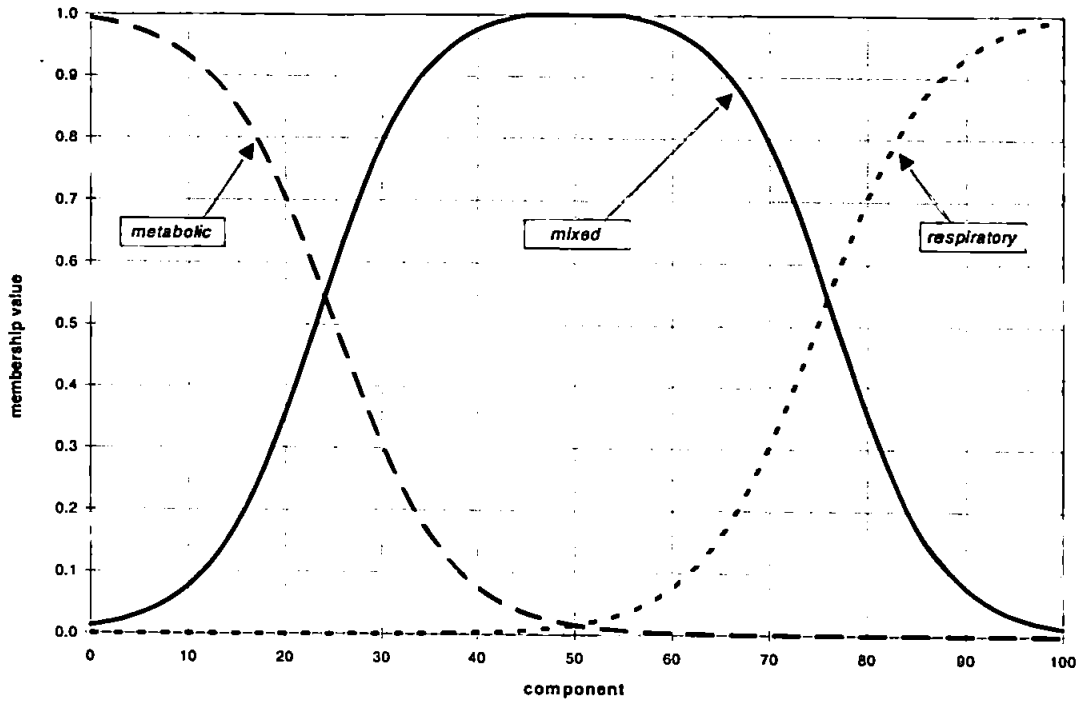


Figure 6.8: *term sets of* component *fuzzy output variable*

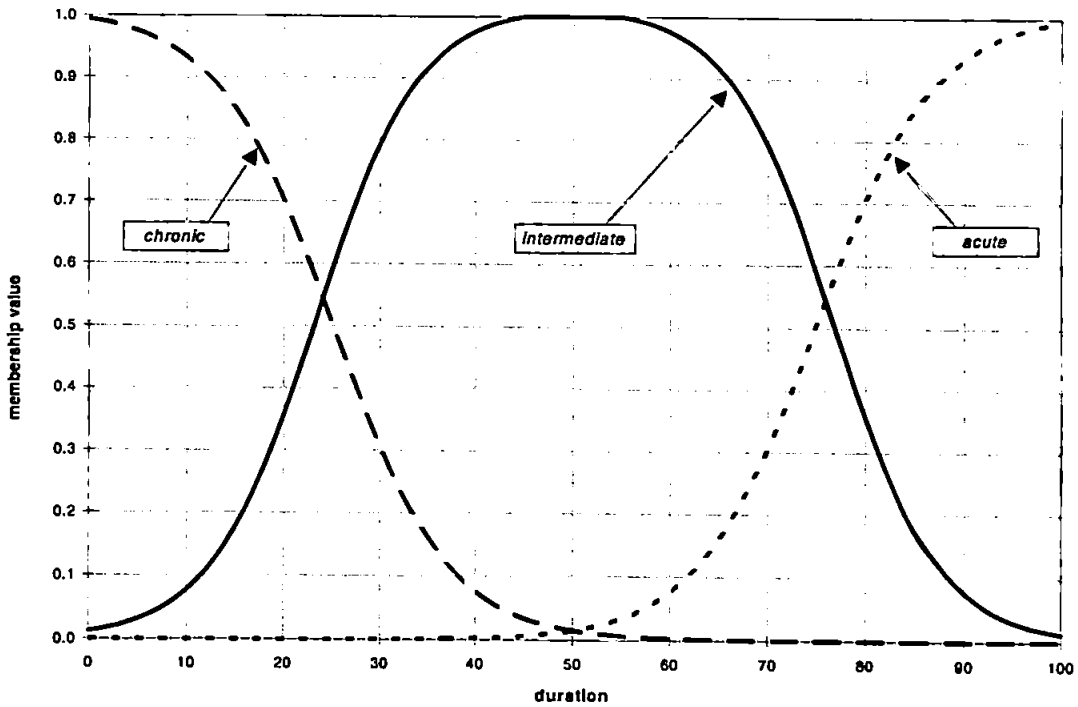


Figure 6.9: *term sets of* duration *fuzzy output variable*

It can be seen from Figures 6.3 and 6.5 that, in contrast to the preliminary system, the slopes of the left-hand *low* and the right-hand *high* base deficit term were different, and also from Figure 6.7 that widths of the terms in the *acidemia* variable differ. These fine tuning changes resulted from expert feedback on the output sets. It may also be noted from Figure 6.7 that the fuzzy set for the *alkalotic* term of the *acidemia* variable is entirely subsumed within the *normal* term. This is contrary to usually accepted practice in fuzzy modelling in which it is generally recommended that for two terms, $\mathbf{x}_1$ and $\mathbf{x}_2$, the function $\sqrt[p]{(x_1^p + x_2^p)}$ is minimised for all x_1 and x_2, where p is an application dependent parameter [96]. However, this condition is usually cited in the context of fuzzy controllers, and again may not be appropriate for fuzzy expert systems.

It was found that having the *alkalotic* term subsumed within the *normal* term caused the centroid output of the *acidemia* variable to remain more constant throughout the period when the pH's traversed the right-hand edges of their *normal* terms and the left-hand edges of their *high* terms. This was because the decreasing *normal* term of *acidemia* caused by output from rule 24 (Appendix H) and the increasing *alkalotic* term caused by output from rule 25 did not alter the centroid until the height of the *alkalotic* term **exceeded** the height of the *normal* term. This behaviour is illustrated in Figure 6.10. The same effect could probably be achieved through the use of asymmetric *normal* terms in the input pH variables and the output *acidemia* variable, but asymmetric terms were not available within the model.

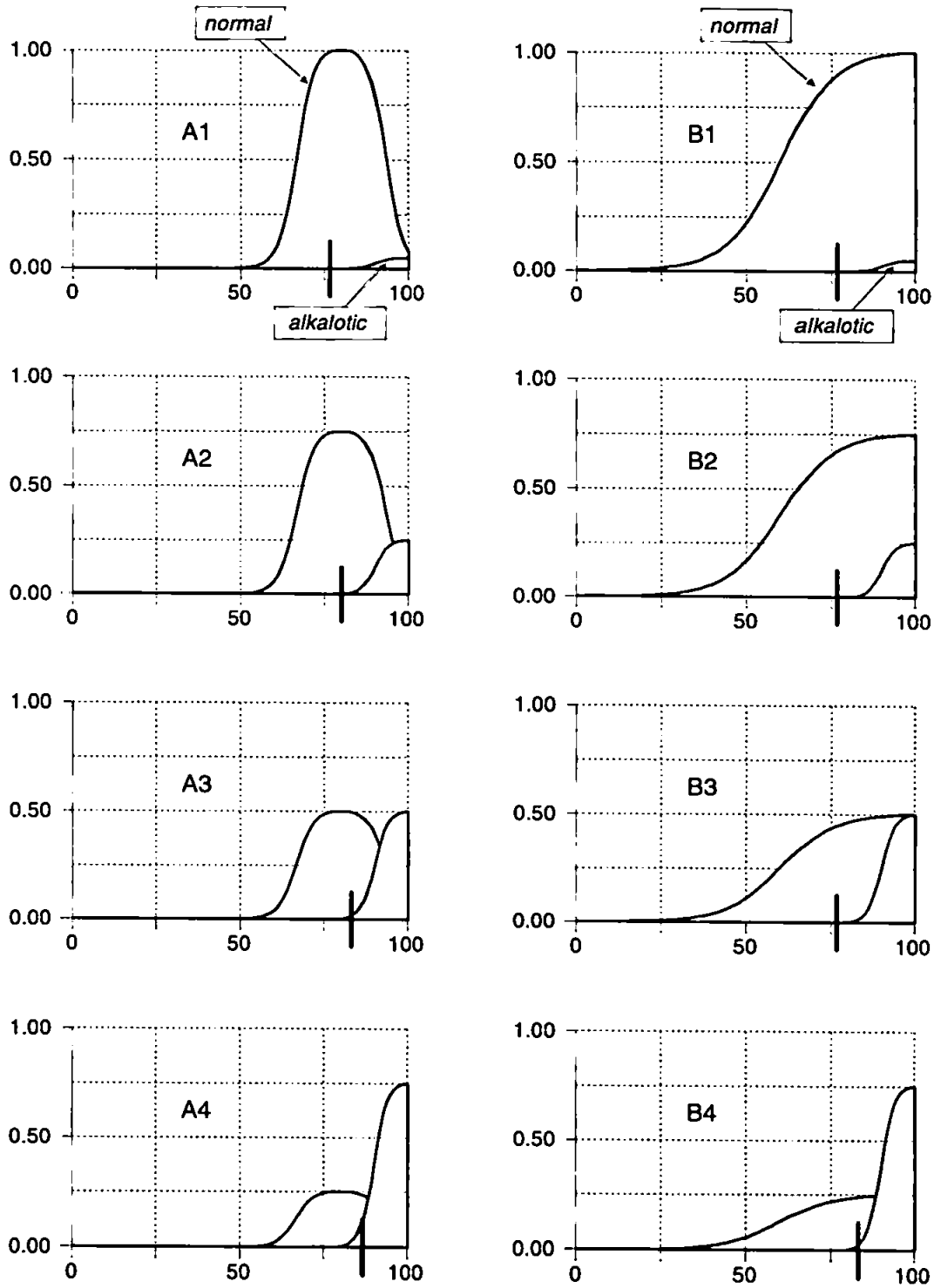


Figure 6.10: *as the* normal *term decreases in height and the* alkalotic *term increases, the centroid gradually shifts with the overlapping terms shown in the left hand sequence (A1 – A4), but remains constant in the subsumed terms shown in the right hand sequence (B1 – B4), until the height of the* alkalotic *term* **exceeds** *the* normal *term*

6.3.6 Relational Operators

Rather than determine the differences between each parameter to form three *delta* fuzzy variables, which would then be related to a specifically generated set of terms, the rules for vessel identification were slightly unusual for a fuzzy expert system, in that they consisted of relationships between two fuzzy input variables,

IF X_A *relop* X_V

THEN *origin* is B_i

where X_A is the arterial variable, X_V is the venous variable and *relop* is a relational operator, rather than the usual

IF ΔX is A_i

THEN *origin* is B_i

Three main relational operators were implemented: *is equal to*, *is greater than* and *is less than*. The *A is equal to B* operator was equivalent to the usual fuzzy membership operator in which the maximum intersection of sets A and B is found, but *A is greater than B* and *A is less than B* require further explanation. Assuming that the sets A and B are regular fuzzy sets with a single maximum point, then *A is greater than B* can be modelled by finding the maximum intersection of A with *not B* for all B to the right of its maximum, i.e. $1-\mu_B(x)$, $\forall x > x_m \mid \mu_B(x_m) = \max_x(\mu_B)$; 0 otherwise. Similarly *A is less than B* can be modelled by finding the maximum intersection of A with *not B* for all B to the left of its maximum, i.e. $1-\mu_B(x)$, $\forall x < x_m$; 0 otherwise. These are illustrated in Figure 6.11 to 6.13.

Each of these three main relational operators can be negated to implement *is not equal to*, *is not greater than* and *is not less than*. For example, *is not less than* is illustrated in Figure 6.14. From comparison of Figures 6.12 and 6.14, it can be seen than *is not less than* is **not** the same as *is greater than*. The relation *is not less than* will be wholly *true* (1.0) for any normalised set A that is superimposed on or to the right of another set B, whereas *is greater than* will be only be wholly *true* (1.0) when set A has its maximum entirely to the right of set B.

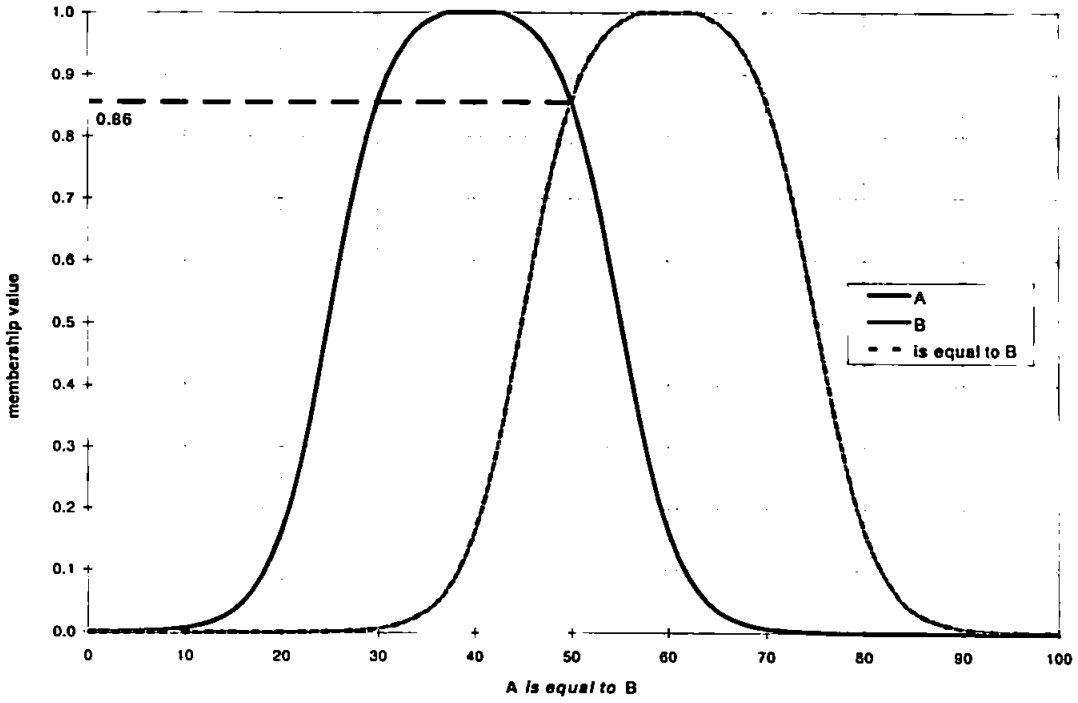


Figure 6.11: *illustration of fuzzy relational operator* is equal to

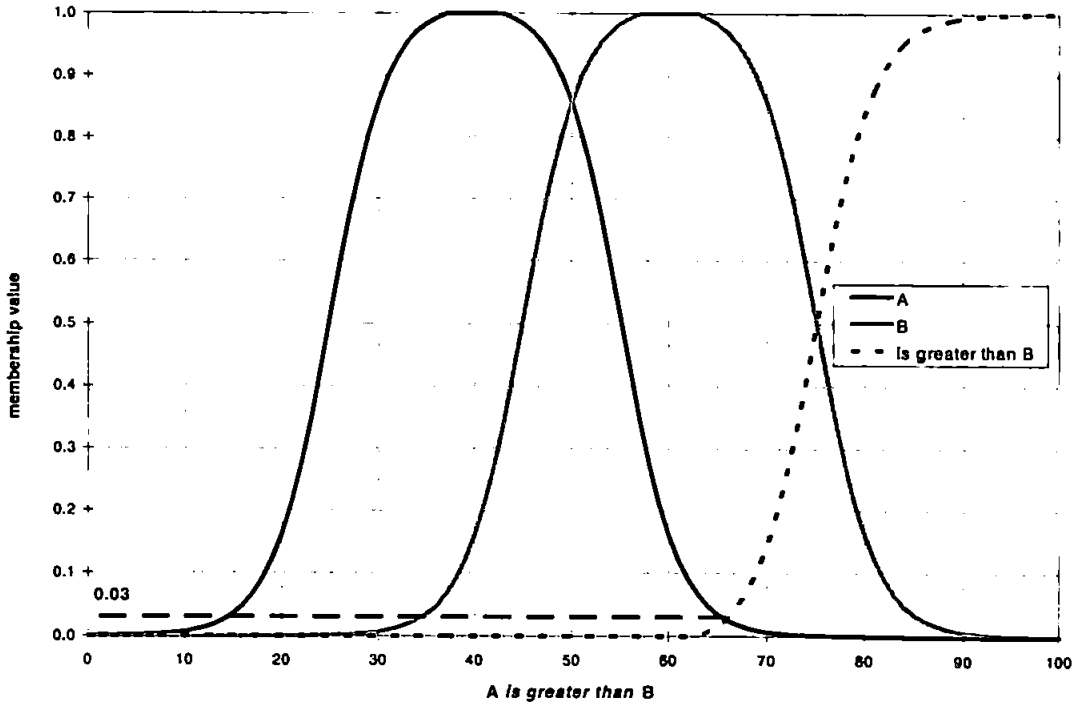


Figure 6.12: *illustration of fuzzy relational operator* is greater than

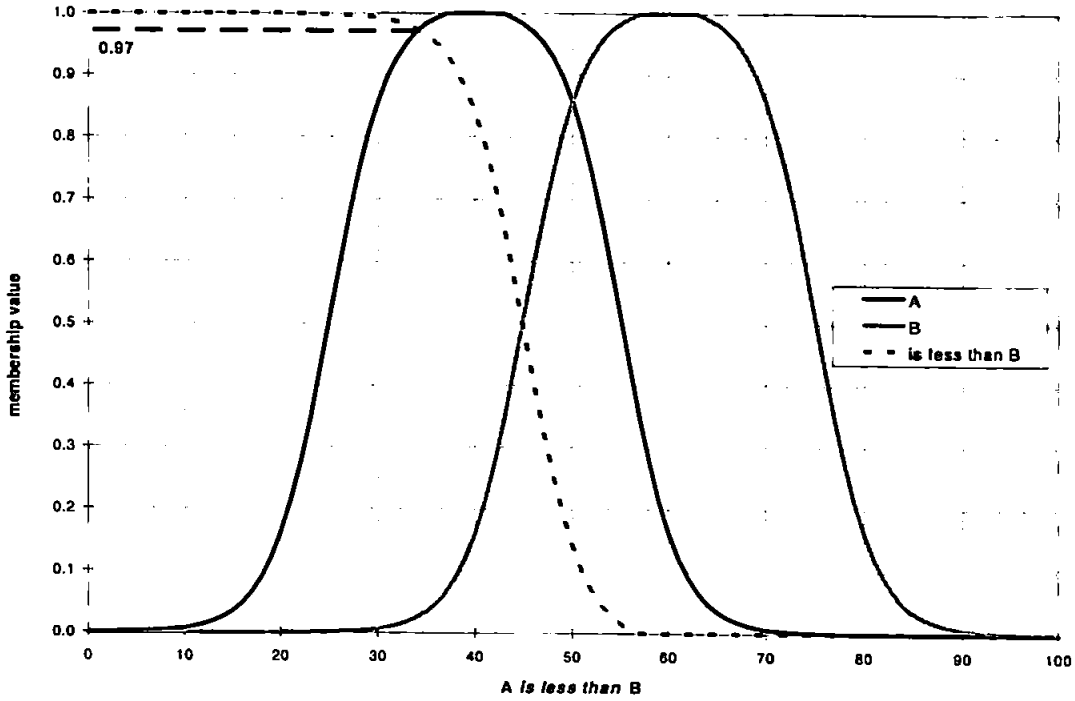


Figure 6.13: *illustration of fuzzy relational operator* is less than

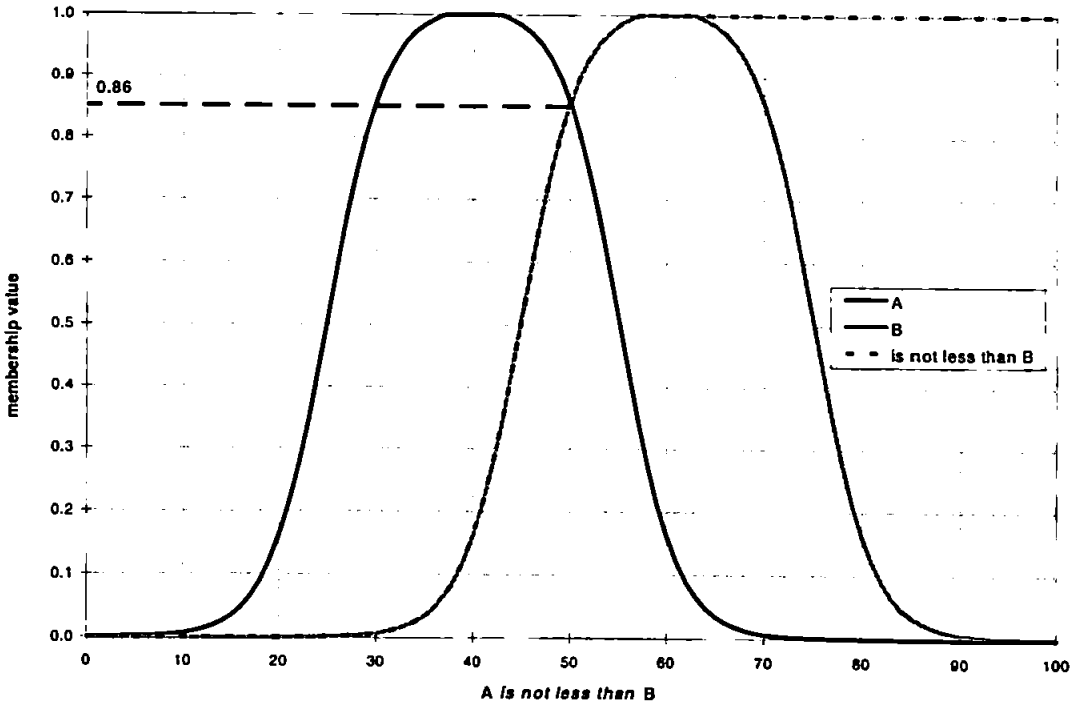


Figure 6.14: *illustration of fuzzy relational operator* is not less than

6.3.7 Parameter Weighting and Operator Precedence

As the relational operators described in Section 6.5 were implemented through an algorithm which located the point where the membership function of the right-hand variable of the relation reached unity, the normality of membership functions was required for fuzzy variables involved in such relationships. This was coupled with the fact that weighting of the pCO_2 and pO_2 parameters in the vessel identification rules was required, as the extent to which this additional information was utilised by 'experts' was unclear. Hence, a method of weighting parameters that did not involve sub-normal membership functions was developed.

The main property required for parameter weighting was that a term which received a weighting of one should make its normal full contribution to the fuzzy rule condition in which it appeared, while a term which received a weighting of zero should make no contribution to the rule condition. This property implied that the weighting factor must depend on the connective form of the rule condition in which the term appeared. For example, if the weight was zero and the term appeared within a disjunction (*or*), then the condition should evaluate to zero to have no effect, as x *or* $0 = x$. Alternatively, if the weight was zero and the term appeared within a conjunction (*and*), then the condition should evaluate to one to have no effect, as x *and* $1 = x$. To provide this behaviour, the following weighting scheme was adopted:

$$\textit{or} \qquad \mathbf{Wt}^{\text{or}}(\textit{truth}) = \textit{truth} * \textit{weight}$$

$$\textit{and} \qquad \mathbf{Wt}^{\text{and}}(\textit{truth}) = 1 - (1 - \textit{truth}) * \textit{weight}$$

$$= \textit{truth} * \textit{weight} + 1 - \textit{weight}$$

It can be seen that in both cases the *truth* is unaffected by a *weight* of 1.0, but a *weight* of zero gives a *truth* of 0.0 for $\mathbf{Wt}^{\text{or}}$, and a *truth* of 1.0 for $\mathbf{Wt}^{\text{and}}$. A weight of a half (0.5) would give $\textit{truth}/2$ for $\mathbf{Wt}^{\text{or}}$, and $\textit{truth}/2 + 0.5$ for $\mathbf{Wt}^{\text{and}}$.

The application of the above strategy was straightforward in the case of rules with exactly two antecedents — the weighting equation appropriate to the joining connective was applied to both terms *before* the operator was applied. However, in the case of one, or more than

two antecedents the application was less straightforward. In the case of one antecedent, the solution lies in the type of inference methodology adopted. With *Mamdani* type inference, in which the output term-set is initialised to zero across the universe of discourse, and then each rule that fires *adds* information to the output set, the intention must be to *not* fire a rule where the antecedent term has a zero weight. In the case of multiple (> 2) antecedents, the solution lies in the precedence of the surrounding operators, such that weighting appropriate to the *innermost* operator is applied. For example, consider a rule antecedent of *A and B or C* in which weights are applied to *A*, *B*, and *C*. If the intention of the antecedent is $((A \textit{ and } B) \textit{ or } C)$, then $\mathbf{Wt}^{\text{and}}$ is applied to *A* and *B*, and $\mathbf{Wt}^{\text{or}}$ is applied to *C*, whereas if the intention of the antecedent is $(A \textit{ and } (B \textit{ or } C))$, then $\mathbf{Wt}^{\text{and}}$ is applied to *A*, and $\mathbf{Wt}^{\text{or}}$ is applied to *B* and *C*.

For the fuzzy expert system developed here, a simplistic scheme was adopted whereby the truth of the rule was initialised with the value of its first antecedent weighted by the immediately following operator. Each antecedent was then applied in sequence to evaluate the overall truth, with each antecedent having the weight equation of its immediate operator applied. Hence, the ordering of the antecedents in the rule governed the precedence of operators. Thus

IF	*A*	would be coded as:	IF	$\mathbf{Wt}^{\text{or}}(A)$
THEN	...		THEN	...

while

IF	(*A and B*) *or C*	would be coded as:	IF	$\mathbf{Wt}^{\text{and}}(A)$
THEN	...		AND	$\mathbf{Wt}^{\text{and}}(B)$
			OR	$\mathbf{Wt}^{\text{or}}(C)$
			THEN	...

and

IF	*A and (B or C)*	would be coded as:	IF	$\mathbf{Wt}^{\mathrm{or}}(B)$
THEN	...		OR	$\mathbf{Wt}^{\mathrm{or}}(C)$
			AND	$\mathbf{Wt}^{\mathrm{and}}(A)$
			THEN	...

An example of such coding can be seen in rules 3 and 4 of the vessel identification rules in Appendix G.

6.4 Linguistic Approximation

Linguistic approximation of the output sets was implemented for each of the fuzzy output variables, as described in Section 2.4.10. The metric defined by Equation 2.65 was chosen, simply because the calculation of area of fuzzy sets was already implemented, and thus no additional programming was required.

In order to limit the search of linguistic terms which might best match the output, an algorithm was employed to search only:

- the primitive terms,
- the union of two adjacent primary terms,
- the negation of primitive terms, and
- the level sets *undefined*, *indeterminate* and *unknown*.

Even the application of the basic hedges *very* and *slightly* was excluded from consideration. This obviously seriously limited the search of possible target matches, but had the advantage that complexity was limited, so that the linguistic output remained comprehensible. Thus for example, when performing linguistic approximation on the *origin* output variables, the only target sets evaluated were: *same*, *mixed*, *different*, *same/mixed*, *mixed/different*, *not same*, *not mixed*, *not different*, *undefined*, *indeterminate* and *unknown*. In fact for the validation process described later, the linguistic target sets considered were even further restricted. Details of which target outputs were actually allowed for each variable are given later.

6.5 Vessel Identification Rules

6.5.1 Rule Development

The crisp rule for vessel identification:

> IF (venous − arterial pH) < 0.025
>
> THEN mark both samples as 'same vessel: probably vein'
>
> ELSE mark lowest pH as 'arterial vessel'; highest pH as 'venous vessel'

presented in Section 3.2.3 was derived from the physiological expectation of a difference between vessels, combined with a plausibility argument based on the existence of apparent humps around zero on the vessel differences graphs (Figures 3.2 to 3.4).

At the time of creation of the expert system rule, there had been discussion amongst the experts as to whether the combination of all three parameters should have been used to confirm the existence of a paired arterial-venous sample. Subsequently, it appeared that clinicians did seem to use the combination of all three parameters, if asked to perform vessel identification manually. Having quantified the likely errors in each of the parameters, the possibility arose of developing a more sophisticated fuzzy rule-based identification scheme. It may be noted that the uncertainty in arterial pH, δpH in Table 6.1, was found to be the same as the 0.025 cut-off used in the crisp vessel identification rule above. Although this is probably coincidental, it does add an additional plausibility to the crisp cut-off.

If two good venous samples were obtained, then each parameter should differ by amounts close to, or less than, the venous values shown in Table 6.1. As two samples may both be accidentally obtained from the vein, both from the arteries, one may be mixed arterial-venous, or both may be mixed, a 'safe' vessel identification rule may be that if all parameters differ by more than the largest uncertainties in Table 6.1, then the samples can definitely be taken as a true arterial-venous pair. Given that the lowest pH is initially labelled as the artery, or if the pH's are the same, the highest pCO_2 is labelled as the artery, or if the pH's and pCO_2's are the same, the lowest pO_2 is labelled as the artery, then the list of possible

sample differences and their proposed vessel identification are shown in Table 6.5. A '0' indicates that $\Delta parameter$ is zero, a '−' indicates that $\Delta parameter$ is negative, and a '+' indicates that $\Delta parameter$ is positive.

ΔpH	ΔpCO_2	ΔpO_2	*Origin*
0	0	0	definitely same
0	0	+	probably same
0	+	0	probably same
0	+	−	probably mixed
0	+	+	probably different
+	0	0	probably same
+	0	−	probably mixed
+	0	+	probably different
+	−	0	probably mixed
+	−	−	definitely mixed
+	−	+	definitely mixed
+	+	0	probably different
+	+	−	definitely mixed
+	+	+	definitely different

Table 6.5: *list of possible sample differences and the proposed vessel identification*

A fuzzy rule-base was designed to produce the target behaviour shown in Table 6.5, with smooth transitions between each of the categories. The rule-base consisted of a set of five rules relating the differences in fuzzy input parameters (the pH, pCO_2, and pO_2 in both samples) to a single fuzzy output variable, the *origin* of samples. Each input parameter was first fuzzified to possess a width equal to the largest (arterial) uncertainties in Table 6.1, by Equation 6.8. The fuzzified input variables were then passed through the vessel identification rule-set (see Appendix G). Four hedge operators were implemented in the fuzzy model: *very*, *roughly*, *about*, and *vaguely*. These were modelled as: $very \equiv \mu^2$, $roughly \equiv \mu^{\frac{1}{2}}$, $about \equiv \mu^{\frac{1}{4}}$, and $vaguely \equiv \mu^{\frac{1}{8}}$. These hedges were used to modify the width of individual parameters within rule clauses, until the output of the FES was found to behave as specified in Table 6.5. In the final rule set only three out of the four possible hedges were actually utilised: *very*, *about*, and *vaguely*.

6.5.2 Performance Evaluation

The output from the vessel identification rules obtained for values of ΔpH from 0.00 to 0.09, ΔpCO_2 from -1.2 kPa to $+1.2$ kPa, and ΔpO_2 from -1.2 kPa to $+1.2$ kPa is shown in Figures 6.15 to 6.24. The graphs show the centre-of-gravity of the *origin* fuzzy output variable obtained for each combination of the input parameters shown as a three-dimensional surface where the surface colour is governed by the actual membership value of the resultant output set at the centre-of-gravity. The surface colour is dark for lower membership values (black for $\mu = 0.0$), gradually getting lighter as the membership value increases (white for $\mu = 1.0$), giving an indication of confidence in the result (see Section 2.4.10). For example, if the input values $\Delta\text{pH} = 0.01$, $\Delta pCO_2 = 0.4$ kPa, and $\Delta pO_2 = 0.0$ kPa resulted in a centre-of-gravity of 55 with a membership of 0.6, then the surface in Figure 6.16 would be plotted at height 55 and shaded at 0.6 brightness.

With a low value of the *origin* variable indicating samples from the same vessel, a mid value (50) indicating mixed samples, and a high value indicating samples from different (arterial and venous) vessels, it can be seen that the surface behaves as required. There is a dip in values when $\Delta = 0$ for all parameters, which gradually disappears as ΔpH increases, whilst simultaneously a high plateau appears at positive ΔpCO_2 and ΔpO_2. All contradictory results produce an output near the mid point.

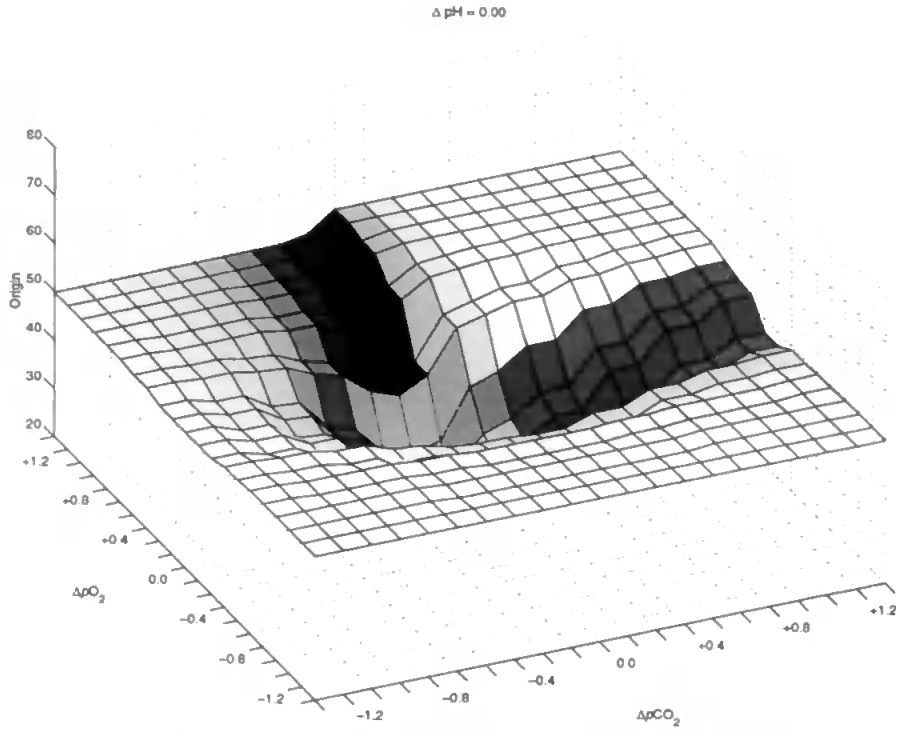


Figure 6.15: *surface plot of the centre-of-gravity of* origin *fuzzy output variable for various* ΔpCO_2 *and* ΔpO_2*, at constant* $\Delta pH = 0.00$

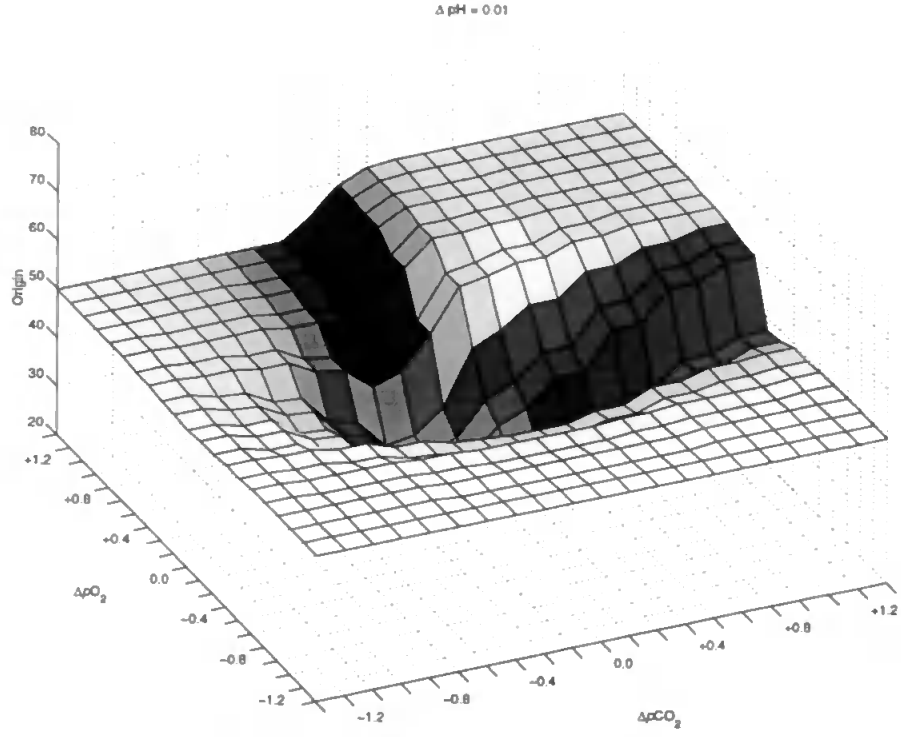


Figure 6.16: *surface plot of* origin *at* $\Delta pH = 0.01$

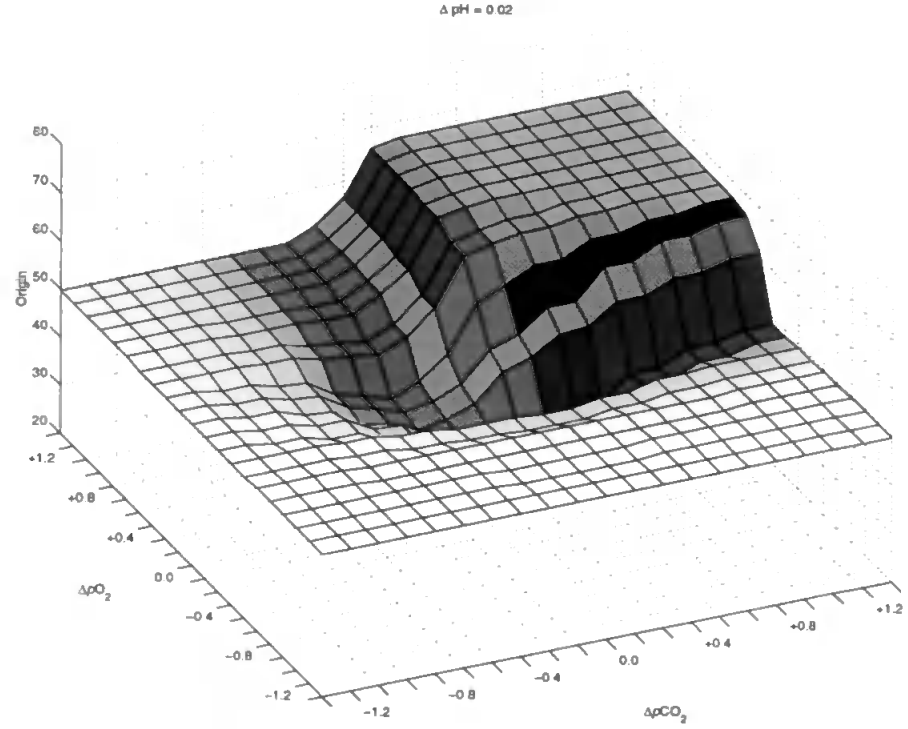


Figure 6.17: *surface plot of* origin *at* $\Delta pH = 0.02$

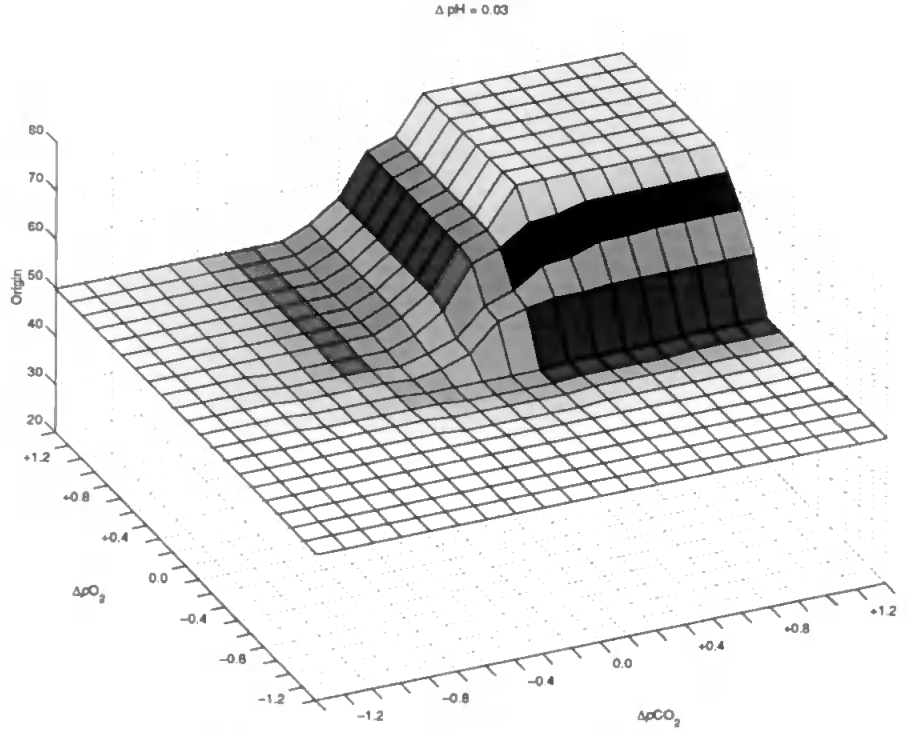


Figure 6.18: *surface plot of* origin *at* $\Delta pH = 0.03$

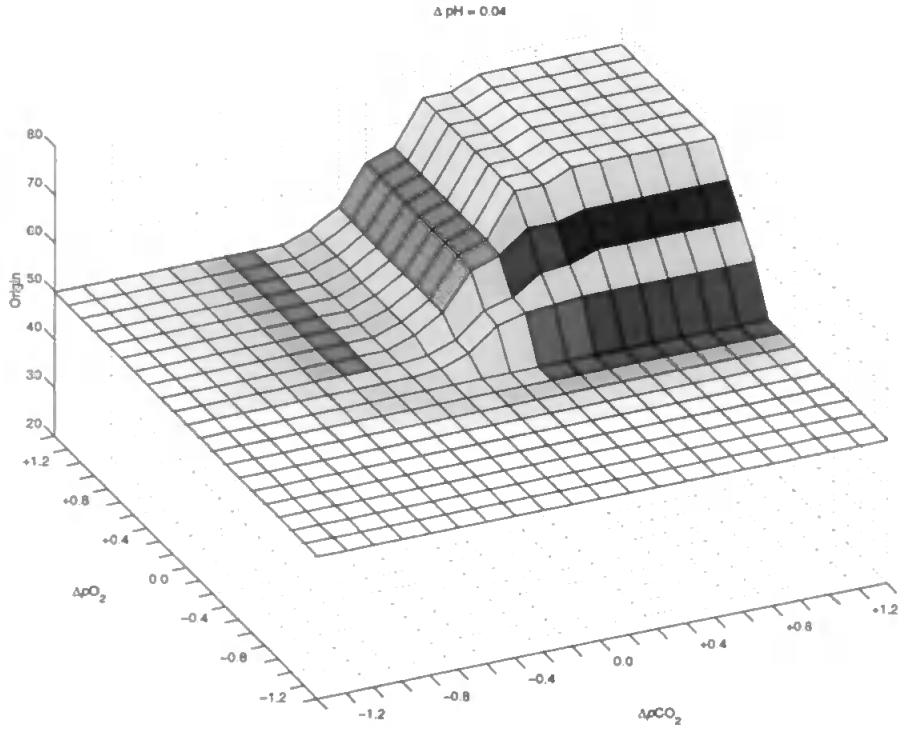


Figure 6.19: *surface plot of* origin *at* $\Delta pH = 0.04$

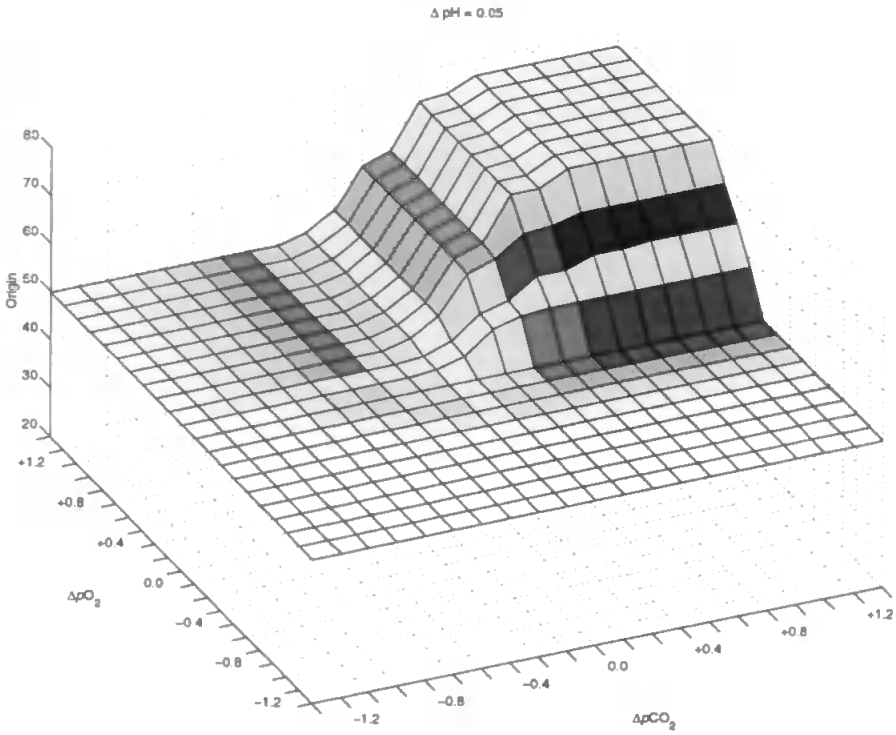


Figure 6.20: *surface plot of* origin *at* $\Delta pH = 0.05$

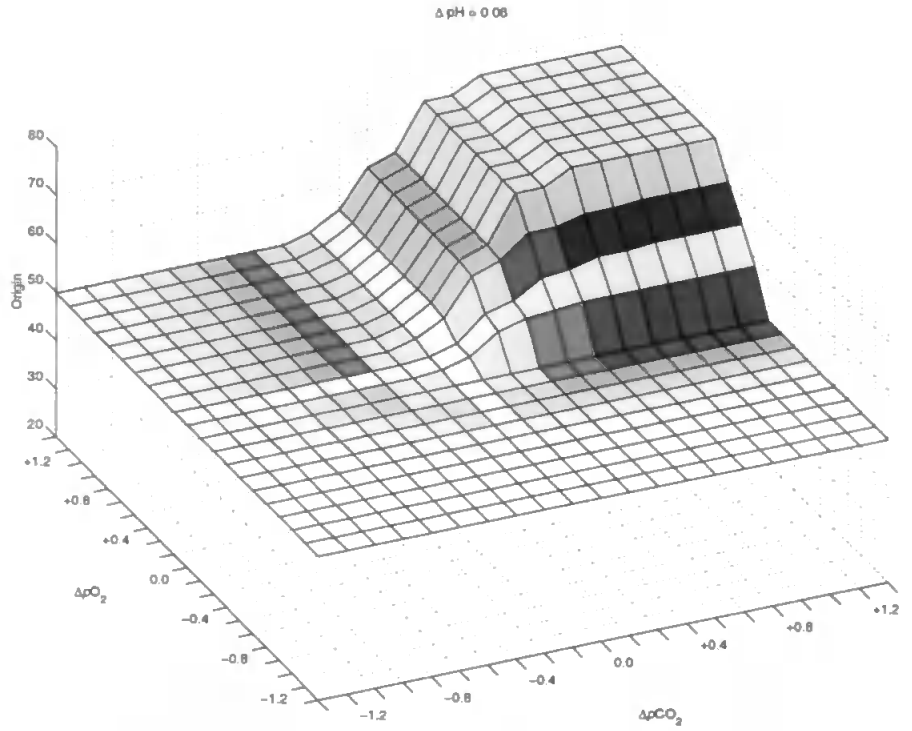


Figure 6.21: *surface plot of* origin *at* $\Delta pH = 0.06$

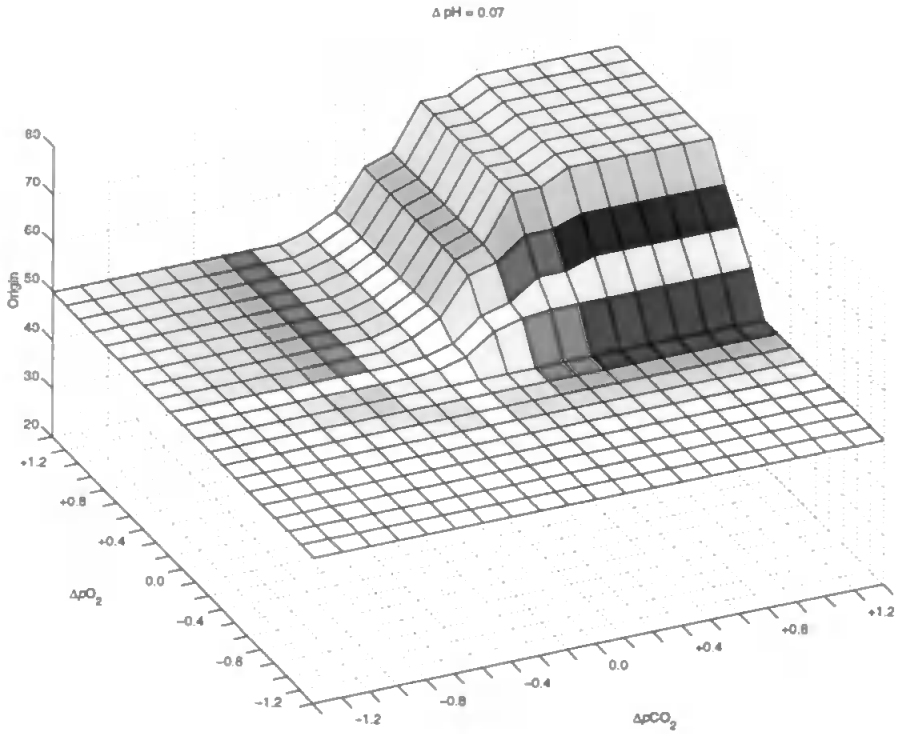


Figure 6.22: *surface plot of* origin *at* $\Delta pH = 0.07$

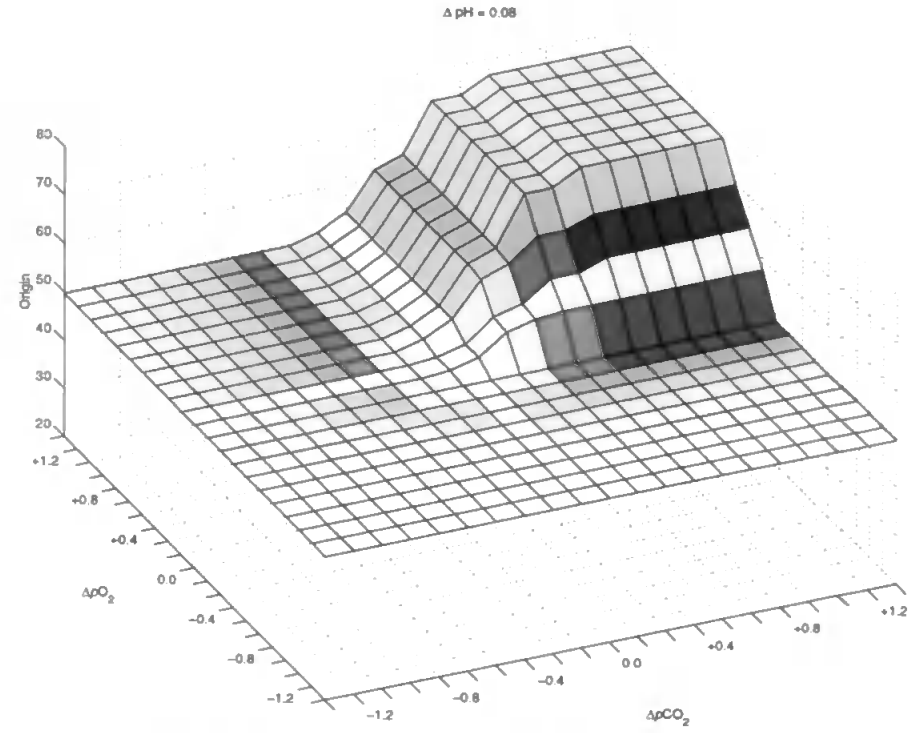


Figure 6.23: *surface plot of* origin *at* $\Delta pH = 0.08$

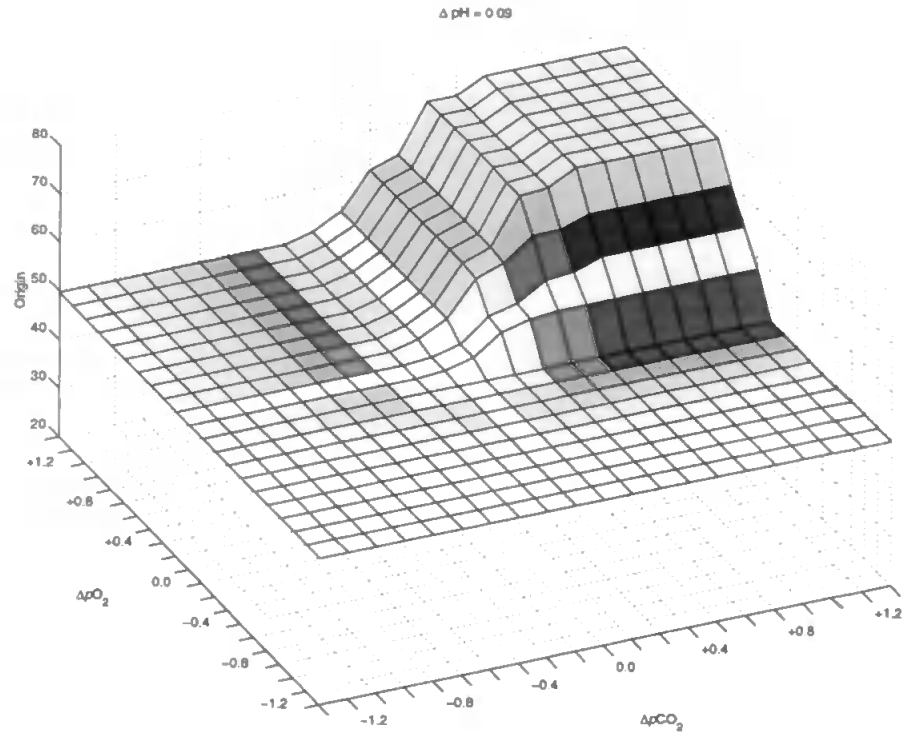


Figure 6.24: *surface plot of* origin *at* $\Delta pH = 0.09$

6.6 Interpretation Rules

6.6.1 Rule Development

As in the crisp expert system, the integrated fuzzy expert system was developed to interpret the results *after* the physiological plausibility checks and vessel identification had been performed.

The vessel identification scheme was slightly more complex than previously. If two input samples were received, they were passed through the vessel identification rules, and the linguistic approximation output of the *origin* variable was used to determine the appropriate vessel labelling. A linguistic output corresponding to *different*, *mixed/different* or *not same* was considered sufficient for the vessels to be labelled as an arterial-venous pair. Any other linguistic output, or the presence of only one input sample, caused the sample to be labelled as a single venous vessel. An arterial-venous pair would then have its input variables re-initialised with the crisp values of the input parameters, fuzzified to have a width equal to those in Table 6.1. The BD_{ecf} for each vessel is then calculated from Equation 1.13 and then fuzzified to the width given by Equation 6.5.

If both vessels were missing, both the arterial and venous parameters would be initialised with *unknown* across the universe of discourse. In such a situation all rules fire with maximum strength and the output of all variables tends to *unknown*. In practice such a situation is very rare (1 case out of > 10000), and the much more common occurrence is the single vessel. With such a single vessel labelled as venous for reasons already stated in Chapter 3, it might be thought that the arterial parameters would simply be initialised with *unknown*. However, this ignores the fact that, physiologically, the arterial parameters would be such as to maintain positive Δ's — i.e. if the arterial pH was not known, it could still be assumed to be lower than the venous pH. Thus, the actual procedure was to initialise the arterial pH and pCO_2 input variables with a fuzzy set consisting of the inverted left-hand edge of the venous fuzzy set, and to initialise the pCO_2 and BD_{ecf} input variables with a fuzzy set consisting of the inverted right-hand edge of the venous fuzzy set. This is demonstrated in

Figure 6.25, in which missing arterial values have been set to *unknown* relative to venous values of pH = 7.10, pCO_2 = 8.0 kPa, pO_2 = 5.0 kPa and BD_{ecf} = 9.7 mmol.l^{-1}.

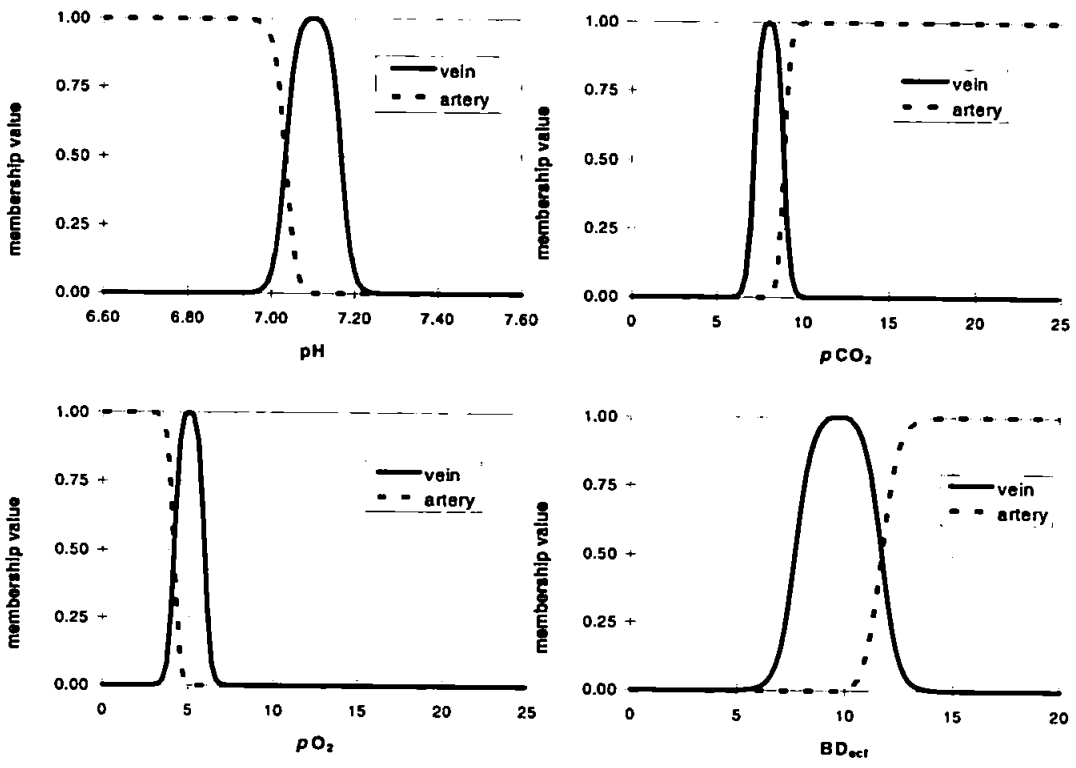


Figure 6.25: *illustration of* unknown *arterial values relative to venous values of pH = 7.10, pCO_2 = 8.0 kPa, pO_2 = 5.0 kPa, and BD_{ecf} = 9.7 mmol.l^{-1} (note that parameter widths have been exaggerated for clarity)*

The basic principles of acid-base analysis elicited from the experts were that:

- *acidemia* is based on the absolute value of arterial pH (lower arterial pH implies worse *acidemia*), refined by the value of the venous pH;
- *component* is based on arterial BD_{ecf} (high BD_{ecf} implies *metabolic* component, low BD_{ecf} implies *respiratory* component), refined by venous BD_{ecf}; and
- *duration* is based on pH and BD_{ecf} differences (smaller differences imply *chronic* duration, larger differences imply *acute* duration), refined by absolute arterial values.

These basic principles were encapsulated in the fuzzy rules such that there was smooth transition over all input and output sets. This ensured that, as far as possible, continuous changes

in input parameters resulted in continuous changes in the fuzzy output sets. A synopsis of the fuzzy rule set is shown in Table 6.6 in which this behaviour can be seen; the actual fuzzy rule set is included in Appendix H. Note that the term *normal* pH also includes *high* pH in all rules except rules 24 and 25. The symbol '—' indicates any value of the parameter in the context of input variables, and no consequence result in the context of output variables — i.e. the variable is not utilised in the rule.

	Input Variables				*Output Variables*		
Rule	pH_A	BD_A	pH_V	BD_V	*acidemia*	*component*	*duration*
1	low	high	low	high	very severe	metabolic	very chronic
2	low	high	low	mid	severe	metabolic	chronic
3	low	high	low	low	severe	metabolic	intermediate
4	low	high	mid	—	severe	metabolic	intermediate
5	low	high	normal	high	severe	metabolic	intermediate
6	low	high	normal	not high	severe	metabolic	acute
7	low	mid	low	not low	significant	mixed	chronic
8	low	mid	low	low	significant	mixed	intermediate
9	low	mid	mid	—	significant	mixed	intermediate
10	low	mid	normal	high	significant	mixed	acute
11	low	mid	normal	not high	significant	mixed	acute
12	low	low	low	not high	moderate	respiratory	chronic
13	low	low	mid	not high	moderate	respiratory	intermediate
14	low	low	normal	not high	moderate	respiratory	acute
15	mid	high	mid	not low	moderate	metabolic	chronic
16	mid	high	mid	low	moderate	metabolic	intermediate
17	mid	high	normal	—	moderate	metabolic	acute
18	mid	mid	mid	high	mild	mixed	chronic
19	mid	mid	mid	not high	mild	mixed	intermediate
20	mid	mid	normal	high	mild	mixed	intermediate
21	mid	mid	normal	not high	mild	mixed	acute
22	mid	low	mid	not high	mild	respiratory	intermediate
23	mid	low	normal	not high	mild	respiratory	—
24	normal	—	normal	—	normal	—	—
25	high	—	high	—	alkalotic	—	—

Table 6.6: *a synopsis of the fuzzy interpretation rule set*

6.6.2 The Possibility of Intrapartum Asphyxial Damage

During development of the preliminary fuzzy expert system the experts' opinions were obtained on the ordering of fifty cases, in terms of the *probability that the infant had suffered intrapartum asphyxial damage*, where this opinion was formed on the basis of the acid-base results only. Having successfully tuned the preliminary fuzzy expert system to this ordering, it was realised that an output from the fuzzy expert system that provided such an ordering could form the basis of a novel interpretation of umbilical cord acid-base balance.

The centroids of the integrated fuzzy expert system were combined into a single index by:

$$condition = acidemia + \frac{component}{20} + \frac{duration}{10} \tag{6.9}$$

where the relative weighting of the three terms was determined empirically. Given that the three output variables are arranged in such a way that low scores indicate a worsening condition for the infant, to the extreme *severe, metabolic, chronic acidemia*, this index can be thought of as indicating the *health* of the infant as represented by its acid-base balance at birth. As explained in the medical introduction, there are many other factors, such as congenital abnormality, that can cause an infant to be in very poor health, despite suffering no asphyxial damage during labour and having a 'healthy' acid-base balance. Hence the acid-base *condition* variable can only be used to reflect the chance that the infant suffered asphyxial damage during labour — henceforth referred to as *intrapartum asphyxial damage*. Although this variable should reflect the chance of *intrapartum asphyxial damage*, it could not be used as a probability without knowledge of other intrapartum clinical factors, such as the gestational age of the fetus, the birth weight, and the intrapartum CTG, *and* other immediate neonatal factors such as Apgar scores and the need for resuscitation.

However, the *condition* variable could be interpreted as indicating the compatibility of the acid-base balance at birth with *intrapartum asphyxial damage* — the lower the *condition* variable is, the 'worse' the acid-base results were, and consequently, the more compatible the state of the infant was with it having suffered *intrapartum asphyxial damage*. The concept of *compatibility* is captured by the mathematical formalisation of *possibility theory*

described in Section 2.3.2. It is suggested that a novel interpretation of acid-base balance is to directly relate the severity of acid-base results with the *possibility* of the infant having suffered *intrapartum asphyxial damage*. With the most severe ('worst') acid-base results, the *possibility* of the infant having suffered *intrapartum asphyxial damage* will be one; the results are entirely compatible with *intrapartum asphyxial damage*. With entirely normal acid-base results, the *possibility* of the infant having suffered *intrapartum asphyxial damage* will be zero; the results are completely incompatible with *intrapartum asphyxial damage*.

On the assumption that the *condition* output of the fuzzy expert system does provide a true and accurate reflection of the severity of acid-base results, a relationship is required to map the *condition* variable to a *possibility* measure. With no other information available to support a more complex mapping, a simple linear mapping was implemented. Therefore the *condition* index, subtracted from its maximum value and normalised to its range, gave the *possibility of intrapartum asphyxial damage* (PIAD) index:

$$PIAD = \frac{condition_{\max} - condition}{condition_{\max} - condition_{\min}} \tag{6.10}$$

It is interesting to note how naturally this interpretation maps onto the conventional clinical understanding of acid-base balance. A *possibility* measure is of little or no use for individual prediction, because the associated *probability* may lie anywhere between 0 and *possibility*. This equates to the clinical situation whereby one fetus may tolerate abnormal acid-base because it is adequately grown with good reserves, whereas another growth-retarded fetus with no reserves may be damaged by the same of abnormal acid-base. However, the *possibility* measure can be used as a group statistic due to its relationship to *probability*. With large numbers so that the other factors average out, a group with a higher mean *possibility* would be expected to have a higher mean *probability*. Thus the mean *possibility* could be used to compare clinical practice between centres, such as the study in Section 4.8. As the PIAD index captures the overall contribution of all parameters to the acid-base severity, only a single comparison rather than four (pH_A, BD_A, pH_V and BD_V) would need to be performed.

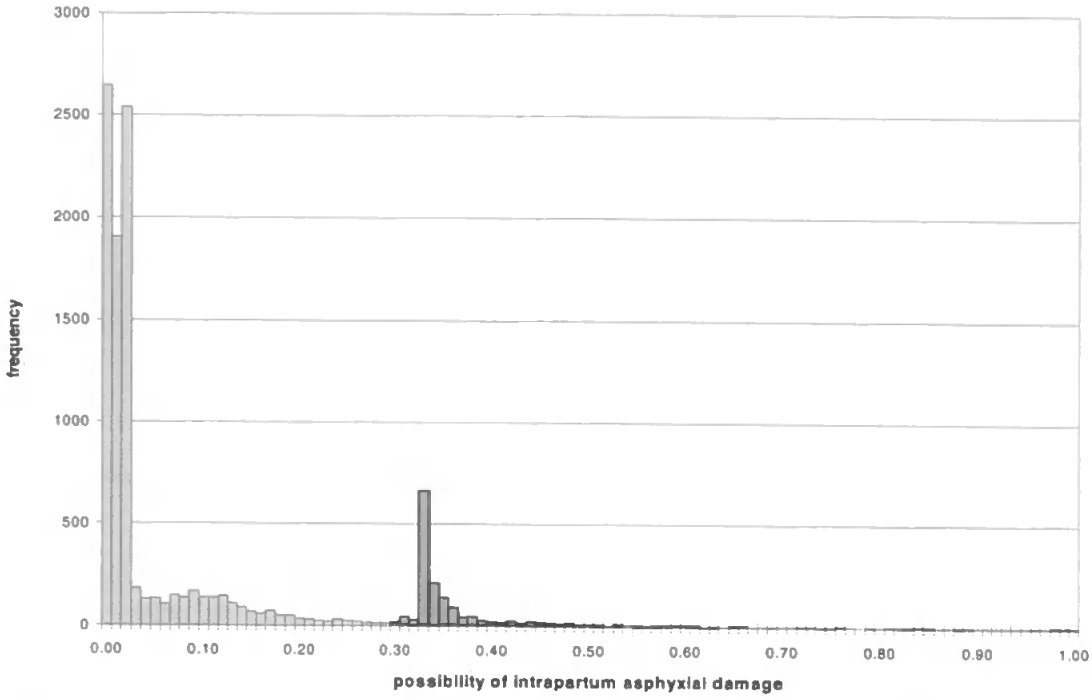


Figure 6.26: *distribution of the* possibility of intrapartum asphyxial damage *(PIAD) index, for 9 630 paired cord samples and 1 438 single cord samples*

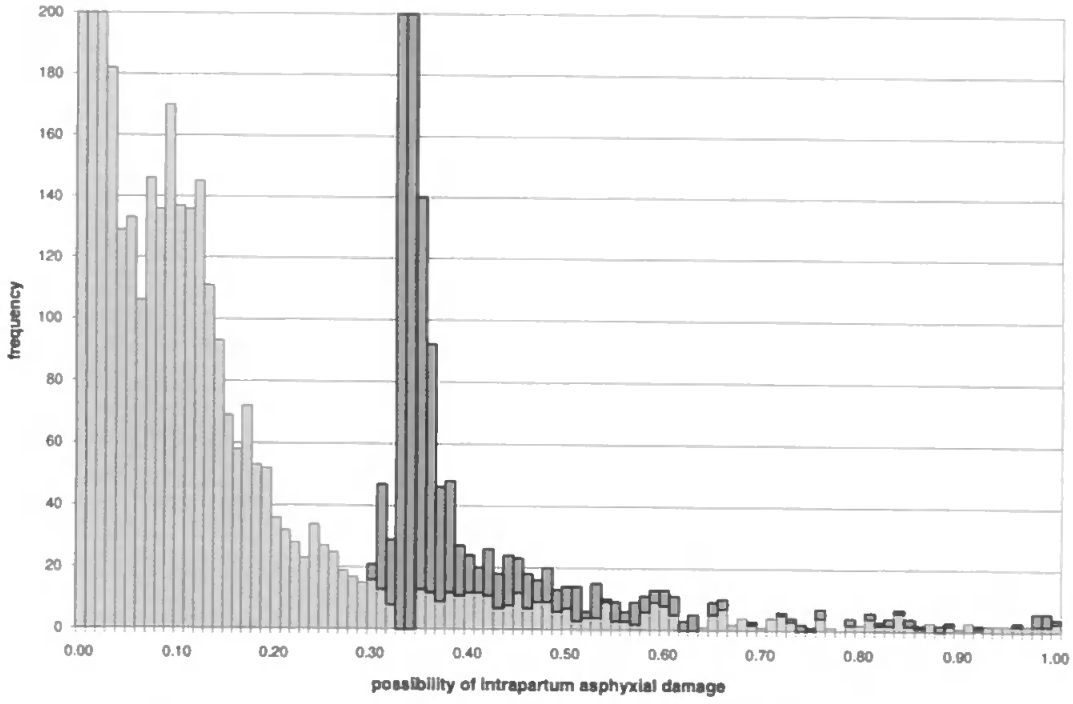


Figure 6.27: *expanded distribution of the PIAD index showing the tail of the distribution in more detail, for 9 630 paired cord samples and 1 438 single cord samples*

Due to the non-linearity of the fuzzy expert system, with the fact that the PIAD index would only reach one if all four variables were at the extreme ($pH_A \approx 6.60$, $BD_A \approx 20$, $pH_V \approx 6.70$, and $BD_V \approx 20$), the PIAD index is very heavily skewed towards zero. The histogram of PIAD index obtained for the 11 608 cases available on the database is shown in Figure 6.26. As the frequencies were found to be very high for just four points, the same distribution is shown in Figure 6.27 with the y-axis expanded, so that the detail at the tail can be seen. The large hump at around 0.30 to 0.40 is due to the presence of single vessel (venous) samples. The minimum PIAD obtainable for a normal venous result is around 0.30 because the arterial values are *unknown*, and hence arterial acidemia cannot be excluded. Only 231 (2.4%) cases out of 9 630 paired cord samples gave a PIAD of > 0.40, with an overall expectation value of 0.052 for paired samples, and 0.372 for single samples.

6.6.3 Performance Evaluation

The PIAD index was used to rank the output of the fuzzy expert system as before (as the PIAD is a linear transform of *condition*, this is equivalent to ranking directly on the *condition* variable). This ranking was then compared against the rankings obtained from six clinicians for fifty cases, as described in the previous chapter, to evaluate the performance of the integrated fuzzy expert system. Initial performance was acceptable, and then increased to be reasonably good immediately after the introduction of the 'alkalotic' terms and the associated rule. However the performance was still below that of the tuned preliminary system (*fuzzy*^{1}). Rather than tune this fuzzy model automatically as before, a further session of knowledge elicitation was undertaken.

The raw fuzzy output sets for all the fifty cases were examined by an 'expert' clinician, and in each case the clinician indicated whether the output set was considered reasonable. In each case the clinician's opinion of the appropriate centre-of-gravity was marked for each of the three output variables, in addition to any other comments about the shape information. Minor modifications were then made to the membership functions of the input and output variables to alter the fuzzy outputs to as close to the indication as possible. This process

resulted in the final membership functions specified in Table 6.4. At this stage the 'extra' rule that had been introduced during the tuning of the preliminary system was also removed, as it was found to be responsible for undesirable output. At the end of this manual tuning process the shape and centre-of-gravity of each case was closer to the clinician's indications.

The performance of the system was subsequently re-evaluated by comparison with the rankings, and was found to have increased to an extremely encouraging level, as shown in Table 6.7. Indeed the average performance of 0.94 for the integrated fuzzy expert system, $fuzzy^2$, exceeded the average performance of 0.93 for the tuned preliminary fuzzy expert system, $fuzzy^1$, as shown in Table 5.9. Although the clinician who provided the opinions was one of the clinicians who had previously ranked these cases, it is notable that this result was by no means inevitable.

The clinician was given no information as to how the fuzzy expert system calculated its final ranking score from the fuzzy output sets, and thus had no knowledge of how the fuzzy system's rankings would be affected by alteration of the centre-of-gravities as indicated. Also, as the original ranking of the fifty cases had taken place approximately one year previously, it is extremely doubtful that the clinician could have remembered them anyway. Thus, the two processes could be considered independent. The fact that alteration of the individual centre-of-gravities produced increased agreement with all clinicians was taken as both an indication of the clinician's expertise (specifically, consistency in acid-base interpretation) and as an indication that the fuzzy model was successful. Despite this, it was clear that the same fifty cases had been used both for inspiring the ammendments to the system and for evaluating the performance of the system. Hence a fresh validation was required to test the generalisation of the model.

Agreement	*Correlation*
$fuzzy^2 \Leftrightarrow \text{clinician}^1$	≈ 0.969
$fuzzy^2 \Leftrightarrow \text{clinician}^2$	≈ 0.894
$fuzzy^2 \Leftrightarrow \text{clinician}^3$	≈ 0.943
$fuzzy^2 \Leftrightarrow \text{clinician}^4$	≈ 0.963
$fuzzy^2 \Leftrightarrow \text{clinician}^5$	≈ 0.927
$fuzzy^2 \Leftrightarrow \text{clinician}^6$	≈ 0.958
Average	≈ 0.942

Table 6.7: *the performance of the integrated fuzzy expert system in terms of its agreements with previously obtained clinical opinions on fifty cases*

6.7 Validation of the Integrated System

6.7.1 Methods

Having developed the fuzzy model through expert knowledge elicitation and evaluated its performance via the agreement of the PIAD index with previously obtained expert ordering of fifty cases, the integrated fuzzy expert system model was validated by means of a fresh validation study. This fuzzy model was fixed to the specification already described, but included a total of five tunable weights which could be adjusted if required. There were three weights for the vessel identification rules:

- the ΔpH weight
- the ΔpCO_2 weight
- the ΔpO_2 weight

Each of these was applied to every rule clause that included the corresponding input variables. Thus the ΔpO_2 weight was applied to every clause which compared arterial pO_2 to venous pO_2, so that if the ΔpO_2 weight was zero then ΔpO_2 had no effect on vessel identification. There were two weights for the interpretation rules:

- the pH weight
- the BD_{ecf} weight

Each of these was applied to every rule clause that included either the corresponding arterial or venous input variable. Thus the BD_{ecf} weight was applied equally to every clause which referred to arterial BD_{ecf} and venous BD_{ecf}, so that if the BD_{ecf} weight was zero then BD_{ecf} had no effect on interpretation. The *default* system refers to the fuzzy expert system with all weight parameters set to one (1.0).

A study was designed to validate the numeric and linguistic outputs of both the vessel identification rules and the interpretation rules. An independent engineer was recruited to select cases from the collected database for the study, and to analyse the performance of the expert system in comparison with the experts, including any alteration of the expert system weights. Four clinical experts were recruited for the study; two Professors, one Consultant and one Research Midwife. Three of these clinicians had also taken part in the previous study, described in Chapter 5, and obtained a high level of inter-agreement. The fourth was an internationally acknowledged authority on acid-base balance. Considering this and the fact that all have published papers on acid-base interpretation, they will henceforth be referred to as *experts A – D*.

Three separate validation sub-tasks were designed:

- *vessel identification* — the experts were given two sets of pH, pCO_2 and pO_2 parameters from each of fifty cases, and were asked to indicate their opinion of the closest description of vessel origin from a choice of:

    1. definitely the same vessel
    2. probably the same vessel
    3. mixed vessel(s)
    4. probably different vessels
    5. definitely different vessels

- *linguistic interpretation* — the experts were given two sets of pH and BD_{ecf} parameters from each of fifty cases, and were asked to indicate their opinion of the closest linguistic interpretation for three linguistic variables; *acidemia*, *component*, and *duration*.

The choices for *acidemia* were:

1. severe
2. significant
3. moderate
4. mild
5. normal

The choices for *component* were:

1. metabolic
2. mixed
3. respiratory

The choices for *duration* were:

1. chronic
2. intermediate
3. acute

For each variable they were instructed to mark *zero*, *one* or *two* terms to indicate the closest match. This was specifically designed to allow the expert to mark two adjacent labels if they felt a result fell in-between two labels, or to mark no label if there was insufficient information, or no label was appropriate;

- *ordering (ranking) of cases* — as previously, the experts were asked to rank fifty cases from 'worst' to 'best', in terms of likelihood that the infant may have suffered intra-partum asphyxial damage, on the basis of the acid-base information alone.

6.7.2 Selection of Cases

The cases for each task were selected by the independent engineer from the database of over 10 000 results (approximately 400 abnormals). The selection of cases provided serious prob-

lems. Cases could not be selected from the entire database on a uniform random basis, as this would have resulted in approximately 75% paired arterial-venous samples, and approximately 98% *normal* interpretations. In essence it was desired to uniformally span the *target* outputs, so that a roughly even spread across the various output sets would have been obtained from the combined experts (and expert system). However, this pre-supposed that the output was known — which it obviously wasn't for the validation study. Other studies [60] have used an in-house expert to select difficult and/or representative cases, but this approach requires the availability of an expert.

Due to the restricted number of experts available to perform acid-base interpretation to the level required for this study, it was not feasible to have an expert select the cases. The problem was solved by adopting a data driven approach. For the vessel identification task, a set of results were selected with Δ's that were concentrated around the uncertainties in Table 6.1 and the cut-offs encoded into the crisp expert system rules. Roughly twenty five of the fifty cases were uniformly spread around the three-dimensional Δ space from the origin at $\Delta\text{pH} = 0$, $\Delta pCO_2 = 0$, and $\Delta pO_2 = 0$ to a surface of size $\Delta\text{pH} = 2\,\delta\text{pH}$, $\Delta pCO_2 = 2\,\delta pCO_2$, and $\Delta pO_2 = 2\,\delta pO_2$. The remaining twenty five or so cases had all Δ's positive, increasing from $\Delta\text{pH} = 0$, $\Delta pCO_2 = 0$, and $\Delta pO_2 = 0$ to roughly the median values $\Delta\text{pH} = 0.08$, $\Delta pCO_2 = 2kPa$, and $\Delta pO_2 = 3kPa$.

For the linguistic interpretation tasks and the ranking task, the crisp expert system categorisation already obtained on the data was used to guide the selection of cases. Three sets of twenty five cases were randomly selected to roughly span the crisp expert system categorisations. This ensured that a few cases were obtained from a variety of conditions, including results that had parameter errors, results from a single vessel, and results ranging from metabolic acidemia to normal. The three sets of twenty five cases were then combined to give two sets of fifty, by having one set of twenty five cases duplicated in the two sets of fifty. Thus each set of fifty had twenty five unique cases and twenty five duplicated cases. This was done to allow the potential for some sort of cross-validation to be carried out between the linguistic interpretation and the ranking (numeric) interpretation obtained from the experts. When presented to the experts, this duplication of data was not explicitly stated, so that the

experts *may* not have noticed.

During development, the expert system had been run over the entire database to ensure that it ran without software failure, and, for example, to generate gross statistics on the number of results identified as *same* vessels. However, all cases selected for validation were previously 'unseen' by the integrated fuzzy expert system, in the sense that no expert system output had ever been examined on an individual basis, in any attempt to tune the system, or to ensure that the output was 'correct' in any way. All cases were different from the fifty cases used in the previous chapter.

6.7.3 Categorical Agreement

To measure the agreement between two expert's linguistic categorisation a measure of (nominal) categorical agreement was required. The χ^2 statistic can be used to measure the degree of *association* between two categorical variables, but this statistic makes no distinction between departure from chance association due to *agreement* or *disagreement*. In 1960, Cohen introduced a measure of *agreement* between two categorical variables termed the *kappa* coefficient [22]. If two judges place N cases into one of k different (nominal) categories, then the agreement between the two judges can be represented in an agreement matrix as shown in Table 6.8, where f_{ij} represents the number of judgements that judge A placed in category i and judge B placed in category j [38].

	Judge B				
Judge A	1	2	$\cdots$	k	Total
1	f_{11}	f_{12}	$\cdots$	f_{1k}	f_{1A}
2	f_{21}	f_{22}	$\cdots$	f_{2k}	f_{2A}
$\vdots$					$\vdots$
k	f_{k1}	f_{k2}	$\cdots$	f_{kk}	f_{kA}
Total	f_{B1}	f_{B2}	$\cdots$	f_{Bk}	N

Table 6.8: *joint frequencies of judgements by two judges on a scale with k categories*

The observed proportion of agreement, p_o, is given by summation over the probabilities of

the diagonals:

$$p_o = \frac{1}{N}\sum_{i=1}^{k} f_{ii} \tag{6.11}$$

and the expected proportion of agreement due to chance, p_e, is given by summation over the joint probabilities of the marginals:

$$p_e = \frac{1}{N^2}\sum_{i=1}^{k} f_{iA} f_{Bi} \tag{6.12}$$

The kappa coefficient is then defined by the excess of observed agreement above chance:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{6.13}$$

to give a value of $+1$ when there is perfect agreement, 0 when there is chance agreement and < 0 for less agreement than chance. Although this basic kappa statistic measures agreement rather than association as it takes into account only the probabilities on the diagonals, this implies that it is only measuring *exact* agreement. To overcome this problem, Cohen later introduced *weighted kappa* to allow for partial agreement [23].

To calculate weighted kappa, a weight matrix is introduced of size $k \times k$, where each cell represents the relative credit of the corresponding cell in the agreement matrix, so that:

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{21} & w_{22} & \cdots & w_{2k} \\ \vdots & & & \vdots \\ w_{k1} & w_{k2} & \cdots & w_{kk} \end{pmatrix}$$

If the weights are normalised relative to the maximum credit to lie between 0 and 1, such that the maximum weight $w_{\max} = 1$, then the observed proportion of weighted agreement, p'_o, is given by:

$$p'_o = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{k} f_{ij} w_{ij} \tag{6.14}$$

the expected proportion of weighted agreement, p'_e, is given by:

$$p'_e = \frac{1}{N^2}\sum_{i=1}^{k}\sum_{j=1}^{k} f_{iA} f_{Bj} w_{ij} \tag{6.15}$$

and the weighted kappa coefficient, κ_w, by:

$$\kappa_w = \frac{p'_o - p'_e}{1 - p'_e} \tag{6.16}$$

It is usual, though not necessary, to assign the maximum weight $w_{\max} = 1$ to all diagonal *exact* agreements so that $w_{ii} = 1$, and to assign the weight matrix symmetrically so that $w_{ij} = w_{ji}$, where $0 \leq w_{ij} < 1$. If $w_{ii} = 1$ and $w_{ij} = 0$ for all $i \neq j$, then $\kappa_w = \kappa$.

The significance of unweighted kappa can be tested by approximation to the normal:

$$z = \frac{\kappa}{\sigma(\kappa)} \tag{6.17}$$

where

$$\sigma(\kappa) \approx \sqrt{\frac{\sum_i f_{iA} f_{Bi}}{N(N^2 - \sum_i f_{iA} f_{Bi})}} = \sqrt{\frac{p_e}{N(1 - p_e)}} \tag{6.18}$$

and weighted kappa, similarly, by:

$$z = \frac{\kappa_w}{\sigma(\kappa_w)} \tag{6.19}$$

where

$$\sigma(\kappa_w) \approx \sqrt{\frac{N^2 \sum_{ij} f_{ij} w_{ij}^2 - \left(\sum_{ij} f_{ij} w_{ij}\right)^2}{N\left(N^2 - \sum_{ij} f_{ij} w_{ij}\right)^2}} \tag{6.20}$$

Note that this simply tests the significance that κ or κ_w is greater than zero, which does not necessarily demonstrate that substantial agreement has been found. Various labels have been assigned to arbitrary ranges of kappa [69, 38]; the proposed nomenclature of Landis and Koch will be adopted, as shown in Table 6.9.

Although it might be thought that the introduction of partial agreements must necessarily *increase* the agreements found, Cohen pointed out in his original paper [23] that this is actually not the case as the weights affect both p'_o and p'_e. The weighted kappa is thus *fully* chance corrected, but the weight matrix does affect both the absolute value of κ_w obtained and its significance, and therefore must be determined prior to any study and any significance testing. However, as there were no external criteria by which to determine the weights for this study, they were determined arbitrarily. Consequently in the analysis that follows, both weighted and unweighted kappa, and their respective significances, are always quoted.

Kappa Statistic	*Strength of Agreement*
< 0.00	poor
$0.00 \ldots 0.20$	slight
$0.21 \ldots 0.40$	fair
$0.41 \ldots 0.60$	moderate
$0.60 \ldots 0.80$	substantial
$0.80 \ldots 1.00$	almost perfect

Table 6.9: *arbitrary ranges of the kappa statistic and the associated label for strength of agreement, from Landis and Koch [69]*

6.7.4 Results of Numeric Vessel Identification

The numeric vessel identification output was evaluated by considering the task a dichotomous process of identifying each case as representing *paired* vessels or the *same* vessel. As stated in Section 6.6, a linguistic output corresponding to *different*, *mixed/different* or *not same* from the expert system was taken as sufficient to consider the samples *paired*. To correspond to this, a categorisation of *probably different* or *definitely different* by an expert was taken to indicate *paired* vessels, and all other categorisations were taken to indicate the *same* vessel.

The fuzzy expert system output was taken as the centroid defuzzification of the *origin* variable. In order to compare this to the experts' categorisation, a variable cut-off point was taken, above which was taken to indicate *paired* vessels and below which was taken to indicate *same* vessel. This produced a 2×2 contingency table, for which plain kappa was calculated to give the *agreement*. Thus a kappa agreement score was obtained for each expert, $\kappa_A \ldots \kappa_D$ at each cut-off point, from the minimum output of 30 to the maximum output of 80 in steps of 1. At each point the *root-mean-square error* in the kappa for all experts was then calculated by:

$$RMS\ error = \sqrt{\frac{(1-\kappa_A)^2 + (1-\kappa_B)^2 + (1-\kappa_C)^2 + (1-\kappa_D)^2}{4}} \tag{6.21}$$

Figure 6.28 shows the overall *RMS error* in kappa plotted at each cut-off point. The lowest overall *RMS error* in kappa was found to be at cut-off point 63 — i.e. when an *origin* centroid > 63 was taken to indicate *paired* vessels. The kappa (and its significance) obtained for each

expert at this point is shown in Table 6.10. As these results were considered somewhat disappointing, although they represented *moderate* to *substantial* agreements, the individual kappa obtained for each expert was examined, as shown in Figure 6.29. The maximum value of kappa (and its significance) obtained for each expert is shown in Table 6.11. It can be seen that *almost perfect* ($\kappa \approx 0.874$) and *substantial* ($\kappa \approx 0.775$) agreement was obtained for expert *A* and expert *B* respectively, but the maximum agreement was still only *moderate* for experts *C* and *D*.

Expert	κ	Z	p
A	0.676	6.434	$\approx 10^{-10}$
B	0.550	4.588	$\approx 10^{-6}$
C	0.572	5.025	$\approx 10^{-7}$
D	0.567	4.916	$\approx 10^{-7}$

Table 6.10: *agreement between* fuzzy^2 *and experts obtained at minimum* RMS error *cut-off point*

Expert	*Cut-off*	κ	Z	p
A	65–71	0.874	12.434	$\approx 10^{-36}$
B	74	0.775	8.108	$\approx 10^{-16}$
C	50–52	0.593	5.025	$\approx 10^{-7}$
D	62	0.590	5.091	$\approx 10^{-7}$

Table 6.11: *maximum agreement between* fuzzy^2 *and experts obtained at various cut-off points*

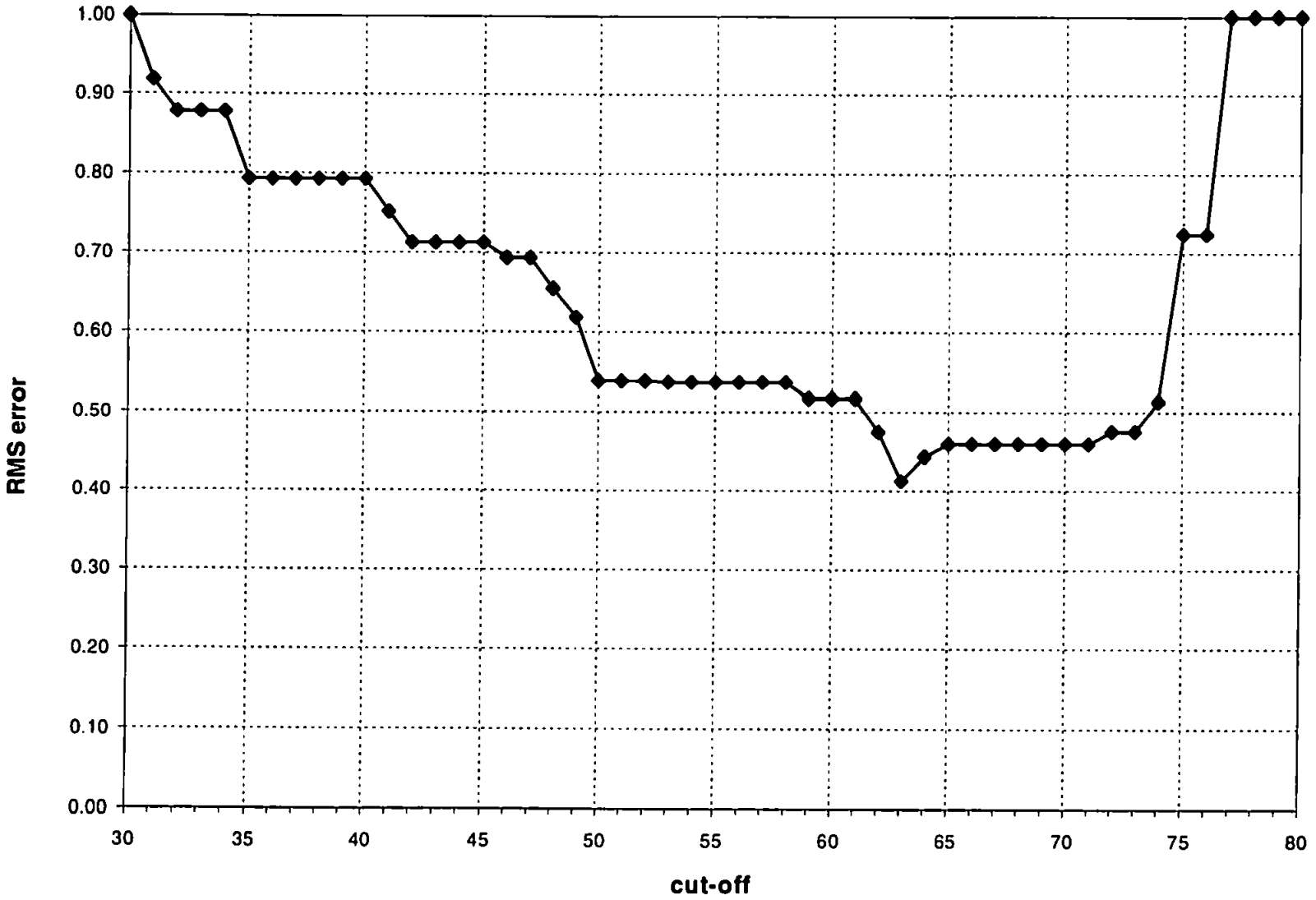


Figure 6.28: *graph of the overall* RMS error *in kappa agreement between the four experts and the expert system, for vessel identification by simple cut-off in* origin *centroid*

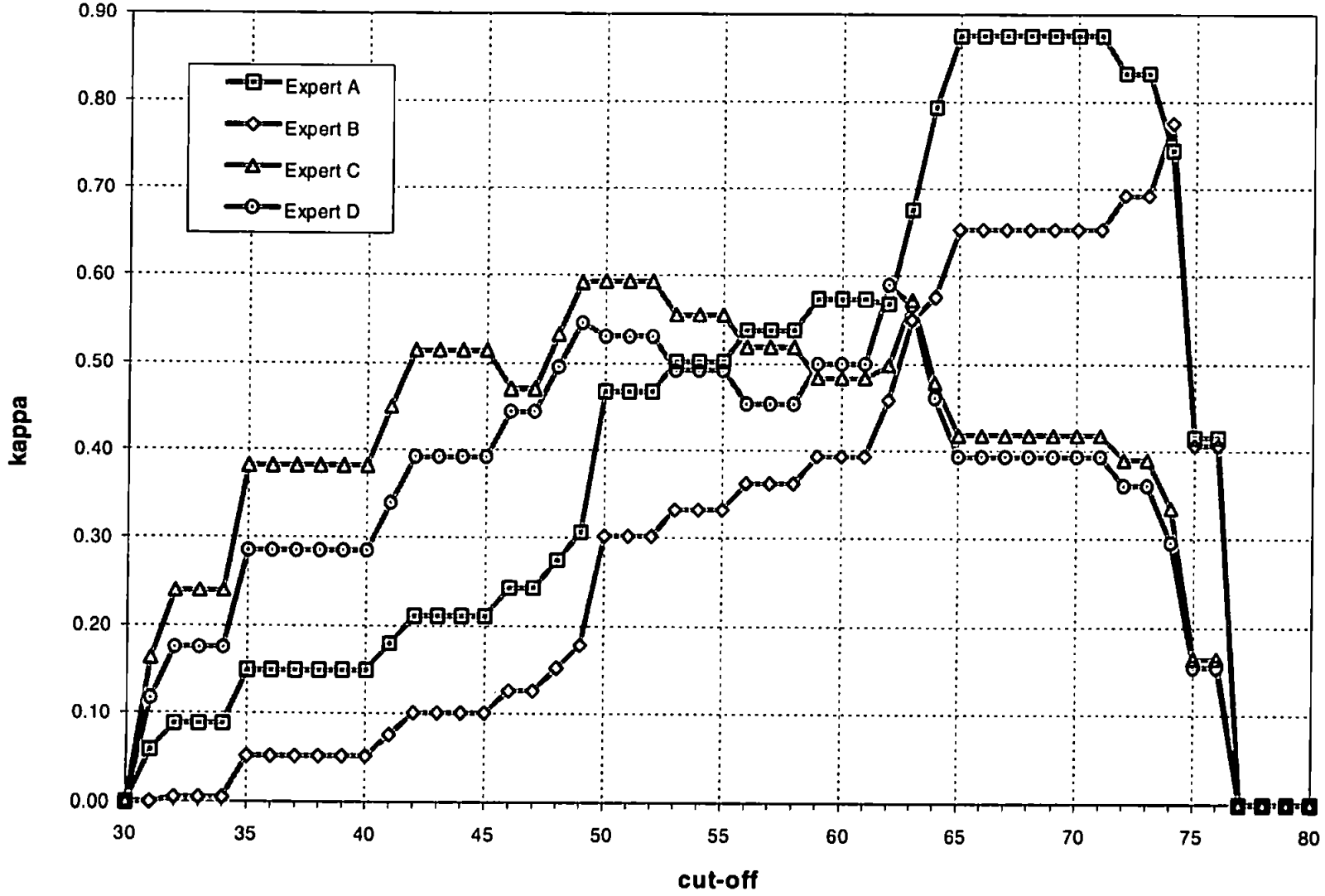


Figure 6.29: *graph of the kappa agreement between each of the four experts and the expert system, for vessel identification by simple cut-off in* origin *centroid*

6.7.5 Results of Linguistic Vessel Identification

The results of the vessel identification task were then further investigated by means of comparison of the linguistic output of the *origin* variable with the categorisations of the experts. For this the *origin* linguistic output was limited in the defuzzification process to produce an output of *same*, *same/mixed*, *mixed*, *mixed/different* or *different*. These outputs were taken to map directly to the categories *definitely same*, *probably same*, *mixed*, *probably different* and *definitely different*, as used by the experts. Thus a 5×5 agreement matrix was obtained for each comparison, and agreement was calculated by plain kappa and weighted kappa using the arbitrary weights matrix:

$$\mathbf{W} = \begin{pmatrix} 1 & 0.667 & 0 & 0 & 0 \\ 0.667 & 1 & 0.667 & 0.333 & 0 \\ 0 & 0.667 & 1 & 0.677 & 0 \\ 0 & 0.333 & 0.667 & 1 & 0.677 \\ 0 & 0 & 0 & 0.677 & 1 \end{pmatrix}$$

Table 6.12 shows the results obtained for plain kappa. The kappa value obtained for each inter-expert agreement is shown, with its associated Z value underneath in parentheses. For each expert the average inter-expert kappa was calculated by taking the average kappa against the *other three* experts. The kappa agreement obtained for the fuzzy expert system, *fuzzy*^{2}, against each of the four experts is also shown. For the expert system the average agreement kappa was calculated by taking the average kappa against the *four* experts. Table 6.13 shows the results obtained in the same way for weighted kappa. As a large number of significance tests were carried out, a relatively cautious value of $\alpha = 0.001$ was taken as the significance level, corresponding to a Z-value greater than ≈ 3.09. Kappa values that are significantly above zero by this criterion are shown in bold type.

It can be seen from Tables 6.12 and 6.13 that the agreements obtained were all fairly low. For unweighted kappa, they ranged from *slight* agreement ($\kappa = 0.043$) between experts B and D, up to only *fair* agreement ($\kappa = 0.370$) between expert C and *fuzzy*^{2}. Although the agreement improved for weighted kappa, they still only ranged from *slight* agreement ($\kappa_w = 0.146$)

Expert	A	B	C	D	*fuzzy*^{2}
A	—	**0.251**	**0.277**	**0.313**	**0.356**
		(3.565)	(3.939)	(4.649)	(4.701)
B	**0.251**	—	0.167	0.043	**0.225**
	(3.565)		(2.300)	(0.568)	(3.191)
C	**0.277**	0.167	—	0.277	**0.370**
	(3.939)	(2.300)		(2.606)	(4.309)
D	**0.313**	0.043	0.277	—	0.166
	(4.649)	(0.568)	(2.606)		(2.128)
Average	0.280	0.154	0.240	0.211	0.279

Table 6.12: *inter-expert agreement and expert-*fuzzy^2 *agreement for linguistic vessel identification calculated by plain kappa*

Expert	A	B	C	D	*fuzzy*^{2}
A	—	**0.452**	**0.518**	**0.413**	**0.595**
		(4.727)	(5.278)	(4.430)	(5.627)
B	**0.452**	—	0.250	0.146	0.281
	(4.727)		(2.602)	(1.483)	(2.985)
C	**0.518**	0.250	—	**0.490**	**0.607**
	(5.278)	(2.602)		(4.181)	(5.316)
D	**0.413**	0.146	**0.490**	—	0.273
	(4.430)	(1.483)	(4.181)		(2.640)
Average	0.461	0.283	0.419	0.350	0.439

Table 6.13: *inter-expert agreement and expert-*fuzzy^2 *agreement for linguistic vessel identification calculated by weighted kappa*

between experts *B* and *D*, up to *moderate* agreement ($\kappa_w = 0.607$) between expert *C* and *fuzzy*^{2}. In an attempt to improve the agreement obtained for the fuzzy expert system, the relative weightings of the input variables were altered by maintaining the ΔpH weight at 1.0 and decreasing the ΔpCO_2 weight and the ΔpO_2 weight, from the default values of 1.0 down to 0.0 for both. The fuzzy expert system was re-run on the fifty cases with the new weights, and the linguistic output of the *origin* variable was compared against the experts. The weighted kappa agreement was calculated individually for each expert at each value of the ΔpCO_2 and ΔpO_2 weights, and the results obtained are shown in Figures 6.30 to 6.33.

Note carefully that Figures 6.30 and 6.31 have a different rotation to Figures 6.32 and 6.33, so that the ΔpCO_2 and ΔpO_2 axes are reversed, and than the kappa (vertical) axis is individually scaled.

From Figures 6.30 and 6.31 it can be seen that the maximum weighted kappa for experts *A* and *B* was obtained very close to the default weights of $\Delta pCO_2 = 1.0$ and $\Delta pO_2 = 1.0$, whereas from Figures 6.32 and 6.33 it can be seen that the maximum weighted kappa for experts *C* and *D* was obtained very close to the extreme weights of $\Delta pCO_2 = 0.0$ and $\Delta pO_2 = 0.0$. For experts *A* and *B* the actual maxima were very similar to the default values, at $\kappa_w = 0.602$ compared to the default $\kappa_w = 0.595$ and $\kappa_w = 0.307$ compared to the default $\kappa_w = 0.281$ respectively. For experts *C* and *D* the maximal kappas were higher, at $\kappa_w = 0.684$ compared to the default $\kappa_w = 0.607$ and $\kappa_w = 0.493$ compared to the default $\kappa_w = 0.273$ respectively. Thus agreement with experts *C* and *D* could be improved through the use of ΔpH alone to identify vessel origin.

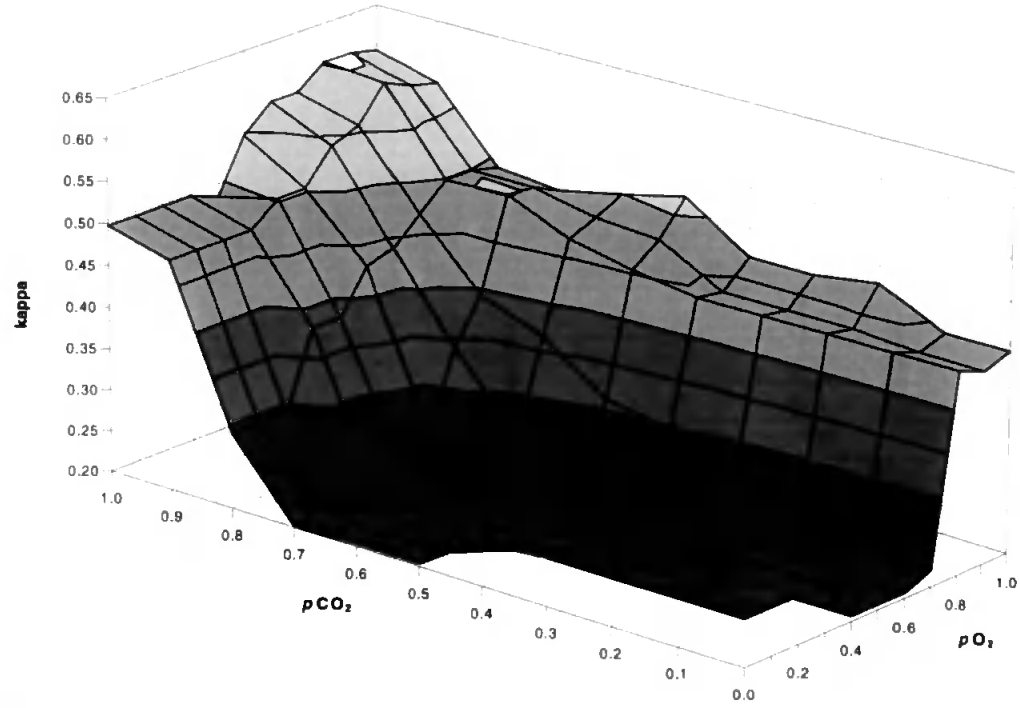


Figure 6.30: *surface plot of weighted kappa obtained from agreement between expert A and* fuzzy^2*, for various values of* ΔpCO_2 *weight and* ΔpO_2 *weight at a constant* ΔpH *weight of 1.0*

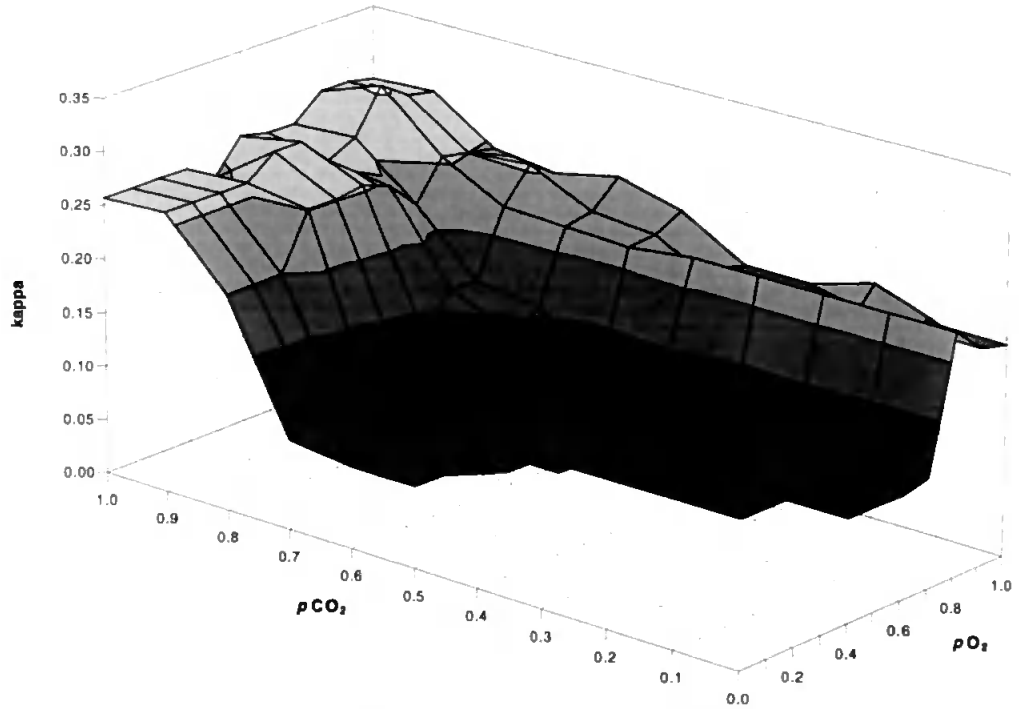


Figure 6.31: *surface plot of weighted kappa obtained from agreement between expert B and* fuzzy^2*, for various values of* ΔpCO_2 *weight and* ΔpO_2 *weight at a constant* ΔpH *weight of 1.0*

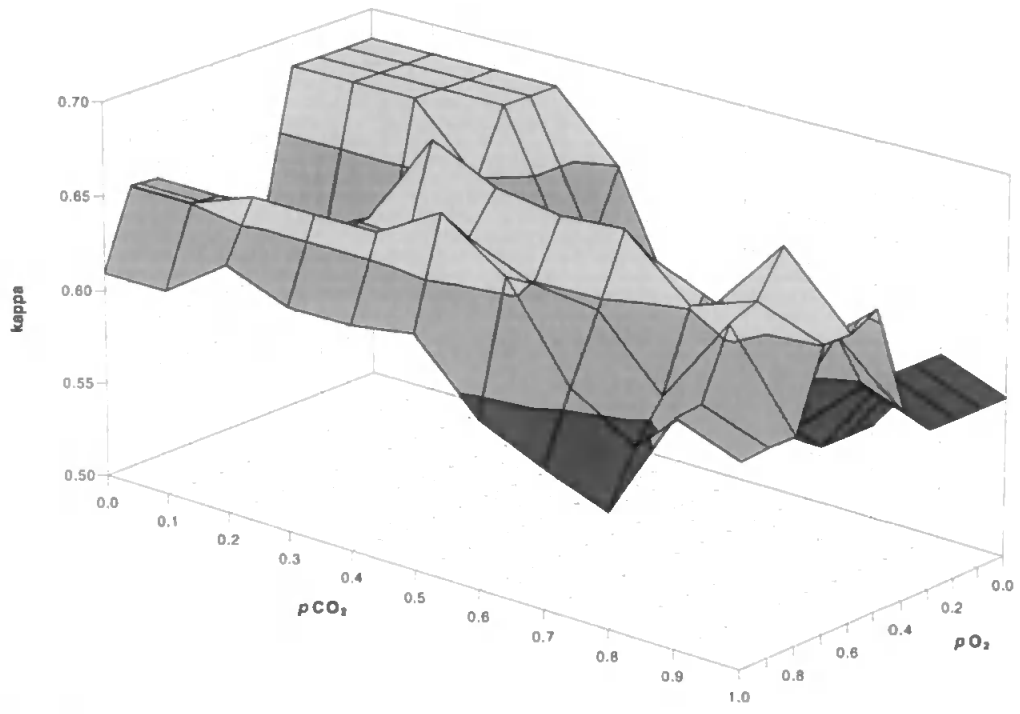


Figure 6.32: *surface plot of weighted kappa obtained from agreement between expert C and* fuzzy^2, *for various values of* ΔpCO_2 *weight and* ΔpO_2 *weight at a constant* ΔpH *weight of 1.0*

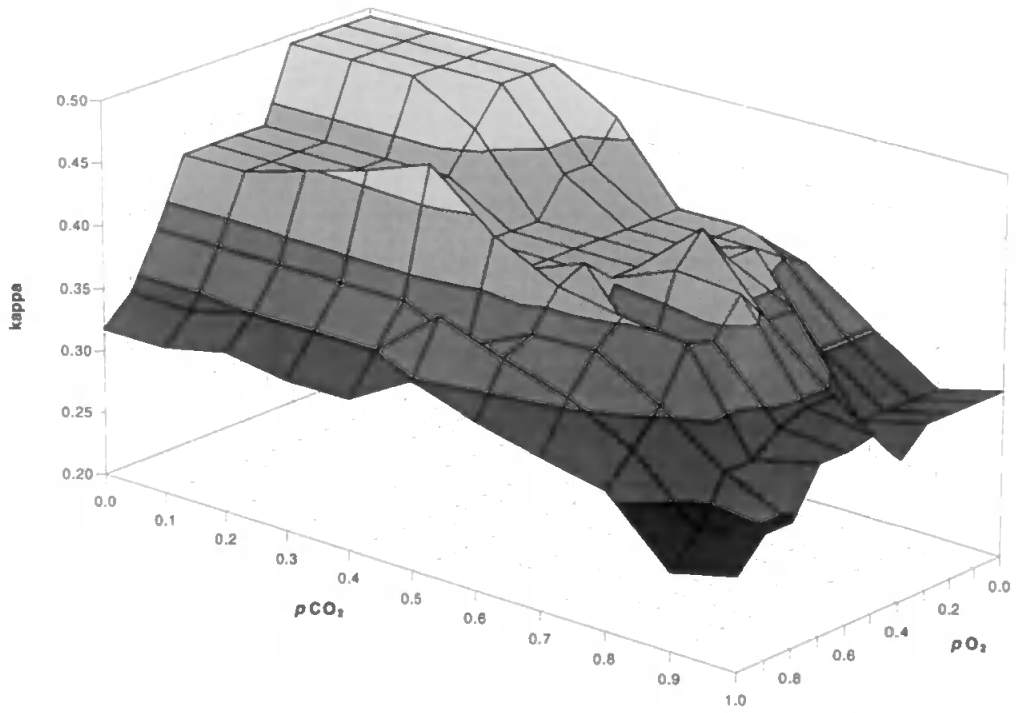


Figure 6.33: *surface plot of weighted kappa obtained from agreement between expert D and* fuzzy^2, *for various values of* ΔpCO_2 *weight and* ΔpO_2 *weight at a constant* ΔpH *weight of 1.0*

6.7.6 Results of Numeric Interpretation

The Spearman rank order correlation coefficient was used in an identical manner to that specified in the previous chapter to evaluate the performance of the numeric interpretation output. The centroids of the *acidemia*, *component* and *duration* variables were combined by Equation 6.9, and this was used to rank the cases from 'worst' to 'best'. This is exactly the same test as carried out previously, with the exception that cases were included with parameter errors and/or single vessel samples. The individual inter-expert and expert-*fuzzy*^{2} Spearman rank order correlation coefficients obtained are shown in Table 6.14 and the agreements are displayed in Figure 6.34. As before, the average inter-expert agreement is calculated by taking the average of each expert against the other *three* experts, and the average *fuzzy*^{2} agreement by taking the average of agreement with all *four* experts.

Expert	A	B	C	D	*fuzzy*^{2}
A	—	0.899	0.888	0.577	0.950
B	0.899	—	0.908	0.701	0.931
C	0.888	0.908	—	0.537	0.925
D	0.577	0.701	0.537	—	0.606
Average	0.788	0.836	0.777	0.605	0.853

Table 6.14: *inter-expert agreement and expert-*fuzzy^2 *agreement for numeric interpretation calculated by Spearman rank order correlation*

As can be seen from Table 6.14, the fuzzy expert system performed exceptionally well against experts *A*, *B*, and *C*. These three experts had taken place in the previous study, and the average expert system agreement with these three is 0.94 — as high as its performance on those previous cases which had been used as training data — as shown in Table 6.7. Lower correlation was achieved against expert *D* (who had not been previously involved in any development or validation), although the expert system achieved the highest overall average correlation.

Further investigation was therefore carried out to discover whether the correlation could be improved through alteration of the pH or BD_{ecf} weights of the input variables for interpretation. Each weight was decreased from the default of 1.0 down to 0.0, whilst the other weight

was kept constant, and the fuzzy expert system with these weights was re-run on the fifty cases. Thus, for example, with a pH weight of 1.0 and a BD_{ecf} weight of 0.0, the expert system placed full emphasis on pH and effectively ignored BD_{ecf}. The results obtained are shown in Table 6.15, where the default weightings of pH and $BD_{ecf} = 1.0$ are shown in bold.

Weight		*Expert*			
pH	BD_{ecf}	*A*	*B*	*C*	*D*
1.0	0.0	0.938	0.898	0.955	0.491
1.0	0.1	0.939	0.903	0.952	0.501
1.0	0.2	0.944	0.915	0.950	0.528
1.0	0.3	0.947	0.922	0.949	0.548
1.0	0.4	0.945	0.925	0.947	0.554
1.0	0.5	0.946	0.928	0.945	0.564
1.0	0.6	0.947	0.931	0.938	0.580
1.0	0.7	0.947	0.931	0.938	0.584
1.0	0.8	0.947	0.931	0.935	0.593
1.0	0.9	0.947	0.930	0.934	0.597
1.0	**1.0**	**0.950**	**0.931**	**0.925**	**0.605**
0.9	1.0	0.953	0.929	0.928	0.599
0.8	1.0	0.942	0.916	0.916	0.584
0.7	1.0	0.936	0.901	0.899	0.571
0.6	1.0	0.920	0.886	0.871	0.576
0.5	1.0	0.900	0.869	0.851	0.578
0.4	1.0	0.884	0.853	0.819	0.589
0.3	1.0	0.865	0.824	0.764	0.600
0.2	1.0	0.812	0.786	0.696	0.634
0.1	1.0	0.716	0.696	0.572	0.683
0.0	1.0	0.466	0.477	0.293	0.730

Table 6.15: *Spearman rank order correlation between the fuzzy expert system and each expert obtained through alteration of the input variable weights*

It can be seen that maximum correlation with experts *A* and *B* was achieved at the default weighting; in fact a minute improvement was obtained for expert *A* at pH = 0.9 and BD_{ecf} 1.0. However it was found that correlation was maximum with expert *C* when the BD_{ecf} weight was 0.0 — i.e. when the expert system used pH alone for interpretation — although the improvement over the default was slight. Maximum correlation with expert *D*, on the other hand, was found when the pH weight was 0.0 — i.e. when the expert system used BD_{ecf} alone for interpretation. The maximum correlation of 0.730 obtained at this point was a good improvement, and was higher than any of the other experts' correlation with expert *D*.

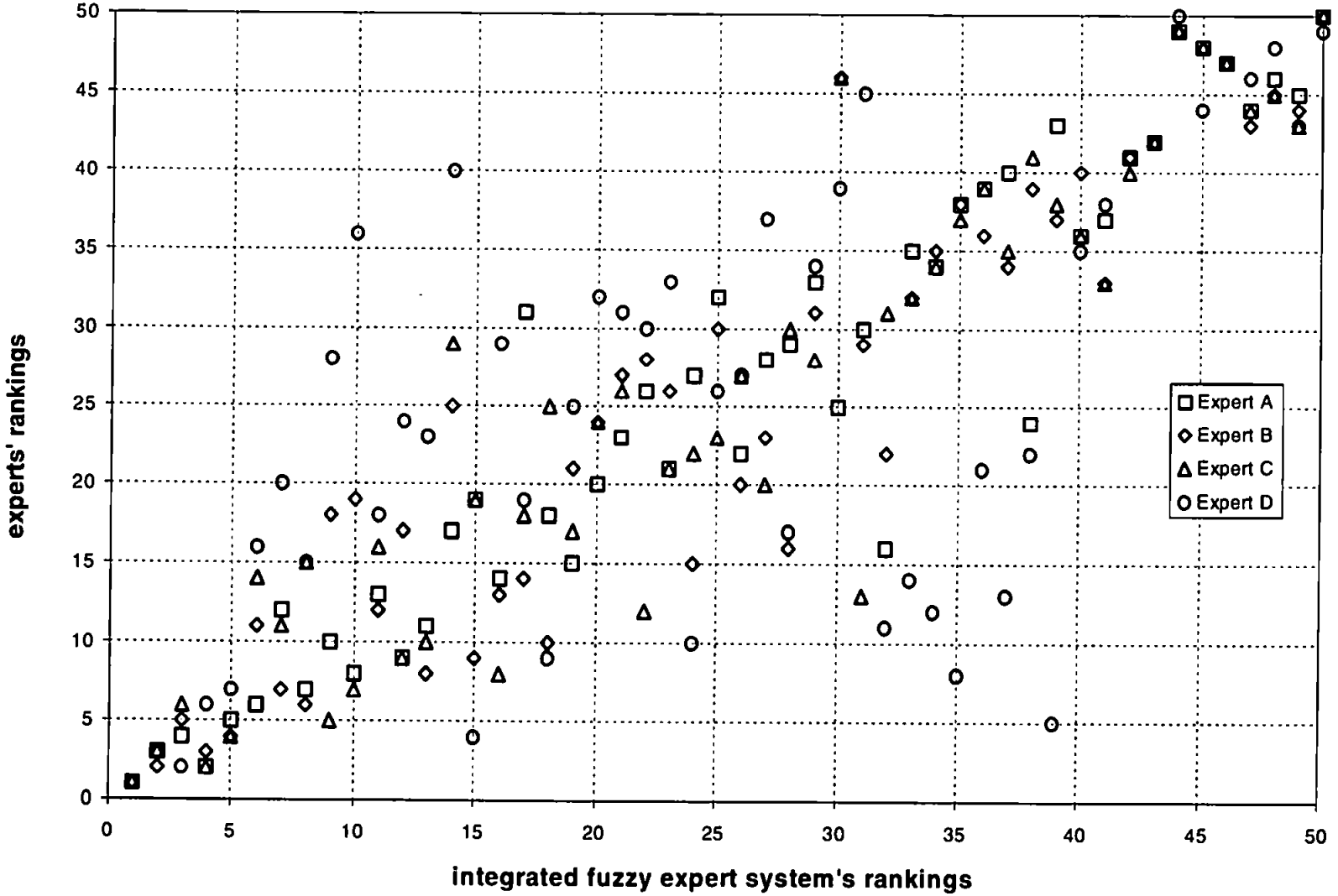


Figure 6.34: *graph of four experts rankings against the integrated fuzzy expert system*

6.7.7 Results of Linguistic Interpretation

The results of the linguistic interpretation were investigated by means of comparison of the linguistic output of the *acidemia*, *component* and *duration* variables with the categorisations of the experts. Although the experts had been instructed that they could mark *two* categories for a case if they considered that necessary, in fact it was found that *no* expert had done this. The experts had, however, frequently left a category blank to indicate *unknown* or *inappropriate* information. Consequently, the linguistic approximation algorithm of the fuzzy system was restricted to output the closest single fuzzy term, or a blank if one of the level sets (*unknown*, *indeterminate*, or *undefined*) was the best match, for each of the output variables. There were no *alkalotic* results in this section of the validation study, so that the *acidemia* variable was restricted to only the terms *severe* to *normal*. Thus, for *acidemia* a 6×6 agreement matrix was used (five primary terms plus one blank), and for *component* and *duration* a 4×4 agreement matrix was used (three primary terms plus one blank). Agreement was calculated by plain kappa and weighted kappa using the arbitrary weights matrices:

$$\mathbf{W}_6 = \begin{pmatrix} 1 & 0.667 & 0 & 0 & 0 & 0.333 \\ 0.667 & 1 & 0.667 & 0.333 & 0 & 0.333 \\ 0 & 0.667 & 1 & 0.677 & 0 & 0.333 \\ 0 & 0.333 & 0.667 & 1 & 0.677 & 0.333 \\ 0 & 0 & 0 & 0.677 & 1 & 0.333 \\ 0.333 & 0.333 & 0.333 & 0.333 & 0.333 & 1 \end{pmatrix}$$

and

$$\mathbf{W}_4 = \begin{pmatrix} 1 & 0.5 & 0 & 0.333 \\ 0.5 & 1 & 0.5 & 0.333 \\ 0 & 0.5 & 1 & 0.333 \\ 0.333 & 0.333 & 0.333 & 1 \end{pmatrix}$$

where the right-hand column and bottom row represents the weights for agreements involving one or more blanks. So, for example, if both experts left a category blank, this would

be assigned a full agreement, and if only one expert left a blank, this would receive a partial credit of 0.333 against any non-blank response from the other expert.

The three interpretation variables should have been treated as dependent variables, as they are inter-related for any particular case. However, there are ninety six ($6 \times 4 \times 4$) possible response combinations, and so a fully dependent agreement matrix would have been of size 96×96. As only fifty cases were included in the study (more was considered too onerous on behalf of the experts), this was not possible to implement, as the observed frequencies of most cells of the agreement matrix would have been zero or one. To overcome this problem, the three output variables were initially treated as independent. This would overstate the agreements found, but the intention was to correct for this artificially high agreement at a later stage. The results of plain kappa and weighted kappa agreements for each of the output variables are shown in Tables 6.16 to 6.21, where the format of each Table is similar to Tables 6.12 and 6.13.

As can be seen from these Tables, both the inter-expert agreement and the agreements are generally relatively low, even for weighted kappa. The best results for expert-*fuzzy*^{2} agreement were obtained on the *acidemia* variable with experts *A* and *B*, where the weighted kappa reached 0.786 and 0.633 respectively. Indeed, the expert system achieved the best average for *acidemia*, but conversely performed the worst on *component* and *duration* (particularly the weighted kappa). From Tables 6.18 and 6.19 it can be seen that the expert system's agreement with expert *A* on *component* actually decreased when kappa was weighted. However, although the expert system's performance was disappointing, it can also be noted that the inter-expert agreement is not particularly high. An attempt was made to investigate the effect of different pH and BD_{ecf} weights on these linguistic agreements, but in general it was found that performance was not significantly increased above the results achieved with default weights, as shown here.

Expert	A	B	C	D	*fuzzy*^{2}
A	—	**0.642**	**0.224**	0.208	**0.671**
		(7.398)	(3.448)	(3.026)	(7.764)
B	**0.642**	—	0.175	**0.312**	**0.567**
	(7.398)		(2.674)	(3.591)	(6.732)
C	**0.224**	0.175	—	**0.253**	**0.271**
	(3.448)	(2.674)		(3.604)	(4.126)
D	0.208	**0.312**	**0.253**	—	0.124
	(3.026)	(3.591)	(3.604)		(1.877)
Average	0.358	0.376	0.217	0.258	0.408

Table 6.16: *inter-expert agreement and expert-*fuzzy^2 *agreement for linguistic* acidemia *interpretation calculated by plain kappa*

Expert	A	B	C	D	*fuzzy*^{2}
A	—	**0.750**	**0.321**	**0.340**	**0.786**
		(6.162)	(3.580)	(3.428)	(6.635)
B	**0.750**	—	0.246	**0.436**	**0.633**
	(6.162)		(2.829)	(3.991)	(5.352)
C	**0.321**	0.246	—	**0.345**	**0.390**
	(3.580)	(2.829)		(3.963)	(4.309)
D	**0.340**	**0.436**	**0.345**	—	0.238
	(3.428)	(3.991)	(3.963)		(2.470)
Average	0.470	0.477	0.304	0.373	0.512

Table 6.17: *inter-expert agreement and expert-*fuzzy^2 *agreement for linguistic* acidemia *interpretation calculated by weighted kappa*

Expert	A	B	C	D	*fuzzy*^{2}
A	—	**0.345**	**0.625**	**0.492**	**0.323**
		(4.040)	(6.651)	(5.994)	(3.321)
B	**0.345**	—	0.165	**0.392**	0.170
	(4.040)		(1.987)	(4.892)	(1.637)
C	**0.625**	0.165	—	**0.402**	0.210
	(6.651)	(1.987)		(4.757)	(2.464)
D	**0.492**	**0.392**	**0.402**	—	0.173
	(5.994)	(4.892)	(4.757)		(2.439)
Average	0.487	0.301	0.398	0.429	0.219

Table 6.18: *inter-expert agreement and expert-*fuzzy^2 *agreement for linguistic* component *interpretation calculated by plain kappa*

Expert	A	B	C	D	*fuzzy*^{2}
A	—	**0.456**	**0.652**	**0.548**	0.320
		(5.487)	(7.742)	(6.599)	(3.063)
B	**0.456**	—	**0.276**	**0.429**	0.226
	(5.487)		(3.278)	(4.746)	(2.206)
C	**0.652**	**0.276**	—	**0.498**	0.216
	(7.742)	(3.278)		(6.136)	(2.270)
D	**0.548**	**0.429**	**0.498**	—	0.188
	(6.599)	(4.746)	(6.136)		(2.264)
Average	0.552	0.387	0.475	0.492	0.238

Table 6.19: *inter-expert agreement and expert-*fuzzy^2 *agreement for linguistic* component *interpretation calculated by weighted kappa*

Expert	A	B	C	D	*fuzzy*^{2}
A	—	**0.365**	**0.379**	**0.347**	**0.314**
		(3.882)	(4.522)	(4.090)	(3.165)
B	**0.365**	—	**0.310**	0.136	0.265
	(3.882)		(3.825)	(1.983)	(1.736)
C	**0.379**	**0.310**	—	**0.423**	0.187
	(4.522)	(3.825)		(4.496)	(2.227)
D	**0.347**	0.136	**0.423**	—	0.161
	(4.090)	(1.983)	(4.496)		(2.181)
Average	0.364	0.270	0.370	0.302	0.232

Table 6.20: *inter-expert agreement and expert-*fuzzy^2 *agreement for linguistic* duration *interpretation calculated by plain kappa*

Expert	A	B	C	D	*fuzzy*^{2}
A	—	**0.396**	**0.429**	**0.408**	**0.323**
		(4.035)	(4.717)	(4.492)	(3.183)
B	**0.396**	—	**0.346**	0.171	0.297
	(4.035)		(4.044)	(2.360)	(1.920)
C	**0.429**	**0.346**	—	**0.497**	0.205
	(4.717)	(4.044)		(4.956)	(2.369)
D	**0.408**	0.171	**0.497**	—	0.162
	(4.492)	(2.360)	(4.956)		(2.137)
Average	0.411	0.304	0.424	0.358	0.247

Table 6.21: *inter-expert agreement and expert-*fuzzy^2 *agreement for linguistic* duration *interpretation calculated by weighted kappa*

6.8 Discussion

This chapter has described the design, development and validation of the integrated fuzzy expert system. The final fuzzy expert system incorporates a vessel identification rule-base and a results interpretation rule-base, and can deal with missing and invalid data. In contrast to the preliminary fuzzy expert system, no automatic tuning (as described in the previous chapter) was carried out. This was simply because a high-performance system was discovered serendipitously, after the modifications to the fuzzy model were made to deal with alkalotic samples. Expert opinion was being used to guide manual 'hill-climbing' of the fuzzy model performance, through the fine-tuning of input and output membership functions, when a decision was taken to incorporate additional terms. After the pH input variables had their *high* term split into *normal* and *high*, and the *acidemia* output variable had the *alkalotic* term added, the performance of the system on the previous validation data improved to exceed that of the previously tuned fuzzy expert system. The observed performance was very close to the optimal performance calculated for that particular data set. This discovery obviated the need to perform automatic fuzzy model tuning.

Another difference from the preliminary fuzzy expert system was that normalised membership functions, for both input and output fuzzy sets, were utilised throughout the system. Although the integrated system did feature parameter weighting, as described in Section 6.3.7, the default configuration of the fuzzy model performed well. This emphasised the point made in the discussion of the previous chapter, that sub-normalisation was only required within the preliminary model. The more sophisticated model developed in this chapter is an example of an alternative fuzzy model that did *not* require sub-normalisation.

An interesting observation resulting from the knowledge elicitation process was that direct elicitation of input membership functions did not achieve good results. However, elicitation of expert opinion on the output sets combined with the intuitive alteration of input membership functions did work. That is, when the expert modified the input functions directly, worse performance resulted — when the engineer modified the input functions to produce output sets that the expert specified, better performance resulted. This may have been be-

cause changes to the input functions have a complex and model dependent effect on the output of the fuzzy model. Often in the fuzzy literature, expert opinion is used to guide the construction of fuzzy input terms, but this may be problematic without detailed knowledge of the fuzzy model to which they are applied. Again, the point is made that neither the fuzzy sets used for variables nor the fuzzy rules themselves have any extrinsic meaning outside the fuzzy model in which they appear.

In general, the validation results of the numeric fuzzy outputs were extremely good. Very high agreement was obtained with experts *A* and *B* for the numeric vessel identification output. Not surprisingly, these are the two experts whose opinions have been used most extensively in the development of the system. As expert opinion is probably the *only* method of validating vessel identification (unless some other chemical marker could be found to distinguish arterial blood from venous blood), this demonstration of high agreement with two experts is probably sufficient. The correlations obtained for the numeric interpretation (ranking) task were equally encouraging. Excellent correlations were obtained against experts *A*, *B* and *C*, and the fuzzy expert system performed better against expert *D* than experts *A* and *C* did. Overall the expert system achieved higher average correlation than any of the experts. Although experts *A*, *B* and *C* were involved in the previous ranking task, it is still an excellent achievement to obtain an average correlation of 0.94 with these three experts. This average correlation achieved on the fresh validation data set was the same as that achieved on the previous data set used in development — a result which imparts a high degree of confidence in the fuzzy model.

However, in contrast, the validation results of the linguistic fuzzy outputs was generally poor. The agreements obtained were usually *poor*, *slight* or *moderate* — even when what might be considered generous weight matrices were used for weighted kappa. This also holds true for the linguistic interpretation variables, where the kappas have been artificially raised through considering the variables independent. Probably the most positive statement that can be made is that the fuzzy expert system is on a par with the experts. Its average agreements were within the range of experts for the vessel identification, higher than the experts for *acidemia*, but slightly lower than the experts for *component* and *duration*. It is not a simple

case that the experts agreed, and the fuzzy expert system didn't. It is more a case that neither the experts nor the system obtained good agreements. This may be due to the 'artificial' nature of the linguistic validation task, in that experts were forced to categorise results by labels, and were unable to express a fuller clinical opinion of the meaning of the data. It may leave scope for enhancing the fuzzy linguistic approximation, but this is unlikely to achieve much unless higher inter-expert agreement could be achieved through a different design of validation task.

The performance of expert *D* warrants some comment. This expert has not been used before in either the development or validation of either the crisp expert system or the fuzzy expert systems. The expert explicitly stated before taking part in the study that they probably placed much more emphasis on the base deficit than either other experts or the fuzzy expert system. Indeed, the expert expressed the opinion that base deficit should be used *almost exclusively* for interpretation of acid-base data. It was no surprise, then, to find lower agreement between expert *D* and the other experts, or between expert *D* and the fuzzy expert system in its default state (in which pH and BD_{ecf} are weighted equally). It was gratifying to find that the fuzzy expert system could be adjusted to achieve good agreement for the numeric interpretation (ranking) task, by decreasing the pH weight to zero (Table 6.15). It is possible that this numeric agreement with expert *D* may be improved still further through alteration of the combination Equation 6.9, to place more emphasis on the metabolic component of the acidemia.

Although the numeric interpretation output used to generate the PIAD index obtained excellent agreement with experts *A*, *B*, and *C*, it important to note that the PIAD index has not been validated against external clinical data. The lower agreement achieved with expert *D* reinforces the point that the experts differ as to their interpretation of what constitutes 'bad' acid-base results. True evaluation of the PIAD index (and expert opinion) would require clinical studies involving the collection of huge numbers of cases. The conjecture made here is that without other antepartum, intrapartum and postpartum data the *possibility* that the infant may have suffered intrapartum asphyxial damage is the most comprehensive assessment that can be made from acid-base data. A conventional *probability* that the infant had

suffered intrapartum asphyxial damage would be the ultimate goal of an expert system for the assessment of neonatal outcome.

Finally, it must be noted that the various uncertainty outputs produced by the fuzzy expert system, as described in Section 6.3.1, remain un-utilised and un-validated at present. It was envisaged that these uncertainty outputs would be combined with the linguistic approximation of the output variables to produce complete sentences indicating the interpretations and their confidence. Given the relatively poor performance of the linguistic outputs from the fuzzy system, the uncertainty in these outputs would probably not be useful at present. The results from this validation study suggest that an amalgamation of the crisp expert system and the fuzzy expert system might be the best overall expert system solution. The crisp expert system would be used for categorisations to produce linguistic output, and the fuzzy expert system would be used to identify paired vessels by utilising the numeric output from the vessel identification rules, and to produce the PIAD index from the numeric output of the interpretation rules.

Chapter 7

Review, Future Work and Conclusions

7.1 Review

The successful development of a crisp expert system for umbilical acid-base assessment was the first achievement of this work. The title of this thesis is *Intelligent Techniques for Handling Uncertainty in the Assessment of Neonatal Outcome*. It is usual in the context of expert systems to consider *uncertainty handling* to be the explicit representation of uncertainty in the form of certainty factors, probability, fuzzy logic, etc. However, the crisp expert system developed here *does* also handle uncertainty in umbilical acid-base assessment — but in an implicit form. The junior, or inexperienced clinician does not, in general, have the knowledge to accurately assess umbilical acid-base information. Thus, any interpretation that such a person performs is implicitly uncertain. The crisp expert system decreases this uncertainty by producing a consistent and reliable interpretation based on rules that embody expert knowledge.

The crisp expert system technology has been licensed to a commercial company, and has been placed at over twenty hospitals in the United Kingdom. Expert systems reaching routine clinical use have been rare, largely as a result of difficulties in clinical validation. It appears to be the case that higher standards are required of expert systems, than are required of the human clinicians. It is generally accepted, or even expected, that humans make

mistakes, but as long as the human clinician is not negligent, then occasional mistakes are tolerated. This does not appear to be so for expert systems. An expert system that offers advice for clinical intervention, albeit as advice intended for decision support such that the final responsibility for decision remains with the clinician, is expected to be infallible [39]. The concept of an *interpretation support* expert system as a distinct class of expert system, that is *less interventionist* than a traditional *decision support* expert system, has been introduced in this work. The design of the expert system as an interpretation support system is believed to have greatly contributed to its swift transfer to clinical use, through the reduction in validation requirements.

Although the crisp expert system represents a major advance to the current clinical assessment of umbilical acid-base, its lack of explicit uncertainty handling is a limitation. The preliminary fuzzy expert system was developed to overcome this limitation through the introduction of explicit uncertainty in the knowledge base, but it was initially found to perform worse than the crisp expert system when compared to human experts. This observation led to the development of the fuzzy model tuning algorithm described in Chapter 5. The tuning algorithm, based on the method of simulated annealing to perform large dimensional function optimisation, was used to improve the performance of the preliminary fuzzy expert system to match the clinicians. During application of the tuning technique, it was found that adjustment of the height of membership functions such that they were no longer normalised was a successful method of altering the relative importance of input parameters within the inference process. The use of sub-normal membership functions within a fuzzy expert system is a novel contribution to fuzzy knowledge engineering.

The integrated fuzzy expert system incorporated explicit uncertainty in the input data, in addition to the uncertainty in the knowledge base. Modification of the linguistic variables, fuzzy membership functions, and the fuzzy rule base in response to fresh knowledge elicitation sessions resulted in a fuzzy model that performed as well as the previously tuned system, but without the need for sub-normal membership functions. This integrated fuzzy expert system also incorporated a fuzzy rule base for performing vessel identification, in addition to the rule base for performing interpretation. The integrated fuzzy expert system was tested in

a validation study and was found to perform favourably compared to the human experts.

The *possibility of intrapartum asphyxial damage* (PIAD) index introduced in Chapter 6 is a novel formulation of the interpretation of umbilical acid-base. The use of a formal *possibility* maps naturally onto the conventional clinical usage of umbilical acid-base information, in that it is only really useful in the negative sense when used individually, but may be used in the positive sense for group statistics. The fact that the fuzzy expert system can rank acid-base information with the PIAD index extremely closely compared to the human experts indicates that this single index could be used for comparison of acid-base data. Umbilical acid-base assessment has been used as an outcome in its own right for past clinical studies on obstetric care. This has been done on the justification that worse acid-base status is associated with worse neonatal outcome. However, as no single acid-base variable correlates with bad outcome, this has caused a problem. Given that the PIAD index produced by the fuzzy expert system does at least correlate highly with expert opinion, it would be a suitable choice for an objective single variable outcome measure.

The achievements of this work can be summarised as having introduced engineering rigour into the assessment of umbilical acid-base information. The necessity for basic data validation of acid-base parameters prior to interpretation has been largely accepted clinically. The requirement to interpret the pH **and** BD_{ecf} from *both* arterial and venous samples in order to reach an accurate assessment has also been accepted, although it is often still not done in practice. The introduction of the crisp expert system is widely regarded as a significant clinical achievement. The subsequent development of the fuzzy expert system, incorporating explicit uncertainty handling, has increased the embedded intelligence within the expert system to a level which is indistinguishable from the best clinical experts.

There are a number of aspects of this work that could also be of benefit to other areas of medical expert systems research.

1. *Incremental development cycle.* Breaking down the project into the development of first the crisp expert system, then the preliminary fuzzy system and, finally, the integrated fuzzy expert system is believed to have greatly contributed to the overall success

of the project. The crisp system was released to clinical use and enabled the collection of data which was invaluable to the subsequent development of the fuzzy systems. It is believed that both the author and the collaborating clinical experts gained knowledge from the development of the crisp expert system that enabled the successful development of the integrated fuzzy expert system.

2. *Knowledge driven approach.* Although the availability of data has been invaluable in this project, the interpretation of the data has always been guided by clinical expertise. The acquisition of clinical knowledge of acid-base by the author has also been an extremely important aspect of the project. This was most clearly demonstrated during the knowledge driven tuning of the integrated fuzzy expert system described in Chapter 6. When the clinical expert attempted to tune the fuzzy input sets directly, the performance of the system deteriorated. However, when the clinical expert directed the fuzzy outputs and the author modified the fuzzy inputs using knowledge of the desired outputs, knowledge of the internal fuzzy model and knowledge of acid-base physiology to guide the changes, the performance of the system improved. Whilst the author would by no means claim to be a clinical expert, the acquisition of detailed knowledge of the very specialist 'sub-domain' of this project has been of huge benefit.

3. *Clinical co-operation.* The success of the crisp expert system in achieving commercial release was greatly influenced by close co-operation with the clinical experts and with the clinical staff in the local hospital. The expert system was designed specifically to incorporate ancillary features which would decrease the workload of the clinical staff, by providing printed labels of results to obviate transcription into notes. An expert system that decreases workload is far more likely to receive clinical approval than an expert system that causes staff extra work. In addition, the design of the user-interface (input screens, input buttons, output screens and output labels) was guided by clinical requirements of ease of use and comprehensibility, rather than 'traditional' requirements of functionality.

4. *Simulated annealing model tuning and sub-normal membership functions.* The simulated annealing fuzzy model tuning algorithm is generally applicable to *any* fuzzy model. The only requirement is to find a suitable cost function to minimise. A suitable cost function should somehow evaluate the *error* of the fuzzy model — preferably in a form of mean-square error. The use of sub-normal membership functions is also generally applicable, and may have use in any fuzzy expert system (rather than fuzzy control) application, where theoretical considerations of correctness are less important than the pragmatic mimicking of human expert performance.

7.2 Limitations

The fundamental problem with umbilical cord acid-base assessment is that it currently remains clinically unevaluated [73]. Specifically, it has yet to be proven that performing umbilical acid-base assessment, either on individual infants perceived to be 'at risk' or routinely on every infant delivered, has any direct or indirect clinical benefit. The American College of Obstetricians and Gynecologists continue to recommend assessment of the umbilical arterial acid-base, and *only* on infants with Apgar[5] of three or less [7]. Indeed, there are some midwives, obstetricians and other clinicians who are of the opinion that umbilical acid-base assessment is a waste of time and money. Given knowledge of the physiology of acid-base balance in the blood, and knowledge of the causes of asphyxia and eventual brain damage, it is undeniable that umbilical acid-base assessment does provide a snapshot of information on the infant's condition at birth [77]. It has also been shown that the presence of an arterial metabolic acidemia is associated with a higher prevalence of neonatal complications [81].

The question remains, however, as to whether this information is of any use. For the clinician responsible for the delivery of the infant, the information is retrospective and therefore useless in guiding the delivery just performed. For the neonatologist or paediatrician responsible for the ongoing treatment of a damaged infant, much more precise information is obtainable directly from the freshly accessible infant. It was stated in the discussion of Chapter 4 that there are four main uses for umbilical acid-base information:

1. individual feedback,
2. group audit statistics,
3. neonatal guidance, and
4. medicolegal protection.

The null (or sceptical) hypotheses to refute each of these uses may be as follows:

1. individual feedback — the individuals don't look at the results, or don't change their practice even if they get bad results;
2. group audit statistics — even though worse acid-base status is associated with bad outcome, it is possible that superior neonatal care or other compensatory effects may cause a group that have worse mean acid-base status to have no more brain damage: in any case, it would be better to measure the prevalence of neonatal complications directly;
3. neonatal guidance — the neonatologists and paediatricians don't look at umbilical acid-base results, or if they do require acid-base status, they should just take fresh blood and measure it directly on the neonate;
4. medicolegal protection — the courts will reach the same decision regardless of umbilical acid-base status.

True substantiation of each of these hypotheses (or their alternatives) would require further studies, almost certainly involving clinical randomised control trials, that would inevitably entail high cost — to date this has not been done. With this in mind, it can be seen that true *evaluation* of an expert system for umbilical acid-base assessment is effectively impossible at present.

The validation of the fuzzy expert system has been carried out exclusively against expert opinion. This expert opinion is time consuming to obtain, and requires the availability and co-operation of suitable experts. Given the difficulty in obtaining such expert opinion, it

would have been desirable to utilise the two sets of fifty rankings described in Chapters 5 and 6 to directly compare the performance of all three versions of the fuzzy expert system (*fuzzy*^{0}, *fuzzy*^{1}, and *fuzzy*^{2}). Unfortunately, however, this was not possible. The first set of fifty cases was selected from the restricted set of previously validated full paired arterial-venous results. This set was then used to tune the performance of *fuzzy*^{0}, resulting in the *fuzzy*^{1} system, and was also used to guide the development of the integrated *fuzzy*^{2} system. Thus it was effectively used as training data for both the preliminary and integrated fuzzy systems. The second set of fifty cases was selected from the entire database of results, to ensure that it contained cases with single samples, missing parameters and parameter errors. This enabled the set to be used to validate the integrated fuzzy system (*fuzzy*^{2}), but unfortunately prevents the data set from being used to additionally compare the performance of the preliminary *fuzzy*^{0} and *fuzzy*^{1}.

Even within the limitations discussed above, the fuzzy expert system needs to be validated more thoroughly against other clinical data. Unfortunately, suitable *other clinical data* is very hard to find. Without additional clinical validation, however, it remains unclear at present whether there would be any benefit in releasing the fuzzy expert system to clinical use. Certainly though, without additional validation, there is no information available as to whether the transformation of the numeric interpretation fuzzy output into the PIAD index has any basis. This is due to the use of a ranking statistic to compare the expert system output against expert opinion. Ranking is invariant across the transform, and thus the same ranking statistic would have been obtained whether the plain numeric output or the PIAD index was used.

The linguistic approximation algorithm implemented in the integrated fuzzy expert system did not produce good results in the validation against expert opinion. It was also found that the inter-expert agreement was not good either. As already pointed out in the discussion of Chapter 6, this may have been due to the artificial nature of the validation task that was designed. However, it remains a limitation of the present work that a validated linguistic output from the fuzzy system has yet to be found. The original intention was to include explicit uncertainty information in the final output of the fuzzy expert system, but this was not done

despite the fact that several measures of confidence in the fuzzy outputs were implemented. The introduction of explicit uncertainty into the output of an expert system may simply cause too much confusion or uncertainty in the minds of the clinical users.

The final form of the output of any expert system is an integral part of its function, which impacts on how the users perceive the system [33]. At one stage direct output of the fuzzy sets of the output variables was considered, as one expert found these relatively easy to relate to. After more consideration, however, it was felt that inexperienced users would probably find these sets more confusing than even the primitive input parameters, and thus the idea was dropped. A single linguistic output (single sentence) might be the best output for the inexperienced clinical user, with additional information on the confidence in the conclusion available to more experienced users. The single numeric PIAD index would also be output to provide a single continuous variable. The fuzzy output sets of the *acidemia*, *component* and *duration* variables would then be passed to subsequent modules of the comprehensive neonatal outcome expert system.

The creation of a comprehensive expert system for the assessment of neonatal outcome, as illustrated in Figure 1.7 of Chapter 1, has turned out to be a far larger task than originally envisaged. The work described in this thesis has effectively developed the umbilical acid-base assessment module of the complete expert system, leaving the remaining four modules outstanding. Although additional labour information had been collected, attempts to incorporate additional variables into the fuzzy expert system were hampered by the lack of available clinical expertise necessary to formulate a knowledge base. There appears to be insufficient data on the additional umbilical cord blood variables such as glucose, lactate, or co-oximetry information, and a lack of clinical co-operation between the obstetricians responsible for labour and the neonatologists or paediatricians responsible for subsequent care. The experts who collaborated with the developments described herein have subsequently moved to other international posts, such that continued collaboration would be much more difficult.

7.3 Future Work

The software for the crisp expert system has recently been updated to allow connection to the latest generation of blood gas analysers. These analysers can measure the glucose and lactate levels of the blood, which may prove to be a more direct measure of the metabolic state than the pH and base deficit (which has to be derived from the pH and pCO_2). They can also measure additional metabolites such as sodium, potassium, calcium, and chloride ions, and can also perform co-oximetry to obtain the haemoglobin levels, and the oxygenated and deoxygenated ratios. The data collected with these blood gas analysers and the crisp expert system could provide the large databases required for either data-mining or for the formulation of additional clinical models of interpretation. As maternity information systems become more established in the UK, the availability of computerised records for labour information and neonatal follow-up may make such data much more readily accessible.

Formulating the rules for any expert system is usually a difficult problem. The process of 'data-mining' in which rules are extracted from large databases has been applied successfully to the formation of crisp rules for expert systems in the past [55, 84]. Recently, these methods of crisp rule induction have been extended into the domain of fuzzy theory. Fuzzy rule induction is a process whereby a large quantity of real example data is searched for patterns and correlations in order to provide fuzzy rules of interpretation [10, 71]. Fuzzy rule induction combined with expert knowledge elicitation may allow the formulation the rules required for the other modules.

To be useful, an expert system must produce output in a similar form to that produced by experts in the domain. In this domain, as in most medical expert systems, some form of linguistic output is considered essential. Accurate and efficient linguistic approximation is a challenging and as yet unsolved problem. Products of linguistic approximation tend to be complex and unnatural combinations of terms that would not be used by human experts. Future work could be undertaken to improve the linguistic approximation outputs of the current fuzzy expert system, and to design and carry out another linguistic validation study with experts. Alternatively, the linguistic output from the crisp expert system and the numeric

outputs from the fuzzy expert system could be combined to form a single expert system.

More work needs to be carried out on the clinical evaluation of the crisp expert system, in order to try to quantify its effect on clinical practice, and to gauge whether the fuzzy system would provide any additional clinical benefit. It is hoped that work can be undertaken to follow up the current clinical installations, some of which have been in place for over two years now. This may lead to the availability of more maternal and neonatal data, if any of these clinical sites have computerised database systems, and may enable the opinions of the clinical users to be ascertained to provide data on the usage and subjective impact of the system.

A new version of the monitor for intrapartum care featuring analysis of the ST segments of the ECG waveform of the fetus during labour has been developed, known as the STAN2 monitor [101]. It is planned that a large randomised trial will take place in Scandinavia from 1998 in which the STAN2 monitor is compared against the conventional CTG monitor. This will feature the collection of antepartum, intrapartum and postpartum data in a controlled and uniform manner, for up to around 30 000 deliveries. There would be additional follow up of any damaged infants delivered during the trial, which might continue for a number of years subsequently. The data will be used to validate a fuzzy-logic based expert system for ST waveform analysis [56], and it is hoped that the crisp expert system could be used in the trial for the collection and validation of the umbilical cord acid-base data.

There are also plans for a similar sized multi-centre trial to take place in Britain, co-ordinated from Plymouth, to test a CTG interpretation expert system [60]. If either of these trials takes place, the data collected could prove an invaluable source for evaluating more accurately the existing crisp and fuzzy expert systems, and for the development of the additional modules required for the comprehensive system.

7.4 Conclusions

The original goal of this work was to create a comprehensive expert system for the assessment of immediate neonatal outcome, incorporating explicit uncertainty handling. Although this ambitious goal has not been fully achieved, a major contribution has been made through the development of a crisp and fuzzy expert system for the interpretation of umbilical cord acid-base information, which would form a significant part of a comprehensive expert system. The crisp expert system has been released to routine clinical use, and the fuzzy expert system has been validated against expert opinion to perform to a level comparable to the foremost experts in the field.

A combination of the crisp expert system and fuzzy expert system developed here would form the ultimate expert system for umbilical acid-base assessment. The crisp expert system would produce broad linguistic categorisations to advise the inexperienced user. The fuzzy expert system would produce two numeric outputs: the numeric vessel identification output to provide an indication of vessel origin with confidence in the labelling, and the numeric interpretation output to provide a single continuous index representing the severity of the infant's acid-base status. This work has effectively taken the interpretation of umbilical acid-base information to its limit — the enlargement of the expert system to the ideal comprehensive expert system will have to await further clinical developments.

Although it is true that umbilical acid-base assessment has not been clinically evaluated, it *is* an objective measurement of immediate neonatal outcome. It was stated in Chapter 1 that "*the art for the clinician in charge of childbirth is to achieve delivery before the infant's condition has deteriorated enough to cause permanent damage*". It may be suggested that:

> *observation is art, measurement is science.*

Up to now, the assessment of immediate neonatal outcome has been based on observation. The widespread introduction of routine umbilical cord acid-base assessment, with expert analysis performed by the expert system described in this work, would provide the objective measurement necessary to enable the transformation of obstetrics from an art to a science.

References

[1] E. Aarts and J. Korst. *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*. John Wiley & Sons, New York, 1989.

[2] J.B. Adams. A probability model of medical reasoning and the MYCIN model. *Mathematical Biosciences*, 32:177–186, 1976.

[3] K.P. Adlassnig. Fuzzy set theory in medical diagnosis. *IEEE Transactions Systems Man Cybernetics*, 16(2):260–265, 1986.

[4] K.P. Adlassnig and W. Horak. Development and retrospective evaluation of Hepaxpert-I: A routinely-used expert system for interpretive analysis of hepatitis A and B serologic findings. *Artificial Intelligence Medicine*, 7(1):1–24, 1995.

[5] K.P. Adlassnig and G. Kolarz. CADIAC-2: Computer-assisted medical diagnosis using fuzzy subsets. In MM Gupta and E Sanchez, editors, *Approximate Reasoning in Decision Analysis*, pages 219–247. Noth-Holland, New York, 1982.

[6] American Academy of Pediatrics. Use and abuse of the Apgar score. *Pediatrics*, 78:1148–1149, 1986.

[7] American College of Obstetricians and Gynecologists. Utility of Umbilical Cord Blood Acid-Base Assessment; Committee on Obstetric Practice (138). *International Journal Gynecology Obstetrics*, 45:303–304, 1994.

[8] V. Apgar. A proposal for a new method of evaluation of the newborn infant. *Current Researches in Anesthesia and Analgesia*, 32:260–267, 1953.

[9] V. Apgar and L.S. James. Further observations on the newborn scoring system. *American Journal Dis Child*, 104:419–428, 1962.

[10] J.F. Baldwin. Knowledge from data using fuzzy methods. *Pattern Recognition Letters*, 17(6):593–600, 1996.

[11] C.F. Ball and D.A. Mlynski. Fuzzy bi- and multi-partitioning for circuits represented by hypergraphs. *Journal Circuits Systems Computers*, 6:503–526, 1996.

[12] W. Bandler and L. Kohout. Fuzzy power sets and fuzzy implication operators. *Fuzzy Sets Systems*, 4:13–30, 1980.

[13] M. Black. Vagueness: An exercise in logical analysis. *Philosophy of Science*, 4:427–455, 1937.

[14] E. Blair and F.J. Stanley. Intrapartum asphyxia: A rare cause of cerebral palsy. *Journal Pediatrica*, 112:515–519, 1988.

[15] J.N. Blechner. Maternal-fetal acid-base physiology. *Clinics Obstetrics Gynecology*, 36:3–12, 1993.

[16] I.O. Bohachevsky, M.E. Johnson, and M.L. Stein. Generalized simulated annealing for function optimization. *Technometrics*, 28(3):209–217, 1986.

[17] D. Brahams and J. Wyatt. Decision-aids and the law. *Lancet*, 2:632–634, 1989.

[18] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds, and D.B. Rosen. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions Neural Networks*, 3(5):698–713, 1992.

[19] S.W.K. Chan, K.S. Leung, and W.S.F. Wong. An expert-system for the detection of cervical-cancer cells using knowledge-based image analyzer. *Artificial Intelligence Medicine*, 8(1):67–90, 1996.

[20] D.A. Clark and D.O. Hakanson. The inaccuracy of Apgar scoring. *Journal Perinatology*, 8:203–205, 1988.

[21] L. Clark, M. Mugford, and C. Paterson. How does the mode of delivery affect the cost of maternity care? *British Journal Obstetrics Gynaecology*, 98:519–523, 1991.

[22] J. Cohen. A coefficient of agreement for nominal scales. *Educational Psychological Measurement*, 20:37–46, 1960.

[23] J. Cohen. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70:213–220, 1968.

[24] M.E. Cohen and D.L. Hudson. The use of fuzzy variables in medical decision making. In M.M. Gupta and T. Yamakawa, editors, *Fuzzy Computing*, pages 263–271. Elsevier Science, North Holland, 1988.

[25] A. Corana, M. Marchesi, C. Martini, and S. Ridella. Minimizing multimodal functions of continuous variables with the 'simulated annealing' algorithm. *ACM Transactions Mathematical Software*, 13(3):262–280, 1987.

[26] B. de Finetti. *Theory of Probability: A Critical Introductory Treatment (Vol 1)*. John Wiley & Sons, New York, 1974.

[27] Department of Health. *Changing Childbirth - Report of the Expert Maternity Group (The Cumberlidge Report)*. HMSO, London, 1993.

[28] Department of Health. *Report on Confidential Enquiries into Maternal Deaths in the United Kingdom, 1988-1990*. HMSO, London, 1994.

[29] M.J. Dijxhoorn, G.H.A. Visser, H.J. Huisjes, V. Fidler, and B.C.L. Touwen. The relationship between umbilical pH values and neonatal neurological morbidity in full term appropriate-for-dates infants. *Early Human Development*, 11:33–42, 1985.

[30] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York, 1980.

[31] D. Dubois and H. Prade. Fuzzy logics and the generalized modus ponens revisited. *Cybernetics Systems*, 15:293–331, 1984.

[32] D. Dubois and H. Prade. Fuzzy sets - a convenient fiction for modeling vagueness and possibility. *IEEE Transactions Fuzzy Systems*, 2(1):16–21, 1994.

[33] E. Edmonds. Human-computer interface evaluation: Not user-friendliness but design for operation. *Medical Informatics*, 15(3):253–260, 1990.

[34] R. Eisberg and R. Resnick. *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles*. John Wiley & Sons, New York, 1974.

[35] R. Engelbrecht, A. Rector, and W. Moser. Verification and validation. In E.M.S.J. van Gennip and J.L. Talmon, editors, *Assessment and Evaluation of Information Technologies*, pages 51–66. IOS Press, 1995.

[36] T.K. Eskes, H.W. Jongsma, and P.C. Houx. Percentiles for gas values in human umbilical cord blood. *European Journal Obstetrics Gynecology Reproductive Biology*, 14:341–346, 1983.

[37] N.N. Finer, C.M. Robertson, R.T. Richards, L.E. Pinnell, and K.L. Peters. Hypoxic-ischemic encephalopathy in term neonates: Perinatal factors and outcome. *Journal Pediatrics*, 98:112–117, 1981.

[38] J.L. Fleiss. *Statistical Methods for Rates and Proportions (2nd edn)*. John Wiley & Sons, New York, 1981.

[39] J. Fox. On the soundness and safety of expert systems. *Artificial Intelligence Medicine*, 5:159–179, 1993.

[40] L.C. Gilstrap, K.J. Leveno, J. Burris, M.L. Williams, and B.B. Little. Diagnosis of birth asphyxia on the basis of fetal pH, Apgar score, and newborn cerebral dysfunction. *American Journal Obstetrics Gynecology*, 161:825–830, 1989.

[41] K.G. Goldaber and L.C. Gilstrap. Correlations between clinical events and umbilical cord blood acid-base and blood gas values. *Clinics Obstetrics Gynecology*, 36:47–59, 1993.

[42] K.G. Goldaber, L.C. Gilstrap, K.J. Leveno, J.S. Dax, and D.D. McIntyre. Pathologic fetal acidemia. *Obstetrics Gynecology*, 78:1103–1107, 1991.

[43] J. Graham and P. Llewelyn-Jones. *Expert Systems: Knowledge, Uncertainty and Decision*. Chapman and Hall, London, 1988.

[44] K.R. Greene and K.G. Rosén. Intrapartum asphyxia. In M.I. Levene, R. Lilford, M.J. Bennett, and J. Punt, editors, *Fetal and Neonatal Neurology and Neurosurgery (2nd edn)*. Churchill Livingstone, Edinburgh, 1994.

[45] G.R. Grimmet and D.R. Stirzaker. *Probability and Random Processes (2nd edn)*. Oxford Univeristy Press, Oxford, 1992.

[46] S. Haack. *Philosophy of Logics*. Cambridge University Press, Cambridge, 1978.

[47] M. Halim, K.M. Ho, and A. Liu. Fuzzy logic for medical expert systems. *Annals Academy Medicine*, 19(5):672–683, 1990.

[48] A. Halligan, M. Connolly, T. Clarke, R.P. Gleeson, M. Holohan, T. Matthews, M. King, and M.R.N. Darling. Intrapartum asphyxia in term and post term infants. *Irish Medical Journal*, 85:97–100, 1992.

[49] W.L. Hays. *Statistics (4th edn)*. Holt, Rinehart and Winston, New York, 1988.

[50] H.M. Hersh and A. Caramazza. A fuzzy set approach to modifiers and vagueness in natural language. *Journal Experimental Psychology*, 105(3), 1976.

[51] E. Hisdal. Interpretative versus prescriptive fuzzy set theory. *IEEE Transactions Fuzzy Systems*, 2(1):22–26, 1994.

[52] D.L. Hudson and M.E. Cohen. Fuzzy logic in a medical expert system. In M.M. Gupta and T. Yamakawa, editors, *Fuzzy Computing*, pages 273–284. Elsevier Science, North Holland, 1988.

[53] C. Hughes. The representation of uncertainty in medical expert systems. *Medical Informatics*, 14:269–279, 1989.

[54] H.J. Huisjes and J.G. Aarnoudse. Arterial or venous umbilical pH as a measure of neonatal morbidity? *Early Human Development*, 3:155–161, 1979.

[55] A. Hutchinson. *Algorithmic Learning*. Clarendon Press, Oxford, 1994.

[56] E.C. Ifeachor and N.J. Outram. A fuzzy expert system to assist in the management of labour. In *Proceedings of the International ICSC Symposium on Fuzzy Logic*, pages 97–102. Zurich, Siwtzerland, 1995.

[57] P. Jackson. *Introduction to Expert Systems (2nd edn)*. Addison-Wesley, Reading, MA, 1990.

[58] J.S.R. Jang. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions Systems Man Cybernetics*, 23:665–685, 1993.

[59] B. Jennett. Assessment of clinical technologies. *International Journal Technology Health Care*, 4:435–445, 1988.

[60] R.D.F. Keith, S. Beckley, J.M. Garibaldi, J.A. Westgate, E.C. Ifeachor, and K.R. Greene. A multicentre comparative study of 17 experts and an intelligent computer system for managing labour using the cardiotocogram. *British Journal Obstetrics Gynaecology*, 102:688–700, 1995.

[61] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[62] I. Kjellmer. Prenatal and intrapartum asphyxia. In M.I. Levene, M.J. Bennett, and J. Punt, editors, *Fetal and Neonatal Neurology and Neurosurgery*. Churchill Livingstone, London, 1988.

[63] I. Kjellmer. Etiology and pathophysiology of postasphyxial brain damage. *International Journal Technology Assessment in Health Care*, 7(1):106–109, 1991.

[64] G.J. Klir. On the alleged superiority of probabilistic representation of uncertainty. *IEEE Transactions Fuzzy Systems*, 2(1):27–31, 1994.

[65] G.J. Klir and T.A. Folger. *Fuzzy Sets, Uncertainty and Information*. Prentice-Hall, London, 1988.

[66] B. Kosko. *Fuzzy Thinking: The New Science of Fuzzy Logic*. Harper-Collins, London, 1994.

[67] B. Kosko. The probability monopoly. *IEEE Transactions Fuzzy Systems*, 2(1):32–33, 1994.

[68] L.I. Kuncheva. Pattern-recognition with a model of fuzzy neuron using degree of consensus. *Fuzzy Sets Systems*, 66(2):241–250, 1994.

[69] J.R. Landis and G.G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174, 1977.

[70] M. Laviolette and J.W. Seaman. The efficacy of fuzzy representations of uncertainty. *IEEE Transactions Fuzzy Systems*, 2(1):4–15, 1994.

[71] D.H. Lee and Kim M.H. Database summarization using fuzzy ISA hierarchies. *IEEE Transactions Systems Man Cybernetics*, 27(4B):671–680, 1997.

[72] K.M. Lee, D.H. Kwak, and H. Lee-Kwang. Fuzzy inference neural network for fuzzy model tuning. *IEEE Transactions Systems Man Cybernetics*, 26(4B):637–645, 1996.

[73] S.C. Leeson, L. Edozien, and A.M. Mander. There are alternatives to Apgar. *Hospital Update*, May:212–254, 1995.

[74] M.I. Levene, J. Kornberg, and T.H.C. Williams. The incidence and severity of post-asphyxial encephalopathy in full-term infants. *Early Human Development*, 11:21–26, 1985.

[75] M.I. Levene, C. Sands, H. Grindulis, and J.R. Moore. Comparison of two methods of predicting outcome in perinatal asphyxia. *Lancet*, i:67–69, 1986.

[76] C.T. Lin and C.S.G Lee. Neural-network based fuzzy logic control and decision making. *IEEE Transactions Computers*, 40(12):1320–1336, 1991.

[77] J.A. Low. The role of blood gas and acid-base assessment in the diagnosis of intrapartum fetal asphyxia. *American Journal Obstetrics Gynecology*, 159:1235–1240, 1988.

[78] J.A. Low, R.S. Galbraith, D.W. Muir, H.L. Killen, E.A. Pater, and E.J. Karchmar. The relationship between perinatal hypoxia and newborn encephalopathy. *American Journal Obstetrics Gynecology*, 152:256–260, 1985.

[79] J.A. Low, R.S. Galbraith, D.W. Muir, H.L. Killen, E.A. Pater, and E.J. Karchmar. Motor and cognitive deficits after intrapartum asphyxia in the mature fetus. *American Journal Obstetrics Gynecology*, 158:356–361, 1988.

[80] J.A. Low, E.J. Karchmar, L. Broekhoven, T. Leonard, M.J. McGrath, S.R. Pancham, and W.N. Piercy. The probability of fetal metabolic acidosis during labor in a population at risk as determined by clinical factors. *American Journal Obstetrics Gynecology*, 141:941–951, 1981.

[81] J.A. Low, C. Panagiotopoulos, and E.J. Derrick. Newborn complications after intrapartum asphyxia with metabolic acidosis in the term fetus. *American Journal Obstetrics Gynecology*, 170:1081–1087, 1994.

[82] E.H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal Man-Machine Studies*, 7:1–13, 1975.

[83] F. Mandl. *Statistical Physics (2nd edn)*. John Wiley & Sons, New York, 1988.

[84] D. Michie, D.J. Spiegelhalter, and C.C. Taylor, editors. *Macine Learning, Neural and Statistical Classification*. Ellis Horwood Limited, New York, 1994.

[85] P.L. Miller and D.F. Sittig. The evaluation of clinical decision support systems: What is necessary versus what is interesting. *Medical Informatics*, 15(3):185–190, 1990.

[86] R.A. Miller, H.E. Pople, and J.D. Myers. Internist-1: An experimental computer-based diagnostic consultant for general internal medicine. *New English Journal Medicine*, 307(8):468–476, 1982.

[87] R.E. Neapolitan. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. John Wiley & Sons, New York, 1989.

[88] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[89] K.B. Nelson and J.H. Ellenberg. Apgar scores as predictors of chronic neurological disability. *Pediatrics*, 68:36–44, 1981.

[90] K.B. Nelson and J.H. Ellenberg. Antecedents of cerebral palsy. *New English Journal Medicine*, 315:81–86, 1986.

[91] K.B. Nelson and A. Leviton. How much of neonatal encephalopathy is due to birth asphyxia? *American Journal Dis Child*, 145:1325–1331, 1991.

[92] A.M. Norwich and I.B. Turksen. A model for the measurement of membership and the consequences of its empirical implementation. *Fuzzy Sets Systems*, 12:1–25, 1984.

[93] R.M. O'Keefe, O. Balci, and E.P. Smith. Validating expert system performance. *IEEE Expert*, 2(4):81–90, 1987.

[94] T.J. O'Leary, M. Goul, K.E. Moffitt, and A.E. Radwan. Validating expert systems. *IEEE Expert*, 5(3):51–58, 1990.

[95] W. Pedrycz and J. Valente de Oliveira. An algorithmic framework for development and optimization of fuzzy models. *Fuzzy Sets Systems*, 80:37–55, 1996.

[96] W. Pedrycz and J. Valente de Oliveira. Optimization of fuzzy models. *IEEE Transactions Systems Man Cybernetics*, 26(4B):627–636, 1996.

[97] L.L.H. Peeters, R.E. Sheldon, M.D. Jones, E.L. Makowski, and G. Meschia. Blood flow to fetal organs as a function of arterial oxygen content. *American Journal Obstetrics Gynecology*, 135:637–643, 1979.

[98] W.H. Press, S.A. Teukolsky, W.T. Vettering, and B.P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing (2nd edn)*. Cambridge University Press, Cambridge, 1992.

[99] C.M.T. Robertson, N.N. Finer, and M.G.A. Grace. School performance of survivors of neonatal encephalopathy associated with birth asphyxia at term. *Journal Pediatrics*, 114:753–760, 1989.

[100] G. Rooth. *Perinatal Acid-Base Balance*. Studentlitteratur, Lund, Sweden, 1988.

[101] K.G. Rosén and R. Luzietti. The fetal electrocardiogram: ST waveform analysis during labour. *Journal Perinatal Medicine*, 22:501–512, 1994.

[102] K.G. Rosén and K.W. Murphy. How to assess fetal metabolic acidosis from cord samples. *Journal Perinatal Medicine*, 19:221–226, 1991.

[103] R. Rovatti and Guerrieri. Fuzzy sets of rules for system identification. *IEEE Transactions Fuzzy Systems*, 4:89–102, 1996.

[104] Royal College of Obstetricians and Gynaecologists. Recommendations Arising from the 26th RCOG Study Group. In J.A.D. Spencer and R.H.T. Ward, editors, *Intrapartum Fetal Surveillance*, page 392. RCOG Press, London, 1993.

[105] V.J. Ruth and K.O. Raivio. Perinatal brain damage: Predictive value of metabolic acidosis and the Apgar score. *British Medical Journal*, 297:24–27, 1988.

[106] W.C. Salmon. *Zeno's Paradoxes*. Bobbs-Merrill Company, New York, 1970.

[107] H.B. Sarnat and M.S. Sarnat. Neonatal encephalopathy following fetal distress. *Archives Neurology*, 33:696–705, 1976.

[108] D.M. Schreck, D. Zacharias, and C.F.V. Grunau. Diagnosis of complex acid-base disorders: Physician performance versus the microcomputer. *Annals Emergency Medicine*, 15(2):164–170, 1986.

[109] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, New Jersey, 1976.

[110] K. Shimomura, H. Shono, M. Kohara, A. Uchiyama, Y. Ito, and H. Sugimori. Neonatal assessment using the Apgar fuzzy expert system. *Computers Biology Medicine*, 24(3):171–178, 1994.

[111] H. Shono, M. Oga, K. Shimomura, M. Yamasaki, Y. Ito, M. Muro, and H. Sugimori. Application of fuzzy logic to the Apgar scoring system. *International Journal Biomedical Computing*, 30:113–123, 1992.

[112] E.H. Shortliffe. *MYCIN: A Rule-Based Computer Program for Advising Physicians Regarding Microbial Therapy Selection*. PhD dissertation, Stanford University, Stanford, 1974.

[113] S. Siegel and N.J. Castellan. *Nonparametric Statistics for the Behavioural Sciences (2nd edn)*. McGraw-Hill, New York, 1988.

[114] O. Siggaard-Andersen. An acid-base chart for arterial blood with normal and pathophysiological reference areas. *Scandanavian Journal Clinical Laboratory Investigation*, 27:239–245, 1971.

[115] O. Siggaard-Andersen. *The Acid-Base Status of the Blood (4th edn)*. Munksgaard, Copenhagen, 1976.

[116] C.H. Sing and B. Postlethwaite. Fuzzy relational model-based control applying stochastic and iterative methods for model identification. *Chemical Engineering Research Design*, 74:70–76, 1996.

[117] F.J. Stanley and E. Blair. Why have we failed to reduce the frequency of cerebral palsy? *Medical Journal Australia*, 154:623–626, 1991.

[118] K.A. Stroud. *Further Engineering Mathematics (3rd edn).* Macmillan, London, 1996.

[119] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions Systems Man Cybernetics*, 15(1):116–132, 1985.

[120] J.A. Thorp, G.A. Dildy, E.R. Yeomans, B.A. Meyer, and V.M. Parisi. Umbilical cord blood gas analysis at delivery. *American Journal Obstetrics Gynecology*, 175:517–522, 1996.

[121] J.A. Thorp, J.E. Sampson, V.M. Parisi, and R.K. Creasy. Routine umbilical cord blood gas determinations? *American Journal Obstetrics Gynecology*, 161:600–605, 1989.

[122] R.M. Tong and J. Efstathiou. A critical assessment of truth functional modification and its use in approximate reasoning. *Fuzzy Sets Systems*, 7:103–108, 1982.

[123] M.E. Towell. The rationale for biochemical monitoring of the fetus. *Journal Perinatal Medicine*, 16:55–70, 1988.

[124] I.B. Turksen and Y. Tian. Combination of rules or their consequences in fuzzy expert systems. *Fuzzy Sets Systems*, 58:3–40, 1993.

[125] United Kingdom Office of Population Censuses and Surveys. *Mortality Statistics: Perinatal and Infant.* HMSO, London, 1993.

[126] J. Valente de Oliveira. A design methodology for fuzzy system interfaces. *IEEE Transactions Fuzzy Systems*, 3:404–414, 1995.

[127] D. Vanderbilt and S.G. Louie. A Monte Carlo simulated annealing approach to optimization over continuous variables. *Journal Computational Physics*, 56:259–271, 1984.

[128] H. Watanabe, W.J. Yakowenko, Y.M. Kim, J. Anbe, and T. Tobi. Application of a fuzzy discrimination analysis for diagnosis of valvular heart disease. *IEEE Transactions Fuzzy Systems*, 2(4):267–276, 1994.

[129] J.A. Westgate, J.M. Garibaldi, and K.R. Greene. Umbilical cord blood gas analysis at delivery: A time for quality data. *British Journal Obstetrics Gynaecology*, 101:1054–1063, 1994.

[130] J.A. Westgate, M. Harris, J.S.H. Curnow, and K.R. Greene. Plymouth randomised trial of cardiotocogram only versus ST waveform plus cardiotocogram for intrapartum monitoring in 2400 cases. *American Journal Obstetrics Gynecology*, 169:1151–1160, 1993.

[131] T. Whalen and B. Schott. Alternative logics for approximate reasoning in expert systems: A comparative study. *International Journal Man-Machine Studies*, 22:327–346, 1985.

[132] N. Wilson. Vagueness and bayesian probability. *IEEE Transactions Fuzzy Systems*, 2(1):34–36, 1994.

[133] C.L. Winkler, J.C. Hauth, J.M. Tucker, J. Owen, and C.G. Brumfield. Neonatal complications at term as related to the degree of umbilical artery acidemia. *American Journal Obstetrics Gynecology*, 164:637–641, 1991.

[134] H.W. Winston. *Artificial Intelligence (2nd edn).* Addison-Wesley, Reading, MA, 1984.

[135] J. Wyatt and D. Spiegelhalter. Evaluating medical expert systems: What to test and how? *Medical Informatics*, 15(3):205–217, 1990.

[136] R.R. Yager. On a general class of fuzzy connectives. *Fuzzy Sets Systems*, 4:235–242, 1980.

[137] E.R. Yeomans, J.C. Hauth, L.C. Gilstrap, and D.M. Strickland. Umbilical cord pH, pCO2, and bicarbonate following uncomplicated term vaginal deliveries. *American Journal Obstetrics Gynecology*, 151:798–800, 1985.

[138] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[139] L.A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions Systems Man Cybernetics*, 3:28–44, 1973.

[140] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning - I. *Information Sciences*, 8:199–249, 1975.

[141] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning - II. *Information Sciences*, 8:301–357, 1975.

[142] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning - III. *Information Sciences*, 9:43–80, 1975.

[143] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *International Journal Fuzzy Sets Systems*, 1:3–28, 1978.

[144] L.A. Zadeh. The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets Systems*, 11:199–227, 1983.

[145] G. Zarkadakis, E.R. Carson, D.G. Cramp, and L. Finkelstein. ANABEL: Intelligent blood-gas analysis in the intensive care unit. *International Journal Clinical Monitoring*, 6:167–171, 1989.

Appendix A

Specification of Crisp Expert System

Functional Specification For DataCare V2.2.1

AMENDMENTS TO:
FUNCTIONAL SPECIFICATION FOR BLOOD GAS ANALYSIS ADVISOR, 14-Aug-92

1. GENERAL SPECIFICATION

1.1 The software will link an IBM-PC compatible (PC) to a CIBA-CORNING Blood Gas Analyser (BGA) Model 238.

1.2 The software will be written to run under Microsoft Windows 3.1 and will run on any true IBM compatible running Microsoft MS-DOS 5.0 and Microsoft Windows 3.1.

1.3 The PC to BGA communications will be via an RS-232 link connecting the PC serial port COM1 or COM2 to the BGA Data Output port.

1.4 **Two levels of user are recognised: USER and TECHNICAL. USER users are those using the system for measurement and logging of blood and quality control samples; TECHNICAL users are those with priviledged access to advanced options such as Database Download, etc.**

1.5 **Ten databases are maintained:**

SAMPLES	**Master samples, recording every sample that is performed**
PAIRED	**Paired Cord Samples**
SINGLE	**Single Cord Samples**
FETAL	**Fetal Blood Samples**
NEONATAL	**Neonatal Blood Samples**
ADULT	**Adult Blood Samples**
OTHER	**Miscellaneous Other Samples**
QUALITY CONTROL	**Quality Control Samples**
CALIBRATIONS	**One- and two-point Calibrations**
UNKNOWNS	**Samples of unknown origin**

Detailed descriptions of the contents and format of each database file are given in the DataCare user manual.

1.6 **A Windows Private Profile File contains all information relevant to a particular configuration of DataCare. The options available are described in PROFILE.DOC.**

2. PROGRAM SPECIFICATION

2.1 **The Private Profile File contains the option to turn the optional auditing features ON or OFF.** With AUDIT ON extra information relevant to the auditing of all use of the system will be gathered and stored to the SAMPLES database.

2.2 The Baud rate will be automatically sensed, the connection between PC and BGA established, and the machine ID of the BGA will be read.

2.3 An initial screen will be displayed consisting of a menu-bar menu, a status line and the GENERAL user options. The menu-bar will contain additional options relevant only to MAINTENANCE users. The status line will display details of the communications link and program state. **Two menu-bars are available. The User menu-bar is initially displayed and contains a restricted number of options available. The Technical menu-bar is enabled by entering a password and contains the full number of options available (see section 3).**

2.4 An *endless* loop will then be entered.

2.4.1 Three main options are displayed:

Paired Artery/Vein Cord Sample	(PAIRED)
Single Cord Sample	(SINGLE)
Other Single Sample	(OTHER)

The user selects an option by keyboard or mouse.

2.4.2 If OTHER the user will be prompted to choose the type of sample from a pre-set list of options: **FETAL, NEONATAL, ADULT, OTHER or QUALITY CONTROL.**

2.4.3 If AUDIT ON the user will be prompted for Personal Identification Number (PIN).

2.4.4 If PAIRED, SINGLE or certain OTHER types the user will be prompted for Patient Name and Hospital Identification Number.

2.4.5 If PAIRED the user will be prompted to enter two samples into the BGA and the information will be recorded. **During sampling, a Sampling In Progress dialog box appears giving the user an approximate indication of the remaining time.**

2.4.6 If SINGLE or OTHER the user will be prompted to enter one sample into the BGA and the information will be recorded. **During sampling, a Sampling In Progress dialog box appears giving the user an approximate indication of the remaining time.**

2.4.7 The sample results will be displayed to the screen, and printed out **on multiple sticky labels via the parallel port to an 80-column dot matrix printer. The existing 'ticker-tape' results are still available** via the existing BGA printer.

2.4.8 If PAIRED or SINGLE the information will be analysed and the user given feedback as to the implications and accuracy of the samples.

i. All results will be checked to ensure the BGA has not identified an error in any measurement. If so the user will be notified **immediately after the sample is received. The user will be given the option to Retry the sample, Ignore the error or Abandon the operation.**

ii. The pH and pCO2 will be checked to ensure separate samples are available (PAIRED only) and that the readings are reasonable.

iii. The base deficit of the extracellular fluid will be calculated.

iv. A summary of results will be given with *expert* advice as to the implications. **Additional detailed explanations of the meanings and derivations of the summary is available on request.**

2.4.9 The sample(s) information received from the BGA will be stored to the SAMPLES database **and to a separate database specific to each type of sample.**

2.5 When the BGA performs a calibration, a dialog box will be placed on the screen informing the users that the BGA is off-line, and giving an approximate indication of the remaining time. The calibration information received from the BGA will be stored to the CALIBRATIONS database.

2.6 The **Technical** menu-bar will provide additional (password-protected) options allowing MAINTENANCE users to off-line databases to backup media (floppy disk), to attempt reconnection to a machine if communications are interrupted and to quit the program **(see section 3).**

2.7 An on-line context-sensitive help system will be available to advise users on how to use the system ~~or to provide additional explanations of results and reasoning behind summaries~~.

3. MENU-BAR COMMANDS

3.1 The User menu-bar provides access to the following commands:

View Quality Controls	Allows the user to view past QC results
View Calibrations	Allows the user to view past Calibration results
Help Menu	Allows the user access to the help system
Switch to Technical Menu	Allows the user to gain access to the Technical menu-bar (this option is password protected)

3.2 The Technical menu-bar provides access to the following commands:

Set Password	Allows the user to change the menu-bar password
Set Database Viewer	Allows the user to set the databae package used to view databases
Set Minimum Field Widths	Allows the user to set minimum required characters for each field
Set Miscellaneous Data	Allows the user to set hospital name, timeout period and number of blank labels
Maintain Audit Database	Allows the user to maintain the database of audit ID's
Database Download	Allows the user to copy and clear-down ALL databases
Database Copy	Allows the user to copy ALL databases
Printer Setup	Allows the user to set the printer type and label copies
Exit Windows	Allows the user to exit Windows
Exit	Allows the user to exit the DataCare system
Sample Labels Enabled	Allows the user to enable sample printing
Sample Labels Disabled	Allows the user to disable sample label printing
QC Labels Enabled	Allows the user to enable QC label printing
QC Labels Disabled	Allows the user to disable QC label printing
Reprint Last Label	Allows the user to reprint the last label
Print Test Label	Allows the user to print a test label
Copy Each Database	Allows the user to copy each database
View Each Database	Allows the user to view each database
Help Menu	Allows the user access to the help system
Switch to User Menu	Allows the user to gain access to the User menu-bar

Appendix B

Operational Sequence Document for the Crisp Expert System

DataCare Operational Sequence Data

The DataCare application runs under the Microsoft Windows 3.1 operating system (henceforth referred to as WINDOWS).

This document assumes that the reader is fully aware of the internal workings of WINDOWS. For details see the Microsoft Windows 3.1 Software Development Kit (SDK) References (Volumes 1 to 4).

Notation

Actual variable names are given in bold text e.g. **State**.

Procedure(function) names are given in bold text as *module file name*::*procedure name* e.g. **BGAMAIN::WinMain**.

If an action causes a WINDOWS message to be generated that will be dealt with immediately, the notation '[send **WM_MESSAGE** -> ***procedure***]' is used. If the message is posted (i.e. placed in the message queue but not dealt with immediately), the notation '[post **WM_MESSAGE** -> ***procedure***]' is used.

WINDOWS SDK internal functions are in bold text.

BGA238 refers to the Ciba-Corning 238 Blood Gas Analyser.

Overview

The DataCare application consists of the main application window, which includes the main caption-bar, menu-bar, the status windows and a blank surface client area, and a number of modeless dialog boxes which have no visible caption-bar, menu-bar or border. The main window handle is always held in the **hMainWnd** global variable.

At any time the program is considered to be in a certain state which is held in the **State** global variable. A modeless dialog box for each state is created at start up and the window handles are held in the **hStateWnd** global array. The modeless dialog box corresponding to the current state is made visible and its handle is held in the **hCurrentWnd** global variable. At the same time, all other modeless dialogs are made invisible and are therefore inactive.

The main window's message routine is **BGAMAIN::WndProc**, whilst all the modeless dialog box routines are in **BGADLG**. The modeless dialogs and their respective message routines are summarised below and several of them share the same routines (see **BGAMAIN::CreateChildWindows** for details). This adds to the complexity of some of the routines but greatly cuts down on code duplication. If necessary the **State** variable can be used to distinguish between different states using the same dialog routine.

State	Name	Dialog Name	Dialog Procedure
0	Main Menu	MainMenu	MenuProc
1	Miscellaneous Menu	OtherMenu	MenuProc
2	Sample Out of Sequence Menu	SampleMenu	MenuProc
3	Paired Cord Sample Patient Information	PatientInfo	PatientProc
4	Single Cord Sample Patient Information	PatientInfo	PatientProc
5	Other Sample Patient Information	PatientInfo	PatientProc
6	Paired Cord First Sample Prompt	Prompt	PromptProc
7	Paired Cord Second Sample Prompt	Prompt	PromptProc
8	Single Cord First Sample Prompt	Prompt	PromptProc
9	Other Cord First Sample Prompt	Prompt	PromptProc
10	Paired Cord Sample Results	PairedData	DataProc
11	Single Cord Sample Results	SingleData	DataProc
12	Other Cord Sample Results	OtherData	DataProc
13	Quality Control Batch Information	BatchInfo	PatientProc
14	Quality Control First Sample Prompt	Prompt	PromptProc
15	Quality Control Sample Results	BatchData	DataProc
16	Sampling Error Prompt	SampleError	SampleErrorProc

As all screens are modeless dialogs, WINDOWS controls the keyboard and mouse interfaces internally without the need for the application to contain any user-interface logic. Because of this, the behaviour of the TAB/SHIFT+TAB, ENTER and ESCAPE keys adopt default behaviour and conform to the standard WINDOWS dialog box interface.

Any communication occuring at the port causes a **WM_COMMNOTIFY** message to be sent to the main window procedure, and two timers are set up to send **WM_TIMER** messages to the main window; the user timer controls the timing-out of modeless and modal dialog boxes to return to the main menu and the calibration timer controls the detection of a calibration event occuring on the BGA238.

Whilst the current modeless is displayed, WINDOWS automatically diverts all user input messages (e.g. mouse or keyboard actions) to the modeless dialog procedure, but the main window continues to receive all menu command messages, the communications messages and the user and calibration timer messages. Thus the main window procedure may process such a message at any point in time.

Any modal dialog that is displayed (for example a **MessageBox**) appears over the top of the current modeless dialog and grabs all user input messages and menu command messages; the modal dialog handle is held in the **hModalDlg** global variable. All commands selected from the main window menu-bar (or via keyboard accelerators) are passed through **BGAMAIN::WndProc** via a **WM_COMMAND** message.

The current state can be altered by user input (e.g. menu selections) or communications events (e.g. receiving a sample) and the control of which state is moved to, which depends on the current state and the precipitating event (context), is governed by the **BGALIB::MoveState** procedure.

The 'sampling in progress' or 'calibration in progress' progress bars and the 'waiting for data' dialogs are implemented as modeless dialogs that overlay and disable the currently displayed modeless dialog screen.

The type of the sampling selected is indicated by the **Origin** global variable, which is set as soon as the user selects the type of sample from the 'main' menu, 'miscellaneous' menu or 'sample out of sequence' menu.

Four global variables are used to hold the main application information; **Patient** holds the patient information (including operator audit information), **Sample0** holds the first sample of a paired-cord or the only sample of any single and **Sample1** holds the second sample of a paired-cord (only), **Calibration** holds the last 1 point or 2 point calibration data.

Due to the fact that the BGA238 sends no code to indicate the beginning of an automatic calibration, the detection of calibration sequences is performed by a method of continual handshaking between the BGA238 and the application. At a frequency specified by the calibration timer, the application sends the BGA238 a calibration data request- if the BGA238 is NOT calibrating it will respond with the last calibration data block, but if it has started a calibration the BGA238 will ignore the request. Thus if the calibration timer goes off and no reply from the BGA238 has been received since the last calibration timer, then the application assumes a calibration is taking place.

Unfortunately the BGA238 also goes offline as soon as the probe door is opened, so the application attempts to cope with this by disabling the calibration timer whenever it is expecting a sample to be received. In case of such 'false' calibrations the application sends the BGA238 a continual stream of calibration data requests once per second whilst the 'calibration in progress dialog' is displayed. If a response is obtained, the dialog is immediately terminated.

On startup, **BGAMAIN::InitApplication** registers the main application window class and **BGAMAIN::InitInstance** creates the main window. However, before **CreateWindow** in **InitInstance** returns, a **WM_CREATE** message is sent to **BGAMAIN::WndProc** which causes **BGAMAIN::OnCreate** to run. This in turn calls **BGAMAIN::CreateChildWindows** to create all the child windows and modeless dialogs, followed by **BGADAT::OpenConnection** to open the connection to the BGA238. Only if both these calls are successful does the original **CreateWindow** return success to **InitInstance** and the application continue to the main message loop.

PROCEDURES BY MODULE

PROCEDURE **BGAMAIN::WinMain (Program entry point)**

```
IF the application is already running
   call BGALIB::ErrorBox to message "application is already running"
   EXIT application immediately

IF the current screen is NOT VGA mode (or compatible)
   call BGALIB::ErrorBox to message "display is not supported"
   EXIT application immediately

allocate space for 32 messages in the incoming message queue
IF message allocation fails
   prompt "message allocation failed: RETRY or CANCEL?"
   IF reply is RETRY
      repeat message allocation
   IF reply is CANCEL
      EXIT application immediately

call BGAMAIN::InitApplication to register the main window class
IF registration failed
   EXIT application immediately

call BGAMAIN::InitInstance to create windows, initialise data and connect to BGA238
IF initialisation failed
   EXIT application immediately

REPEAT get a message from the application message queue
   IF the message is destined for a modeless dialog box
      send message to BGADLG::'routine for currently active dialog box'
   ELSE (its for the main window)
      send message to BGAMAIN::WndProc
UNTIL a WM_QUIT message is received
```

END PROCEDURE

PROCEDURE **BGAMAIN::InitApplication**

```
register the main application window class
```

END PROCEDURE

PROCEDURE **BGAMAIN::InitInstance**

```
get current screen paramters
position the cursor in the middle of the screen

call BGALIB::ReadProfile to read the current configuration parameters from INI file

create the main application window (send WM_CREATE -> BGAMAIN::WndProc)
IF creation failed
   END PROCEDURE immediately

load the menu accelerators
initialise miscellaneous data for WINDOWS 3.1 dialog boxes
set the calibration block size according to the BGA238 software version number:
   173 for version 1 calibrations
   158 for version 2 calibrations with kPa gas units selected
   156 for version 2 calibrations with mmHg gas units selected
set the tick marks on the label printing menu according to saved settings

show the initial window
start the user timer and the calibration timer
```

END PROCEDURE

PROCEDURE **BGAMAIN::WndProc**

```
IF the message is one of
```

```
   WM_ACTIVATE
   WM_ACTIVATEAPP
   WM_CLOSEMODELESS (private to this application)
   WM_COMMAND
   WM_COMMNOTIFY
   WM_CREATE
   WM_CTLCOLOR
   WM_DESTROY
   WM_ENABLE
   WM_SETFOCUS
   WM_SETTIMER (private to this application)
   WM_SYSCOLORCHANGE
   WM_SYSCOMMAND
   WM_TIMER
THEN
   pass the message WM_XXXX to BGAMAIN::OnXxxx
   (e.g. WM_ACTIVATEAPP is sent to BGAMAIN::OnActivateApp)
ELSE (the message is not used by the application)
   pass the message to the default WINDOWS message handler
END PROCEDURE


PROCEDURE BGAMAIN::OnActivate

IF the main window is made inactive by a MessageBox
   store the window handle of the MessageBox in hModalDlg
ELSE IF the main window is being made active
   set the input focus to the currently displayed modeless dialog
END PROCEDURE


PROCEDURE BGAMAIN::OnActivateApp

IF the application is being made active for any reason
   call BGALIB::ProtectAllFiles to set all the data files to read-write
ELSE IF the application is being made inactive AND ...
   ... the database viewer command is activating
   call BGALIB::ProtectAllFiles to set all the data files to read-only
END PROCEDURE


PROCEDURE BGAMAIN::OnCloseModeless

IF the progress-bar dialog is currently displayed
   call BGALIB::CancelProgress to stop it
END PROCEDURE


PROCEDURE BGAMAIN::OnCommand

call BGALIB::CancelModal to stop any modal dialog on display
reset the user time-out period [send WM_SETTIMER -> BGAMAIN::WndProc]

IF command Technical Menu! (code IDM_ENABLE)
   run modal dialog "Password" with BGADLG::PasswordProc to get the password
   IF the user got the password right
      display the full technical user menu-bar

IF command User Menu! (code IDM_DISABLE)
   display the limited normal user menu-bar

IF command File|Set Password (code IDM_FILE_PASSWORD)
   run modal dialog "Password" with BGADLG::PasswordProc to set the password

IF command File|Set Database Viewer (code IDM_FILE_VIEWER)
   run modal dialog "Viewer" with BGADLG::ViewerProc to set the database viewer

IF command File|Set Minimum Field Widths (code IDM_FILE_MINIMUMS)
   run modal dialog "Minimums" with BGADLG::MinimumsProc to set minimums
```

```
IF command File|Set Miscellaneous Data (code IDM_FILE_MISC)
   run modal dialog "MiscData" with BGADLG::MiscProc to set misc data

IF command File|Database Download (code IDM_FILE_DOWNLOAD)
   call BGAMAIN::CmdFileSave to download all files

IF command File|Database Copy (code IDM_FILE_COPY)
   call BGAMAIN::CmdFileSave to copy all files

IF command File|Maintain Audit Database (code IDM_FILE_AUDIT)
   run modal dialog "Audit" with BGADLG::AuditProc to edit the audit IDs in database

IF command File|Printer Setup (code IDM_FILE_PRINTSET)
   run WINDOWS 3.1 common dialog box 'Printer Setup' to configure the printer

IF command File|Exit Windows (code IDM_FILE_EXITWIN)
   prompt "are you sure: OK or CANCEL?"
   IF reply is OK
      EXIT windows immediately

IF command File|Exit (code IDM_FILE_EXIT)
   EXIT application immediately (post WM_CLOSE -> BGAMAIN::WndProc)

IF command Label|Sample Labels Enabled (code IDM_LABEL_SMPON)
   set sample label printing flag to TRUE
   place check (tick) mark on menu item 'enabled' and remove from 'disabled'
   set Sample Labels entry in the [Labels] section of the INI file to ON

IF command Label|Sample Labels Disabled (code IDM_LABEL_SMPOFF)
   set sample label printing flag to FALSE
   place check (tick) mark on menu item 'disabled' and remove from 'enabled'
   set Sample Labels entry in the [Labels] section of the INI file to OFF

IF command Label|Quality Control Labels Enabled (code IDM_LABEL_QCON)
   set quality control label printing flag to TRUE
   place check (tick) mark on 'QCs enabled' and remove from 'QCs disabled'
   set Quality Control Labels entry in the [Labels] section of the INI file to ON

IF command Label|Quality Control Labels Disabled (code IDM_LABEL_QCOFF)
   set quality control label printing flag to TRUE
   place check (tick) mark on 'QCs disabled' and remove from 'QCs enabled'
   set Quality Control Labels entry in the [Labels] section of the INI file to OFF

IF command Label|Reprint Last Label (code IDM_LABEL_REPRINT) - accelerator F5
   IF the last sample was NOT a QC AND sample label printing is enabled OR ...
      the last sample was a QC AND QC label printing is enabled
   THEN
      IF the data is still available in the global variables
         call BGALIB::PrintData to reprint the label
      ELSE (the data has been lost, corrupted or overwritten)
         call BGALIB::ErrorBox to message "unable to reprint: no data"
   ELSE (the last sample type had label printing disabled)
      call BGALIB::ErrorBox to message "unable to reprint: printing switched off"

IF command Label|Print Test Label (code IDM_LABEL_TESTPRINT)
   IF either sample printing or QC printing is enabled
      set up test data in global variables
      call BGALIB::PrintData to print the test label
      set the 'global data available for reprinting' flag to FALSE
   ELSE (no printing is enabled)
      call BGALIB::ErrorBox to message "unable to test print: printing switched off"

IF command is from the Copy menu (code IDM_SAVE_XXXX)
   call BGAMAIN::CmdFileSave to copy the single database indicated

IF command is from the View menu (code IDM_VIEW_XXXX)
   call BGAMAIN::CmdFileView to view the single database indicated

IF command Help|Contents (code IDM_HELP_INDEX)
   run application help with the contents page displayed
```

```
IF command Help|Keyboard (code IDM_HELP_KEYBOARD)
   run application help with the 'how to use the keyboard' page displayed

IF command Help|Commands (code IDM_HELP_COMMANDS)
   run application help with the 'commands' page displayed

IF command Help|Procedures (code IDM_HELP_PROCS)
   run application help with the 'procedures' page displayed

IF command Help|Help on Help (code IDM_HELP_HELP)
   run WINDOWS 'how to use help' with the 'contents' page displayed

IF command Help|About (code IDM_HELP_ABOUT)
   run modal dialog "AboutBox" with BGADLG::AboutProc to display version info

IF command 'display diagnosis detail' (code IDM_HELP_DETAIL) - accelerator F9
   IF paired or single cord results modeless dialog is currently being displayed
      send a command message to the modeless dialog box ...
         ... [send WM_COMMAND with code IDD_DETAIL -> BGADLG::DataProc]

IF command is unrecognised
   message "command not implemented" - IMPOSSIBLE!

reset the user time-out period [send WM_SETTIMER -> BGAMAIN::WndProc]
END PROCEDURE


PROCEDURE BGAMAIN::OnCommNotify

read the character received by the comms port
IF the character received is DC1, DC2 or DC3
   call BGADAT::ReadCommBlock to read and throw away one character (ETX)

IF the character received is DC1 (sampling about to commence)
   call BGALIB::CancelProgress to stop any previous progress bar
   call BGADAT::BeginSequence to start sampling sequence

IF the character received is DC2 (sample data available)
   call BGALIB::CancelProgress to stop previous sampling progress bar
   call BGADAT::GetData to receive a sample data block
   IF sample block received successfully
   THEN
      call BGALIB::WriteData to write the sample data to master sample database
      call BGAEXP::SampleCheck to check the sample for measurement errors
      call BGALIB::MoveState to change the program state with context ...
         CTX_SAMPLEOK if the sample has no measurement errors
         CTX_SAMPLEBAD if the sample has errors
   ELSE (a data transmission occured)
      call BGALIB::MoveState to change the program state with context ...
         CTX_GOTOMAINMENU to move straight back to the main menu

IF the character received is DC3 (calibration data available)
   call BGALIB::CancelProgress to stop previous calibration progress bar
   call BGADAT::GetData to receive a calibration data block
   IF calibration block received successfully
      call BGALIB::WriteData to write the calibration data to database
   ELSE (a data transmission occured)
      call BGALIB::MoveState to change the program state with context ...
         CTX_GOTOMAINMENU to move straight back to the main menu

IF the character received is CR (BGA238 reply to 'are you online?' request)
   call BGALIB::CancelProgress to stop any progress bar
   call BGADAT::ReadCommBlock to eat the transmission block
   IF there was an error in the received block
      call BGADAT::ResetCommState to clean up the comms port
   reset the calibration in process flag

IF the character received is anything else (something has gone wrong)
   call BGALIB::CancelProgress to stop any progress bar
   call BGADAT::ResetCommState to clean up the comms port
```

```
reset comms event to ensure that next received character generates a WM_COMMNOTIFY
set the 'reply received from BGA238' flag
END PROCEDURE


PROCEDURE BGAMAIN::OnCreate

CALL BGALIB::LoadResources to load the summary and detail text resources
create a button-face brush to colour status bar and detail button
load the full and limited menu-bars from the resource file

IF the auditing features are switched off
   delete the Maintain Audit Database option from the File menu
display the restricted menu

set message filter to catch the dialog box F1 help key in BGAMAIN::MessageFilter
install a sub-class to stop edit-controls highlighting text when they gain focus
call BGAMAIN::CreateChildWindows create the status bar windows and modeless dialogs

IF the child windows were created successfully
   call BGADAT::OpenConnection to open communications line to BGA238
END PROCEDURE


PROCEDURE BGAMAIN::OnCtlColor

IF the message is from a child static control (a status line window)
   alternately colour the text normal and system highlight colour
END PROCEDURE


PROCEDURE BGAMAIN::OnDestroy

remove the message filter for the F1 help key
remove the edit-control sub-class
call BGATXT::FreeResources to free memory allocated to summary and detail text
remove the button-face coloured brush
remove the full and restricted menu-bars
disable the WM_COMMNOTIFY comms message notification
close the communications port
request the end of the application (post WM_QUIT -> BGAMAIN::WinMain)
END PROCEDURE


PROCEDURE BGAMAIN::OnEnable

IF the main window gets enabled for any reason
   set the modal dialog box handle hModal global variable to NULL ...
      ... (because if so, there cannot possibly be a modal on screen)
END PROCEDURE


PROCEDURE BGAMAIN::OnSetFocus

IF a modal dialog is displayed (hModal is not NULL)
   set the input focus to the modal dialog box
ELSE (the is no modal dialog on display)
   set the input focus to the main window
END PROCEDURE


PROCEDURE BGAMAIN::OnSetTimer

IF the message is to switch ON the timer
   call BGALIB::StopTimer to stop the timer
   call BGALIB::StartTimer to re-start the timer with the required period
ELSE (the message is to switch OFF the timer)
   call BGALIB::StopTimer to stop the timer
```

```
END PROCEDURE


PROCEDURE BGAMAIN::OnSysColorChange

remove the (possibly) out of date button-face brush
re-create the button-face brush to colour status bar and detail button

END PROCEDURE


PROCEDURE BGAMAIN::OnSysCommand

IF the system command is NOT a menu selection
   prevent the system command from taking place

END PROCEDURE


PROCEDURE BGAMAIN::OnTimer

IF the timer has a minute conuter that is greater than zero
   decrement the timer minute conuter
   IF the minute counter is still greater than zero
      END PROCEDURE immediately

IF the timer is the calibration timer
   IF a reply has been received since the last calibration timer
      send a request for calibration data from the BGA238 ...
         ... to see if the BAG238 is still online
   ELSE (no reply has been received so the BGA238 is offline calibrating)
      call BGADAT::BeginSequence to display calibration progress bar

   reset the 'reply received from BGA238' flag

IF the timer is the user timer (to time out screens to return to the main menu)
   IF the main menu is already displayed
      switch off the timer
   ELSE (a non main menu screen has been left unattended)
      send a cancel command to the currently displayed modeless dialog ...
         ... [send WM_COMMAND (IDCANCEL) -> BGADLG::'current modeless dialog proc']

END PROCEDURE


PROCEDURE BGAMAIN::CmdFileSave

IF a single database is being copied
   IF the database selected to copy does not exist
      message "the file does not exist"
      END PROCEDURE immediately
   set the dialog box title to "<FILENAME> - Save As"
ELSE (all the databases are being copied)
   IF it's a Database Download command
      set the dialog title to "Database Download"
   ELSE (it's a Database Copy command)
      set the dialog title to "Database Copy"

call BGALIB::ProtectAllFiles to set all the databases to read-only
run the WINDOWS common dialog box to get a file name to save as ...
   ... with BGAMAIN::CheckFileName called when the user selects a file name

IF the user completed the dialog with OK
   IF all the databases are being copied
      FOR each the database file in turn
         make the destination file name
         call BGALIB::CopyFile to copy the database
         IF the copy succeeded
            go on to the next file
         ELSE IF the copy failed and the user selected CANCEL
            message "the download/copy failed"
            stop copying files and exit loop without deleting files
         ELSE the copy failed and the user selected RETRY
            re-process the same database file
```

```
      END FOR

      IF all the databases were copied successfully and Download was selected
         FOR each database file in turn
            call BGALIB::DeleteFile to 'roll-back' the database
            IF the delete failed
               call BGALIB::AbortApp to exit the application immediately (IMPOSS.?)
         END FOR

   ELSE (only a single database is being copied)
      call BGALIB::CopyFile to copy the database
      IF the copy failed and the user selected RETRY
         re-process the same database file

call BGALIB::ProtectAllFiles to set all the databases to read-write
IF the copy succeeded
   call BGALIB::ErrorBox to message "download completed successfully"

END PROCEDURE


PROCEDURE BGAMAIN::CheckFileName

IF a WM_FileOkString message is received from the File SaveAs dialog box
   get the current name of the selected directory from the dialog box variables
   IF the directory is read-only
      message "this directory is read-only"
      do not allow the dialog box to terminate
   ELSE IF a Download command was selected
      prompt "are you sure: OK or CANCEL?"
      IF reply is OK
         allow the dialog box to terminate
      ELSE (the reply was CANCEL)
         do not allow the dialog box to terminate
IF a WM_INITDIALOG is received
   set the initial input focus to the CANCEL button

END PROCEDURE


PROCEDURE BGAMAIN::CmdFileView

IF a database viewer task is already running
   re-awaken the previous database viewer task

IF the database selected to view does not exist
   message "the file does not exist"
   END PROCEDURE immediately

set the selected database file to read-only
get the current database viewer application from the [DB Viewer] section of INI file
run the database viewer application with selected database as command line argument
IF the database viewer application fails to run
   set the selected database file to read-write
   call BGALIB::ExecErrorBox to display reason for failure

END PROCEDURE


PROCEDURE BGALIB::MoveState

IF the move is due to a menu-selection (context CTX_MENU)
   look up the new value for State based on the Origin of the sample

IF the move is due to the patient screen being completed (CTX_PATIENTOK)
   look up the new value for State based on the current state
   IF a sample had been received out of sequence previously
      call BGAEXP::SampleCheck to re-check the previously received sample ...
         ... with the by now correctly specified Origin
      IF the sample has no measurement errors
         set new state instead to the second sample prompt or results screen
      ELSE (the sample has errors)
         set new state to the sampling error screen
```

```
      reset the out of sequence flag

IF the move is due to a sample having been received without errors (CTX_SAMPLEOK)
   IF the current state is the sampling error state
      look up the new value for State based on the state before the error
   ELSE (the current state is a normal state)
      look up the new value for State based on the current state

IF move is due to the user requesting retry after sampling error (CTX_SAMPLERETRY)
   set the new state to the state before the sampling error

IF the move is due to a sample having been received with errors (CTX_SAMPLEERROR)
   IF the current state is NOT already the sampling error state
      remember the current state as the state before the error occurred
      set the new state to the sampling error state

IF the move is due to the user requesting ignore second sample (CTX_SINGLESAMPLE)
   IF the current state is the paired cord second sample prompt
      change the origin of the sample to single cord sample
      set the new state to the single cord sample results
   ELSE (IGNORE second was selected when not on paired second prompt - IMPOSSIBLE!)
      set the new state to the main menu

IF the move is due to a move back to main menu instruction (CTX_GOTOMAINMENU)
   set the new state to the main menu

now perform some housekeeping tasks relating to the new state

IF new state is paired second sample prompt
   set the global sample index number to 1

IF new state is the main menu or other menu
   reset the sample received flag
   reset the sample out of sequence flag
   reset the origin to unknown

IF the new state is the sample out of sequence menu
   set the sample out of sequence flag
   reset the origin to unknown

IF new state is anything other than paired second sample prompt or sampling error
   set the global sample index number to 0

IF the new state is different to the old state
   hide the currently displayed modeless dialog box
   set the hCurrentWnd global variable to the handle of the new dialog box
   call BGALIB::SetMainWindowText to set the title caption of the main window
   show the modeless dialog box for the new state
   set the input focus to the newly displayed modeless dialog box

   IF the new state is sample out of sequence menu OR sampling error
      call BGALIB::WarningBeep to sound the warning alarm

reset the user timer to restart time-out [send WM_SETTIMER -> BGAMAIN::WndProc]
END PROCEDURE


PROCEDURE BGALIB::AbortApp

EXIT the application immediately with a brief death message (IMPOSSIBLE?)
END PROCEDURE


PROCEDURE BAGLIB::ErrorBox

load a string from the resource file corresponding to the current error condition
display a MessageBox with the error message string
END PROCEDURE
```

```
PROCEDURE BGALIB::ExecErrorBox

load a string with a message correspongding to the last WinExec error condition
display a MessageBox with the error message string

END PROCEDURE


PROCEDURE BGALIB::CancelProgress

IF there is currently a progress bar on display
   force it to close [send WM_CLOSE -> BGADLG::'current modeless dialog proc']

   IF the progress bar is closing because it went on too long
      reset both sequence-in-progress flags
      set the status line 'status' box to "awaiting user input"
   ELSE (the progress bar is closing because the data ready signal was received)
      average the dialog elapsed time and expected time ...
         ... to maintain a rolling average of progress bar expected time

   IF the timers are to be switched back on again ...
      ... (a data transmission is not expected to follow)
      restart the user and calibration timers ...
         ... [send WM_SETTIMER (user timer) -> BGAMAIN::WndProc]
         ... [send WM_SETTIMER (calibration timer) -> BGAMAIN::WndProc]

END PROCEDURE


PROCEDURE BGALIB::CancelResults

IF there is currently any one of the results screens on display
   send it a cancel message [send WM_COMMAND (IDCANCEL) -> BGADLG::DataProc]

END PROCEDURE


PROCEDURE BGALIB::CancelModal

IF there is currently a modal dialog box on display
   send it a cancel message [send WM_COMMAND (IDCANCEL) -> 'dialog procedure']

END PROCEDURE


PROCEDURE BGALIB::ReadProfile

read the software configuration flags from the [Config] section of the INI file
read the minimum data field widths from the [Minimums] section of the INI file
read the label printing flags from the [Labels] section of the INI file

END PROCEDURE


PROCEDURE BGALIB::StartTimer

IF the requested timer period is zero
   END PROCEDURE immediately

store the number of requested minutes into global timer minutes count
IF the internal timer is already running with the correct period
   END PROCEDURE immediately

set the internal timer period to requested period or ...
   ... one minute for requested timer periods of greater the one minute
IF the internal timer was NOT successfully set
   prompt "too many timers: ABORT, RETRY or IGNORE?"
   IF reply is ABORT
      call BGALIB::AbortApp to exit the application immediately (IMPOSSIBLE?)
   IF reply is RETRY
      try to set the timer again
   IF reply is IGNORE
      leave the timer unset

END PROCEDURE
```

```
PROCEDURE BGALIB::StopTimer

IF the specified timer is actually set
   attempt to stop the specified timer
   IF the running timer refused to stop
      call BGALIB::AbortApp to exit the application immediately (IMPOSSIBLE!)
   reset the global timer identification and timer minute count variables
END PROCEDURE


PROCEDURE BGALIB::SetProfileData

FOR each of the vessel groups on the results screen
   set the pH units text to "pH" or "H+"
   set the gas units text to "(kPa)" or "(mmHg)"
   set the HCO3 units text to "(ACT)" or "(STD)"
END FOR

END PROCEDURE


PROCEDURE BGALIB::SetSampleData

FOR each of the vessel groups on the results screen
   set the vessel label to "Artery", "Vein" or "Either" as assigned by expert system
   call BGALIB::SampleStr to set the sample number
   call BGALIB::DoubleStr to set floating point results (pH, pCO2, pO2, HCO3, BD)
   call BGALIB::DateStr to set the sampling date
   call BGALIB::TimeStr to set the sampling time
END FOR

IF the current screen is Other Data or Quality Control Data
   set the BE units text to "(VT)" or "(VV)"
END PROCEDURE


PROCEDURE BGALIB::DoubleStr

IF the data field has its measurement error flag set
   set the display field to "******"
ELSE (the data field has no errors)
   set the display field to the character representation of the floating point ...
         ... number displaying the requested number of decimal places
END PROCEDURE


PROCEDURE BGALIB::SampleStr

set the display field to the character representation of the long integer number
END PROCEDURE


PROCEDURE BGALIB::DateStr

set the display field to the character representation of the date fields
END PROCEDURE


PROCEDURE BGALIB::TimeStr

set the display field to the character representation of the time fields
END PROCEDURE


PROCEDURE BGALIB::GetDateTime

get and store the current computer date and time
END PROCEDURE
```

```
PROCEDURE BGALIB::ProtectAllFiles

IF the database files are to be made read-only
   FOR each database file in turn (including the database directory)
      set the read-only and hidden attributes on the file
   set the BLOODGAS.INI file to read-only
   set the PROGMAN.INI file to read-only
ELSE (the database files are to be made read-write)
   FOR each database file in turn
      get the current file attributes
      IF the current file is the database directory
         reset all attributes except the directory and volume ID attributes
      ELSE (the current file is an ordinary database file)
         reset just the read-only and hidden attributes
   set the BLOODGAS.INI file to read-write
   set the PROGMAN.INI file to read-write

END PROCEDURE


PROCEDURE BGALIB::WriteData

(see application user manual for detailed layout of each database)

IF the database files are currently supposed to be read-only
   call BGALIB::ErrorBox to message "cannot write: files are protected"
   END PROCEDURE immediately

IF the database to write to is out-of-range
   call BGALIB::AbortApp to exit the application immediately

open the destination database file
IF the file cannot be opened
   call BGALIB::AbortApp to exit the application immediately

get the record number of the database from the [DB Record#] section of the INI file
increment the record number
IF the database is the calibration
   store the record number in the calibration sample number field
ELSE the database is a sample database
   store the record number in the computer DateTime global number field

start building the record by writing the record number into the buffer
write the incremented record number back to the [DB Record#] section of the INI file
get the current computer date and time into the DateTime global
add the current date and time to record buffer

IF the record is for database ...
   ... Fetal, Neonatal, Adult, Quality Control or Other
   add the Patient information to the record buffer
   add the Sample0 information to the record buffer

IF the record is for database Master Sample or Unknown
   add the Sample0 information to the record buffer

IF the record is for database Paired Cord or Single Cord
   add the Patient information to the record buffer
   add the Sample0 information to the record buffer
   IF the record is for database Paired Cord
      add the Sample1 information to the record buffer
      add the Patient diagnosis information to the record buffer
   ELSE (the record is for database Single Cord)
      add the Patient diagnosis information to the record buffer

IF the record is for database Calibration
   add the Calibration information to the record buffer

add a CR+LF pair to the end of the record
move the file pointer to the end of the open file
store the position moved to in the open file

FOR up to DATA_RETRIES (3) times
```

```
   IF the constructed record is of the correct size AND ...
      ... the position in the open file is correct AND ...
      ... the number of bytes written to the file is correct AND ...
      ... the file is then successfully closed
   THEN
      END PROCEDURE immediately

the file write has failed, call BGALIB::AbortApp to end the application immediately
END PROCEDURE


PROCEDURE BGALIB::CopyFile

IF the file to be copied does not exist
   END PROCEDURE immediately with success status

IF the destination file already exists
   message "destination file already exists"
   END PROCEDURE immediately with failure status

open the source file
IF the source file could not be opened
   message "source file open failed"
   END PROCEDURE immediately with failure status

create the destination file
IF the destination file could not be created
   close the source file
   message "destination file create failed"
   END PROCEDURE immediately with failure status

allocate a global memory buffer to hold files transfer bytes
IF the memory allocation failed
   close the source and destination files
   call BGALIB::ErrorBox to message "memory allocation failed"
   END PROCEDURE immediately with failure status

REPEAT
   read a block of bytes from the source file into the transfer buffer
   IF the source read failed
      close the source and destination files
      message "could not read source file"
      END PROCEDURE immediately with failure status

   IF the TAB-separated to Comma-separated file translation flag is on
      FOR each byte in the transfer buffer
         IF the byte is a TAB
            change it to a comma

   write the transfer buffer to the destination file
   IF the destination write failed
      close the source and destination files
      call BGALIB::GetDriveFree to find the free space on destination drive
      IF the destination drive is a floppy disk AND the disk is full
         delete the incomplete destination file
         call BGALIB::ErrorBox to prompt "change disks: OK or CANCEL?"
         END the procedure immediately with the reply code
      ELSE (some sort of more serious error has occurred)
         message "could not write destination file"
         END PROCEDURE immediately with failure status

UNTIL all the source file is copied

call BGALIB::EGlobalFreePtr to free memory for the transfer buffer
set success flag if the source file is the same size as the destination file AND ...
   ... both files close successfully

END PROCEDURE
```

```
PROCEDURE BGALIB::DeleteFile

make a destination file name with the "BAK" extension

IF the backup already exists
   set the backup file attribute to read-write
   IF the set attribute worked
      delete the old backup file
      IF the delete failed
         END PROCEDURE immediately with failed status

IF the source file exists
   rename it to the backup file name
   IF the rename failed
      END PROCEDURE immediately with failed status

set success status

END PROCEDURE


PROCEDURE BAGLIB::GetDriveFree

run DOS system call to report destination drive statistics
calculate number of bytes free

END PROCEDURE


PROCEDURE BGALIB::SetMainWindowText

set the main window caption text to the ...
   ... concatenation of the application name, origin title and state title

END PROCEDURE


PROCEDURE BGALIB::SetWindowString

set the specified window text to the specified string resource ...
   ... with the specified number of leading spaces

END PROCEDURE


PROCEDURE BGALIB::PrintData

get the number of blank labels to print from the [Labels] section of the INI file
call BGALIB::GetPrinterDC to obtain a device context for the printer

IF the device context was successfully obtained
   call BGALIB::GetDateTime to get the current computer date and time
   get the font information of the label printer
   call BGATXT::LoadInterpretation to load the current ...
      ... expert system summary interpretation(s) into memory

   FOR the number of label copies to be printed
      start a new document
      start a new page
      call BGALIB::SummaryTextLines to initialise the interpretation summary
variables
      open the label template file

      FOR each line of the label template file
         call BGALIB::TranslateFields to substitute actual values into label fields
         print the line on the printer

      close the label template file
      end the page
      end the document

   FOR the number of blank labels to be printed
      start a new document
      start a new page
      end the page
```

```
      end the document

   call BGATXT::FreeInterpretation to free the memory allocated to ES summary
   delete the printer device context
ELSE (a device context could not be obtained for the printer)
   call BGALIB::ErrorBox to message "could not print"

END PROCEDURE


PROCEDURE BGALIB::GetPrinterDC

obtain a device context for the currently selected printer
(for example see "Microsoft Windows Guide to Programming pp. 271-272)

END PROCEDURE


PROCEDURE BGALIB::SummaryTextLine

IF initialising the variables
   set the internal current text position to the start of text
   END PROCEDURE immediately

REPEAT for each character in the text in turn
   IF the character is a carriage return (CR)
      END REPEAT loop immediately
   IF the character is a space
      store the position of the space character
UNTIL the width of the label is reached OR the end of summary text is reached

IF the full width of the label was reached AND a space character was found
   break the line at the space character
ELSE
   break the line at the full width of the label

store the position at which the break was made to be the start of the next line
return the position of the start of the current line

END PROCEDURE


PROCEDURE BGALIB::TranslateFields

initialise the label output line to blank

REPEAT for each character in the label template line in turn

   IF the character is a substitute-field character (%)
      IF the next character is a substitute-field character (%) also
         add a single % character to the output line
      ELSE (this is a true substitute-field sequence)
         substitute the actual current variable value into the output line ...
            ... with the width and format as specified by the substitute-field ...
            ... in the template line
   ELSE IF the character is a special-code character (\)
         substitute a special-code character into the output line ...
            ... as specified by the special-code in the template line
   ELSE (the character is an ordinary character)
      copy the character from the template line into the output line

UNTIL the end of the template line is reached

END PROCEDURE


PROCEDURE BGALIB::WarningBeep

get the current sound level from the [Config] section of the INI file
IF the warning sound requested is of a higher level than the current sound level
   play a warning sound of requested type through the computer speaker

END PROCEDURE
```

```
PROCEDURE BGALIB::EGlobalAllocPtr

allocate a block of global memory of the requested size from the WINDOWS pool
IF the allocation was unsuccessful (IMPOSSIBLE?)
   call BGALIB::AbortApp with message "unable to allocate memory"

END PROCEDURE


PROCEDURE BGALIB::EGlobalReAllocPtr

re-allocate a block of global memory of the requested size from the WINDOWS pool
IF the re-allocation was unsuccessful (IMPOSSIBLE?)
   call BGALIB::AbortApp with message "unable to re-allocate memory"

END PROCEDURE


PROCEDURE BGALIB::EGlobalFreePtr

free a block of allocated global memory and return it to the WINDOWS pool
IF the free was unsuccessful (IMPOSSIBLE!)
   call BGALIB::AbortApp with message "unable to free memory"

END PROCEDURE


PROCEDURE BGALIB::lstrfind

look for the last occurrence of the specified character in the specified string
IF the character was found
   return the string following the character
ELSE (the character was not found)
   return the originally specified string

END PROCEDURE


PROCEDURE BGADAT::OpenConnection

display the hour-glass cursor while the connection is being attempted
open a new blank boot-diagnostics file
write the current date and time to the [Boot] section of the INI file

FOR each serial port COM1 to COM4
   write "attempting to open port" to log file
   IF the port is successfully opened AND set to the default comms settings
      call BGADAT::GetMachineResponse to attempt to get a response from the BGA238
      IF a successful response was obtained
         write the comms parameters to the [Comms] section of the INI file
         write the comms parameters to the log file
         set the success flag
         END FOR loop immediately
   ELSE (the port was not opened)
      write "failed to open port" to log file
   close the comms port to clear any errors
END FOR

reset the original cursor count and cursor display
IF the connection failed
   write "all connection failed" to log file
close the log file
write "SUCCESS" or "FAILURE" to the Flag entry of the [Boot] section of the INI file

END PROCEDURE


PROCEDURE BGADAT::GetMachineResponse

FOR each baud rate 7200,2400,1200,300 in turn
   set the port to the new baud rate
   write "attempting connection at baud rate" to log file
   clear any comms error from the port
   transmit an ID request sequence
   call BGADAT::ReadCommBlock to read an ID block from the comms port
```

```
   IF an ID block was read successfully
      write "obtained response" to log file
      call BGADAT::CheckResponse to check the response is as expected

      IF the response was as expected
         write "response obtained" to log file
         set the status line display to show machine ID and baud rate
         enable WM_COMMNOTIFY messages to be sent to BGAMAIN::WndProc
         IF WM_COMMNITIFY notification was successful
            END PROCEDURE immediately with success status
         ELSE (WM_COMMNOTIFY notification failed)
            write "notification failed" to log file
      ELSE (the response was NOT as expected)
         write "response unknown" to log file
   ELSE (no ID block could be read)
      write "no response" to log file

END FOR

END PROCEDURE


PROCEDURE BGADAT::CheckResponse

IF the first three characters are digits AND are followed by CR-LF
   set success status

END PROCEDURE


PROCEDURE BGADAT::ReadCommBlock

call BGALIB::StartTimer to start a comms time-out timer
WHILE the number of characters read is less than the number expected
   read any characters waiting at the comms port and add them to buffer
   call BGADAT::CheckMessage to check the message queue for a WM_TIMER message

   IF an ETX character was received OR ...
      ... no characters were received AND a WM_TIMER message has arrived
      END WHILE loop immediately
   ELSE IF some characters were read from the comms port
      call BGALIB::StopTimer to stop the comms time-out timer
      call BGALIB::StartTimer to restart the comms time-out timer
END WHILE

call BGALIB::StopTimer to stop the comms time-out timer
IF the number of characters read was as expected AND the last character is an ETX
   set success status

END PROCEDURE


PROCEDURE BGADAT::CheckMessage

use PeekMessage to check the message queue for the specified message
IF an hour-glass is being displayed
   reset the cursor to the hour-glass in case it was changed by PeekMessage

END PROCEDURE


PROCEDURE BGADAT::ResetCommState

flush the port of any waiting characters
clear the port of any error status

END PROCEDURE


PROCEDURE BGADAT::BeginSequence

call BGALIB::CancelModal to cancel any modal dialog box on display
call BGALIB::CancelResults to cancel the results screen
turn off the main window timers ...
```

```
   ... [send WM_SETTIMER (UserTimer) -> BGAMAIN::WndProc]
   ... [send WM_SETTIMER (CalTimer) -> BGAMAIN::WndProc]
set the sampling sequence in progress flag (fSampling or fCalibrating)
load the user status line with "sequence in progress" message
run dialog "Progress" with BGADLG::ProgressProc to display progress bar

END PROCEDURE


PROCEDURE BGADAT::GetData

call BGALIB::CancelModal to cancel any modal dialog box on display
call BGALIB::CancelResults to cancel the results screen
turn off the main window timers ...
   ... [send WM_SETTIMER (UserTimer) -> BGAMAIN::WndProc]
   ... [send WM_SETTIMER (CalTimer) -> BGAMAIN::WndProc]
load the user status line with "waiting for data" message
clear the master error list

FOR up to DATA_RETIES (3) times
   clear the current error list
   transmit a data request sequence to the BGA238
   call BGADAT::ReadCommBlock to read the data block from the BGA238
   IF the data read was successful
      IF this is a sampling sequence
         call BGADAT::MakeDateTime to extract date and time info from the data
buffer
         call BGADAT::MakeSample to extract sample info from the data buffer
      ELSE (this is a calibration sequence)
         call BGADAT::MakeDateTime to extract date and time info from data buffer
         call BGADAT::MakeCalibration to extract calibration info from data buffer
      IF no errors were found during data extraction
         END FOR loop immediately
   ELSE
      add transmission error to current error list
   END IF

   add the current error list to master error list
END FOR

restore the user status line to "waiting for user input"
IF the current error list is not empty
   run dialog "DataError" with BGADLG::MessageProc to display list of all errors
turn the user timer back on [send WM_SETTIMER (UserTimer) -> BGAMAIN::WndProc]
IF a sampling prompt screen is NOT currently displayed
   turn calibration timer back on [send WM_SETTIMER (CalTimer) -> BGAMAIN::WndProc]

END PROCEDURE


PROCEDURE BGADAT::MakeDateTime

IF the data buffer contains "HH:MM   MMM DD/YY" at positions 9..25
   set DATA_DTE_ERROR as the BGA238 date has not been set
call BGADAT::ReadInt with hour template to read data from the buffer
call BGADAT::ReadInt with minute template to read data from the buffer
call BGADAT::ReadInt with day template to read data from the buffer
call BGADAT::ReadStr with month template to read data from the buffer
call BGADAT::ReadInt with year template to read data from the buffer
call BGADAT::ReadInt with sample number template to read data from the buffer

END PROCEDURE


PROCEDURE BGADAT::MakeSample

IF pH units are being used
   call BGADAT::ReadDbl with pH template to read data from the buffer
ELSE (H+ units are being used)
   call BGADAT::ReadDbl with H+ template to read data from the buffer

IF kPa gas units are being used
   call BGADAT::ReadDbl with pCO2 kPa template to read data from the buffer
```

```
   call BGADAT::ReadDbl with pO2 kPa template to read data from the buffer
ELSE (mmHg gas units are being used)
   call BGADAT::ReadDbl with pCO2 mmHg template to read data from the buffer
   call BGADAT::ReadDbl with pO2 mmHg template to read data from the buffer

IF HCO3 ACT is being used
   call BGADAT::ReadDbl with HCO3 ACT template to read data from the buffer
ELSE (HCO3 STD is being used)
   call BGADAT::ReadDbl with HCO3 STD template to read data from the buffer

IF BE VT is being used
   call BGADAT::ReadDbl with BE VT template to read data from the buffer
ELSE (BE VV is being used)
   call BGADAT::ReadDbl with BE VV template to read data from the buffer

calculate the pH, pCO2(mmHg) and pO2(mmHg) for unit-dependant expert system
END PROCEDURE


PROCEDURE BGADAT::MakeCalibration

IF pH units are being used
   call BGADAT::ReadStr with pH template to read CAL/SLP data from the buffer
   call BGADAT::ReadDbl with pH template to read old data from the buffer
   call BGADAT::ReadDbl with pH template to read new data from the buffer
ELSE (H+ units are being used)
   call BGADAT::ReadStr with H+ template to read CAL/SLP data from the buffer
   call BGADAT::ReadDbl with H+ template to read old data from the buffer
   call BGADAT::ReadDbl with H+ template to read new data from the buffer

IF kPa gas units are being used
   call BGADAT::ReadDbl with pCO2 kPa template to read old data from the buffer
   call BGADAT::ReadDbl with pCO2 kPa template to read new data from the buffer
   call BGADAT::ReadDbl with pO2 kPa template to read old data from the buffer
   call BGADAT::ReadDbl with pO2 kPa template to read new data from the buffer
   IF BGA238 version 2 software is being used
      call BGADAT::ReadDbl with BP kPa V2 template to read data from the buffer
   ELSE (BGA238 version 1 software is being used)
      call BGADAT::ReadDbl with BP kPa V1 template to read data from the buffer
ELSE (mmHg gas units are being used)
   call BGADAT::ReadDbl with pCO2 mmHg template to read old data from the buffer
   call BGADAT::ReadDbl with pCO2 mmHg template to read new data from the buffer
   call BGADAT::ReadDbl with pO2 mmHg template to read old data from the buffer
   call BGADAT::ReadDbl with pO2 mmHg template to read new data from the buffer
   IF BGA238 version 2 software is being used
      call BGADAT::ReadDbl with BP mmHg V2 template to read data from the buffer
   ELSE (BGA238 version 1 software is being used)
      call BGADAT::ReadDbl with BP mmHg V1 template to read data from the buffer
END PROCEDURE


PROCEDURE BGADAT::ReadInt

call BGADAT::MatchField with template supplied to match data characters in buffer
IF any data characters did NOT match those expected
   set DATA_CHR_ERROR to indicate unexpected character in data sequence
   END PROCEDURE immediately

convert the text characters in data buffer to integer variable
IF the integer lies outside the minimum and maximum data limits
   set DATA_VAL_ERROR to indicate data out of range in data sequence
   END PROCEDURE immediately
END PROCEDURE


PROCEDURE BGADAT::ReadLng

identical to BGADAT::ReadInt above with long variables instead of integer variables
END PROCEDURE
```

```
PROCEDURE BGADAT::ReadDbl

identical to BGADAT::ReadInt above with double variables instead of integer
variables

END PROCEDURE


PROCEDURE BGADAT::ReadStr

call BGADAT::MatchField with template supplied to match data characters in buffer
IF any data characters did NOT match those expected
   set DATA_CHR_ERROR to indicate unexpected character in data sequence
   END PROCEDURE immediately

copy the string characters into the destination buffer
FOR each string in the set of allowable string values array
   IF the string found matches an allowable string
      END PROCEDURE immediately with success

set DATA_VAL_ERROR to indicate data out of range in data sequence

END PROCEDURE


PROCEDURE BGADAT::MatchField

FOR each character in the supplied template
   IF the template character is 'dont-care' (@)
      move immediately on to the next character

   IF the template character is 'no-endpoint' (*)
      IF the buffer character is '*' or ' '
         set the data-measurement-error flag to the buffer character
      ELSE (the buffer character is illegal)
         set the error-position global variable
         END PROCEDURE immediately with failure status

   IF the template character is 'leading-minus' (-)
      IF the buffer character is '-' or ' '
         IF the start-of-data position in not set
            set the start-of-data position to the current position
      ELSE (the buffer character is illegal)
         set the error-position global variable
         END PROCEDURE immediately with failure status

   IF the template character is 'numeric-field' (#)
      IF the start-of-data position in not set
         set the start-of-data position to the current position
      continue processing as if 'numeric-ignore'

   IF the template character is 'numeric-ignore' (=)
      IF the buffer character is a digit
         reset the leading spaces allowed flag
      ELSE IF the buffer character is a dash (-) or a star (*)
         IF the template character is 'numeric-field' (#)
            set the data-measurement-error flag to the buffer character
      ELSE IF the buffer character is a space ( ) AND leading spaces are allowed
         do nothing (go on to the next character)
      ELSE (the buffer character is illegal)
         set the error-position global variable
         END PROCEDURE immediately with failure status

   IF the template character is 'alpha-field' ($)
      IF the start-of-data position in not set
         set the start-of-data position to the current position
      continue processing as if 'alpha-ignore'

   IF the template character is 'alpha-ignore' (&)
      IF the buffer character is a letter
         do nothing (go on to the next character)
      ELSE (the buffer character is illegal)
         set the error-position global variable
         END PROCEDURE immediately with failure status
```

```
   IF the template character is anything else
      IF the buffer character matches the template character
         set the leading spaces allowed flag
      ELSE (the buffer character is illegal)
         set the error-position global variable
         END PROCEDURE immediately with failure status
END FOR
END PROCEDURE


PROCEDURE BGADLG::MenuProc

IF the message is WM_INITDIALOG
   store the default focus window handle for future use (the top button normally)
IF the message is WM_SETFOCUS
   get the stored default window handle
   set the focus to the stored default
IF the message is WM_COMMAND
   IF the command is NOT CANCEL
      set the origin of the sample to the command index number
      call BGALIB::MoveState with context CTX_MENU to move to the next screen
   ELSE (the command is CANCEL)
      call BGALIB::MoveState with context CTX_GOTOMAINMENU to return to main menu
END PROCEDURE


PROCEDURE BGADLG::PatientProc

IF the message is WM_INITDIALOG
   IF the dialog is for a patient information screen
      limit the maximum number of characters in each patient field to ...
         .. the maximum field width of the database files
   ELSE (the dialog is for a quality control batch information screen)
      limit the maximum number of characters in the batch number field to ...
         .. the maximum field width of the database files

   IF auditing features are enabled
      limit the maximum number of characters in the audit ID field to 4
   ELSE (auditing features are disabled)
      set the patient audit ID global variable to "*****
      set the patient audit name global variable to "UNKNOWN"
      hide the auditing fields on the dialog box

IF the message is WM_SHOWWINDOW
   IF the window is being shown (as opposed to being hidden)
      initialise the OK button to enabled and highlighted ...
         ... [send WM_NEXTDLGCTL -> BGADLG::PatientProc]
      set all the internal dialog edit-fields to blank ...
         ... IDD_AUDITID, IDD_AUDITNAME ...
         ... IDD_PATIENTSURNAME, IDD_PATIENTNAME, IDD_PATIENTID
      for each field [send WM_COMMAND (IDD_XXXXXXX, EN_CHANGE) ->
BGADLG::PatientProc]

IF the message is WM_COMMAND
   IF the control is IDD_AUDITID
      IF auditing is switched on
         IF command is EN_SETFOCUS
            disable the OK button
         ELSE IF command is EN_CHANGE
            set the audit name field to blank

   IF the control is IDD_PATIENTSURNAME, NAME or ID
      IF command is EN_SETFOCUS AND auditing is switched on
         get the audit id from the dialog field into temporary variable
         get the matching audit name from [Audit IDs] section of the INI file ...
            ... into temporary variable (OR "UNKNOWN" if not found)
         set the dialog audit name field to the temporary audit name
         IF the audit name is "UNKNOWN"
            call BGALIB::ErrorBox to message "audit ID not known"
            set the input focus back to the audit ID field
```

```
         ELSE (the audit name is valid)
            enable the OK button
      ELSE IF the command is EN_CHANGE
         IF the dialog is for a patient information screen
            IF patient surname, name and ID fields are all greater than the minimums
               enable the OK button
            ELSE
               disable the OK button
         ELSE (the dialog is for a quality control batch information screen)
            IF QC batch field is greater than the minimum
               enable the OK button
            ELSE
               disable the OK button

   IF the command is OK
      IF auditing is switched on
         get the audit ID and name from dialog fields into global variables
      get the patient surname, name and ID from dialog fields into global variables
      call BGALIB::MoveState with context CTX_PATIENTOK to change to the next screen
      disallow the reprinting of labels

   IF the command is CANCEL
      call BGALIB::MoveState with context CTX_GOTOMAINMENU to return to main menu
END PROCEDURE


PROCEDURE BGADLG::PromptProc

IF the message is WM_INITDIALOG
   call BGADLG::DisplayIconText to display the icon and text in the dialog fields
   IF the prompt screen is NOT the paired sceond prompt
      disable and hide the INGORE second sample button

IF the message is WM_SETFOCUS
   set the input focus to the IGNORE second sample prompt

IF the message is WM_SHOWWINDOW
   IF the window is being shown
      stop the calibration timer [send WM_SETTIMER (FALSE) -> BGAMAIN::WndProc]
   ELSE (the window is being hidden)
      re-start the calibration timer [send WM_SETTIMER (TRUE) -> BGAMAIN::WndProc]

IF the message is WM_COMMAND
   IF the command is IDIGNORE
      call BGALIB::MoveState with context CTX_SINGLESAMPLE to move to single results
   IF the command is IDCANCEL
      call BGALIB::MoveState with context CTX_GOTOMAINMENU to return to main menu
END PROCEDURE


PROCEDURE BGADLG::DetailProc

perform the default WINDOWS processing for the message and store the result
IF the message is WM_PAINT or BM_SETSTATE
   IF the message is BM_SETSTATE
      set the button state according to the parameter
   get the button text
   re-paint the button text in the system highlight color

return the original message result
END PROCEDURE


PROCEDURE BGADLG::DataProc

IF the message is WM_INITDIALOG
   IF the dialog has a detail button (paired or single cord results)
      sub-class the detail button to use BGADLG::DetailProc

   IF the dialog is for a patient information screen
```

```
      limit the maximum number of characters in each patient field to ...
         ... the maximum field width of the database files
   ELSE (the dialog is for a quality control batch information screen)
      limit the maximum number of characters in the batch number field to ...
         ... the maximum field width of the database files

   call BGALIB::SetProfileData to set the profile data (pH/H+, kPa/mmHg etc)

IF the message is WM_SHOWWINDOW
   IF the window is being shown
      enable the OK button
      IF the dialog is for a quality control batch information screen
         enable the CANCEL button
      set the dialog fields to the values stored in the patient global variables
      IF the origin of the sample is paired or single cord
         call BGAEXP::Diagnosis to run the expert system diagnosis module
      ELSE
         set the patient diagnosis global variable to zero

IF the message is WM_SETFOCUS
   set the input focus and highlight to the OK button

IF the message is WM_COMMAND
   IF the control is IDD_PATIENTSURNAME, NAME or ID
      IF the command is EN_CHANGE
         IF the dialog is for a patient information screen
            IF patient surname, name and ID fields are all greater than the minimums
               enable the OK button
            ELSE
               disable the OK button
         ELSE (the dialog is for a quality control batch information screen)
            IF QC batch field is greater than the minimum
               enable the OK button
            ELSE
               disable the OK button

   IF the command is IDD_DETAIL
      run the modal dialog "Detail" with procedure BGADLG::BlankProc

   IF the command is IDOK
      get the dialog fields into the patient global variables
      continue with processing as for the IDCANCEL command

   IF the command is IDCANCEL
      IF the dialog is for a quality control batch information screen
         set the patient name global variable to "Quality Control"
         IF the command is IDOK
            set the patient surname global variable to "PASSED"
         ELSE (the command is IDCANCEL)
            set the patient surname global variable to "FAILED"

      write the results to the appropriate database file
      IF the origin is a sample AND sample printing is turned on OR ...
         ... the origin is a quality control AND QC printing is turned on
         call BGALIB::PrintData to print the results out on the label

      allow label reprinting
      call BGALIB::MoveState with context CTX_GOTOMAINMENU to return to main menu
END PROCEDURE


PROCEDURE BGADLG::ProgressProc

IF the message is WM_INITDIALOG
   disable the main window
   call BGALIB::SetWindowString to set the dialog caption
   get the expected duration of the sequence in seconds and display them on the
dialog
   call BGALIB::StartTimer to set a one second interval timer

IF the message is WM_CTLCOLOR
```

```
   IF the control is the expected seconds field
      colour it with the system highlight colour

IF the message is WM_TIMER
   increment the number of seconds elapsed
   IF the number of seconds elapsed is less than 5 times the expected duration
      IF this is a calibration-in-progress
         send the BGA238 a calibration data request to see if it's back online
      call BGADLG::PaintProgress to colour in a progress bar
      IF the number of seconds elapsed is greater than the expected duration
         display the actual number of seconds in the dialog box
   ELSE (acutal elapsed duration is greater than 5 times expected duration)
      post a message to ask the main window to close down this progress bar
      (post WM_CLOSEMODELESS -> BGAMAIN::WndProc)

IF the message is WM_CLOSE
   call BGALIB::StopTimer to stop the one second interval timer
   re-enable the main window
   destroy this dialog box

END PROCEDURE


PROCEDURE BGADLG::PaintProgress

IF the actual number of seconds is greater than the expected number of seconds
   set the actual number of seconds to the maximum

paint a coloured-bar in proportion to (actual / maximum) in the progress bar

END PROCEDURE


PROCEDURE BGADLG::DisplayProc

IF the message is WM_INITDIALOG
   disable the main window
   call BGADLG::DisplayIconText to display the icon and text in the dialog fields

IF the message is WM_CLOSE
   re-enable the main window
   destroy this dialog box

END PROCEDURE


PROCEDURE BGADLG::SampleErrorProc

IF the message is WM_PAINT
   set the text dialog field
   set the icon dialog field

IF the message is WM_SETFOCUS
   set the input focus to the RETRY button

IF the message is WM_SHOWWINDOW
   do nothing (necessary for the WM_SETFOCUS processing to take place)

IF the message is WM_COMMAND
   IF the command is IDRETRY
      call BGALIB::MoveState with CTX_SAMPLERETRY to move back to previous state
   IF the command is IDIGNORE
      call BGALIB::MoveState with CTX_SAMPLEOK to move forward to next state
   IF the command is IDABORT or IDCANCEL
      call BGALIB::MoveState with CTX_GOTOMAINMENU to move back to main menu

END PROCEDURE


PROCEDURE BGADLG::MessageProc

IF the message is WM_INITDIALOG
   call BGADLG::DisplayIconText to display the icon and text in the dialog fields
   IF this is the "DataError" dialog
```

```
      call BGADLG::DisplayChecks to display error tick boxes
   IF a warning beep is required on displaying the dialog
      call BGALIB::WarningBeep to sound the alarm

IF the message is WM_COMMAND
   IF the command is IDOK or IDCANCEL
      destroy this dialog box
END PROCEDURE


PROCEDURE BGADLG::DisplayIconText

IF the caption of the destination dialog box is blank
   set the dialog box caption to match the icon type to be displayed
set the text dialog field
set the icon dialog field
END PROCEDURE


PROCEDURE BGADLG::DisplayChecks

FOR the number of types of error paooible
   IF the error list global variable specifies the error has occurred
      display a check-mark (tick) in the dialog field
IF the last error position global variable is not zero
   display the last error position in the dialog field
END PROCEDURE


PROCEDURE BGADLG::BlankProc

IF the message is WM_INITDIALOG
   call BGATXT::LoadInterpretation to load the current diagnosis detail text
   set the dialog caption to the first line of the diagnosis detail
   call BGATXT::JustifyTextPage to calculate the text justification parameters
   call BGATXT::InitScrollBar to set the vertical scroll bar for the window

IF the message is WM_PAINT
   call BGATXT::DisplayTextPage to display the text in the dialog field

IF the message is WM_VSCROLL
   call BGATXT::ScrollTextPage to scroll the text in the dialog field

IF the message is WM_COMMAND
   IF the command is IDOK or IDCANCEL
      destroy this dialog box
      call BGATXT::FreeInterpretation to free the allocated memory
END PROCEDURE


PROCEDURE BGADLG::PasswordProc

IF the message is WM_INITDIALOG
   store the get/set password flag for future use

IF the message is WM_COMMAND
   IF the command is IDOK
      get the password from the dialog field into a temporary variable
      IF the password is being set
         call BGADLG::Encrypt to encrypt the new password
         set the new encrypted password in the [Password] section of the INI file
      ELSE (the password isbeing got)
         get current encrypted password from the [Password] section of the INI file
         call BGADLG::Encrypt to encrypt the input password
         IF the input password matches the current password
            END the dialog with a success flag
         ELSE (the input password does not match the current password)
            END the dialog with a failure flag
END PROCEDURE
```

```
PROCEDURE BGADLG::Encrypt

FOR the number of character in the string to encrypt
   transform the input character into an encrypted (lower-case) character

END PROCEDURE


PROCEDURE BGADLG::AuditProc

IF the message is WM_INITDIALOG
   call BGALIB::EGlobalAllocPtr to allocate memory for edit-text, INI keys and
values
   FOR each audit ID key in the [AuditIDs] section of the INI file
      set an edit-text line to match the audit ID line
   call BGALIB::EGlobalFreePtr to free the allocated memory

IF the message is WM_COMMAND
   IF the command is IDOK
      call BGALIB::EGlobalAllocPtr to allocate memory for the dialog edit-text
      get the edit-text from the dialog into global memory
      FOR each line in the edit-text
         set a matching entry in the [AuditIDs] section of the INI file
      call BGALIB::EGlobalFreePtr to free the allocated memory
      destroy this dialog box

   IF the command is IDCANCEL
      destroy this dialog box

END PROCEDURE


PROCEDURE BGADLG::ViewerProc

IF the message is WM_INITDIALOG
   call BGALIB::EGlobalAllocPtr to allocate memory for INI database viewer entries
   FOR each database viewer key in the [DB Viewers] section of the INI file
      set a dialog list-box line to match the Db viewer line
   get the current database viewer from the [DB Current] section of the INI file
   set the list-box corresponding to the current viewer to be the default list-box
item
   call BGALIB::EGlobalFreePtr to free the allocated memory

IF the message is WM_COMMAND
   IF the command is IDD_DBPROG
      IF the command code specifies a mouse double-click
         process the command as if it was IDOK

   IF the command is IDOK
      get the currently selected list-box item
      call BGALIB::EGlobalAllocPtr to allocate memory for the currently selected
item
      get the list-box item from the dialog box into global memory
      set the current database viewer INI entry to be the selected list-box item
      call BGALIB::EGlobalFreePtr to free the allocated memory
      destroy this dialog box

   IF the command is IDCANCEL
      destroy this dialog box

END PROCEDURE


PROCEDURE BGADLG::MinimumsProc

IF the message is WM_INITDIALOG
   set the dialog fields to the corresponding minimum data widths ...
      ... held in the Profile global structure

IF the message is WM_COMMAND
   IF the command is IDOK
      set the minimum data widths in the Profile global variable to those in the ...
         ... corresponding fields of the dialog box
```

```
      destroy this dialog box

   IF the command is IDCANCEL
      destroy this dialog box
END PROCEDURE


PROCEDURE BGADLG::MiscProc

IF the message is WM_INITDIALOG
   set the dialog fields to the corresponding mscellaneous data ...
      ... held in the Profile global structure

IF the message is WM_COMMAND
   IF the command is IDOK
      set the miscellaneous data in the Profile global variable to those in the ...
         ... corresponding fields of the dialog box
      destroy this dialog box

   IF the command is IDCANCEL
      destroy this dialog box
END PROCEDURE


PROCEDURE BGADLG::AboutProc

IF the message is WM_INITDIALOG
   store the about box type for future use
   IF the about box is a normal about box
      set the picture window handles to NULL
   ELSE (the about box is a picture sub-about box)
      set the picture window handles to the corresponding dialog box control handles

IF the message is WM_PAINT
   IF the about box is a picture sub-about box
      FOR the three pictures
         call BGADLG::DrawBitmap to display the resource file bitmap on dialog box

IF the message is WM_COMMAND
   IF the command is IDD_HIDDEN
      IF the shift and control keys are currently pressed
         run dialog box "AboutBmp" with BGADLG::AboutProc to display team pictures
      IF the shift and control and keypad-5 keys are currently pressed
         run dialog box "AboutBmp" with BGADLG::AboutProc to display silly pictures

   IF the command is IDCANCEL
      IF this about box is a normal about box and a picture sub-dialog is on display
         send the picture sub-about box an IDCANCEL command to close it down ...
            ... [send WM_COMMAND (IDCANCEL) -> BGADLG::AboutProc]
      destroy this dialog box

   IF the command is IDOK
      destroy this dialog box
END PROCEDURE


PROCEDURE BGADLG::SetDlgVersionInfo

get the file version information size from the resource file
IF the file version information size is greater than zero
   call BGALIB::EGlobalAllocPtr to allocate memory to hold the version information
   get the file version information into global memory
   FOR each version information field in the dialog box
      set the dialog field to the corresponding version information
   call BGALIB::EGlobalFreePtr to free the allocated memory

get the last boot data and time entries from the [Boot] section of the INI file
set the corresponding dialog box fields
END PROCEDURE
```

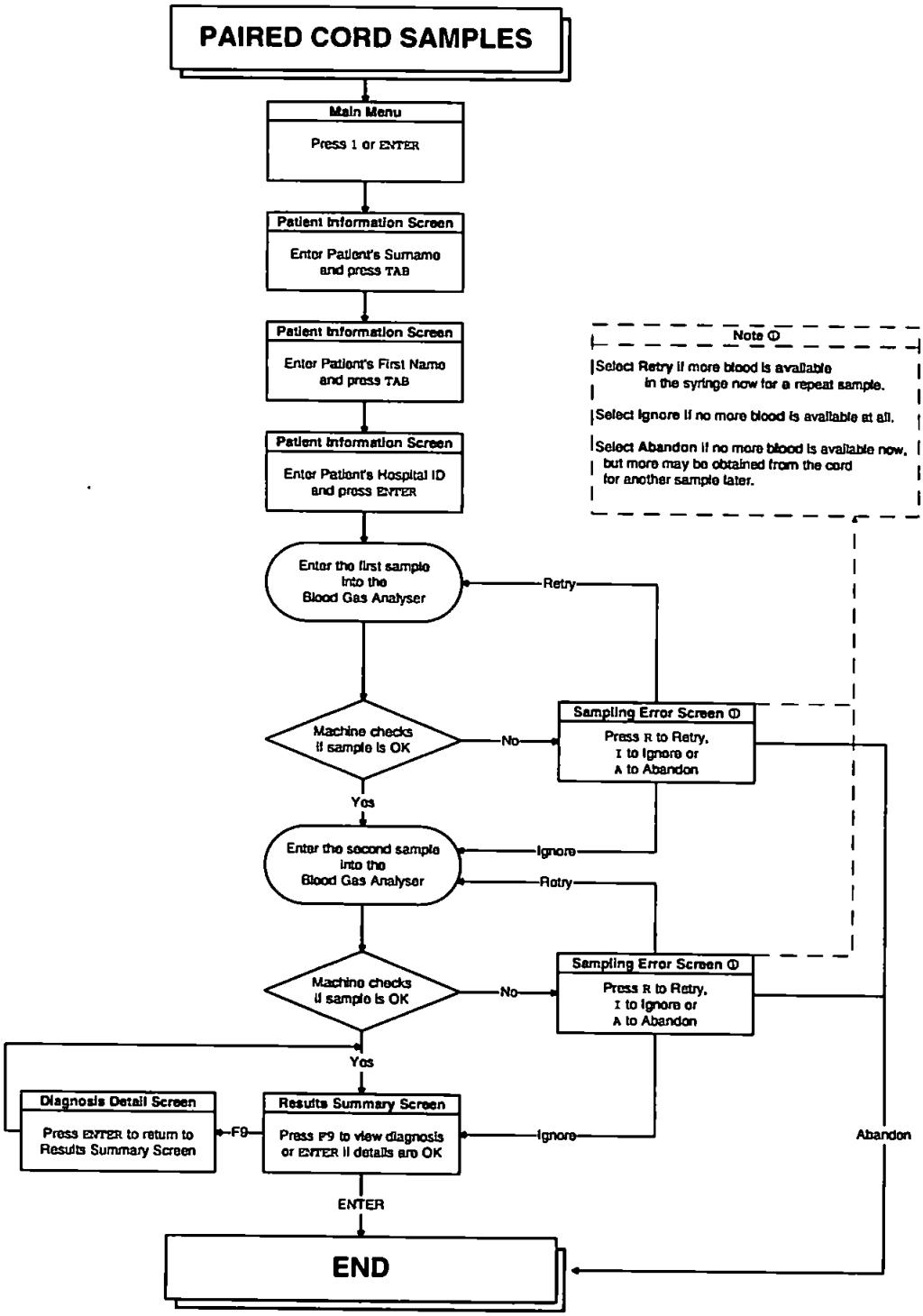
```
PROCEDURE BGADLG::DrawBitmap

draw the bitmap on the surface of the specified device context
(see Microsoft Windows Guide to Programming pp. 221-248, especially pp. 229-230)

END PROCEDURE


PROCEDURE BGATXT::JustifyTextPage

call BGALIB::EGlobalAllocPtr to allocate memory for 256 pointers to start of each
line
get the current window font information to calculate the height and width of chars
get the height and width of the window client area to be written on

FOR all the text pointed to by the lpText global variable
   skip any leading break characters (spaces) or line feeds chars (LF)
   remember the position of the first real character

   REPEAT
      find the next break, carriage-return (CR) or end-of-text (EOT) character
      remember if the line contains a tab character
      calculate the width of the line up to the current position
      IF the width will fit
         remember this as the last position that does fit
         add the number of break characters found to a running total for the line
   UNTIL the width of the line exceeds the width of the client area OR ...
      ... a CR or EOT character was found

   adjust the end of text pointer for a single word that is wider than one line
   call BGATXT::AppendLine to store the variables in global memory array
   IF the line contained a TAB character
      set the indent variable to indent all the following lines in the paragraph
   IF the line ends with a CR or EOT
      reset the indent variable to un-indent the paragraph
END FOR

call BGATXT::AppendLine to add blank lines to pad text to fit exactly on page
boundary

END PROCEDURE


PROCEDURE BAGTXT::AppendLine

IF the end of the 256 pointers initially allocated
   call BGALIB::EGlobalReallocPtr to allocate a further 256 pointers
store the current line information at the end of the global array

END PROCEDURE


PROCEDURE BGATXT::DisplayTextPage

get the first and last line of the currently invalid client rectangle
FOR each line in the invalid rectangle
   get the information for this line from the global array
   IF the line is not and end-of-paragraph
      justify the text to fill the width of a client line
   draw the text on the client area
   reset the text justification overflow variables

END PROCEDURE


PROCEDURE BGATXT::InitScrollBar

show the scroll bar with the calculated number of scroll lines
IF scrolling is required
   set the scroll range and scroll position to correct initial values

END PROCEDURE
```

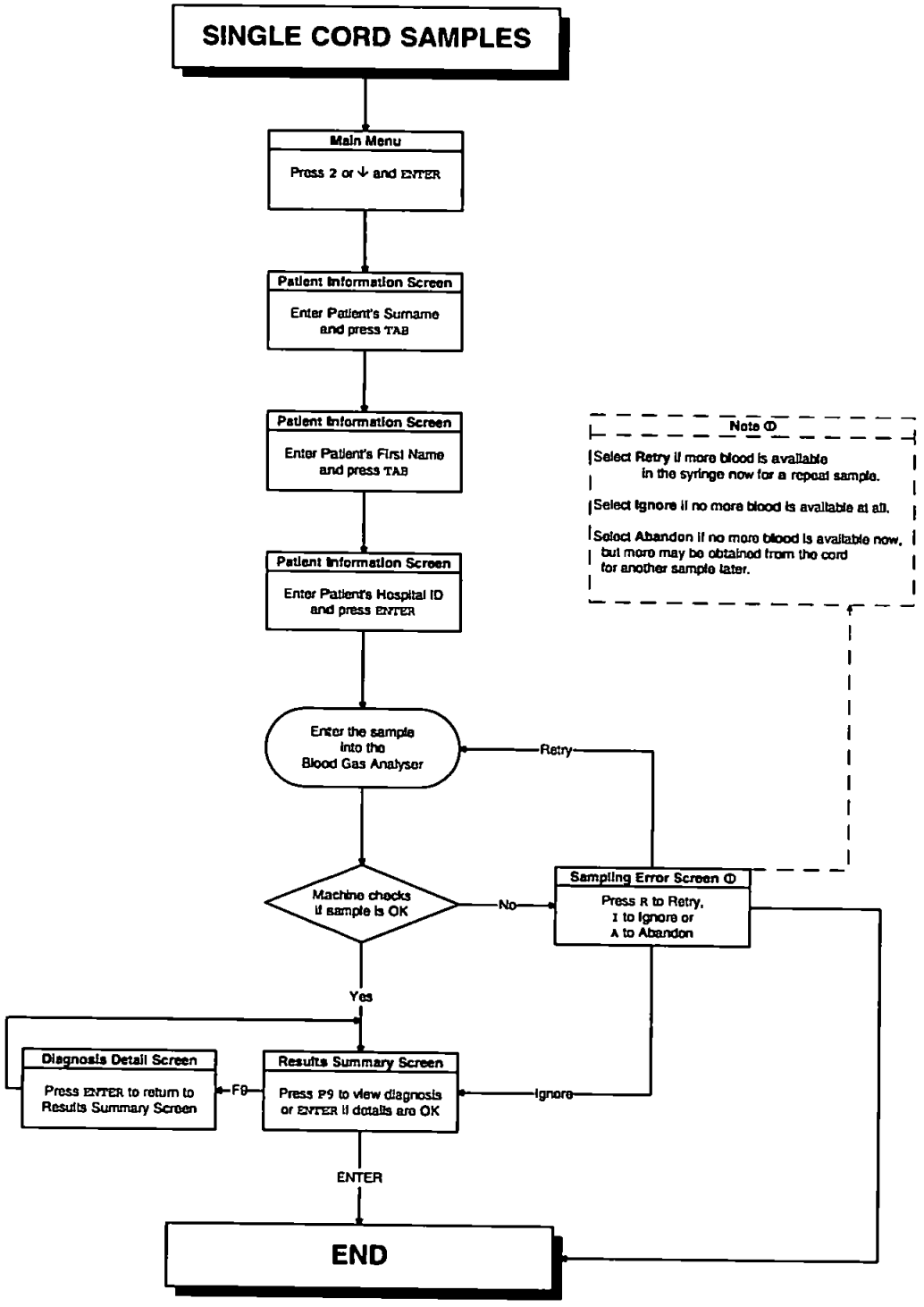
```
PROCEDURE BGATXT::ScrollTextPage

IF the scroll command is move-to-top (SB_TOP)
   set the scroll position to zero
IF the scroll command is move-to-bottom (SB_BOTTOM)
   set the scroll position to scroll maximum
IF the scroll command is move-up-one-line (SB_LINEUP)
   decrement the scroll position
IF the scroll command is move-down-one-line (SB_LINEDOWN)
   increment the scroll position
IF the scroll command is move-up-one-page (SB_PAGEUP)
   subtract one page from the scroll position
IF the scroll command is move-down-one-page (SB_PAGEDOWN)
   add one page to the scroll position
IF the scroll command is move-to-position (SB_THUMBPOSITION)
   set the scroll position to the specified position
IF the scroll command is moving-position (SB_THUMBTRACK)
   set the scroll position to the specified position

IF the display is being scrolled by a non-zero amount
   scroll the client area the specified distance
   call BGATXT:DisplayTextPage to repaint the scrolled client area

END PROCEDURE


PROCEDURE BGATXT::LoadResources

load all the summary and detail text pages into global space

END PROCEDURE


PROCEDURE BGATXT::FreeResources

free all the summary and detail text pages from global space

END PROCEDURE


PROCEDURE BGATXT::LoadInterpretation

call BGATXT::ReadIndex to find the position and size of the requested info in the
file
call BGALIB::EGlobalAllocPtr to allocate memory to hold the requested info
copy the requested info from the resource file into global memory

IF there has been a pH measurement error
   call BGATXT::ReadIndex to find the position and size of the pH error text
   call BGATXT::AppendInterpretation to add the extra text to that in memory

IF there has been a pCO2 measurement error
   call BGATXT::ReadIndex to find the position and size of the pCO2 error text
   call BGATXT::AppendInterpretation to add the extra text to that in memory

IF this was a paired cord sample BUT the samples came from the same vessel
   call BGATXT::ReadIndex to find the position and size of the same-vessel error
text
   call BGATXT::AppendInterpretation to add the extra text to that in memory

END PROCEDURE


PROCEDURE BGATXT::AppendInterpretation

call BGALIB::EGlobalReAllocPtr to re-allocate memory to add the extra text
append the extra text to the end of the current text

END PROCEDURE


PROCEDURE BGATXT::FreeInterpretation

call BGALIB::EGlobalFreePtr to free the allocated text memory
call BGALIB::EGlobalFreePtr to free the allocated line-justification memory
```

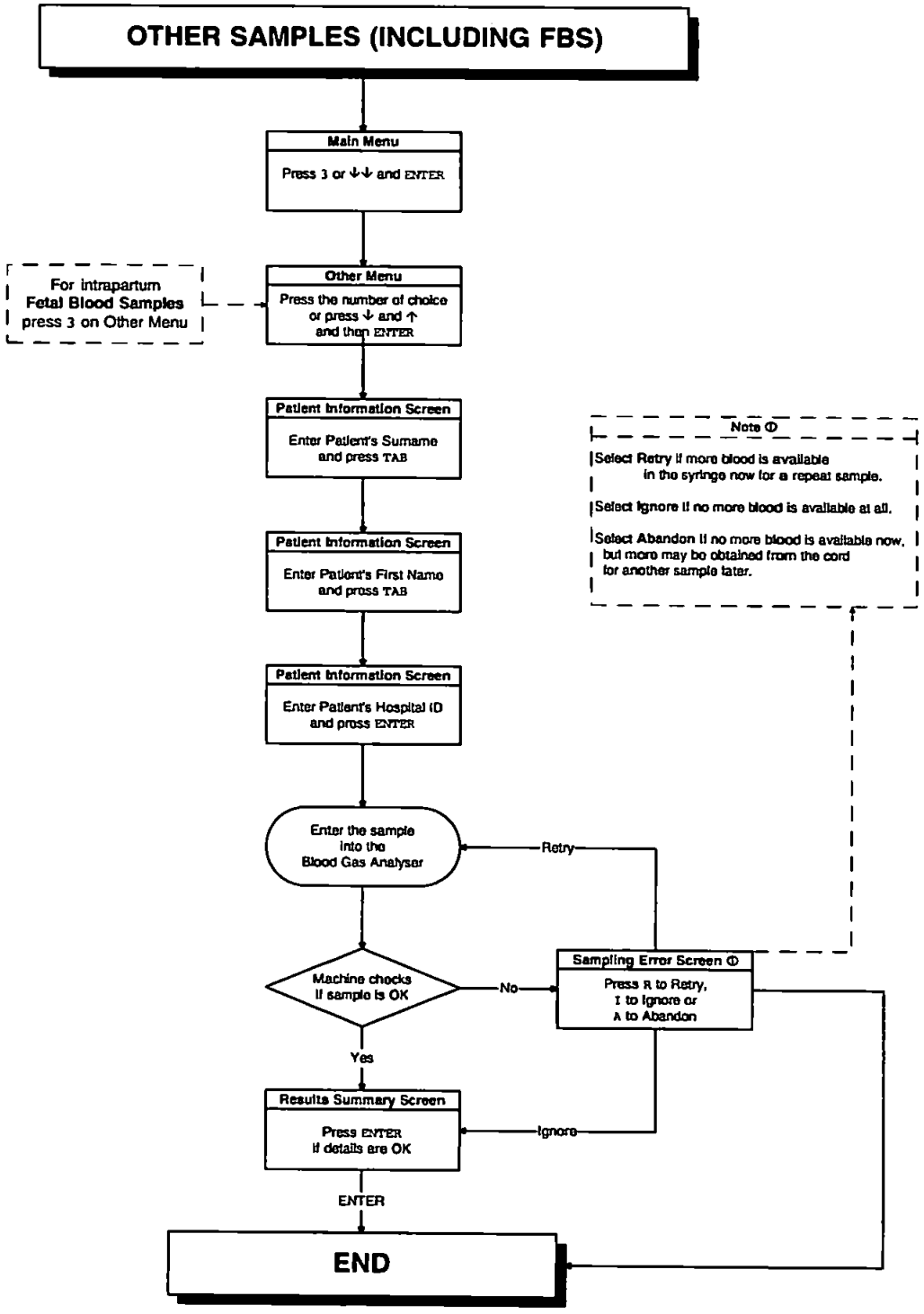
```
END PROCEDURE


PROCEDURE BGATXT::ReadIndex

move to the specified position in the index file
read the position and size of text information from the index file

END PROCEDURE


PROCEDURE BGAEXP::SampleCheck

check the sample for BGA238 measurement errors
check the sample for physiological inconsistencies according to its origin

END PROCEDURE


PROCEDURE BGAEXP::CalcBD

re-calculate the base-deficit using Siggarrd-Andersen's formula to BD(ecf)

END PROCEDURE


PROCEDURE BGAEXP::Diagnosis

remove sample with no pH and no pCO2 as totally useless
IF the origin is paired AND the samples are different
   set the type of the sample to double
ensure that the 'artery' of double samples or the 'best' of same-vessel samples ...
   ... is in the artery position
IF the type is double
   do additional physiological checks to ensure that the pair is jointly reasonable
perform the expert system diagnosis on the results

END PROCEDURE
```

Appendix C

User Flow Diagrams


PAIRED CORD SAMPLES
Main Menu
Press 1 or ENTER
Patient Information Screen
Enter Patient's Surname and press TAB
Patient Information Screen
Enter Patient's First Name and press TAB
Patient Information Screen
Enter Patient's Hospital ID and press ENTER
Note ①
Select Retry if more blood is available in the syringe now for a repeat sample.
Select Ignore if no more blood is available at all.
Select Abandon if no more blood is available now, but more may be obtained from the cord for another sample later.
Enter the first sample into the Blood Gas Analyser
Retry
Machine checks if sample is OK
No
Sampling Error Screen ①
Press R to Retry, I to Ignore or A to Abandon
Yes
Enter the second sample into the Blood Gas Analyser
Ignore
Retry
Machine checks if sample is OK
No
Sampling Error Screen ①
Press R to Retry, I to Ignore or A to Abandon
Yes
Diagnosis Detail Screen
Press ENTER to return to Results Summary Screen
F9
Results Summary Screen
Press F9 to view diagnosis or ENTER if details are OK
Ignore
Abandon
ENTER
END


SINGLE CORD SAMPLES
Main Menu
Press 2 or ↓ and ENTER
Patient Information Screen
Enter Patient's Surname and press TAB
Patient Information Screen
Enter Patient's First Name and press TAB
Patient Information Screen
Enter Patient's Hospital ID and press ENTER
Note ①
Select Retry if more blood is available in the syringe now for a repeat sample.
Select Ignore if no more blood is available at all.
Select Abandon if no more blood is available now, but more may be obtained from the cord for another sample later.
Enter the sample into the Blood Gas Analyser
Retry
Machine checks if sample is OK
No
Sampling Error Screen ①
Press R to Retry, I to Ignore or A to Abandon
Yes
Diagnosis Detail Screen
Press ENTER to return to Results Summary Screen
F9
Results Summary Screen
Press F9 to view diagnosis or ENTER if details are OK
Ignore
ENTER
END


OTHER SAMPLES (INCLUDING FBS)
Main Menu
Press 3 or ↓↓ and ENTER
For intrapartum
Fetal Blood Samples
press 3 on Other Menu
Other Menu
Press the number of choice
or press ↓ and ↑
and then ENTER
Patient Information Screen
Enter Patient's Surname
and press TAB
Patient Information Screen
Enter Patient's First Name
and press TAB
Patient Information Screen
Enter Patient's Hospital ID
and press ENTER
Enter the sample
into the
Blood Gas Analyser
Retry
Machine checks
If sample is OK
No
Sampling Error Screen ①
Press R to Retry,
I to Ignore or
A to Abandon
Yes
Results Summary Screen
Press ENTER
if details are OK
Ignore
ENTER
END
Note ①
Select Retry if more blood is available
in the syringe now for a repeat sample.
Select Ignore if no more blood is available at all.
Select Abandon if no more blood is available now,
but more may be obtained from the cord
for another sample later.

Appendix D

Crisp Expert System Rules

```
[0]
Impossible
The DataCare expert system has malfunctioned.
Please contact the software supplier as soon as possible.
[1]
No results at all
No interpretation is possible.
[10]
Single Vessel, pH only; pH < 7.10
The pH value is < 7.10 which is low for a vein sample.
[11]
Single Vessel, pH only; pH >= 7.10
The pH value is >= 7.10 which is not low for a vein sample.
A low arterial value cannot be excluded.
[20]
Probably severe acidemia; pH < 7.00, BD >= 12
Difficult to determine which vessel this is.
Verification with repeat paired sample is strongly recommended.
[21]
Moderate acidemia; pH < 7.10, BD >= 10
Probably vein sample.
Possible venous metabolic acidemia.
[22]
Non-significant acidemia; pH < 7.10, BD < 10
Probably vein sample.
If this is the vein cannot exclude arterial metabolic acidemia.
[23]
Questionable significance; pH 7.10 - 7.14, BD >= 10
Probably vein sample.
Arterial metabolic acidemia cannot be excluded.
[24]
Questionable significance; pH >= 7.15, BD >= 10
Probably vein sample with a high BD.
Arterial acidemia cannot be excluded.
[25]
Normal range; pH > 7.10, BD < 10
Probably vein sample.
Arterial acidemia cannot be excluded.
```

```
[30]
Significant acidemia; pHa < 7.05, pHv < 7.10
Limited interpretation possible (no BD's).
Both artery and vein have low pH's.
Possibility of neonatal sequelae.
[31]
Questionable significance; pHa < 7.05, pHv 7.10 - 7.24
Limited interpretation possible (no BD's).
Arterial pH is low but venous pH is in physiological range.
[32]
Acute acidemia with large A-V difference; pHa < 7.05, pHv >= 7.25
Limited interpretation possible (no BD's).
Clinical details are necessary to assess significance.
[33]
Questionable significance; pHa >= 7.05, pHv < 7.10
Limited interpretation possible (no BD's).
Vein pH is low and arterial pH is also approaching low levels.
[34]
Results OK; pHa >= 7.05, pHv >= 7.10
Limited interpretation possible (no BD's).
Both results are in the physiological range for pH.
[40]
Moderate - severe acidemia; pHa < 7.05, pHv < 7.10, BDv >= 10
Limited interpretation possible (no artery BD).
Low artery pH, venous metabolic acidemia.
Possibility of neonatal sequelae.
[41]
Significant - moderate acidemia; pHa < 7.05, pHv >= 7.10, BDv >=10
Limited interpretation possible (no artery BD).
Low artery pH, venous metabolic acidemia.
Possibility of neonatal sequelae.
[42]
Significant - moderate acidemia; pHa < 7.05, pHv < 7.10, BDv < 10
Limited interpretation possible (no artery BD).
Low artery pH and low vein pH.
Possibility of neonatal sequelae.
[43]
Questionable significance; pHa < 7.05, pHv 7.10 - 7.24, BDv < 10
Limited interpretation possible (no artery BD).
Low artery pH, venous results satisfactory.
Possibility of neonatal sequelae.
[44]
Acute A-V difference; pHa < 7.05, pHv >= 7.25, BDv < 10
Limited interpretation possible (no artery BD).
Low artery pH, vein normal.
Possibility of neonatal sequelae.
[50]
Moderate acidemia; pHa >= 7.05, pHv < 7.10, BDv >= 10
Limited interpretation possible (no artery BD).
Venous metabolic acidemia.
Possibility of neonatal sequelae.
[51]
Significant acidemia; pHa >= 7.05, pHv >= 7.10, BDv >= 10
Limited interpretation possible (no artery BD).
Venous metabolic acidemia.
Possibility of neonatal sequelae.
```

[52]
Significant acidemia; pHa >= 7.05, pHv < 7.10, BDv < 10
Limited interpretation possible (no artery BD).
No venous metabolic acidemia.
[53]
Results OK; pHa >= 7.05, pHv >= 7.10, BDv < 10
Limited interpretation possible (no artery BD).
Vein in physiological range.
[60]
Moderate or severe acidemia; pHa < 7.05, BDa >= 12, pHv < 7.10
Limited interpretation possible (no vein BD).
Arterial metabolic acidemia, low venous pH.
Possibility of neonatal sequelae.
[61]
Moderate acidemia; pHa < 7.05, BDa >= 12, pHv 7.10 - 7.24
Limited interpretation possible (no vein BD).
Arterial metabolic acidemia, vein pH lowish.
Possibility of neonatal sequelae.
[62]
Acute moderate acidemia with large AV difference; pHa < 7.05, BDa >= 12, pHv >= 7.25
Limited interpretation possible (no vein BD).
Arterial metabolic acidemia, vein pH normal.
Possibility of neonatal sequelae.
[63]
Significant acidemia; pHa < 7.05, BDa < 12, pHv < 7.10
Limited interpretation possible (no vein BD).
Low arterial and venous pH's.
[64]
Non-significant acidemia; pHa < 7.05, BDa < 12, pHv >= 7.10
Limited interpretation possible (no vein BD).
Low arterial pH, venous pH normal.
[70]
Moderate acidemia; pHa 7.05 - 7.09, BDa >= 12, pHv < 7.10
Limited interpretation possible (no vein BD).
Arterial metabolic acidemia, low venous pH.
Possibility of neonatal sequelae.
[71]
Significant acidemia; pHa 7.05 - 7.09, BDa >= 12, pHv >= 7.10
Limited interpretation possible (no vein BD).
Arterial metabolic acidemia, vein pH normal.
Possibility of neonatal sequelae.
[72]
Questionable significance; pHa >= 7.10, BDa >= 12, pHv >= 7.10
Limited interpretation possible (no vein BD).
Artery BD high, vein pH normal. Measurement error?
Possibility of neonatal sequelae.
[73]
Questionable significance; pHa >= 7.05, BDa < 12, pHv < 7.10
Limited interpretation possible (no vein BD).
No arterial metabolic acidemia, but low venous pH.
[74]
Results OK; pHa >= 7.05, BDa < 12, pHv >= 7.10
Limited interpretation possible (no vein BD).
All results are in physiological range.

[80]
Severe acidemia; pHa < 7.05, BDa >= 12, pHv < 7.10, BDv >= 10
Arterial and venous metabolic acidemia.
Possibility of neonatal hypoglycaemia and cardio-pulmonary sequelae.
[81]
Moderate acidemia; pHa < 7.05, BDa >= 12, pHv >= 7.10, BDv >=10
Arterial and venous metabolic acidemia.
Possibility of neonatal hypoglycaemia and perhaps cardio-pulmonary sequelae.
[82]
Moderate acidemia; pHa < 7.05, BDa >= 12, pHv < 7.10, BDv < 10
Arterial metabolic acidemia, vein pH is low.
Possibility of neonatal hypoglycaemia.
[83]
Acute moderate acidemia; pHa < 7.05, BDa >= 12, pHv 7.10 - 7.25, BDv < 10
Arterial metabolic acidemia, vein results normal.
Possibility of neonatal hypoglycaemia.
[84]
Acute moderate acidemia with large AV.difference; pHa<7.05, BDa>=12, pHv>=7.25, BDv<10
Possibility of neonatal hypoglycaemia.
[90]
Non-acute moderate acidemia; pHa < 7.05, BDa 10 - 12, pHv < 7.10, BDv >= 10
Arterial and venous pH's low and BD's high.
Possibility of neonatal sequelae.
[91]
Moderate acidemia; pHa < 7.05, BDa < 10, pHv < 7.10, BDv >= 10
Arterial pH low, venous metabolic acidemia.
Possibility of neonatal sequelae.
[92]
Significant acidemia; pHa < 7.05, BDa < 12, pHv >= 7.10, BDv >= 10
Arterial pH low and venous BD high.
[93]
Significant acidemia; pHa < 7.05, BDa < 12, pHv < 7.10, BDv < 10
Arterial and venous pH's low, but no metabolic acidemia.
[94]
Non-significant acidemia; pHa < 7.05, BDa < 12, pHv >= 7.10, BDv < 10
Arterial pH low, vein results in physiological range.
[100]
Moderate non-acute acidemia; pHa >= 7.05, BDa >= 12, pHv < 7.10, BDv >= 10
Arterial and venous metabolic acidemia.
Possibility of neonatal hypoglycaemia.
[101]
Significant acidemia; pHa 7.05 - 7.10, BDa >= 12, pHv >= 7.10, BDv >= 10
Arterial metabolic acidemia, high venous BD.
Possibility of neonatal hypoglycaemia.
[102]
Questionable significance; pHa >= 7.10, BDa >= 12, pHv >= 7.10, BDv >=10
Arterial and venous pH's not low, but BD's high.
Possibility of measurement error.
[103]
Significant acidemia; pHa >= 7.05, BDa >= 12, pHv < 7.10, BDv < 10
Arterial metabolic acidemia, venous pH low.
Possibility of neonatal hypoglycaemia.
[104]
Acute moderate acidemia; pHa 7.05 - 7.09, BDa >= 12, pHv >= 7.10, BDv < 10
Arterial metabolic acidemia, vein in physiological range.
Possibility of neonatal hypoglycaemia.

```
[105]
Questionable significance; pHa >= 7.10, BDa >= 12, pHv >= 7.10, BDv < 10
Arterial pH not low, but BD high.
Possibility of measurement error.
[110]
Significant acidemia; pHa >= 7.05, BDa < 12, pHv < 7.10, BDv >= 10
Venous metabolic acidemia, arterial pH approaching low level.
Possibility of neonatal hypoglycaemia.
[111]
Significant acidemia; pHa 7.05 - 7.14, BDa < 12, pHv 7.10 - 7.14, BDv >= 10
Venous metabolic acidemia, both pH's approaching low level.
Possibility of neonatal hypoglycaemia.
[112]
Non-acute mixed acidemia; pHa 7.05 - 7.14, BDa 8 - 11.9, pHv >= 7.15, BDv >= 10
Arterial pH lowish with both BD's highish.
Clinical details are necessary to assess significance.
[113]
Questionable significance; pHa >= 7.05, BDa < 12, pHv >= 7.10, BDv >= 10
Neither arterial nor venous pH's low, but high venous BD.
Possibility of measurement error.
[120]
All results OK; pHa >= 7.05, BDa < 12, pHv >= 7.10, BDv < 10
Arterial and venous results within physiological range.
[997]
Error in pH measurement(s).
[998]
Error in pCO2 measurement(s).
[999]
Results probably come from same vessel.
```

Appendix E

Preliminary Fuzzy Expert System Rules

```
// [0] extra rule to weight pH more than BD
IF, "Arterial pH", IS, PLAIN, "Low",
THEN, "Acidemia", IS, PLAIN, "Severe",

// [1] RULE 80
IF, "Arterial pH", IS, PLAIN, "Low",
AND, "Arterial BD", IS, PLAIN, "High",
AND, "Venous pH", IS, PLAIN, "Low",
AND, "Venous BD", IS, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Severe",
THEN, "Duration", IS, PLAIN, "Chronic",

// [2] RULE 81
IF, "Arterial pH", IS, PLAIN, "Low",
AND, "Arterial BD", IS, PLAIN, "High",
AND, "Venous pH", ISNOT, PLAIN, "Low",
AND, "Venous BD", IS, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Moderate",

// [3] RULE 82
IF, "Arterial pH", IS, PLAIN, "Low",
AND, "Arterial BD", IS, PLAIN, "High",
AND, "Venous pH", IS, PLAIN, "Low",
AND, "Venous BD", ISNOT, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Moderate",

// [4] RULE 83
IF, "Arterial pH", IS, PLAIN, "Low",
AND, "Arterial BD", IS, PLAIN, "High",
AND, "Venous pH", IS, PLAIN, "Mid",
AND, "Venous BD", ISNOT, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Moderate",
THEN, "Duration", IS, PLAIN, "Acute",

// [5] RULE 84
IF, "Arterial pH", IS, PLAIN, "Low",
AND, "Arterial BD", IS, PLAIN, "High",
AND, "Venous pH", IS, PLAIN, "High",
AND, "Venous BD", ISNOT, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Moderate",
THEN, "Duration", IS, VERY, "Acute",
```

```
// [6] RULE 90
IF, "Arterial pH", IS, PLAIN, "Low",
AND, "Arterial BD", IS, PLAIN, "Mid",
AND, "Venous pH", IS, PLAIN, "Low",
AND, "Venous BD", IS, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Moderate",
THEN, "Duration", ISNOT, PLAIN, "Acute",

// [7] RULE 91
IF, "Arterial pH", IS, PLAIN, "Low",
AND, "Arterial BD", IS, PLAIN, "Low",
AND, "Venous pH", IS, PLAIN, "Low",
AND, "Venous BD", IS, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Moderate",

// [8] RULE 92
IF, "Arterial pH", IS, PLAIN, "Low",
AND, "Arterial BD", ISNOT, PLAIN, "High",
AND, "Venous pH", ISNOT, PLAIN, "Low",
AND, "Venous BD", IS, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Significant",

// [9] RULE 93
IF, "Arterial pH", IS, PLAIN, "Low",
AND, "Arterial BD", ISNOT, PLAIN, "High",
AND, "Venous pH", IS, PLAIN, "Low",
AND, "Venous BD", ISNOT, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Significant",

// [10] RULE 94
IF, "Arterial pH", IS, PLAIN, "Low",
AND, "Arterial BD", ISNOT, PLAIN, "High",
AND, "Venous pH", ISNOT, PLAIN, "Low",
AND, "Venous BD", ISNOT, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Mild",

// [11] RULE 100
IF, "Arterial pH", ISNOT, PLAIN, "Low",
AND, "Arterial BD", IS, PLAIN, "High",
AND, "Venous pH", IS, PLAIN, "Low",
AND, "Venous BD", IS, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Moderate",
THEN, "Duration", ISNOT, PLAIN, "Acute",

// [12] RULE 101
IF, "Arterial pH", IS, PLAIN, "Mid",
AND, "Arterial BD", IS, PLAIN, "High",
AND, "Venous pH", ISNOT, PLAIN, "Low",
AND, "Venous BD", IS, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Significant",

// [13] RULE 103
IF, "Arterial pH", ISNOT, PLAIN, "Low",
AND, "Arterial BD", IS, PLAIN, "High",
AND, "Venous pH", IS, PLAIN, "Low",
AND, "Venous BD", ISNOT, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Significant",
```

```
// [14] RULE 104
IF, "Arterial pH", IS, PLAIN, "Mid",
AND, "Arterial BD", IS, PLAIN, "High",
AND, "Venous pH", ISNOT, PLAIN, "Low",
AND, "Venous BD", ISNOT, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Moderate",
THEN, "Duration", IS, PLAIN, "Acute",

// [15] RULE 110
IF, "Arterial pH", ISNOT, PLAIN, "Low",
AND, "Arterial BD", ISNOT, PLAIN, "High",
AND, "Venous pH", ISNOT, PLAIN, "Low",
AND, "Venous BD", IS, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Significant",

// [16] RULE 111
IF, "Arterial pH", IS, PLAIN, "Mid",
AND, "Arterial BD", ISNOT, PLAIN, "High",
AND, "Venous pH", IS, PLAIN, "Mid",
AND, "Venous BD", IS, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Significant",

// [17] RULE 112
IF, "Arterial pH", IS, PLAIN, "Mid",
AND, "Arterial BD", IS, PLAIN, "Mid",
AND, "Venous pH", IS, PLAIN, "High",
AND, "Venous BD", IS, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Significant",
THEN, "Acidemia", IS, PLAIN, "Mild",
THEN, "Duration", ISNOT, PLAIN, "Acute",

// [18] RULE 120
IF, "Arterial pH", ISNOT, PLAIN, "Low",
AND, "Arterial BD", ISNOT, PLAIN, "High",
AND, "Venous pH", ISNOT, PLAIN, "Low",
AND, "Venous BD", ISNOT, PLAIN, "High",
THEN, "Acidemia", IS, PLAIN, "Normal",
```

Appendix F

Preliminary Validation Data

P00136	7.07	57.8	12.8	7.17	45.0	21.8	112	F	41	2940	8	9
P00486	7.04	72.0	14.2	7.08	60.0	20.2	91	M	40	3580	6	9
P00583	7.01	54.0	18.0	7.37	26.2	33.8	84	M	40	3640	8	9
P00742	7.07	50.2	21.8	7.13	36.0	29.2	101	M	38	2760	9	9
P00862	6.96	51.0	38.2	7.11	45.8	32.2	81	F	42	2895	1	5
P00865	6.66	78.8	36.0	6.75	77.2	45.8	80	M	37	2210	1	1
P01201	7.21	47.2	20.2	7.31	27.8	36.0	113	M	39	2780	7	8
P01876	7.06	60.8	9.0	7.09	57.8	14.2	110	M	37	2075	5	9
P01932	7.06	53.2	33.8	7.16	42.0	30.0	101	F	39	3000	9	9
P02370	6.92	108.8	9.0	6.95	105.0	12.0	93	F	35	2380	8	9
P02833	7.09	53.2	21.0	7.21	41.2	18.8	101	F	38	2820	7	9
P02844	7.12	36.0	23.2	7.41	24.8	48.0	105	F	41	3320	9	9
P03444	7.02	65.2	20.2	7.14	50.2	30.0	81	M	40	3360	5	8
P03680	7.03	72.8	8.3	7.09	60.8	17.2	90	F	40	3020	9	9
P03977	6.96	90.8	3.0	7.05	66.8	15.0	90	M	42	3820	8	9
P04217	6.99	86.2	18.8	7.09	48.8	26.2	91	M	24	750	-	-
P04234	6.94	89.2	3.8	6.97	80.2	6.8	90	M	41	3820	6	9
P06086	7.24	30.7	15.0	7.45	26.2	29.2	105	F	39	3310	9	9
P06097	6.98	65.2	30.0	7.21	35.2	38.2	81	M	40	3700	7	9
P06463	7.04	81.0	9.0	7.09	72.8	14.2	93	M	39	3400	5	9
P06467	6.86	60.8	135.8	7.32	48.0	23.2	84	M	36	2560	4	9
P06617	7.10	63.8	15.0	7.18	45.0	24.0	112	M	39	2820	9	9
P07194	7.00	72.0	21.0	7.21	42.8	33.0	83	M	41	2920	6	8
P07320	7.01	66.8	11.2	7.06	63.8	14.2	80	M	38	2965	7	8
P07921	7.10	57.8	17.2	7.14	51.0	18.8	111	M	38	0	7	9
P08070	7.21	29.2	21.8	7.40	21.0	24.0	102	M	40	3770	9	9
P08130	6.87	129.8	20.2	7.11	59.2	38.2	94	M	40	3210	2	7
P08376	7.25	36.8	12.8	7.34	21.0	32.2	113	F	41	2600	8	9
P08512	6.61	140.2	8.3	7.18	53.2	39.0	83	M	41	4255	0	0
P08578	7.16	39.0	27.0	7.33	33.8	30.0	105	F	33	2210	9	9
P08587	7.00	54.0	20.2	7.22	47.2	26.2	83	M	42	3320	9	9
P08649	7.11	45.8	21.8	7.32	24.8	35.2	102	F	40	3260	9	9
P08722	7.08	51.0	18.8	7.29	36.8	29.2	104	M	38	2610	9	9
P08759	7.15	36.0	18.0	7.20	27.8	26.2	102	F	39	3050	9	9
P08997	7.05	47.2	14.2	7.09	41.2	17.2	100	M	28	1360	1	7
P09151	7.25	36.0	23.2	7.45	18.0	41.2	113	F	38	2640	9	9
P09200	7.06	66.8	15.0	7.11	54.8	18.8	111	M	41	3680	7	9
P09216	6.93	78.8	18.8	6.96	77.2	18.8	80	F	39	2975	9	9
P09258	7.05	66.0	20.2	7.27	26.2	42.8	112	F	41	3800	9	9
P09308	7.04	69.8	26.2	7.24	35.2	44.2	92	F	40	0	9	9
P09617	7.05	51.0	23.2	7.08	48.8	18.8	100	F	38	3180	6	9
P09677	7.04	80.2	9.0	7.23	51.8	26.2	94	F	41	3750	8	9
P09699	6.97	107.2	6.0	7.42	32.2	32.2	94	F	40	3610	6	9
P09865	7.04	75.8	12.0	7.18	45.0	24.0	92	F	38	4440	8	9
P09905	6.97	84.0	29.2	7.20	32.2	41.2	92	F	40	3680	7	8
P10034	7.06	42.8	23.2	7.36	30.0	48.0	104	M	34	2595	8	9
P10297	7.03	69.0	12.0	7.09	65.2	12.0	93	M	40	3640	8	9
P10364	7.04	62.3	9.8	7.30	39.0	21.8	84	M	40	3070	6	9
P10538	7.09	54.8	12.8	7.14	48.0	17.2	111	F	40	3200	9	9
P10828	7.06	57.8	17.2	7.28	42.0	24.0	104	F	40	3560	7	9

Appendix G

Integrated Vessel Identification Rules

```
// [0] all results similar => same vessel
IF,     ___,    VERY,   "Arterial pH",  EQ, ___,    VERY,   "Venous pH",
AND,    ___,    ABOUT,  "Arterial pCO2",EQ, ___,    ABOUT,  "Venous pCO2",
AND,    ___,    ABOUT,  "Arterial pO2", EQ, ___,    ABOUT,  "Venous pO2",
THEN,   ___,    PLAIN,  "Origin",       EQ, ___,    PLAIN,  "Same"

// [1] pCO2 OR pO2 results wrong way around => mixed
IF,     ___,    PLAIN,  "Venous pCO2",  GT, ___,    VAGUE,  "Arterial pCO2",
OR,     ___,    PLAIN,  "Venous pO2",   LT, ___,    VAGUE,  "Arterial pO2",
THEN,   ___,    PLAIN,  "Origin",       EQ, ___,    PLAIN,  "Mixed"

// [2] gas results very similar, but pHs different => mixed
IF,     ___,    PLAIN,  "Arterial pCO2",EQ, ___,    PLAIN,  "Venous pCO2",
OR,     ___,    PLAIN,  "Arterial pO2", EQ, ___,    PLAIN,  "Venous pO2",
AND,    ___,    PLAIN,  "Arterial pH",  LT, ___,    VAGUE,  "Venous pH",
THEN,   ___,    PLAIN,  "Origin",       EQ, ___,    PLAIN,  "Mixed"

// [3] pHs similar, but with correct gas values => mixed
IF,     ___,    PLAIN,  "Arterial pH",  EQ, ___,    PLAIN,  "Venous pH",
AND,    ___,    PLAIN,  "Arterial pCO2",GT, ___,    PLAIN,  "Venous pCO2",
AND,    ___,    PLAIN,  "Arterial pO2", LT, ___,    PLAIN,  "Venous pO2",
THEN,   ___,    PLAIN,  "Origin",       EQ, ___,    PLAIN,  "Mixed"

// [4] all results correct => different
IF,     ___,    PLAIN,  "Arterial pH",  LT, ___,    PLAIN,  "Venous pH",
AND,    ___,    PLAIN,  "Arterial pCO2",GT, ___,    PLAIN,  "Venous pCO2",
AND,    ___,    PLAIN,  "Arterial pO2", LT, ___,    PLAIN,  "Venous pO2",
THEN,   ___,    PLAIN,  "Origin",       EQ, ___,    PLAIN,  "Different"
```

Appendix H

Integrated Interpretation Rules

```
// [1]
IF,     ___,   PLAIN,  "Arterial pH",  EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD",  EQ, ___,   PLAIN,  "High",
AND,    ___,   PLAIN,  "Venous pH",    EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Venous BD",    EQ, ___,   PLAIN,  "High",
THEN,   ___,   PLAIN,  "Acidemia",     EQ, ___,   VERY,   "Severe",
THEN,   ___,   PLAIN,  "Component",    EQ, ___,   PLAIN,  "Metabolic",
THEN,   ___,   PLAIN,  "Duration",     EQ, ___,   VERY,   "Chronic",

// [2]
IF,     ___,   PLAIN,  "Arterial pH",  EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD",  EQ, ___,   PLAIN,  "High",
AND,    ___,   PLAIN,  "Venous pH",    EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Venous BD",    EQ, ___,   PLAIN,  "Mid",
THEN,   ___,   PLAIN,  "Acidemia",     EQ, ___,   PLAIN,  "Severe",
THEN,   ___,   PLAIN,  "Component",    EQ, ___,   PLAIN,  "Metabolic",
THEN,   ___,   PLAIN,  "Duration",     EQ, ___,   PLAIN,  "Chronic",

// [3]
IF,     ___,   PLAIN,  "Arterial pH",  EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD",  EQ, ___,   PLAIN,  "High",
AND,    ___,   PLAIN,  "Venous pH",    EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Venous BD",    EQ, ___,   PLAIN,  "Low",
THEN,   ___,   PLAIN,  "Acidemia",     EQ, ___,   PLAIN,  "Severe",
THEN,   ___,   PLAIN,  "Component",    EQ, ___,   PLAIN,  "Metabolic",
THEN,   ___,   PLAIN,  "Duration",     EQ, ___,   PLAIN,  "Intermediate",

// [4]
IF,     ___,   PLAIN,  "Arterial pH",  EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD",  EQ, ___,   PLAIN,  "High",
AND,    ___,   PLAIN,  "Venous pH",    EQ, ___,   PLAIN,  "Mid",
THEN,   ___,   PLAIN,  "Acidemia",     EQ, ___,   PLAIN,  "Severe",
THEN,   ___,   PLAIN,  "Component",    EQ, ___,   PLAIN,  "Metabolic",
THEN,   ___,   PLAIN,  "Duration",     EQ, ___,   PLAIN,  "Intermediate",

// [5]
IF,     ___,   PLAIN,  "Arterial pH",  EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD",  EQ, ___,   PLAIN,  "High",
AND,    ___,   PLAIN,  "Venous pH",    NL, ___,   PLAIN,  "Normal",
AND,    ___,   PLAIN,  "Venous BD",    EQ, ___,   PLAIN,  "High",
```

```
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Severe",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Metabolic",
THEN,   ___,   PLAIN,  "Duration",    EQ, ___,   PLAIN,  "Intermediate",

// [6]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD", EQ, ___,   PLAIN,  "High",
AND,    ___,   PLAIN,  "Venous pH",   NL, ___,   PLAIN,  "Normal",
AND,    ___,   PLAIN,  "Venous BD",   EQ, NOT,   PLAIN,  "High",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Severe",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Metabolic",
THEN,   ___,   PLAIN,  "Duration",    EQ, ___,   PLAIN,  "Acute",

// [7]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous pH",   EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Venous BD",   EQ, NOT,   PLAIN,  "Low",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Significant",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Mixed",
THEN,   ___,   PLAIN,  "Duration",    EQ, ___,   PLAIN,  "Chronic",

// [8]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous pH",   EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Venous BD",   EQ, ___,   PLAIN,  "Low",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Significant",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Mixed",
THEN,   ___,   PLAIN,  "Duration",    EQ, ___,   PLAIN,  "Intermediate",

// [9]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous pH",   EQ, ___,   PLAIN,  "Mid",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Significant",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Mixed",
THEN,   ___,   PLAIN,  "Duration",    EQ, ___,   PLAIN,  "Intermediate",

// [10]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous pH",   NL, ___,   PLAIN,  "Normal",
AND,    ___,   PLAIN,  "Venous BD",   EQ, ___,   PLAIN,  "High",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Significant",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Mixed",
THEN,   ___,   PLAIN,  "Duration",    EQ, ___,   PLAIN,  "Intermediate",

// [11]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous pH",   NL, ___,   PLAIN,  "Normal",
AND,    ___,   PLAIN,  "Venous BD",   EQ, NOT,   PLAIN,  "High",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Significant",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Mixed",
THEN,   ___,   PLAIN,  "Duration",    EQ, ___,   PLAIN,  "Acute",
```

```
// [12]
IF,     ___,   PLAIN,  "Arterial pH",  EQ,  ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD",  EQ,  ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Venous pH",    EQ,  ___,   PLAIN,  "Low",
THEN,   ___,   PLAIN,  "Acidemia",     EQ,  ___,   PLAIN,  "Moderate",
THEN,   ___,   PLAIN,  "Component",    EQ,  ___,   PLAIN,  "Respiratory",
THEN,   ___,   PLAIN,  "Duration",     EQ,  ___,   PLAIN,  "Chronic",

// [13]
IF,     ___,   PLAIN,  "Arterial pH",  EQ,  ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD",  EQ,  ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Venous pH",    EQ,  ___,   PLAIN,  "Mid",
THEN,   ___,   PLAIN,  "Acidemia",     EQ,  ___,   PLAIN,  "Moderate",
THEN,   ___,   PLAIN,  "Component",    EQ,  ___,   PLAIN,  "Respiratory",
THEN,   ___,   PLAIN,  "Duration",     EQ,  ___,   PLAIN,  "Intermediate",

// [14]
IF,     ___,   PLAIN,  "Arterial pH",  EQ,  ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Arterial BD",  EQ,  ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Venous pH",    NL,  ___,   PLAIN,  "Normal",
THEN,   ___,   PLAIN,  "Acidemia",     EQ,  ___,   PLAIN,  "Moderate",
THEN,   ___,   PLAIN,  "Component",    EQ,  ___,   PLAIN,  "Respiratory",
THEN,   ___,   PLAIN,  "Duration",     EQ,  ___,   PLAIN,  "Acute",

// [15]
IF,     ___,   PLAIN,  "Arterial pH",  EQ,  ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Arterial BD",  EQ,  ___,   PLAIN,  "High",
AND,    ___,   PLAIN,  "Venous pH",    EQ,  ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous BD",    EQ,  NOT,   PLAIN,  "Low",
THEN,   ___,   PLAIN,  "Acidemia",     EQ,  ___,   PLAIN,  "Moderate",
THEN,   ___,   PLAIN,  "Component",    EQ,  ___,   PLAIN,  "Metabolic",
THEN,   ___,   PLAIN,  "Duration",     EQ,  ___,   PLAIN,  "Chronic",

// [16]
IF,     ___,   PLAIN,  "Arterial pH",  EQ,  ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Arterial BD",  EQ,  ___,   PLAIN,  "High",
AND,    ___,   PLAIN,  "Venous pH",    EQ,  ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous BD",    EQ,  ___,   PLAIN,  "Low",
THEN,   ___,   PLAIN,  "Acidemia",     EQ,  ___,   PLAIN,  "Moderate",
THEN,   ___,   PLAIN,  "Component",    EQ,  ___,   PLAIN,  "Metabolic",
THEN,   ___,   PLAIN,  "Duration",     EQ,  ___,   PLAIN,  "Intermediate",

// [17]
IF,     ___,   PLAIN,  "Arterial pH",  EQ,  ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Arterial BD",  EQ,  ___,   PLAIN,  "High",
AND,    ___,   PLAIN,  "Venous pH",    NL,  ___,   PLAIN,  "Normal",
THEN,   ___,   PLAIN,  "Acidemia",     EQ,  ___,   PLAIN,  "Moderate",
THEN,   ___,   PLAIN,  "Component",    EQ,  ___,   PLAIN,  "Metabolic",
THEN,   ___,   PLAIN,  "Duration",     EQ,  ___,   PLAIN,  "Acute",

// [18]
IF,     ___,   PLAIN,  "Arterial pH",  EQ,  ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Arterial BD",  EQ,  ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous pH",    EQ,  ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous BD",    EQ,  ___,   PLAIN,  "High",
```

```
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Mild",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Mixed",
THEN,   ___,   PLAIN,  "Duration",    EQ, ___,   PLAIN,  "Chronic",

// [19]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Arterial BD", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous pH",   EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous BD",   EQ, NOT,   PLAIN,  "High",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Mild",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Mixed",
THEN,   ___,   PLAIN,  "Duration",    EQ, ___,   PLAIN,  "Intermediate",

// [20]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Arterial BD", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous pH",   NL, ___,   PLAIN,  "Normal",
AND,    ___,   PLAIN,  "Venous BD",   EQ, ___,   PLAIN,  "High",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Mild",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Mixed",
THEN,   ___,   PLAIN,  "Duration",    EQ, ___,   PLAIN,  "Intermediate",

// [21]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Arterial BD", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Venous pH",   NL, ___,   PLAIN,  "Normal",
AND,    ___,   PLAIN,  "Venous BD",   EQ, NOT,   PLAIN,  "High",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Mild",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Mixed",
THEN,   ___,   PLAIN,  "Duration",    EQ, ___,   PLAIN,  "Acute",

// [22]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Arterial BD", EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Venous pH",   EQ, ___,   PLAIN,  "Mid",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Mild",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Respiratory",
THEN,   ___,   PLAIN,  "Duration",    EQ, ___,   PLAIN,  "Intermediate",

// [23]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "Mid",
AND,    ___,   PLAIN,  "Arterial BD", EQ, ___,   PLAIN,  "Low",
AND,    ___,   PLAIN,  "Venous pH",   NL, ___,   PLAIN,  "Normal",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Mild",
THEN,   ___,   PLAIN,  "Component",   EQ, ___,   PLAIN,  "Respiratory",

// [24]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "Normal",
AND,    ___,   PLAIN,  "Venous pH",   EQ, ___,   PLAIN,  "Normal",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Normal",

// [25]
IF,     ___,   PLAIN,  "Arterial pH", EQ, ___,   PLAIN,  "High",
AND,    ___,   PLAIN,  "Venous pH",   EQ, ___,   PLAIN,  "High",
THEN,   ___,   PLAIN,  "Acidemia",    EQ, ___,   PLAIN,  "Alkalotic",
```

Appendix I

Integrated Fuzzy Expert System Validation Data

Vessel Identification Data

P00007	7.32	33.8	21.0	7.33	42.8	17.2	25	M	40	3540	9	9
P00040	7.29	45.0	24.8	7.30	45.0	24.8	25	F	38	2980	9	9
P00373	7.16	57.0	15.8	7.21	54.8	18.8	120	F	38	2910	9	9
P00583	7.01	54.0	18.0	7.37	26.2	33.8	84	M	40	3640	8	9
P00615	7.38	26.2	35.2	7.44	24.0	36.8	120	M	41	3130	9	9
P00633	7.37	38.2	39.0	7.38	39.8	38.2	25	M	40	3000	9	9
P01006	7.40	39.0	21.0	7.41	33.8	26.2	25	F	39	3980	9	9
P01009	7.23	41.2	35.2	7.25	41.2	35.2	25	F	41	2400	9	9
P01426	7.33	48.0	18.8	7.36	47.2	21.8	34	F	39	3420	9	9
P01726	7.45	32.2	30.0	7.47	33.8	27.8	25	F	41	3820	9	9
P01780	7.33	39.8	23.2	7.37	36.8	24.0	120	M	38	3000	8	9
P01802	7.41	39.0	15.8	7.45	33.8	21.0	120	F	33	2495	9	9
P01927	7.29	45.8	27.8	7.30	36.8	20.2	25	M	38	3410	9	9
P01980	6.94	105.0	12.0	7.27	51.8	23.2	94	M	41	3380	5	8
P02000	7.26	45.8	33.0	7.27	42.8	30.7	25	M	38	2320	9	9
P02118	7.17	60.8	21.8	7.17	56.2	18.0	25	M	37	2900	9	9
P02233	7.16	54.8	21.0	7.22	48.8	24.0	120	F	38	2660	9	9
P02319	7.30	50.2	20.2	7.42	35.2	35.2	120	F	37	2720	9	9
P02442	7.36	41.2	18.8	7.37	39.0	21.0	25	F	32	2065	9	9
P02516	7.30	39.8	18.0	7.34	36.0	21.8	120	F	41	3380	9	9
P02638	7.29	44.2	15.0	7.33	42.8	20.2	34	F	36	2400	9	9
P03351	7.33	39.8	21.8	7.35	42.0	24.8	25	F	41	3500	9	9
P03692	7.28	38.2	20.2	7.32	33.8	24.8	120	M	41	3790	9	9
P04013	7.36	45.0	27.0	7.38	42.8	29.2	25	M	38	3440	9	9
P04107	7.38	35.2	14.2	7.41	30.0	15.8	120	M	40	3680	9	9
P04195	7.02	78.0	5.3	7.03	71.2	12.0	22	M	40	3260	9	9
P04375	7.48	27.8	15.0	7.50	27.0	18.8	25	F	40	3980	9	9
P04846	7.31	42.0	17.2	7.36	38.2	26.2	120	F	42	4020	9	9
P05517	7.24	48.0	17.2	7.29	47.2	20.2	34	F	41	3720	9	9
P05594	7.31	41.2	20.2	7.34	38.2	21.0	120	F	40	3800	9	9
P05953	7.26	39.8	33.8	7.32	39.0	29.2	34	M	38	3475	9	9
P06370	7.32	36.8	32.2	7.33	39.0	23.2	25	M	38	2960	9	9
P06485	7.21	42.0	18.0	7.28	39.0	21.0	34	M	39	3920	9	9
P06536	7.28	51.8	12.0	7.33	42.8	15.8	120	F	37	2700	9	9
P06672	7.28	39.8	8.3	7.31	36.0	15.0	120	M	42	3730	9	9
P06693	7.31	41.2	12.8	7.36	33.0	20.2	120	F	40	3430	8	9
P07190	7.37	39.8	26.2	7.39	36.0	27.8	25	M	38	3540	9	9
P07234	7.39	32.2	35.2	7.39	30.7	33.8	25	F	40	4120	9	9
P07500	7.29	45.0	29.2	7.37	44.2	29.2	34	F	37	3050	9	9
P07692	7.36	45.8	21.8	7.39	39.0	36.8	120	F	37	3320	9	9
P07725	7.19	36.8	21.0	7.23	41.2	21.8	34	F	40	2880	8	9
P07932	7.35	38.2	33.0	7.39	38.2	33.0	34	F	40	3380	9	9
P07959	7.35	39.8	17.2	7.38	36.0	21.0	120	M	34	2355	9	9
P08060	7.37	27.0	23.2	7.43	26.2	26.2	34	M	35	2160	9	9
P08084	7.14	51.8	20.2	7.17	54.8	17.2	34	M	38	4100	9	9
P08166	7.42	23.2	38.2	7.43	21.8	36.0	25	F	40	2700	9	9
P08221	7.23	48.0	8.3	7.45	27.0	51.8	120	M	39	3460	9	9
P09341	7.24	51.0	24.0	7.30	53.2	21.0	34	F	40	4340	8	9
P09823	7.31	32.2	23.2	7.37	29.2	29.2	34	M	40	2950	9	9
P09989	7.25	56.2	23.2	7.28	42.8	30.7	120	F	41	2975	9	9

Numeric Interpretation Data

P00300	7.03	62.3	21.0	7.28	33.8	42.0	84	M	41	3830	7	9
P00775	6.97	96.0	3.8	7.03	80.2	14.2	93	F	40	3310	2	9
P00881	7.20	38.2	20.2	7.22	33.8	21.0	24	F	41	4340	6	7
P00951	7.01	77.2	15.0	7.25	35.2	33.0	32	M	41	3320	9	9
P01012	7.06	56.2	20.2	7.13	42.0	32.2	101	M	36	3050	3	5
P01757	6.96	56.2	20.2	7.41	32.2	17.2	84	F	38	2800	9	9
P02039	6.95	78.8	21.0	7.19	53.2	20.2	83	F	32	1584	0	5
P02122	7.03	75.8	17.2	7.12	63.8	21.0	94	F	40	3570	9	9
P02261	6.97	78.8	18.0	6.97	72.0	17.2	20	F	42	2770	3	7
P02507	7.03	84.8	5.3	7.07	84.0	12.0	30	M	40	3880	6	9
P02593	7.05	66.0	21.8	7.09	72.0	12.8	33	M	40	3260	9	9
P02697	7.25	45.0	17.2	7.39	35.2	23.2	120	F	40	3930	5	9
P03004	7.13	60.8	8.3	7.32	36.8	24.0	120	M	42	3970	9	9
P03345	7.19	54.0	29.2	7.30	41.2	35.2	120	M	39	3345	9	9
P03541	7.04	53.2	29.2	7.15	44.2	24.8	81	F	39	3020	8	8
P03879	6.82	102.8	3.0	6.88	89.2	12.8	80	F	40	3110	8	9
P04022	6.68	144.0	6.8	6.72	125.2	6.8	80	M	29	1250	0	1
P04112	7.09	77.2	12.8	7.09	74.2	12.8	22	M	37	1980	8	9
P04359	7.07	47.2	17.2	7.24	38.2	21.0	104	M	40	3340	9	9
P04695	7.32	36.0	21.0	7.39	35.2	24.0	34	M	42	3740	9	9
P04943	7.11	47.2	23.2	7.36	27.0	45.0	105	F	36	2680	7	9
P05362	7.02	60.0	9.0	7.23	45.0	18.8	83	M	40	3360	8	9
P05783	7.16	48.8	17.2	7.22	39.8	29.2	113	M	41	3180	5	9
P05859	7.12	9.8	39.8	7.38	38.2	36.0	53	M	38	2970	9	9
P06206	7.06	63.8	17.2	7.14	48.8	18.8	111	F	41	3655	9	9
P06594	7.15	33.0	15.0	7.36	26.2	23.2	105	M	39	3020	9	9
P06786	7.02	69.8	9.8	7.05	63.0	12.0	90	M	39	3250	6	9
P07197	7.51	26.2	17.2	7.60	18.8	30.0	120	M	33	1809	9	9
P07334	6.79	126.7	5.3	6.82	119.2	9.8	80	F	39	2980	6	7
P07855	7.03	72.8	12.8	7.15	44.2	30.0	92	F	42	3380	9	9
P07858	7.18	36.8	15.0	7.23	29.2	36.8	102	F	40	3250	9	9
P07908	7.01	99.8	15.0	7.37	24.8	39.8	64	F	42	4730	7	9
P08182	7.10	62.3	15.8	7.15	48.8	21.0	112	M	38	3120	9	9
P08277	7.28	41.2	30.7	7.28	39.0	48.8	25	M	32	2040	6	9
P08390	7.34	45.0	14.2	7.43	26.2	30.0	74	M	39	4395	9	9
P08509	7.18	32.2	23.2	7.36	21.8	35.2	102	M	40	3500	9	9
P08547	7.02	57.8	21.8	7.11	60.8	21.8	31	M	40	3380	8	9
P08709	7.14	45.0	17.2	7.25	36.8	24.8	105	M	40	3340	9	9
P08905	7.08	60.0	11.2	7.13	53.2	15.8	111	M	41	4070	5	8
P08998	6.89	72.0	12.8	7.25	26.2	33.0	81	M	28	1265	1	5
P09029	6.99	78.8	12.8	7.07	62.3	18.0	90	M	41	3940	5	9
P09113	6.98	83.2	15.0	7.12	54.0	27.8	92	M	41	3300	5	9
P09276	7.17	59.2	21.8	7.30	36.8	39.8	120	F	40	2880	9	9
P09598	7.05	72.0	14.2	7.13	51.8	23.2	111	M	41	2500	8	9
P10358	7.28	51.0	5.3	7.37	44.2	12.8	120	F	40	4260	7	9
P10449	6.93	74.2	30.7	7.30	39.8	32.2	84	M	38	3085	0	7
P10715	7.01	77.2	5.3	7.10	54.8	15.8	92	F	40	3210	7	9
P10961	7.17	42.0	20.2	7.25	30.7	33.0	113	M	38	3000	9	9
P10984	7.24	53.2	11.2	7.28	41.2	15.8	120	M	41	3710	6	9
P11104	7.26	56.2	14.2	7.36	39.0	26.2	120	M	41	3940	5	9

Linguistic Interpretation Data

P00258	7.06	69.8	9.0	7.20	41.2	27.8	112	M	39	3100	8	9
P00775	6.97	96.0	3.8	7.03	80.2	14.2	93	F	40	3310	2	9
P00788	7.09	48.0	21.0	7.28	35.2	29.2	104	F	40	3680	8	9
P00951	7.01	77.2	15.0	7.25	35.2	33.0	32	M	41	3320	9	9
P01012	7.06	56.2	20.2	7.13	42.0	32.2	101	M	36	3050	3	5
P01730	7.01	72.0	17.2	7.04	66.0	20.2	90	M	40	3120	9	9
P01757	6.96	56.2	20.2	7.41	32.2	17.2	84	F	38	2800	9	9
P02122	7.03	75.8	17.2	7.12	63.8	21.0	94	F	40	3570	9	9
P02261	6.97	78.8	18.0	6.97	72.0	17.2	20	F	42	2770	3	7
P02507	7.03	84.8	5.3	7.07	84.0	12.0	30	M	40	3880	6	9
P02593	7.05	66.0	21.8	7.09	72.0	12.8	33	M	40	3260	9	9
P02611	7.08	51.0	15.8	7.14	42.8	18.8	101	M	39	2900	9	9
P02927	6.99	84.0	8.3	7.07	66.8	18.0	93	F	39	3050	5	9
P03037	7.16	54.0	15.8	7.25	42.8	24.0	120	M	40	3910	9	9
P03530	7.25	63.0	12.0	7.31	47.2	32.2	120	M	37	2385	9	9
P03855	7.06	51.0	18.8	7.10	45.8	23.2	101	M	40	3450	6	7
P03879	6.82	102.8	3.0	6.88	89.2	12.8	80	F	40	3110	8	9
P04083	7.22	59.2	6.8	7.30	47.2	12.8	120	F	39	3585	9	9
P04359	7.07	47.2	17.2	7.24	38.2	21.0	104	M	40	3340	9	9
P04695	7.32	36.0	21.0	7.39	35.2	24.0	34	M	42	3740	9	9
P04831	7.31	51.0	15.0	7.36	42.8	23.2	120	M	41	4080	9	9
P05362	7.02	60.0	9.0	7.23	45.0	18.8	83	M	40	3360	8	9
P05559	7.06	48.0	33.0	7.39	32.2	39.0	104	F	42	3480	7	9
P05859	7.12	9.8	39.8	7.38	38.2	36.0	53	M	38	2970	9	9
P06102	7.15	51.8	15.8	7.31	26.2	39.0	113	M	39	3190	8	9
P06300	7.03	80.2	3.8	7.11	68.2	11.2	94	F	41	2990	6	9
P06622	6.93	104.2	6.8	6.93	99.0	18.8	21	M	38	3940	8	9
P07110	7.19	60.8	9.8	7.37	33.0	29.2	120	M	40	2700	8	9
P07282	7.19	35.2	11.2	7.26	24.0	27.0	102	M	41	4410	5	8
P07855	7.03	72.8	12.8	7.15	44.2	30.0	92	F	42	3380	9	9
P07858	7.18	36.8	15.0	7.23	29.2	36.8	102	F	40	3250	9	9
P07908	7.01	99.8	15.0	7.37	24.8	39.8	64	F	42	4730	7	9
P08118	7.11	45.0	18.8	7.13	47.2	17.2	23	M	41	2680	8	9
P08182	7.10	62.3	15.8	7.15	48.8	21.0	112	M	38	3120	9	9
P08277	7.28	41.2	30.7	7.28	39.0	48.8	25	M	32	2040	6	9
P08390	7.34	45.0	14.2	7.43	26.2	30.0	74	M	39	4395	9	9
P08547	7.02	57.8	21.8	7.11	60.8	21.8	31	M	40	3380	8	9
P08709	7.14	45.0	17.2	7.25	36.8	24.8	105	M	40	3340	9	9
P08905	7.08	60.0	11.2	7.13	53.2	15.8	111	M	41	4070	5	8
P08998	6.89	72.0	12.8	7.25	26.2	33.0	81	M	28	1265	1	5
P09029	6.99	78.8	12.8	7.07	62.3	18.0	90	M	41	3940	5	9
P09063	7.27	45.8	18.0	7.35	33.0	29.2	120	M	40	4160	9	9
P09136	6.93	89.2	12.8	7.12	51.0	32.2	81	F	38	3040	9	9
P09475	7.01	66.0	30.7	7.17	48.8	27.0	83	F	40	3300	9	9
P10051	7.26	45.0	24.0	7.31	39.0	29.2	120	M	37	3300	8	9
P10400	7.23	48.8	21.8	7.26	41.2	27.0	120	M	40	3680	9	9
P10586	7.09	63.8	9.0	7.22	39.0	23.2	112	M	41	3420	9	9
P10941	7.04	78.8	15.0	7.08	72.0	18.8	93	M	34	1995	3	8
P10961	7.17	42.0	20.2	7.25	30.7	33.0	113	M	38	3000	9	9
P11030	6.99	81.8	14.2	7.11	68.2	27.0	94	M	39	3050	9	9

Appendix J

MATLAB Ranks Function

```
function [RANK] = ranks(X)
%RANKS Produces a matrix of ranks of the original matrix
%   [RANK] = RANKS(X) returns a matrix RANK of the
%   data assigned a rank of 1 .. m in order from lowest to highest
%   for an input matrix X whose rows are observations and whose
%   columns are variables.
%
%   Tied observations are assigned the average rank.

% Jon Garibaldi: University of Plymouth: 31st Jan 1997
%
% Uses a 'clever' algorithm to set the rankings (including ties)
% without looping through each element of the ranking array,
% to greatly improve efficiency for large arrays.
%
% A matrix is formed where a 1 indicates that sorted elements
% either form the start of a set of tied observations or the
% end of a set of tied observations.  There must be an even
% number (including 0) of non-zero elements of this matrix.
% For example for two sets of twelve observations with no ties
% in the first column and the 3rd & 4th tied and 10th, 11th and 12th
% tied in the second column would produce the matrix:
%   0   0
%   0   0
%   0   1
%   0   1
%   0   0
%   0   0
%   0   0
%   0   0
%   0   0
%   0   1
%   0   0
%   0   1
%
% Then for each column:
% 'find' the ones to produce the null vector for the first
% and [ 3, 4, 10, 12 ] for the second column.
% This vector then provides the indices which can be set
% with array operations. e.g. for column 1 ranks 1:12 = 1:12;
% for column 2 ranks 1:2 = 1:2, 3:4 = 3.5, 6:9 = 6:9 and 10:12 = 11
```

```
%
% 'dead' easy and very much quicker (~10x) than a brute force loop!


if nargin ~= 1
    error('Requires exactly one matrix input argument.');
end

[m, n]= size(X);
if m < 2 | n < 2
    error('Argument must be at least a 2x2 matrix.');
end

% sort to get order and indices
[Y, Z]= sort(X);

% initialise the target rank matrix
RANK= zeros(m, n);

% trickery to get a matrix of rank boundaries in sorted matrix
% i.e. the array E is 1 at each start&end of tied ranks
e= (Y(1:m-1,:) == Y(2:m,:));
E= xor([ e; zeros(1, n) ], [ zeros(1, n); e ]);

for i=1:n
    % get vector of start&end of tied rank pairs
    Ei= [find(E(:,i));m+1;m+1];
    Zi= Z(:,i);

    k1= 1;
    for j=1:2:size(Ei)
        k2= Ei(j) - 1;
        if k1 <= k2
            %disp(sprintf('set ranks %d to %d',k1,k2));
            RANK(Zi(k1:k2),i)= [k1:k2]';
        end
        k1= Ei(j);
        k2= Ei(j+1);
        if k1 < k2
            %disp(sprintf('set tied ranks %d to %d',k1,k2));
            RANK(Zi(k1:k2),i)= ones(k2-k1+1,1) .* (k1 + k2) / 2;
        end
        k1= k2 + 1;
    end
end
```

```
function [RHO,SIG,RANK] = rankcorr(X,tail)
%RANKCORR Spearman rank order correlation.
%   [RHO,SIG,RANK] = RANKCORR(X,TAIL) returns a matrix RHO of the
%   Spearman rank order correlation coefficients calculated from
%   an input matrix X whose rows are observations and whose
%   columns are variables.
%
%   A matrix SIG of the significance of each RHO and
%   a matrix RANK of the calculated rankings are also returned.
%
%   TAIL is an optional parameter to indicate whether the
%   significance of rho should be calculated for one-sided or
%   two-sided hypotheses, with default value 1 (one-sided).

% Jon Garibaldi: University of Plymouth: 31st Jan 1997
%

if nargin < 1
    error('Requires at least one input argument.');
end

if nargin < 2
    tail= 1;
end

if tail ~= 1 & tail ~= 2
    error('Arg "tail" must be 1 (one-sided) or 2 (two-sided).');
end

[m, n]= size(X);
if m < 2 | n < 2
    error('First argument must be at least a 2x2 matrix.');
end

% get the rankings and calc the correlation coefficient for the ranks
RANK= ranks(X);
RHO= corrcoef(RANK);

if nargout > 1
    if m <= 30
        T= RHO .* sqrt((m - 2) ./ (1 - RHO .^ 2 + eps));
        SIG= (0.5 - abs(tcdf(T, m - 2) - 0.5)) * tail
    else
        Z= RHO .* sqrt(m - 1);
        SIG= (0.5 - abs(normcdf(Z) - 0.5)) * tail
    end
end
```